

THÈSE

Pour obtenir le grade de
Docteur

Délivré par **Montpellier SupAgro**

Préparée au sein de l'école doctorale GAIA
Et de l'unité de recherche
**Architecture et Fonctionnement des Espèces
Fruitières (AFEF)**

Spécialité : Biologie, Interactions, Diversité Adaptative des
Plantes

Présentée par **Alix ALLARD**

**Estimation de la valeur génétique de
pommiers hybrides pour plusieurs
caractères d'intérêt agronomique et sur la
base de leurs apparentements**

Soutenue le 22/03/2016 devant le jury composé de

Mr Alain CHARCOSSET, DR, INRA
Mr Andrea PATOCCHI, DR, Agroscope
Mr Jacques DAVID, DR, Montpellier SupAgro
Mme Christelle PITIOT, Ingénieur, Novadi
Mme Evelyne COSTES, DR, INRA
Mr François LAURENS, Ingénieur, INRA
Mr Jean-Jacques KELNER, MC, Montpellier SupAgro

Rapporteur
Rapporteur
Examinateur
Examinatrice
Encadrante
Co-Encadrant
Invité

Montpellier SupAgro

Ecole Doctorale : GAIA

THESE

Pour obtenir le grade de

Docteur de Montpellier SupAgro

Discipline : Biologie, Interactions, Diversité Adaptative des Plantes

Présentée et soutenue par

Alix ALLARD

**Estimation de la valeur génétique de pommiers hybrides sur plusieurs caractères d'intérêt
agronomique sur la base de leurs apparentements**

Thèse dirigée par Evelyne COSTES, François LAURENS, et Jean-Jacques KELNER

Soutenue le 22 Mars 2016

Composition du Jury

Alain CHARCOSSET	INRA, UMR GQE – Le Moulon, Gif-sur-Yvette	Rapporteur
Andrea PATOCCHI	Agroscope, IPS, Wädenswil	Rapporteur
Jacques DAVID	Montpellier SupAgro, UMR DIAPC, Montpellier	Examineur
Christelle PITIOT	NOVADI-CEP INNOVATION, Lyon	Examinatrice
Evelyne COSTES	INRA, UMR AGAP, Montpellier	Encadrante
François LAURENS	INRA, UMR IRHS, Beaucauzé	Co-Encadrant
Jean-Jacques KELNER	Montpellier SupAgro, UMR AGAP, Montpellier	Invité

Résumé

Chez le pommier, la sélection porte principalement sur des traits majeurs tels que la qualité du fruit, la résistance aux maladies, et le rendement. D'autres caractères secondaires prennent aujourd'hui de l'importance car ils sont liés à des risques agronomiques et adaptatifs. Ainsi, la régularité de production est gérée par l'éclaircissage manuel ou chimique, mais la volonté de réduire l'utilisation de produits phytosanitaires requiert de nouvelles méthodes de gestion. La sélection de variétés régulières est une solution envisagée. Par ailleurs, il a été montré que le changement climatique entraîne des perturbations de la floraison chez le pommier dans les climats doux qui pourraient s'étendre à l'avenir à d'autres zones géographiques. Ce risque pourrait être limité par la sélection de variétés adaptées. Les objectifs de sélection chez le pommier sont donc multiples et portent sur des caractères de nature différente. Le premier objectif de la thèse a été d'étudier les déterminismes génétiques de la régularité et de la date de floraison dans un dispositif multi-population. Le deuxième objectif a été d'intégrer plusieurs caractères majeurs et secondaires dans un idéotype prédéfini et d'estimer des valeurs génétiques multi-caractères avec des indices de sélection. Le dispositif expérimental était composé de cinq familles en ségrégation dont les parents sont reliés par leur pedigree. Des cartographies de QTL ont été réalisées grâce à un modèle bayésien permettant de prendre en compte les apparentements, le Pedigree Based Analysis (PBA). Ces analyses ont été réalisées avec d'une carte génétique intégrée composée d'environ 7000 SNPs issus d'une puce à 20 000 SNPs. Le modèle Bayésien de détection de QTLs a aussi permis de calculer des valeurs génétiques basées soit sur l'intégralité du génome ou seulement sur les QTLs détectés. Plusieurs caractères majeurs et secondaires ont été intégrés dans un idéotype prédéfini, et leurs corrélations analysées. Des valeurs génétiques multi-caractères ont été estimées avec des indices de sélection sur la base de l'idéotype afin d'identifier les meilleurs individus. Les analyses QTL sur la phénologie de la floraison ont permis l'identification de plusieurs QTLs robustes sur les groupes de liaison (LG) 7, 10, 12, et 9. Plusieurs gènes candidats ont été cartographiés sous ces QTLs : un homologue du gène *FT* sous le QTL du LG12, un homologue d'*AGL24* sous le QTL du LG9 et deux gènes *DAM* sous deux QTLs mineurs sur les LG8 et LG15. Pour la régularité de production, plusieurs indices mesurés à l'échelle de la pousse annuelle ont été élaborés et utilisés comme prédicteurs du comportement global de l'arbre. De plus, les analyses QTL ont révélé plusieurs QTLs déjà identifiés dans des études précédentes sur les LG4, 8 et 10, mais aussi de nouveaux QTL sur les

LGs 5 et 7. Les corrélations entre traits ont permis de révéler des synergies, entre régularité de production et taille du fruit, et des antagonismes, entre rendement et taille/régularité de production, par apport à l'idéotype fixé. De plus, les valeurs d'index de sélection ont permis de révéler un potentiel différent pour les deux familles utilisées. Finalement, l'estimation de valeurs génétiques multi-caractères des individus sur la base des marqueurs moléculaires vise à ouvrir la perspective d'une stratégie de sélection génomique simultanée sur plusieurs caractères.

Summary

In apple tree, major traits considered in breeding are fruit quality, disease resistance, yield and its regularity. Other secondary characters are becoming important because they are linked to agronomical and adaptive risks. Yield regularity is managed by chemical and hand thinning, but the will to reduce the use of chemical products requires new yield management methods. Breeding regular varieties is considered as one of the solutions. Besides, it has been shown that climate changing is responsible for problematic behavior during flowering in warm climates and that it could be extended to cooler regions. This risk could be limited by breeding of adapted varieties. Breeding goals in apple are numerous and characters of different nature are considered. The first objective of this work was to study the genetic determinism of fruit bearing habit and flowering time in a multi-parental population. The second objective was to integrate several characters, major and secondary, in a pre-defined ideotype and to estimate multi-character breeding values with selection indexes. The experimental design was composed by five segregation families related by their pedigree. QTLs mapping studies were realized with a Bayesian method that takes into account relationships, this method was called Pedigree Based Analysis (PBA). These analyses were realized with an integrated genetic map composed of 7000 SNPs scored from a 20 000 SNP array. The Bayesian method also allowed the computation of breeding values, either based on QTLs detected or on genome wide markers' effect. Several major and secondary characters were integrated in a predefined ideotype, and their correlations were studied. Multi-character genetic values were estimated with selection indexes in order to identify the best individuals. QTL analyses on budbreak and flowering time revealed several strong QTLs on linkage groups (LG) 7, 10, 12, and 9. Several candidates genes were mapped under several QTL: a FT homolog was mapped under QTL on LG12, an AGL24 homolog was mapped under LG9 QTL and two DAM-genes were mapped under two minors QTLs on LG8 and LG15. Concerning fruit bearing habit, several annual shoot scale indexes were elaborated and used as prediction of global tree habit. In addition, QTL mapping analyses revealed already known QTLs on LG4, 8, and 10, and new QTL on LG 5 and 7. Correlations between traits revealed synergies and antagonisms between traits, according to the ideotype. In addition, selection index revealed different genetic potential for the two families used. Finally, the estimation of multi-character breeding values based on molecular markers give perspectives for multi-character genomic selection combining primary and secondary characters.

Remerciements

Le doctorat a été pour moi une grande aventure qui a duré un peu plus de trois ans et pendant laquelle j'ai énormément appris, tant d'un point de vue scientifique que d'un point de vue humain. Pour arriver au bout de ce projet j'ai été entourée de nombreuses personnes qui m'ont permis de trouver la ressource pour avancer.

D'abord, l'encadrement de qualité d'Evelyne Costes m'a permis de bâtir une relation de confiance sur laquelle j'ai pu largement m'appuyer pendant toute la durée de la thèse. Ses critiques rigoureuses et bienveillantes m'ont permis d'approfondir ma réflexion sur le projet, de me remettre en question, d'identifier mes points faibles, et de les surmonter. C'est cette exigence qui m'a permis d'atteindre le niveau requis pour un doctorat.

Les discussions avec François Laurens m'ont amenée à replacer mon travail de recherche dans le cadre plus large de la sélection variétale chez le pommier. Cet exercice fut assez difficile mais très intéressant car nécessite d'envisager des perspectives concrètes d'application de mon travail de recherche.

Dans le cadre d'un séjour à l'université de Wageningen, j'ai eu la chance de travailler sous la supervision d'Eric van de Weg et Marco Bink. Je souhaite les remercier tous les deux pour m'avoir apporté de solides bases théoriques sur les grands principes de la génétique quantitative, d'avoir eu la patience de répondre à mes nombreuses questions et de m'avoir apporté leur aide et leur soutien lorsque c'était nécessaire.

Je souhaite aussi remercier les membres de mon comité de thèse : Mathilde Causse, Charles-Eric Durel, Hélène Muranty, Vincent Segura, et Marie Denis. Leurs conseils avisés qui ont jalonné mon parcours m'ont été précieux pour remplir mes objectifs et identifier les points de progrès. Je souhaite remercier particulièrement Mathilde Causse pour son écoute et pour m'avoir donné l'occasion de présenter mon travail aux membres de son unité de recherche, à l'INRA d'Avignon.

Le contexte de travail de l'UMR AGAP m'a permis d'être en contact avec de nombreux chercheurs, Sébastien Tisne, David Cros, Timothée Flutre, Ingrid Vilms, avec lesquels j'ai pu

avoir des discussions intéressantes et constructives. Parmi les collègues de l'UMR, je souhaite particulièrement remercier Baptiste Guitton qui m'a suivie tout au long de la thèse, dont l'amitié m'a été d'un grand réconfort dans les moments difficiles et qui m'a donné le coup de pouce final et décisif lors de la rédaction de ce manuscrit.

Au quotidien, j'ai eu la grande chance d'être entourée de mes collègues de l'équipe AFEF. Parmi eux, je remercie Sébastien Martinez de m'avoir fait une place dans son bureau et de m'avoir révélé tous les secrets du phénotypage des pommiers. Je remercie Sylvie Bluy, Véronique Prudhon et Isabelle Farrera pour leur présence bienveillante et leur écoute, Jean-Michel Legave pour nos discussions sur le changement climatique vu par les agronomes, Bouchaïb Khadari et Jean-Luc Regnard pour nos discussions scientifiques et musicales, Jean-Jacques Kelner pour son aide à la préparation de la soutenance, ainsi que tous les post-doctorants, doctorants et stagiaires qui m'ont apporté leur amitié et leur grain de folie : Nicolas, Inès, Hicham, Annaelle, Gerardo, David, Muriel, Vincent, Céline, Matteo, Lucille, Lucie, Loriane, Paul, ...

La thèse n'a pas seulement été une expérience professionnelle, mais aussi une expérience personnelle qui a nécessité l'affection de tous mes proches. En particulier, mes parents, mes frères et sœur qui m'ont suivi de très près et dont le soutien sans faille m'a permis d'atteindre mes objectifs. L'amitié aussi a été une source de réconfort abondante. Stéphanie et Christophe m'en ont apporté la plus belle preuve en m'accueillant à bras ouverts au moment le plus critique. Les Marions et Hélène ont répondu à mes appels à l'aide avec beaucoup de douceur, d'intelligence et de compréhension. Antonin a eu la patience de relire une partie de la thèse, de corriger mes nombreuses fautes d'orthographe et de syntaxe sans porter de jugement, et sa présence attentionnée m'a apporté force et motivation. Enfin, je remercie mes amis qui ont tous joué un rôle dans cette aventure qu'a été la thèse : Sandra, Aude, Jérôme, Rémi, Hugo, Mathilde, Harold, Julie, Leslie, Alice, Mario...

Table des matières

INTRODUCTION GENERALE	1
CONTEXTE BIOLOGIQUE ET MÉTHODOLOGIQUE	8
1. Le pommier	9
1.1. Origine, domestication et lieu de culture du pommier	9
1.2. La culture du pommier	10
1.3. La sélection chez le pommier	11
2. Les traits phénotypiques étudiés : Alternance de floraison et dates de floraison	14
2.1. Le cycle phénologique du pommier et son développement architectural	14
2.2. Risques adaptatifs et agronomiques liés à l'induction florale et à la dormance	20
3. Etude des déterminismes génétiques de caractères complexes chez le pommier	24
3.1. Stabilité et transférabilité des QTLs	24
3.2. Elargissement de la diversité génétique	25
3.3. Estimation des héritabilités et des effets génétiques	28
3.4. Apparentements	30
3.5. Les statistiques Bayésiennes au secours de la détection de QTLs	32
4. Intégration de plusieurs caractères	33
4.1. Analyses conjointes pour des traits corrélés	33
4.2. Idéotypes et indices de sélection	35
RESULTS.....	46
CHAPTER 1: New insights into genetic determinisms of budbreak and flowering time in apple through a multi-family QTL discovery and identification of putative candidate genes linked to temperature requirements.....	46
Introduction	49
Materials and Methods	52
Results	58
Discussion	61
Tables and Figures	75
CHAPTER 2: Genetic determinisms of flowering regularity in apple tree: a multi-family QTL detection based on statistical indices extracted from branch sequence analyses.....	88
Introduction	91

Materials and methods.....	93
Results..	100
Discussion.....	106
Tables and Figures.....	121
CHAPTER 3: Identification of segregating QTLs in two families for ‘pack-out’ traits	136
Introduction	137
Material and methods	140
Results..	142
Discussion.....	144
CHAPTER 4: Integration of phenotypic traits into a pre-defined ideotype: A study case using genome wide breeding value and selection index estimation on apple tree	151
Introduction	152
Materials and methods.....	156
Results..	161
Discussion.....	164
DISCUSSION GENERALE	173
Annexes	190

Figures contexte biologique et méthodologique

Figure 1 : Origine et domestication du pommier.	9
Figure 2 : Chromosomes homologues chez le pommier.....	9
Figure 3 : Répartition de la production mondiale de pomme.	10
Figure 4 : Production de pommes par pays en 2014.....	10
Figure 5 : Grandes zones de production de pommes en France.....	10
Figure 6 : Carte des régions climatiques sur la planète	11
Figure 7 : Schéma de sélection pommier.....	13
Figure 8 : Cycle phologique du pommier.....	14
Figure 9 : Bourse et pousse de bourse chez un jeune pommier.	14
Figure 10 : Differentiation morphologique des méristèmes apicaux.	15
Figure 11 : Méristème floral terminal et méristème latéral.	15
Figure 12 : Inflorescence sur méristème terminal des pommiers, poiriers et cognassiers	15
Figure 13 : Idéotypes architecturaux.	16
Figure 14 : Gènes impliqués dans l'Induction florale chez <i>Arabidopsis thaliana</i>	17
Figure 15 : Augmentation de température prévue pour 2081-2100.....	22
Figure 16 : Evolution des dates de floraison avant et après le réchauffement climatique.....	23
Figure 17 : Population MAGIC.	27
Figure 18 : Population NAM	27
Figure 19 : Structure de pedigree pour une analyse de type PBA du projet FruitBreedomics.	28

Introduction Générale

Introduction générale

La pomme est l'un des fruits qui compose le régime alimentaire de l'homme depuis des milliers d'années. Plusieurs siècles avant Jésus Christ, des écrits portaient déjà, en Chine et en Grèce par exemple, sur la description de vergers et de méthodes agronomiques comme le greffage. Comme pour d'autres espèces d'intérêt, la sélection empirique des meilleurs individus est amorcée avec le développement de l'agriculture et concerne principalement des caractères comme le rendement, la qualité gustative du fruit, et la facilité de récolte (Westwood, 1993). Néanmoins, le pommier aurait subi une première phase de sélection avant sa domestication par des mammifères de grande taille, notamment les ours, qui auraient sélectionné des arbres portant les fruits les plus gros et les plus juteux (Harris *et al.*, 2002). La taille du fruit aurait même été fixée lors de cette première étape de sélection, car un goulot d'étranglement dans la diversité génétique pour ce caractère est observé entre le pommier domestique, *Malus x domestica* Borkh, les espèces sauvages les plus proches *Malus siversii*, *Malus orientalis* et *Malus sylvestris*, et les autres espèces de pommier sauvage (Yao *et al.*, 2015). Les hommes ont eux aussi axé la sélection sur des caractères liés au fruit. Aujourd'hui, la qualité du fruit est un caractère primordial en sélection et se décline en de nombreux paramètres relatifs à l'apparence et la texture de la chair, son goût mais aussi la capacité du fruit à maintenir son niveau de qualité lors de sa conservation après récolte. La sélection s'intéresse aussi au rendement, en quantité et régularité, ainsi qu'aux résistances aux nombreuses maladies et ravageurs auxquels le pommier est sensible. Ce dernier caractère ne sera pas abordé dans cette thèse.

Cependant, le contexte de changement climatique d'une part et la volonté de réduction des intrants d'autre part conduisent à envisager d'élargir le panel des caractères pris en compte dans les schémas de sélection (Laurens *et al.*, 2000). L'élévation progressive des températures conduit à porter plus d'attention aux mécanismes de dormance des arbres et aux variations de date de floraison en fonction des zones de culture. Par ailleurs, les arboriculteurs font face à une rude compétition économique qui les conduit à réduire le temps de travail au verger principalement destiné à la taille ou l'éclaircissage des fruits. De plus, la réglementation européenne réduit progressivement le nombre de produits chimiques homologués pour réaliser l'éclaircissage. La maîtrise de la charge en fruits devient donc particulièrement délicate et de nouvelles solutions doivent être explorées, agronomiques mais aussi génétiques. L'étude du déterminisme génétique de la régularité de production, en absence d'éclaircissage ou avec des éclaircissages légers, pourrait donc constituer une voie innovante dans la sélection de nouvelles variétés de pommier.

Introduction générale

Dans ce contexte, le présent travail s'est intéressé à la prise en compte de nouveaux traits, soit d'intérêt agronomique comme la régularité de production ou de traits adaptatifs, tels que la variation des dates de floraison en réponse aux conditions de températures. Cette thèse s'est donc focalisée sur deux étapes du cycle phénologique du pommier en particulier : l'induction florale et la période de reprise d'activité des bourgeons après l'hiver.

Le cycle phénologique d'une plante représente l'enchaînement des événements morphologiques et physiologiques se déroulant au cours d'un cycle annuel de croissance, en réponse aux conditions environnementales qui changent avec les saisons. Chez les plantes pérennes et en particulier les arbres tels que le pommier, il débute au printemps par le développement des bourgeons floraux puis végétatifs et se prolonge par le déploiement du feuillage. Concomitamment, les fleurs s'ouvrent et, suite à leur pollinisation, se transforment en fruits dont le développement et la maturation se dérouleront jusqu'à la fin de l'été. Chez le pommier, l'induction florale se situe environ 50 jours après la floraison (Foster *et al.*, 2003), donc pendant le développement du fruit. La floraison étant terminale, la transition florale doit être coordonnée avec l'arrêt de croissance des rameaux qui peut intervenir à des périodes variables selon le type de rameau : long, moyen, ou court (Costes *et al.*, 2003, 2006). Après les étapes de développement du fruit et de différenciation florale, les bourgeons entrent en dormance à l'arrivée de l'automne. Pendant l'automne et l'hiver, ils passent par différents états de dormance et leur activité reprend de nouveau au printemps.

Parmi ces étapes du cycle phénologique du pommier, l'induction florale joue un rôle important dans l'établissement du rendement ainsi que la qualité des fruits. En effet, chez le pommier, l'induction de la floraison peut être inhibée par une forte charge en fruits, ce qui génère un faible rendement l'année suivante. Ce comportement d'alternance de production est observé chez de nombreuses espèces fruitières (Monselise and Goldschmidt, 1982). Il résulte d'une dépendance entre induction florale et charge en fruit au cours d'un cycle annuel et génère une dépendance inter-annuelle des niveaux de production, se traduisant par des comportements irréguliers ou alternants.

En ce qui concerne le débourrement, la problématique est d'ordre climatique et consiste à obtenir des dates de floraison qui permettent aux fleurs à la fois d'éviter les gelées printanières et

Introduction générale

d'être pollinisées dans des conditions climatiques favorables au travail des abeilles et à la synchronisation avec les variétés pollinisatrices. Le pommier étant originaire de climat tempéré, il a besoin de températures froides pour entrer en dormance, maintenir la dormance puis sortir de la dormance. Une mauvaise satisfaction des besoins en froid perturbe le déroulement des stades phénologiques et peut avoir un impact négatif sur les dates de floraison et par voie de conséquence sur la production de fruits.

Ces deux caractères, débourrement des bourgeons végétatifs et floraux et alternance de production, ont pour point commun de nécessiter un phénotypage long et/ou lourd. En effet, si l'observation des dates de débourrement est aisée, des séries chronologiques longues sont nécessaires pour déterminer par modélisation les besoins en froid et en chaleur à partir des dates de débourrement (Legave *et al.*, 2009) ou alors des moyens expérimentaux conséquents doivent être mis en place. Je reviendrai plus longuement sur ce point. De même, la caractérisation de la régularité de production ne peut se faire qu'après plusieurs années de production. Pourtant, il a été montré chez le pommier que ces deux caractères ont un déterminisme génétique fort, bien que complexe (Celton *et al.*, 2011; Guitton *et al.*, 2012). Ceci suggère qu'ils pourraient être considérés dans les schémas de sélection et qu'il est envisageable d'utiliser une approche de génétique quantitative pour prédire le potentiel génétique des individus pour la sélection.

L'une des façons d'aborder l'étude de déterminismes génétiques complexes est de les décomposer en traits plus élémentaires puis d'identifier les associations statistiques entre variation phénotypique et classes alléliques pour des loci répartis sur tout le génome, c'est à dire la détection de Quantitative Trait Loci (QTLs) (Lander and Botstein, 1989). Cette méthodologie a été développée depuis plusieurs décennies et a été facilitée par le développement de marqueurs moléculaires et de méthodes statistiques fiables. Des limites ont cependant été identifiées, comme le manque de stabilité des QTLs détectés en fonction du fond génétique considéré, ou le manque de précision de la position de ces QTLs, lorsque le nombre de recombinants disponibles est faible (Khan and Korban, 2012). L'élargissement du nombre d'individus recombinés et de la variabilité allélique est une solution pour pallier à ces problèmes. Chez le pommier, il existe un grand nombre de variétés cultivées au sein des pays qui pratiquent cette culture, et certaines des sauvages apparentées sont connues. Cette diversité peut être utilisée pour répondre aux problèmes précédemment cités. De plus, les pedigrees des variétés sont eux aussi connus et inventoriés pour

Introduction générale

la plupart et rendent possible la prise en compte des apparentements de manière précise. Le pommier possède des caractéristiques particulières qui peuvent rendre difficile l'étude des déterminismes génétiques : une forte polymorphie, l'impossibilité de développer des lignées homozygotes en raison de son caractère pérenne et de son auto-incompatibilité et l'homologie de certains de ces chromosomes due à des événements d'allopolyploïdisation récents (Velasco *et al.*, 2010).

Dans ce contexte, j'ai étudié dans cette thèse le déterminisme génétique de deux caractères, la date de débourrement et l'alternance de production, et ai effectué une cartographie de QTLs sur une population composée de cinq descendances apparentées entre elles. Ensuite, sur un dispositif réduit à deux descendances, j'ai réalisé une détection de QTL sur plusieurs traits liés à la qualité du fruit et au rendement. Sur le même dispositif réduit, les déterminismes génétiques des traits abordés dans cette étude ont été comparés, les valeurs génétiques pour chaque caractère ont été estimées, leur corrélation phénotypique et génétique analysée. Enfin, j'ai initié une réflexion sur une stratégie possible pour combiner ces caractères entre eux. Cette stratégie est basée sur des indices de sélection construits à partir d'un idéotype préalablement défini afin de déterminer le potentiel génétique multi-trait des individus de la population considérée.

Au cours de la thèse, j'ai aussi abordé une série de questions scientifiques, liées à la stratégie de phénotypage et aux traits étudiés :

Les dates de débourrement et de floraison sont-elles représentatives de la variabilité des besoins en froid et en chaleur ?

Comment ces dates sont-elles impactées par le site géographique et l'année climatique ?

Quels mécanismes candidats et gènes, potentiellement impliqués dans la perception du froid ou des températures élevées en période de post-dormance, sont présents dans les zones QTLs majeures ?

Comment mesurer et caractériser précisément le niveau de régularité d'un arbre ?

Le phénotypage partiel des arbres, c'est-à-dire par échantillonnage de branches ou de rameaux, permet-il de représenter fidèlement leur comportement à l'échelle de l'arbre entier ?

Introduction générale

Quelle est la relation entre synchronisation de la floraison à l'échelle de la branche et le comportement régulier, alternant ou irrégulier des génotypes étudiés ?

Quelle est l'influence de l'année et du site, ou encore du niveau d'éclaircissage réalisé sur la régularité des génotypes ?

Au cours de l'analyse génétique réalisée sur une population composée de cinq familles en ségrégation, j'ai abordé des questions liées à la modélisation des phénotypes, à l'estimation des valeurs génétiques des parents et des géniteurs utilisés dans ce dispositif mais aussi à la valeur génétique multi-caractère des individus :

Quels modèles et structure de covariance utiliser pour modéliser des phénotypes collectés sur plusieurs années et plusieurs sites géographiques ?

Comment l'architecture génétique des caractères influence t'elle l'estimation de leurs valeurs génétiques ?

Comment estimer un potentiel génétique multi-caractère ?

Présentation du document de thèse

Ce document est composé de six parties. La première est une synthèse bibliographique du contexte biologique et méthodologique dans lequel s'ancre ce projet. Les quatre parties qui suivent sont des chapitres de résultats, et se présentent sous la forme d'articles scientifiques rédigés en anglais. La première est une étude du déterminisme génétique des dates de débourrement et floraison par cartographie de QTLs, et inclue une identification de gènes candidats. Cet article a été accepté avec révisions dans la revue *Journal of Experimental Botany*. La deuxième partie présente une stratégie de simplification du phénotypage de l'alternance de production par l'élaboration de nouveaux indices calculés à l'échelle des branches, évalue leur capacité à représenter le comportement des arbres entiers, suivie d'une cartographie de QTLs sur ces indices. Cette partie sera soumise à publication prochainement. La troisième partie porte sur plusieurs caractères de qualité du fruit et son objectif est d'identifier les QTLs qui ségrégent dans la population utilisée. Par souci d'homogénéité, cette partie est aussi présentée sous forme d'article scientifique, mais elle ne sera pas publiée car elle n'apporte pas de résultats nouveaux.

Introduction générale

Elle permet de compléter le panel de caractères étudiés dans ce travail, afin d'aborder dans la partie suivante la question de la combinaison de traits en sélection, à l'exception de la résistance aux maladies qui n'a pas été abordée dans ce travail. La quatrième partie porte sur l'intégration des caractères étudiés précédemment. Dans cette dernière partie, un idéotype est défini et une stratégie d'estimation du potentiel génétique multi-caractère est mise en place. Enfin, le document se termine sur une dernière partie de discussion générale et de conclusion.

Ma contribution et les collaborations dont cette thèse est issue

Au cours de ce travail, j'ai interagi avec de nombreuses personnes et équipes. J'ai réalisé en propre :

- une partie du phénotypage, et du recueil des données, avec l'aide des personnels techniques des deux équipes AFEF (AGAP) et FruitQual (IRHS)
- l'élaboration de la stratégie de travail, avec mes encadrants et les membres du Comité de Suivi de Thèse
- le traitement statistique des données et leur modélisation, avec mes encadrants et l'aide de plusieurs personnes d'AGAP et l'Univ. de Grenoble
- la détection de QTLs, en étroite collaboration des collègues de WUR (Wageningen University & Research)
- l'interprétation des résultats et la rédaction des articles avec mes encadrants et co-auteurs, pour les chapitres portant sur le déterminisme génétique du débourrement et de la floraison, la qualité du fruit et l'intégration des caractères. Pour la partie sur la régularité de production, j'ai entièrement réalisé la cartographie de QTLs, j'ai rédigé les parties de Matériels & Méthodes et Résultats correspondants à cette analyse et j'ai participé aux discussions, à la relecture et l'amélioration de cet article

Tableau 1 : Récapitulative des familles, de leurs caractéristiques ainsi que des phénotypes mesurés.

Famille	Parents		Site	Nombre d'individus (phénotypés et génotypés)	Répétitions	Conduite	Caractères phénotypés
	♀	♂					
SG	cv ^a 'Starkrimson' (mutant de 'Delicious')	cv ^a 'Granny Smith'	Montpellier	115	2	Un peu taillé Un peu éclairci Irrigué et traité	Phénologie Alternance Qualité du fruit Rendement
XB	X-3263 ^b	cv ^a 'Belrène'		58	1	Pas taillé Pas éclairci Irrigué et traité	
HIVW	X-3263 ^b	X-3259 ^b	Angers	172	1	Taillé, Eclairci Irrigué Traité	Phénologie Alternance
N	X-3305 ^b	X-3259 ^b		42	1		
P	cv ^a 'Rubinette'	X-3305 ^b		45	1		

^a *Cultivars Commerciaux*

^b **Hybride sélectionné à Angers pour sa qualité de fruit**

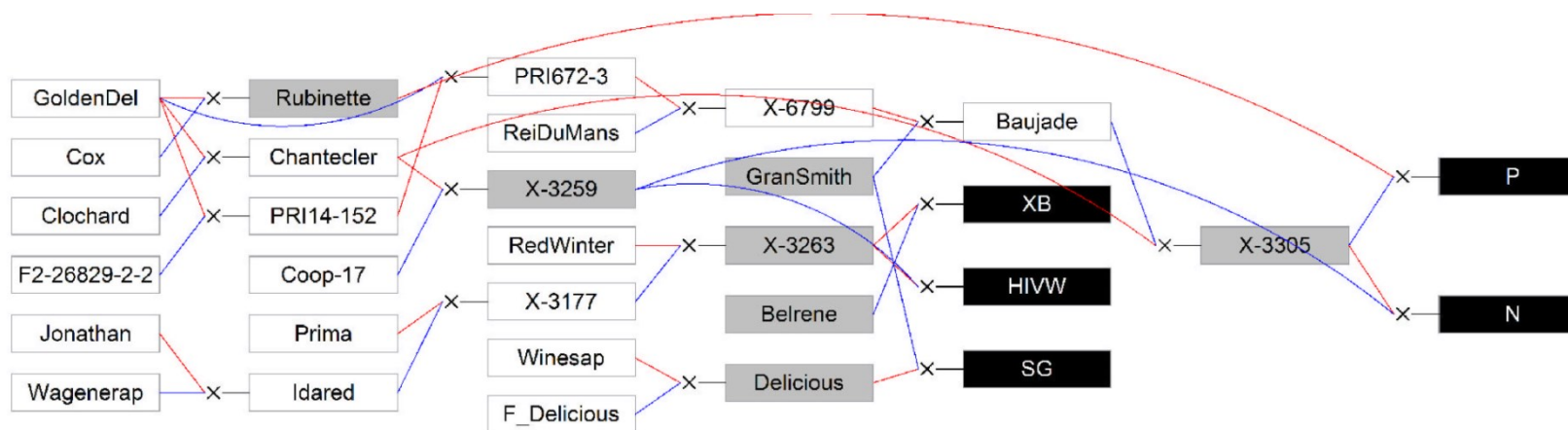


Figure 1 : Illustration des relations d'apparentement entre les familles utilisées dans ce travail. Les familles sont représentées par des encadrés noirs, les parents par des encadrés gris et les autres membres du pedigree par des encadrés blancs.

Tableau 2 : Récapitulatif des caractères mesurés, de leur modélisation ainsi que des variables utilisées pour la détection de QTLs.

Caractère	Phénotypage	Traits	Modélisation	Variable utilisée pour la détection de QTLs
Phénologie	Date de débourrement en jours calendaires	BBcd : date de débourrement en jours calendaires	Modèles mixtes $P=S+Y+G+G*Y+\varepsilon$ P : trait Effet fixes : S : effet Site Y : effet Année (structure de corrélation) Effets aléatoires : G : effet Génotype G*E : effet d'interaction entre Génotype et Année ε : résiduelle du modèle (cf Allard et al, 2016, en révision)	Best Linear Unbiased Predictors (BLUPs) pour les effets Génotype et interaction entre Génotype et Année, extraits des modèles mixtes.
	Date de début de floraison, en jours calendaires et en Growing Degree Hour (GDH)	Delta : temps écoulé entre débourrement et début de floraison, exprimé en GDH		
Alternance	Phénotypage rétrospectif des pousses annuelles par identification de la présence d'événement de floraison pour végétatifs.		Modélisation du nombre d'inflorescence par modèle mixte (Durand et al, à soumettre)	Calcul d'un indice d'alternance (BBI) Extraction du paramètre d'autocorrélation du nombre d'inflorescence en fonction des années (γ)
				Calcul d'indices d'entropie (Ent_g et $Ent_{glmm,g}$)
			Modélisation de la probabilité d'un événement par modèle mixte de floraison (Durand et al, à soumettre)	Extraction des BLUPs d'interaction Année*Génotypes ($\eta_{g,t}$) Mémoire*Génotype ($\theta_{g,m}$)
Qualité	Mesure de la fermeté par pénétrométrie	Fermeté	Moyenne sur l'échantillon prélevé	Phénotypes bruts
	Mesure du taux de sucre par réfractométrie	Taux de Sucre		
	Titrage du taux d'acide	Acidité		
	Poids moyen des fruits	Taille du Fruit		
Rendement	Nombre total de fruits Poids total de fruits			Phénotypes bruts

Contexte biologique et méthodologique

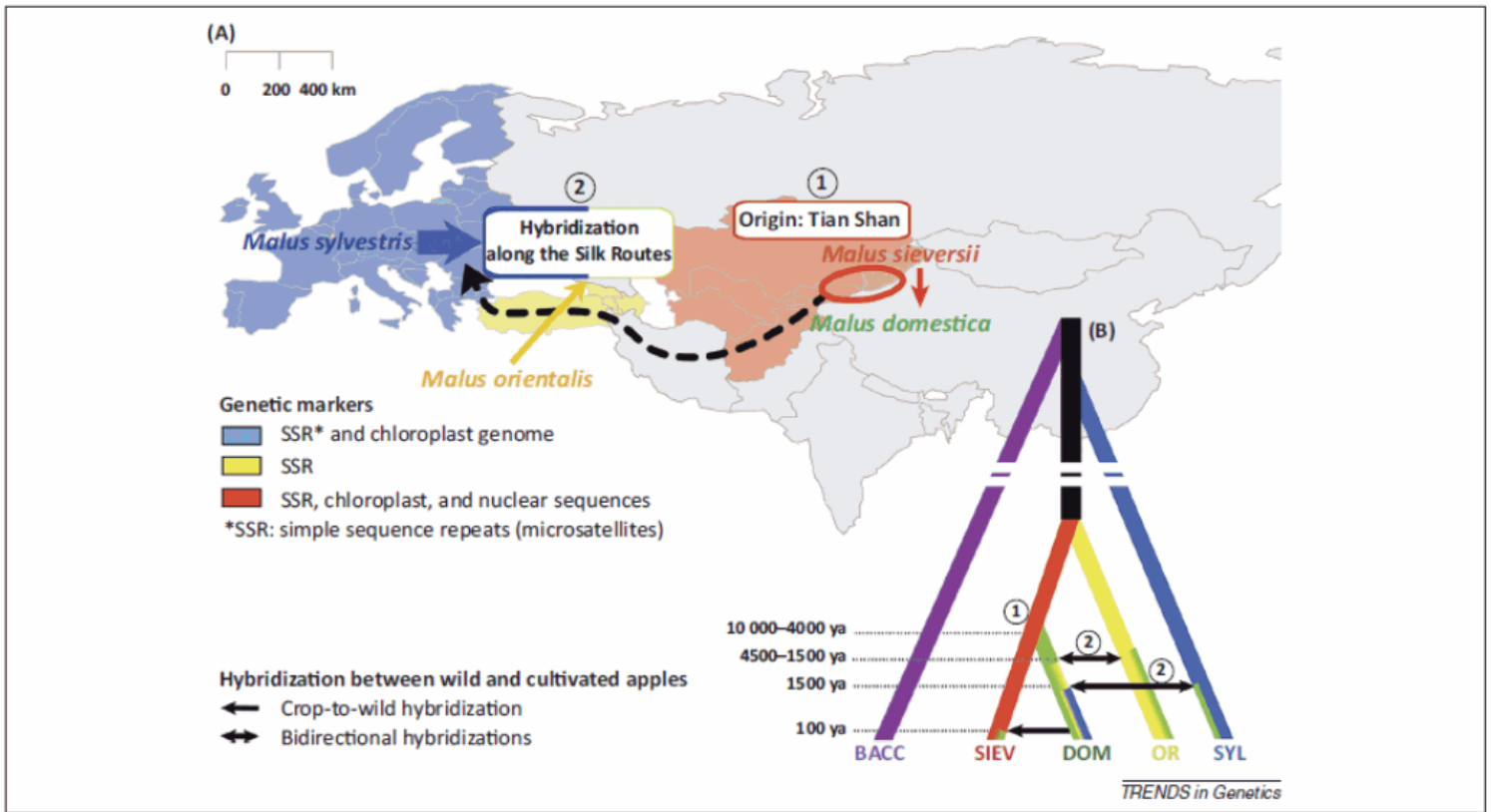


Figure 1 : Origine et domestication du pommier (Cornille et al., 2014).

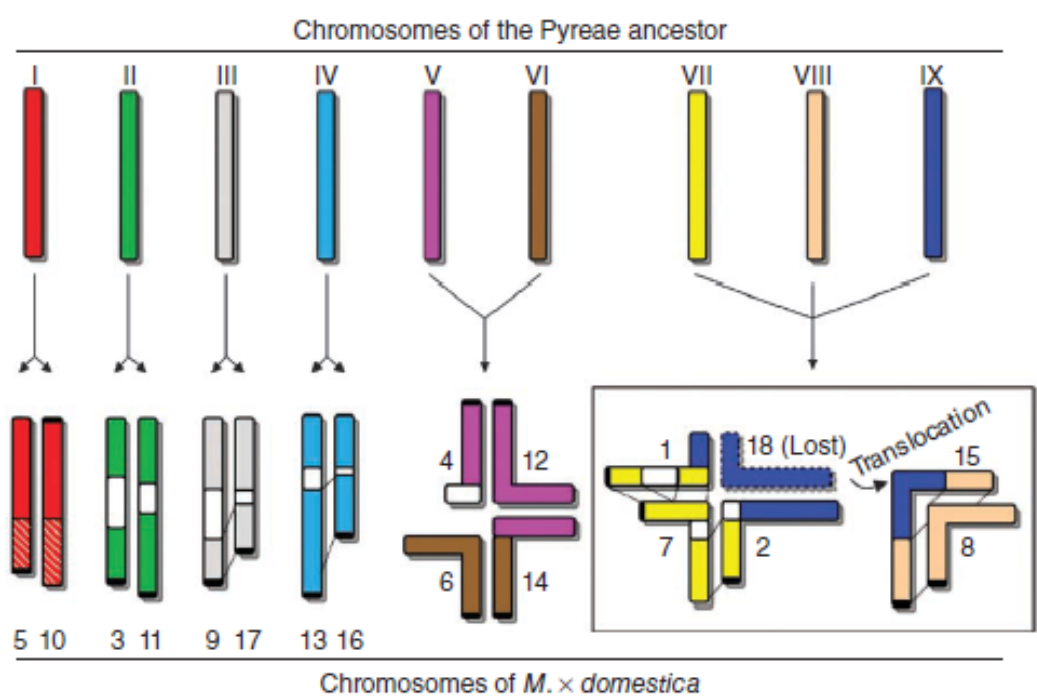


Figure 2 : Figure illustrant l’évolution d’un génome ancestral à 9 chromosomes vers le génome actuel du pommier à 17 chromosomes. Sont aussi représentés, par un jeu de couleur, les chromosome et régions homologues au sein du génome du pommier. Figure issue de Velasco *et al.*, (2010).

1. Le pommier

1.1. Origine, domestication et lieu de culture du pommier

L'ancêtre du pommier aurait été identifié en Asie centrale, au Kazakhstan (Juniper, 2007), par le botaniste et généticien russe Nikolaï Vavilov, dans les années 30. Cette région serait à la fois le lieu d'origine de l'espèce et son centre de diversité (Harris *et al.*, 2002). Le long des routes de commerce entre la Chine et l'Europe de l'est, on retrouve des traces de l'espèce datant du néolithique. Ensuite, il semblerait que ce soit les romains qui aient introduit l'espèce ainsi que les méthodes de bouturage et greffage en Europe Occidentale (Harris *et al.*, 2002) (Figure 1). Les croisements entre l'espèce en voie de domestication et les espèces de pommier sauvage ont été probablement fréquents du fait de l'auto-incompatibilité de la fleur du pommier (Cornille *et al.*, 2014). L'ancêtre du pommier serait l'espèce *Malus sieversii*, mais lors de sa domestication *Malus x domestica* Borkh aurait subi des introgressions de l'espèce sauvage européenne *Malus sylvestris* (Cornille *et al.*, 2014).

Le pommier fait partie de la famille des Rosaceae qui contient les principales espèces fruitières et ornementales des régions tempérées ; cerisier (*Prunus avium* ou *Prunus cerasus*), pêcher (*Prunus persica*), abricotier (*Prunus armeniaca*), amandier (*Prunus dulcis*), prunier (*Prunus domestica* ou *Prunus silicina*) pour le genre *Prunus*, poirier (*Pyrus communis*), pommier (*Malus x domestica*), fraisier (*Fragaria X annanassa*), et rosier (*Rosa* spp) (Dirlewanger *et al.*, 2004). Le génome du pommier a récemment été séquencé (Velasco *et al.*, 2010), ce qui a permis d'émettre des hypothèses sur l'origine de l'espèce. Les espèces de la tribu des *Pyreae*, à laquelle pommiers et poiriers appartiennent ont un génome à 17 chromosomes, alors que les autres espèces de la famille des *Rosaceae* en comptent entre 7 et 9. Ceci s'expliquerait par un événement de duplication du génome relativement récent dans l'histoire de la tribu des *Pyreae*. Cette tribu descendrait d'un ancêtre qui comportait 9 chromosomes, la duplication de son génome suivie de la perte d'un chromosome par fusion avec un autre, aurait donné naissance à un génome à 17 chromosomes. Ainsi, chez le pommier, on retrouve des paires de chromosomes homologues ainsi que des groupes de chromosomes partiellement homologues. Par exemple, les chromosomes 5 et 10 sont homologues et les chromosomes 4, 6, 12, et 14 comportent de larges zones homologues (Figure 2, Velasco *et al.*, 2010). De plus, un haut niveau de syntenie est observé entre le génome des genres *Prunus* et *Malus* (Dirlewanger *et al.*, 2004; Vilanova *et al.*, 2008;

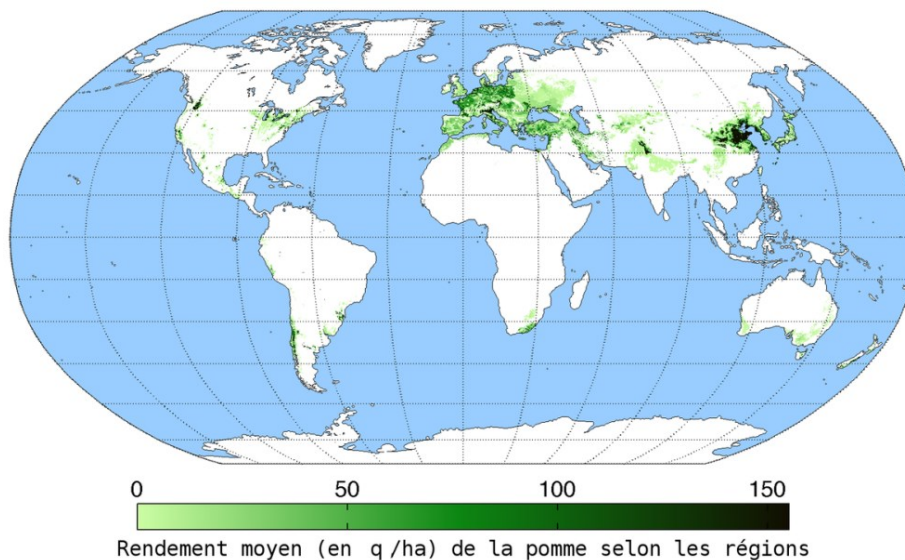


Figure 3 : Répartition de la production mondiale de pomme (Monfreda *et al.*, 2008).

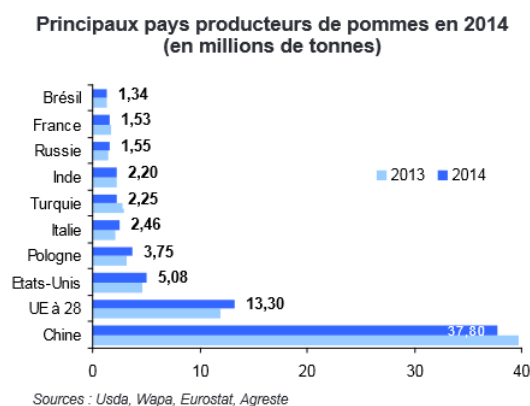


Figure 4 : Production de pommes par pays en 2014 (www.franceagrimer.fr).



Figure 5 : Grandes zones de production de pommes en France, Sud-Est (en rouge), Sud-Ouest (en violet) et Val-de-Loire (en vert) (Sauphanor *et al.*, 2009).

Sargent *et al.*, 2009; Bushakra *et al.*, 2012), ce qui peut permettre d'extrapoler la fonction de certains gènes connus entre ces deux genres. Parmi les plantes dont le génome a été séquencé, *Malus x domestica Borkh* est celle pour laquelle le plus grand nombre de gènes aurait été prédit, 57 386 au total. Ce génome serait aussi caractérisé par un grand nombre de séquences répétées et un haut niveau de polymorphisme (Velasco *et al.*, 2010).

La pomme a été le troisième fruit le plus produit dans le monde en 2013, avec 80 millions de tonnes produites, après la pastèque (108 millions de tonnes produites en 2013), la banane (105 millions de tonnes produites en 2013), et avant le raisin (77 millions de tonnes produites en 2013) (http://faostat3.fao.org/browse/rankings/commodities_by_regions/E, 2016). La production de pomme est principalement localisée dans l'hémisphère nord (Figure 3), une partie est produite dans l'hémisphère sud, au Chili, Brésil, Afrique du Sud, Australie et en Nouvelle-Zélande (Figure 3). La Chine est le plus gros producteur mondial de pommes avec 37.8 millions de tonnes produites en 2014 (Figure 4). La production est située dans la partie Nord-Est du pays, là où le climat y est propice. L'Europe est la deuxième région productrice de pomme (Figure 4) avec 13,3 millions de tonnes produite en 2014 dont 3.75, 2.46, 1.53 millions de tonnes produites en Pologne, Italie et France, respectivement (Figure 4). La production française est répartie sur 3 zones : Bassin Rhône Méditerranée, bassin Val de Loire, et bassin Grand Sud-Ouest (Figure 5). Les Etats-Unis sont aussi un producteur important de pommes avec 5.08 millions de tonnes produites en 2014 (Figure 3).

1.2. La culture du pommier

Le pommier est une espèce originaire du Kazakhstan, une région à climat tempéré continental (Harris *et al.*, 2002; Juniper, 2007). Elle est adaptée au cycle saisonnier de cette région, qui est caractérisé par de grandes amplitudes de température entre été et hiver, une variation de la photopériode ainsi que des précipitations variables selon les saisons (Westwood, 1993). Malgré cela, le pommier est une espèce cultivée dans diverses régions climatiques. On trouve des vergers de pommiers dans des régions tempérées : Chine, Europe, Etats-Unis par exemple, mais aussi dans d'autres types de climat comme au sud du Brésil où les pommiers poussent en altitude où la majorité du territoire a un climat tropical ou en Afrique du sud dont le territoire est divisé entre

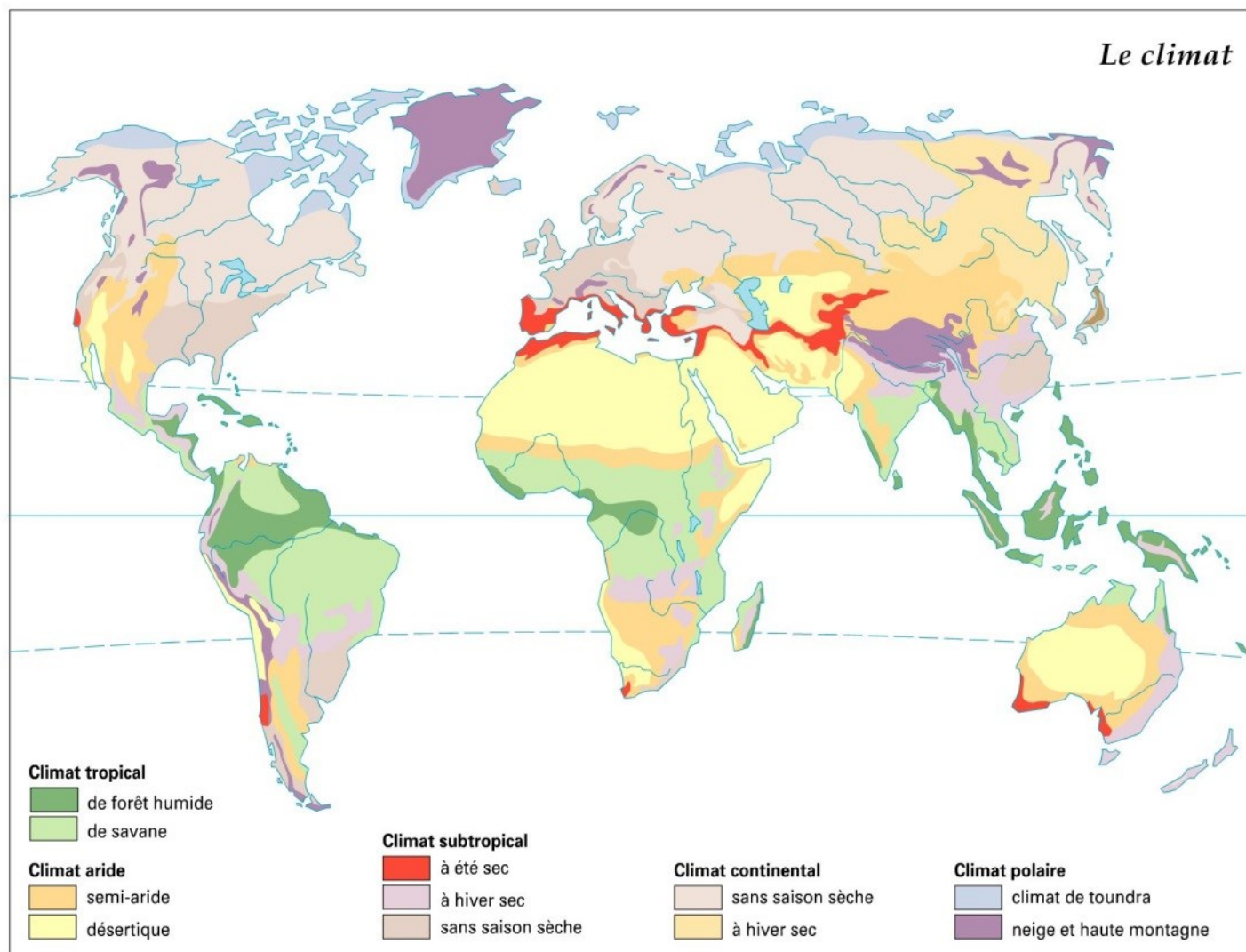


Figure 6 : Carte des régions climatiques sur la planète (www.larousse.fr).

climat méditerranéen, tropical et désertique (Figure 6). Dans ces climats doux, la culture du pommier est compliquée par le manque de températures froides mais est possible en altitude.

Plusieurs freins caractérisent la sélection et la culture du pommier. Le stade juvénile du pommier, pendant lequel l'arbre ne produit pas de fruits, dure plusieurs années. Sa fin est marquée par la première floraison. Le choix du porte greffe peut permettre de réduire la phase juvénile (Seleznova *et al.*, 2008). Différentes méthodes de taille sont utilisées pour optimiser le rapport entre développement végétatif et fructification et pour faciliter la récolte (Lauri and Lespinasse, 2000; Lauri, 2002; Trillot *et al.*, 2002). Ensuite, la fleur du pommier est auto-incompatible. Il est donc nécessaire d'implanter au sein des vergers des individus pollinisateurs dont le pollen est compatible avec la variété cultivée. La pollinisation se fait principalement par les insectes pollinisateurs en particulier par les abeilles (Trillot *et al.*, 2002). Enfin, la plupart des variétés de pommier sont très sensibles à de nombreux agents pathogènes d'origine fongique comme la tavelure, causée par le champignon *Venturia inaequalis*, ou d'origine bactérienne, comme le feu bactérien causé par la bactérie *Erwinia amylovora*. Il peut être aussi sujet à des attaques d'insectes ravageurs comme le carpocapse (*Cydia pomonella* (L.)), et différents types de pucerons. Les maladies comme les attaques d'insectes ravageurs peuvent avoir un impact considérable sur le rendement en fruit. Pour ces raisons, les vergers de pommier font l'objet d'un suivi sanitaire important et les traitements par fongicides et insecticides sont fréquents (Sauphanor *et al.*, 2009 ; Butault *et al.*, 2010).

1.3. La sélection chez le pommier

La création variétale chez le pommier a pour objectif de développer des variétés répondant aux exigences des consommateurs mais également de tous les acteurs de la filière : les producteurs, les coopératives, les metteurs en marché, les distributeurs. La sélection prend donc en compte aujourd'hui un très grand nombre de critères qui peuvent être résumés en trois piliers : productivité, qualité du fruit et résistance aux bioagresseurs. Ce dernier point est particulièrement important car la plupart des variétés aujourd'hui commercialisées sont toutes sensibles voire très sensibles à la plupart des parasites importants du verger ; la conséquence directe de cet état des lieux est que l'Indice de Fréquence de Traitements (IFT) de la pomme de table est de loin le plus élevé parmi toutes les cultures (IFT pomme de table= 36.5 ; viticulture = 12.5 d'après Agreste,

2006). La qualité organoleptique du fruit est l'un des axes principaux d'amélioration dans la plupart des schémas de sélection, il comporte un très grand nombre de caractères qui concernent l'attrait du fruit (calibre, coloration, défauts d'épiderme...), la qualité gustative (qualité de la texture, goût, rapport sucre/acide, parfum, arôme...). La pomme étant un fruit qui se conserve longtemps, ces caractéristiques doivent se maintenir le plus longtemps possible en cours de stockage. Pour la qualité du fruit, la demande de la distribution vise des produits standardisés avec un gros calibre, une forte coloration, sans défaut d'épiderme et de bonnes qualités gustatives (Costes *et al.*, 2009). Pour les raisons évoquées précédemment, la résistance aux maladies en verger est un axe d'amélioration très important, la première maladie sur laquelle une sélection est effectuée est la tavelure (*Venturia inequalis*), mais les individus sont aussi testés sur leur résistance à l'oïdium (*Podosphaera leucotricha*). La résistance aux maladies de conservation est importante car est en partie garante d'une bonne capacité de conservation des fruits. Le troisième caractère très important est la productivité et sa régularité qui peuvent être sélectionnés en se basant sur des caractères liés à l'architecture de l'arbre (port de l'arbre, type de fructification). Les caractères adaptatifs tels que la résistance au stress hydrique ou les besoins en froid et chaleur sont étudiés dans les équipes de recherche mais ils ne sont pas, pour le moment faute de critères faciles à phénotyper, intégrés dans les schémas de sélection. Les attentes de la part des acteurs de la filière étant nombreuses, la définition d'idéotypes chez le pommier doit être placée dans son environnement agronomique, climatique, économique, et commercial (Costes *et al.*, 2009). Les producteurs souhaitent des variétés faciles à cultiver et rentables, les distributeurs veulent des variétés avec de bonnes capacités de conservation et les consommateurs préfèrent les variétés 'reconnaissables' et bonnes.

La première étape de la sélection est le choix des géniteurs, des croisements puis la production deshybrides. La variabilité génétique des géniteurs est en général assez restreinte. Les variétés 'Golden Delicious', 'Red Delicious', 'Jonathan', 'MacIntosh' et 'Cox's Orange Pippin' sont dans les pedigrees de 75% des variétés ((Noiton and Shelbourne, 1992; Noiton and Alspach, 1996). Face à ce constat, certaines équipes de sélection, notamment celle de l'INRA d'Angers développent une stratégie d'élargissement de la base génétique qui fait intervenir dans les croisements soit des espèces sauvages soit des variétés anciennes qui présentent des caractéristiques intéressantes, notamment pour leur rusticité et leur résistance aux bioagresseurs. Le problème avec ce type de démarche est que l'introggression de caractères d'intérêt est souvent

Programme INRA-Novadi (1997-)

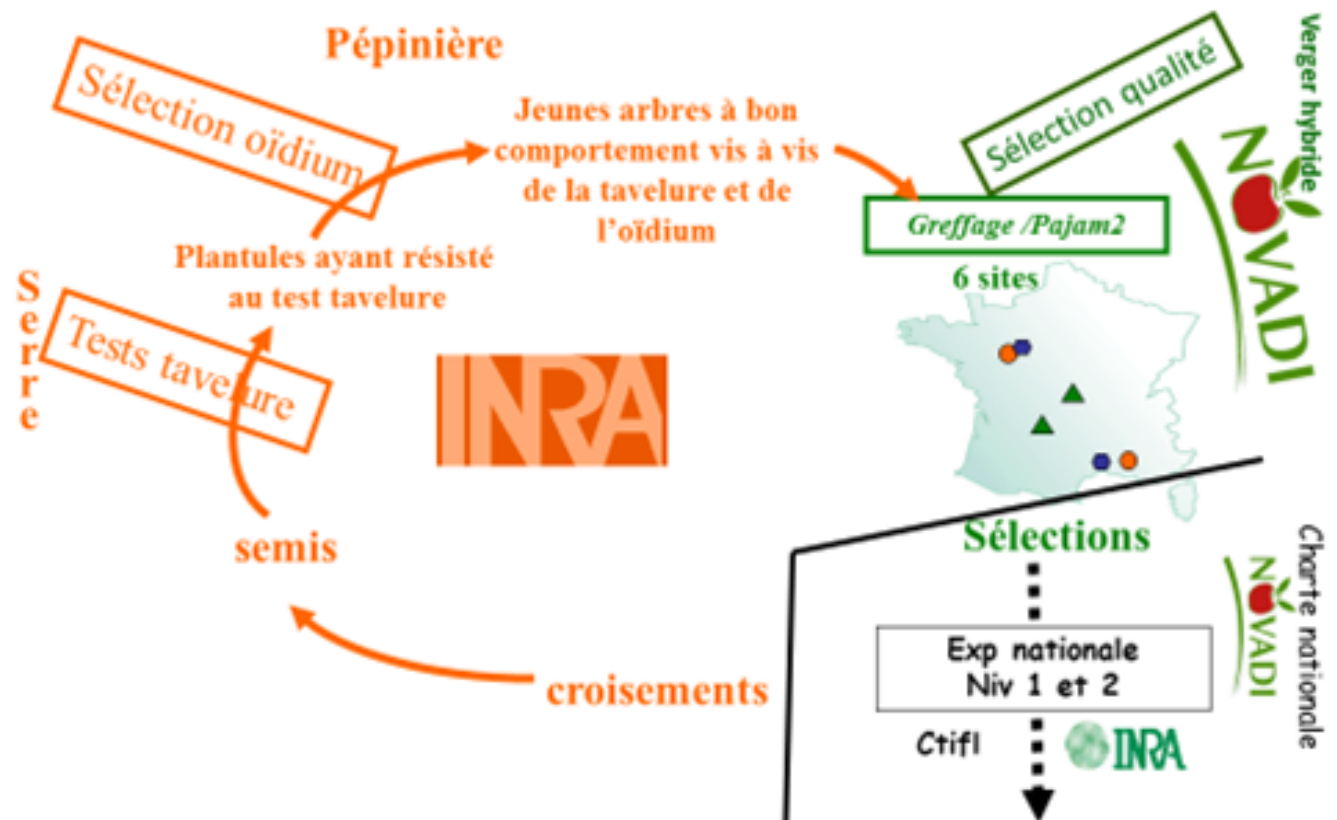


Figure 7 : Schéma de sélection pommier, NOVADI.

associés à l'introggression de caractères rédhibitoires, et plusieurs cycles de sélection sont nécessaires pour les éliminer (Costes *et al.*, 2009).

Dans le cas du programme de sélection mené à Angers, les jeunes semis issus des croisements du programme d'amélioration sont testés pour leur résistance à la tavelure ; les plus sensibles sont éliminés. Les plus résistants sont plantés en pépinière et testés pour leur résistance à l'oïdium. A ce stade et dans certains cas, les semis peuvent être également sélectionnés sur des critères architecturaux. Après ces premières phases, les hybrides sélectionnés sont greffés et implantés chez les pépiniéristes associés à la SARL NOVADI qui travaillent en partenariat avec l'INRA. Ils sont implantés sur cinq sites dans les trois grands bassins de production de la pomme : deux en Pays de la Loire, un dans le Sud Ouest et deux dans le Sud Est. Une fois, leur phase juvénile levée, c'est-à-dire entre 3 et 5 ans après le semis, les hybrides sont sélectionnés par étape : une première qui élimine les individus possédant des caractères de productivité, d'attrait et de taille de fruits rédhibitoires ; les hybrides sélectionnés sont dégustés et notés sommairement pour des caractéristiques organoleptiques de base à la récolte en verger (attrait, calibre, fermeté, goût, ouverture du tube calinical) mais également des critères de sensibilité aux bioagresseurs (oïdium, pucerons, ...). Les meilleurs sont ensuite conservés en chambre froide et évalués deux à trois fois au cours de la conservation sur une gamme plus large et plus précise de critères organoleptiques. Les quelques hybrides ayant passé ces différentes étapes de sélection sont ensuite implantés sur des essais multi-sites chez les pépiniéristes partenaires mais également au Ctifl (Centre Technique Interprofessionnel des Fruits et Légumes) dans le cadre d'une charte nationale d'expérimentation. La sélection est très drastique : il faut en moyenne semer 100000-150000 pépins pour sélectionner une variété commerciale qui cumule l'ensemble des caractères favorables minimum pour être commercialisée. Ce taux de sélection très fort s'explique par le très grand nombre de caractères à prendre en compte (plus de 30 uniquement pour la qualité, à évaluer à plusieurs reprises pendant la conservation), souvent complexes et très subjectifs, donc difficiles à évaluer (au moins pour la qualité). Il faut également que les arbres sélectionnés puissent avoir un bon comportement quelques soient les conditions climatiques : la durée d'un verger varie en moyenne de 10 à 30 ans (Figure 7 pour le schéma de sélection SARL NOVADI).

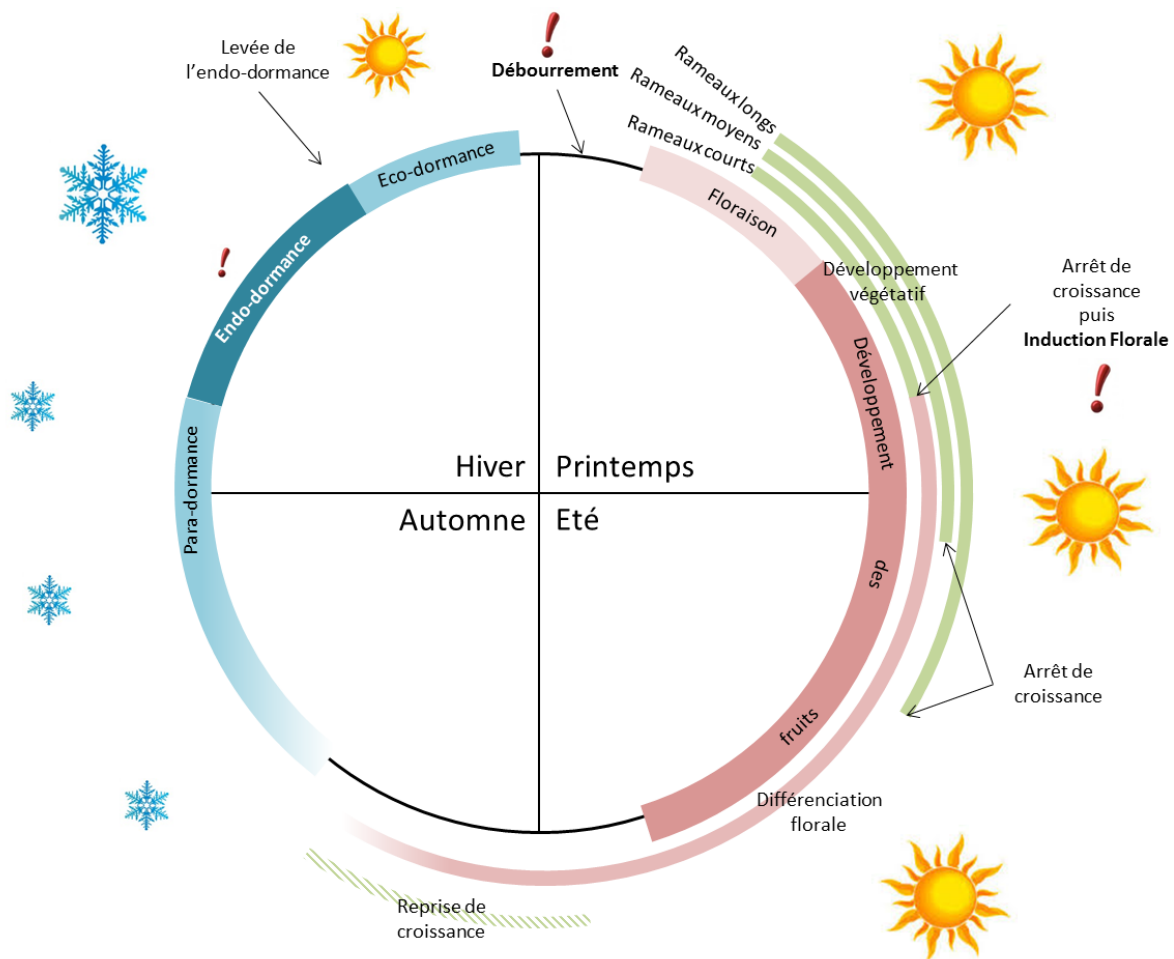


Figure 8 : Cycle phénologique du pommier (d'après Foster *et al.*, 2003, Neilsen et Dennis, 2000 ; Westwood, 1993 ; Cooke *et al.*, 2012 et Lang *et al.*, 1987).



Figure 9 : Bourse et pousse de bourse chez un jeune pommier (photo : S. Martinez, verger unité expérimentale INRA Diascope, Mauguio).

2. Les traits phénotypiques étudiés : Alternance de floraison et dates de floraison

2.1. Le cycle phénologique du pommier et son développement architectural

Le cycle phénologique représente l'enchaînement des événements biologiques au cours d'une année (Figure 8). Il débute avec le débourrement des bourgeons qui est l'un des premiers marqueurs visible de la reprise d'activité de l'arbre après l'hiver, début Avril dans l'hémisphère Nord. Il est suivi directement par la floraison, c'est-à-dire l'ouverture des inflorescences et des fleurs qui la composent, qui en quelques jours, laisse place au développement des fruits. Le développement végétatif de l'arbre se fait en même temps que la floraison et le début du développement des fruits.

Chez le pommier, comme chez de nombreuses espèces fruitières, il existe une hétéroblastie des axes (voir Costes *et al.*, 2006 pour une revue). Généralement trois types de rameaux sont distingués: courts, moyens et longs qui se caractérisent par le nombre et la nature des organes végétatifs différenciés dans les méristèmes terminaux. Les rameaux courts sont constitués d'organes préformés uniquement, c'est à dire différenciés dans le bourgeon hivernal. Par contre, les méristèmes des rameaux longs ont la capacité de produire, suite au déploiement des organes préformés, des organes néo-formés. Enfin les rameaux moyens ont un état intermédiaire, et contiennent des organes préformés et quelques organes néo-formés. Selon les types de rameaux et l'âge de l'arbre, la croissance végétative s'arrête de quelques semaines, de Juin pour les rameaux courts à Aout ou parfois ultérieurement pour les rameaux les plus longs. Dans certaines conditions, les rameaux longs peuvent reprendre une phase de croissance à la fin de l'été.

L'induction florale se situe généralement 50 jours après floraison pour les rameaux courts (Foster *et al.*, 2003), mais se déroule probablement de manière échelonnée en fonction de la longueur des rameaux (Neilsen et Dennis, 2000). Elle intervient dans les méristèmes terminaux de tous les axes. En cas de nouaison, la présence d'un fruit provoque l'épaississement de l'unité de croissance qui le porte, qui est alors appelé une bourse. Cette spécificité morphologique est observée chez les pommiers ainsi que les poiriers qui ont tous les deux une floraison terminale (Westwood, 1993). Les rameaux qui se développent à partir de la bourse sont appelés les pousses de bourse (Figure 9). A la fin de l'été, les fruits arrivent à maturité. Enfin, l'arrivée de l'automne amène l'entrée en dormance des bourgeons (Cooke *et al.*, 2012). Durant l'hiver, les bourgeons

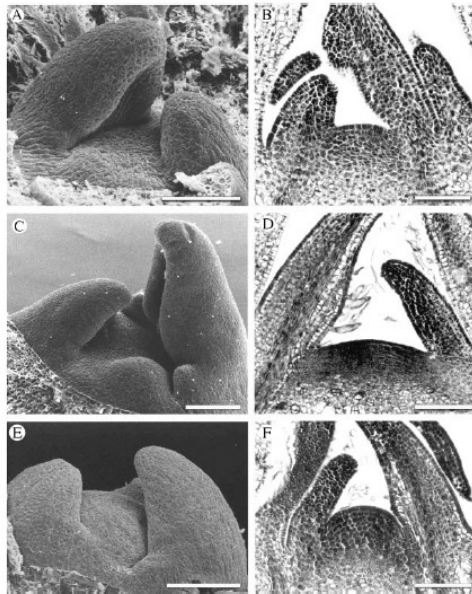


Figure 10 : Différentiation morphologique des méristèmes apicaux chez *Malus x domestica* Borkh. De (A) à (C) Méristèmes végétatifs à différents stades, (E) et (F) méristèmes floraux (Foster *et al.*, 2003).

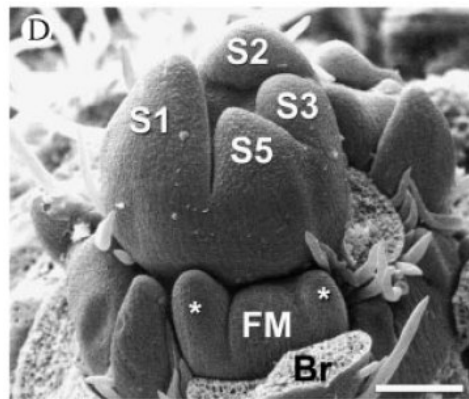


Figure 11 : Méristème floral terminal encadré de sépales (S1-S5), et le méristème latéral (FM) encadré d'une paire de bractées (*) (Foster *et al.*, 2003).

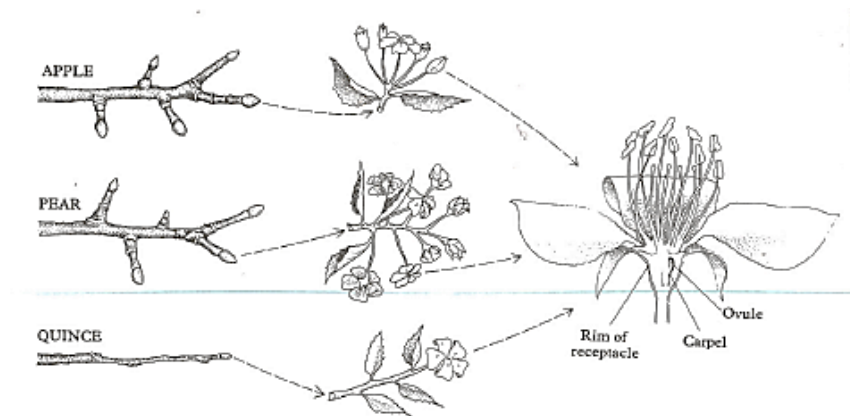


Figure 12 : Inflorescence sur méristème terminal des pommiers, poiriers et cognassiers (Westwood, 1993).

passeront par les différentes étapes de dormance (Lang *et al.*, 1987) avant de reprendre leur développement au printemps.

2.1.1. L'induction florale chez le pommier

L'induction florale est l'étape initiale conduisant à la différenciation morphogénétique des méristèmes terminaux des axes aériens en méristèmes floraux capables de différencier des organes floraux. Les méristèmes ayant subi l'induction florale sont caractérisés par une forme bombée au niveau du centre quiescent, contrairement aux méristèmes végétatifs qui restent plats et étroits (Figure 10). Ces méristèmes induits se différencient en un méristème floral terminal et des méristèmes latéraux qui formeront la future inflorescence (Figure 11) (Foster *et al.*, 2003). Chez les espèces annuelles, la transition florale concerne tous les méristèmes et marque la fin de la vie de la plante. En revanche, la transition florale chez les espèces pérennes est un phénomène quantitatif qui concerne une partie seulement des méristèmes tout en conservant une autre partie des méristèmes à l'état végétatif (Bangerth, 2009).

Comme la floraison est terminale chez le pommier (Figure 12), l'induction florale marque la fin du développement végétatif du méristème terminal. Pour les rameaux courts, l'induction florale se fait environ 3 semaines après floraison (Westwood, 1993; Foster *et al.*, 2003). Malgré une relation étroite entre arrêt de croissance et induction florale, aucune étude histologique n'a été réalisée à ce sujet. L'induction florale est suivie de la différenciation florale qui se déroule en fin d'été, en même temps que la maturation des fruits, et jusqu'à l'entrée en dormance des bourgeons. Dans l'étude conduite sur rameaux courts sur la variété 'Royal Gala', près de quatre mois après la pleine floraison, la plupart des méristèmes induits à floraison ont atteint un stade où les méristèmes de l'inflorescence sont différenciés (Foster *et al.*, 2003). Ce stade correspond à un état avancé mais pas final de la différenciation florale, qui semble être atteint à la fin de l'été seulement.

Chez les espèces pérennes d'origine tempérée, le phénomène d'induction florale est principalement régulé par des mécanismes endogènes, tels que les mécanismes hormonaux, et la compétition entre organes pour les ressources carbonées. Concernant les mécanismes hormonaux, plusieurs hormones seraient impliquées, les acides gibbérélliques, les cytokinines et l'éthylène. L'induction florale serait inhibée par la production d'acides gibbérélliques (GAs) par les graines

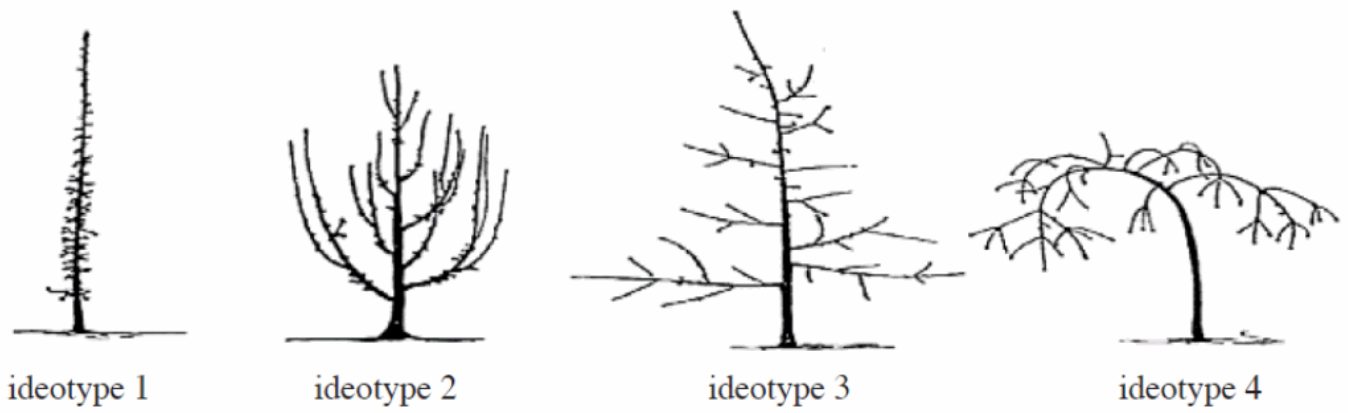


Figure 13 : Idéotypes architecturaux, du port colonnaire (idéotype 1) au port retombant (idéotype 4) (Costes *et al.*, 2006).

des fruits en formation (Bangerth, 2006, 2009). L'effet inhibiteur des graines des fruits a été montré principalement par l'étude de la variété 'Spencer seedless', qui est parthénocarpique, dont les fruits sans graine n'exercent aucune influence sur l'induction florale (Chan and Cain, 1967; Neilsen and Dennis, 2000). Les GAs synthétisés par les graines stimuleraient la production et le transport d'acide indol-acétique (IAA) qui cheminerait jusqu'aux méristèmes terminaux encore indifférenciés et inhiberaient leur transition florale. Au contraire, les cytokinines (CK), principalement d'origine racinaire, auraient une action stimulante sur l'induction florale. Il semblerait que la stimulation de l'induction florale soit possible à une concentration précise en CK car à faible concentration, le méristème est proche d'un état dormant et à forte concentration, c'est le développement végétatif qui est favorisé. L'éthylène aurait aussi un effet stimulateur de l'induction florale par inhibition du transport de l'IAA (Bangerth, 2006).

La compétition entre organes, quant à elle, a été associée aux caractéristiques architecturales de l'arbre. Chez le pommier, quatre types architecturaux ont été identifiés, notés de I à IV (Costes *et al.*, 2006) (Figure 13). Le type 1 est le type le plus extrême, caractérisé par des entre-nœuds très courts et de nombreux rameaux courts axillaires, on l'appelle le type colonnaire. Le type 2 a un port très érigé. Les branches charpentières se développent à partir de la partie basse du tronc et la floraison a lieu sur de rameaux courts qui se développent en grand nombre et directement sur les branches charpentières. Ces deux types sont caractérisés par une séparation spatiale et temporelle du développement végétatif et reproductif, et par une faible capacité d'extinction des méristèmes (i.e. arrêt de l'activité des méristèmes ; Lauri *et al.*, 1995 et 1997). Ces types présentent aussi, en général, une production alternante. Les types 3 et type 4 sont caractérisés respectivement par un port plus ouvert et retombant : les branches charpentières se développent dans la partie supérieure du tronc et les fruits se forment sur les rameaux longs (Costes *et al.*, 2006). Ce type est connu pour sa régularité de production et un équilibre entre développement végétatif et développement reproductif (Lauri *et al.*, 1995). Ces comportements seraient liés à une capacité intrinsèque à gérer l'équilibre entre extinction (i.e. mortalité) de certains méristèmes et réveil d'autres méristèmes latents (Lauri *et al.*, 1997). A l'échelle de l'arbre, ce phénomène est positivement corrélé avec la capacité à retourner à floraison deux années successives (Lauri *et al.*, 1995). A l'échelle du rameau, cette capacité serait liée à la longueur et au nombre de feuilles du rameau (Neilsen and Dennis, 2000) ainsi qu'à la taille de la bourse (Lauri and Laurens, 2005). La longueur du rameau contribuerait donc à réguler l'induction florale, en augmentant la distance

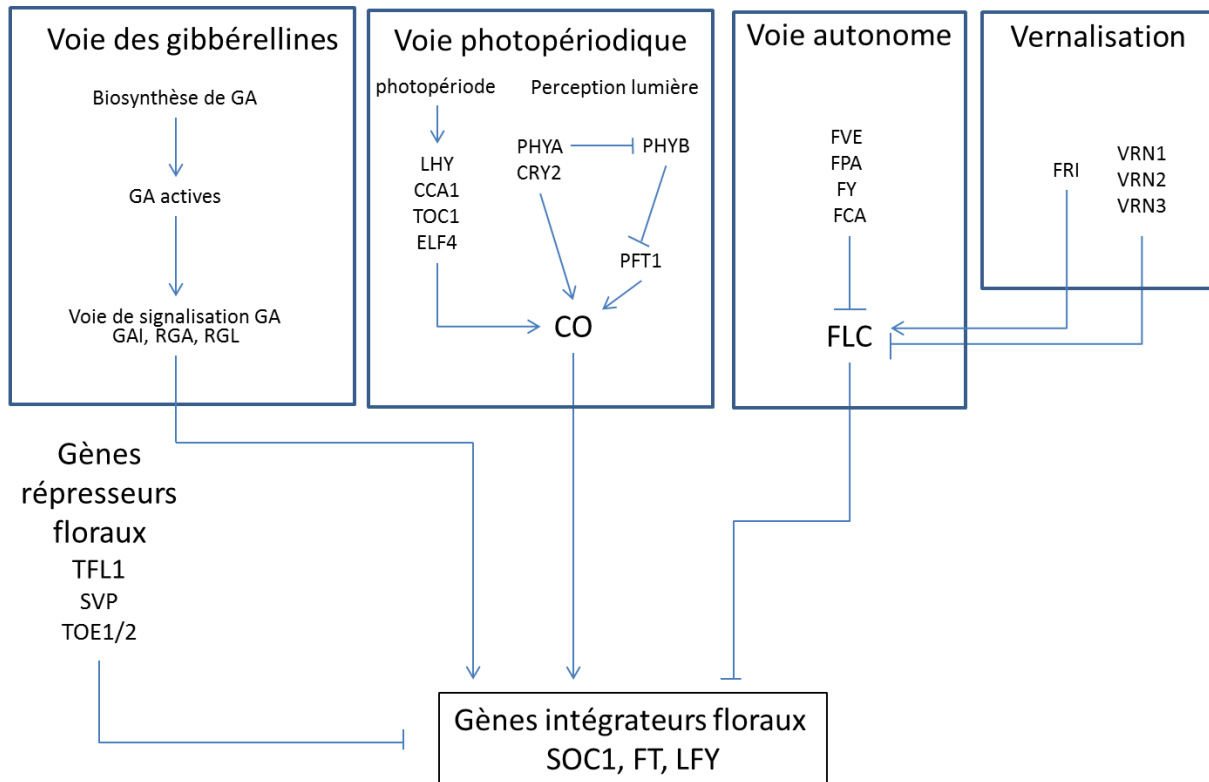


Figure 14 : Schéma simplifié des gènes et voies impliqués dans l'induction florale chez *Arabidopsis thaliana*. L'inhibition est représentée par des flèches en T et l'activation par des flèches. Schéma inspiré de (Mouradov *et al.*, 2002; Sung *et al.*, 2003; Komeda, 2004; Boss *et al.*, 2004).

entre le lieu d'inhibition (les fruits) du lieu d'induction (le méristème terminal de la pousse de bourse). L'effet d'inhibition serait donc particulièrement marqué pour les rameaux très courts et moins pour les rameaux très longs (Nielsen and Dennis, 2000). La plus grande surface foliaire des rameaux longs permettrait une synthèse plus importante de composés carbonés que pour des rameaux courts. En effet, les composés carbonés jouent un rôle important dans le contrôle du développement des plantes (Rolland *et al.*, 2006). Il a été montré chez la vigne (*Vitis vinifera*) qu'ils ont un rôle particulier dans l'induction florale et le développement des pièces florales (Lebon *et al.*, 2008). Ainsi une plus grande quantité de composés carbonés chez les rameaux longs pourrait limiter la compétition entre organes végétatifs et fructifères pour les composés carbonés et ainsi favoriser l'induction florale (Hanke *et al.*, 2007) .

Ainsi, que cela soit dû aux régulations hormonales et/ou à la compétition pour les carbohydrates, une forte charge en fruits se traduit par une forte inhibition de l'induction florale dans les méristèmes terminaux des pousses de bourse. Par conséquent, une faible floraison (et donc charge en fruit) sera observée l'année suivante. Ce phénomène est appelé alternance de production et est commun chez de nombreux arbres fruitiers (Monselise and Goldschmidt, 1982).

Au niveau moléculaire, la régulation de la floraison a été étudiée principalement chez la plante modèle *Arabidopsis thaliana*. Chez cette espèce, la floraison est régulée par plusieurs voies en réponse à différents signaux endogènes et environnementaux : la voie autonome, la voie des gibbérellines, la vernalisation et la voie régulée par la photopériode (Figure 14). Ces voies convergent vers les gènes floraux d'intégration qui eux-même agissent sur les gènes responsables de l'identité florale des méristèmes (Mouradov *et al.*, 2002). Le gène *FLOWERING LOCUS T* (*FT*) est l'un de ces intégrateurs floraux ; il est inhibé par le gène *FLOWERING LOCUS C*, qui réprime aussi les autres intégrateurs floraux tels que *SUPPRESSOR OF OVEREXPRESSION OF CO1* (*SOC1*) et *LEAFY* (*LFY*). Cette répression peut-être levée par la vernalisation mais aussi par la voie autonome (Boss *et al.*, 2004). En plus, de ces deux voies de signalisation, autonomes et vernalisation, il existe plusieurs gènes répresseurs de floraison dont le gène *TERMINAL FLOWER 1* (*TFL1*), homologue de *FT*, qui agit sur *LFY* (Sung *et al.*, 2003). Les gènes intégrateurs floraux, dont *FT*, sont activés par la photopériode et notamment le gène *CONSTANS* (*CO*). L'expression du gène *CO* est régulée par la photopériode et le rythme circadien de la plante

(Mouradov *et al.*, 2002). La voie des gibbérellines est aussi une voie d'activation des intégrateurs floraux *SOC1* et *LFY*, chez *Arabidopsis thaliana*.

Par homologie de séquence, certains gènes identifiés chez la plante modèle ont été étudiés chez le pommier et d'autres *Rosaceae* et semblent aussi jouer un rôle dans l'induction florale. *MdTFL1-1* et *MdTFL1-2*, les gènes *TFL1-like* chez le pommier, seraient impliqués dans la répression de la transition florale chez les *Rosaceae* (Remay *et al.*, 2009; Randoux *et al.*, 2014). En effet, il a été montré que ces gènes jouent un rôle dans le maintien de la croissance végétative et qu'une baisse de leur expression marquerait l'induction florale. Leur rôle serait donc similaire à celui de *TFL1* (Kotoda and Wada, 2005; Flachowsky *et al.*, 2012; Mimida *et al.*, 2013). Ils s'expriment dans le méristème de l'apex des rameaux pendant la période végétative, de Mai à Juin puis son expression décline rapidement lors de l'induction florale (Mimida *et al.*, 2013). Deux gènes homologues de *FT* ont aussi été identifiés chez le peuplier (*Populus spp*) : *PtFT1* qui joue un rôle dans la transition florale (Böhlenius *et al.*, 2006) et *PtFT2* qui favorise le développement végétatif et inhibe la formation de bourgeons en réponse aux températures printanières et estivales (Hsu *et al.*, 2011). Les gènes *MdFT1* et *MdFT2*, homologues de *FT* ont aussi été identifiés chez le pommier. *MdFT1* serait impliqué dans la cessation de croissance, l'induction florale, mais aussi la dormance des bourgeons (Kotoda *et al.*, 2010; Tränkner *et al.*, 2010; Mimida *et al.*, 2013). *MdFT2* jouerait potentiellement un rôle dans la transition florale (Kotoda *et al.*, 2010) mais aurait aussi un rôle dans la division cellulaire et la différenciation des organes reproductifs (Mimida *et al.*, 2013). Les deux gènes s'expriment tous les deux dans l'apex des rameaux et présentent un pic d'expression fin Juin (Mimida *et al.*, 2013).

2.1.2. Débourrement, dormance, et besoins en froid

Pour survivre dans les conditions défavorables de l'hiver, le pommier, comme la plupart des espèces pérennes adaptées aux régions de climat tempéré, a mis en place des mécanismes de protection des méristèmes, dont la dormance des bourgeons. La dormance est composée de trois états différents : (i) l'endodormance régulée par des facteurs physiologiques propres au bourgeon, par exemple la perception des températures froides ; (ii) la paradormance régulée par des facteurs physiologiques externes au bourgeon, par exemple la dominance apicale d'autres bourgeons ; et (iii) l'écodormance régulée par des facteurs environnementaux comme le déficit en éléments

nutritifs par exemple (Lang *et al.*, 1987). Dans les paragraphes suivants, le terme de dormance sera utilisé pour faire référence à l'endodormance.

La dormance se caractérise par un isolement des cellules du méristème du reste de la plante par développement de callose qui bloque les connections symplasmic, empêchant ainsi tout apport extérieur à la cellule (Schoot *et al.*, 2013). Ceci peut expliquer l'activité cellulaire minimale des méristèmes qui sont alors incapables de réaliser des divisions ou de l'expansion cellulaire, facteurs majeurs de la croissance des organes (Cooke *et al.*, 2012). Chez les *Malus*, *Pyrus*, *Prunus* et *Sorbus*, la dormance est régulée par les températures froides qui permettent l'entrée, le maintien et la sortie de dormance (Lang *et al.*, 1987; Heide and Prestrud, 2005; Cooke *et al.*, 2012; Schoot *et al.*, 2013). Dans les zones tempérées de moyenne latitude, dont les températures hivernales peuvent osciller entre températures froides, voire négatives, et températures douces, les espèces se sont adaptées à ces conditions par la mise en place d'un besoin minimal en froid qui assure le maintien de la dormance même lorsque les températures sont momentanément clémentes (Westwood, 1993). Le besoin en froid est déterminé par le temps d'exposition au froid qui assure 100% de débourrement des bourgeons au printemps. Ce temps n'est pas absolu et est d'autant plus grand que la température est élevée pendant la phase d'accumulation de froid (Cooke *et al.*, 2012). Ce temps peut être exprimé de manière absolue par les unités de froid qui représentent une quantité de froid ou un temps d'exposition à une température donnée (Anderson and Richardson, 1986).

Chez les *Rosaceae* des gènes spécifiques à la dormance ont été identifiés, d'abord chez *Prunus persica*, puis *Malus domestica* x *Borkh* et *Pyrus pyrifolia* Nakai. Ces gènes ont aussi été identifiés chez les *Populus tremuloides* (peuplier faux-tremble). On les appelle les DORMANCY ASSOCIATED MADS BOX (gènes DAM). Les gènes DAM régulent l'entrée et la sortie de dormance, la perception de froid et chaleur et ferait parti d'un système d'inhibition du développement du bourgeon (Jiménez *et al.*, 2010; Saito *et al.*, 2013; Falavigna *et al.*, 2013).

Comme la levée de dormance n'est pas un stade visible et que sa détermination aurait nécessité des moyens expérimentaux lourds et destructifs, non compatibles avec le phénotypage de plusieurs caractères sur les mêmes arbres, le choix a été fait, en amont de cette thèse, de porter

l'attention sur le premier stade visible de la reprise de croissance des bourgeons au printemps : le débourrement des bourgeons floraux et végétatifs.

Les deux stades phénologiques précédemment décrits, induction florale et dormance des bourgeons, sont d'une importante capitale à la fois pour l'établissement du rendement en fruit et la survie de l'arbre pendant hiver. Dans ce qui suit, je vais expliquer brièvement comment certaines conditions environnementales ou agronomiques peuvent entraîner des risques pour la production de fruits, comment ces risques sont gérés au plan agronomique et comment ils peuvent éventuellement être quantifiés.

2.2. Risques adaptatifs et agronomiques liés à l'induction florale et à la dormance

2.2.1. Alternance de production

L'alternance de production est le phénomène courant chez les espèces pérennes par lequel un arbre produit une grande quantité de fruits de petite taille une année et seulement quelques fruits de grande taille l'année suivante (Monselise and Goldschmidt, 1982). Suivant l'espèce considérée, l'alternance de production peut être liée à l'induction florale, à la pollinisation, ou au développement des fruits. Chez le pommier, une forte charge en fruit une année a un impact sur l'induction florale par inhibition et ainsi cause un faible rendement l'année suivante (Monselise and Goldschmidt, 1982). Les variétés de pommier diffèrent fortement par leur capacité naturelle à la régularité ou l'irrégularité (Lespinasse, 1977 ; Lauri *et al.*, 1995 et 1997). Par exemple, 'Granny Smith' est réputée très régulière alors que 'Fuji' est plus sensible au phénomène d'alternance de production (www.orangeippin.com/apples/). La génétique de l'alternance de production a été assez peu étudiée chez le pommier. Néanmoins, le déterminisme génétique de ce caractère a fait l'objet de deux études sur une population en ségrégation, Starkrimson x Granny Smith. Elles ont permis d'une part d'identifier plusieurs régions chromosomiques, d'autre part de révéler que les principaux gènes de floraison n'étaient pas situés dans les intervalles de confiance des QTLs contrairement à plusieurs gènes de la voie de biosynthèse de gibbérellines (Guitton *et al.*, 2012; Durand *et al.*, 2013).

Au plan agronomique, l'alternance de production est un problème majeur pour les producteurs, car la qualité des fruits et le rendement sont inférieurs chez un arbre alternant que chez un arbre régulier (Link, 2000). La surcharge en fruit peut causer des dommages physiques

sur l'arbre comme la rupture de branches et les arbres alternants sont aussi plus sensibles aux maladies. Pour atténuer ce phénomène naturel, les producteurs déploient des efforts importants notamment par l'éclaircissage, manuel ou chimique des jeunes fruits de façon à régulariser la charge en fruits des arbres. Chez la variété 'Fuji', l'application de produits d'éclaircissage, Carbaryl, NAA, Ethephon, ainsi que la pratique de l'éclaircissage manuel se sont révélés favorables à une production régulière de fruit sur la taille des arbres (Jones *et al.*, 1992; Waldner and Knoll, 1998). L'éclaircissage par produits chimiques est délicat car une dose trop forte peut être phyto-toxique et le stade d'application est crucial pour garantir des résultats optimums (Dennis, 2000). Or, le produit 'Carbaryl' n'est plus autorisé dans la plupart des pays européens à cause de son impact négatif sur l'environnement. La conduite des arbres peut être aussi un moyen de réguler la charge en fruit des arbres. Il a été montré qu'une taille sévère une année de forte production régule la charge en fruit pour les années suivantes (Jonkers, 1979). De façon plus raisonnée, l'extinction artificielle est une méthode efficace pour la régulation de la charge en fruit qui est issue de l'observation des variétés régulières chez lesquelles l'extinction naturelle des rameaux est positivement corrélée au retour à fleur (Lauri *et al.*, 1997; Lauri and Lespinasse, 2000; Lauri, 2002). Cette méthode est aujourd'hui reconnue et appliquée internationalement (Breen *et al.*, 2015).

L'alternance de production est un caractère complexe à phénotyper car il n'est observable que lorsque l'arbre a atteint son stade adulte et à partir de plusieurs années de production. Le Biennial Bearing Index (BBI) a été introduit en 1936 par Hoblyn *et al.* (1936) sur pommier pour mesurer l'intensité de déviation du rendement sur plusieurs années successives :

$$BBI = \frac{\sum_{i=3}^n (|y_i - y_{i-1}|) / (y_i + y_{i-1})}{n - 1}$$

Avec n le nombre d'années d'observations et y_i le rendement de l'année i . Cet indice a été renommé Biennial Bearing Index par Wilcox *et al.* (1944) et a ensuite été largement utilisé sur plusieurs espèces, dont le pommier (Guitton *et al.*, 2012), le caféier (Cilas *et al.*, 2011) et le pistachier (Rosenstock *et al.*, 2010). D'autres indices ont été proposés pour caractériser la régularité de production, notamment par l'observation de couples d'années successives (Lauri *et al.*, 1995). Plus récemment, des index dérivés du BBI ont été développés, tout d'abord en prenant en compte l'écart du rendement annuel par rapport à un rendement moyen, et aussi en prenant en

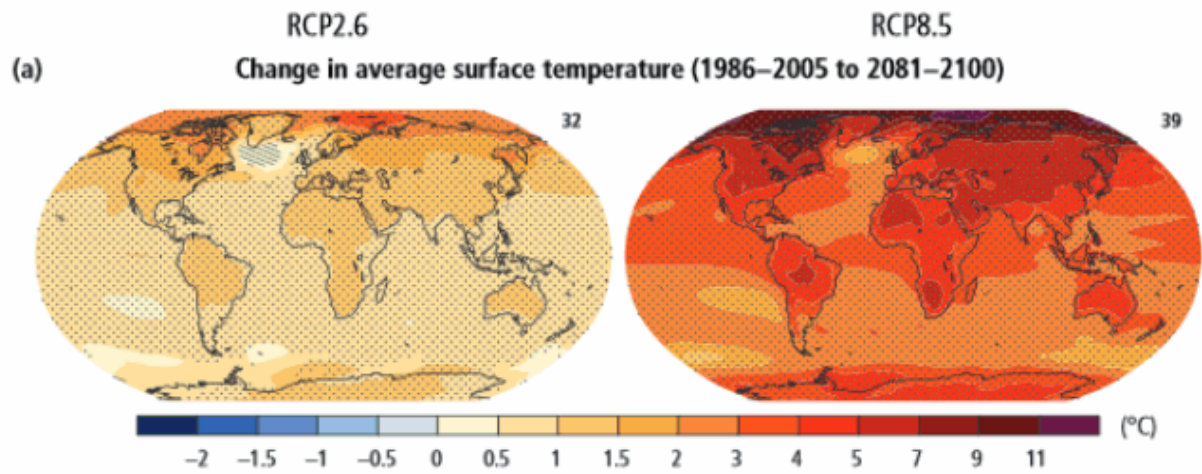


Figure 15 : Changement de température moyenne (°C) de surface prédit selon deux scénarii, RCP2.6 (optimiste) et RCP8.5 (pessimiste), entre les périodes 1986-2005 et 2081-2100 (Kirtman *et al.*, 2013).

compte la tendance à l'augmentation de la production des jeunes arbres, ce qui permet d'évaluer la régularité dès les premières années de floraison sur de jeunes arbres (Durand *et al.*, 2013). Le BBI et ses dérivés permettent de discriminer les individus réguliers des individus irréguliers. Mais, l'introduction d'un index d'auto-corrélation, basé sur la corrélation du rendement entre années successives permet de discriminer les individus irréguliers des individus alternants : des individus avec une forte corrélation négative entre rendements annuels sont identifiés comme alternants. Enfin, un indice mesurant la synchronisation des événements de floraison au sein d'un arbre a aussi été proposé, sur le concept de l'entropie (Durand *et al.*, 2013). L'entropie a révélé que différents comportements sont possibles entre génotypes : les génotypes réguliers ont tendance à avoir une floraison asynchrone entre méristèmes. Il peut être simplement irrégulier, n'ayant qu'une proportion variable de bourgeons floraux chaque année et finalement il peut avoir une production alternante, une année sur deux, due à une synchronisation forte de ces bourgeons une année donnée mais en alternance de phase entre années successives.

Le pommier présente l'avantage de posséder des marqueurs morphologiques clairs des événements de fructification, liés à la présence de bourses et de pousses de bourse (Figure 9). Ceci permet le phénotypage rétrospectif des séquences de floraison/croissance végétative le long des branches et des rameaux. Cet avantage a été mis à profit dans cette thèse pour tester la possibilité d'échantillonner quelques rameaux au sein des génotypes et par modélisation, d'estimer le comportement global de l'arbre. Ce travail sera décrit dans le deuxième chapitre de résultats.

2.2.2. Changement climatique et impact sur la phénologie

Le changement climatique, qui se traduit, entre autre, par une augmentation de la température particulièrement dans l'hémisphère Nord (Figure 15), a un fort impact sur la phénologie des espèces pérennes tempérées et sur le phénomène de dormance. En Europe, cette influence se traduit, pour l'instant, par une avancée des dates de floraison, expliquée par une levée de dormance plus tardive compensée par une forte réduction du temps nécessaire entre levée de dormance et floraison (Legave, 2009). Dans les régions tempérées chaudes comme le nord de la Méditerranée, l'avancée de floraison commence à être contrecarrée par un temps nécessaire pour satisfaire les besoins en froid de plus en plus long qui est de moins en moins compensé par la

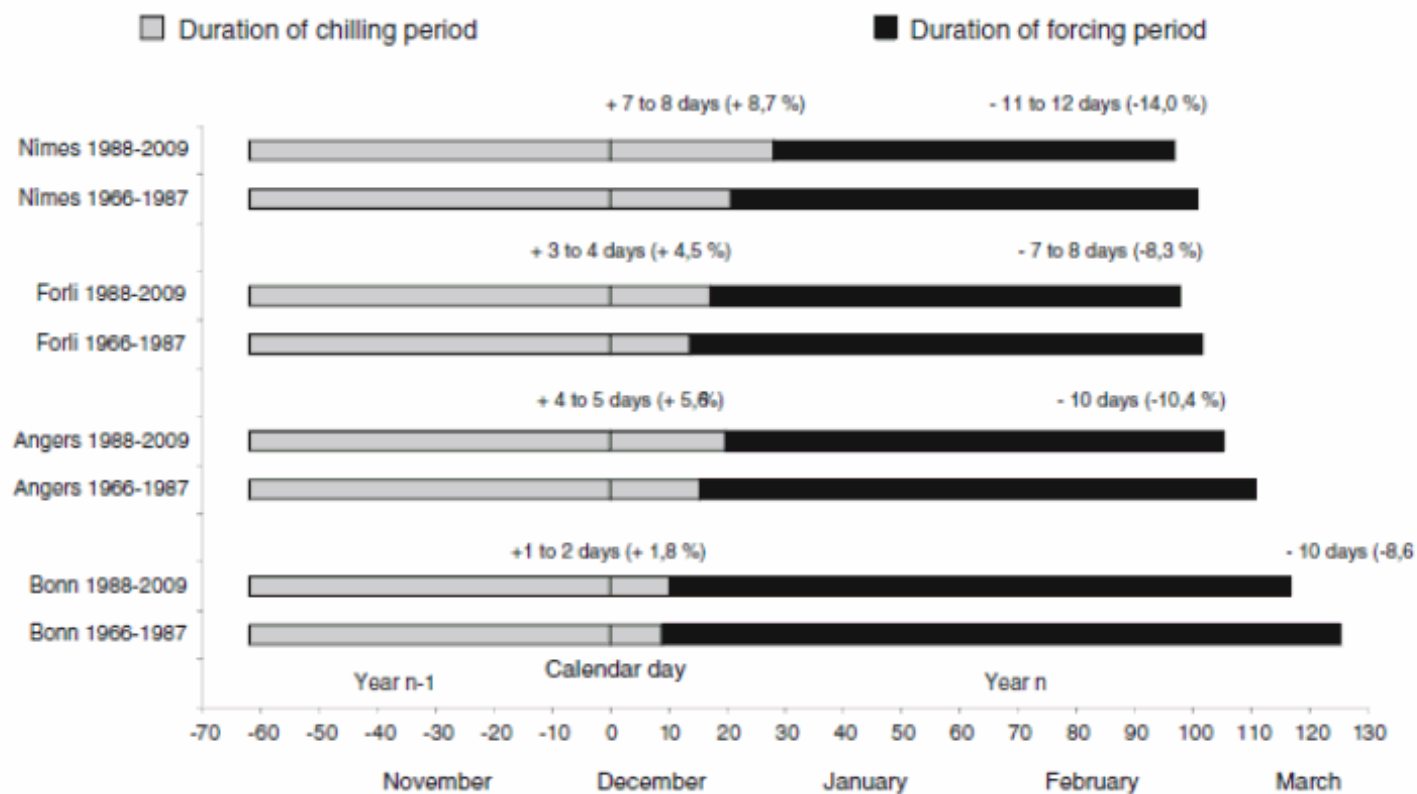


Figure 16 : Durée de satisfaction des besoins en froid et en chaleur, et dates de levée de dormance et de floraison pendant la période 1966-1987 et la période plus chaude 1988-2009, pour la variété ‘Golden Delicious’, dans différentes régions d’Europe à climats contrastés, Nîmes, Forlì en Italie, Angers, et Bonn en Allemagne (Legave *et al.*, 2013).

réduction du temps nécessaire pour satisfaire les besoins en chaleur (Figure 16) (Legave *et al.*, 2013). Au Brésil, des comportements anormaux lors de la floraison sont aussi observés et associés au manque de froid. Dans le contexte de changement climatique, l'augmentation de température prévue pour les décennies à venir pourraient entraîner des phénomènes similaires dans les régions tempérées plus froides (Malagi *et al.*, 2015).

Le manque de froid chez le pommier peut entraîner une mauvaise levée de dormance, être responsable d'un étalement de la floraison, d'une mauvaise pollinisation et d'un mauvais développement des fruits (Atkinson *et al.*, 2013). L'impact économique d'un manque de froid peut donc être important pour un producteur. C'est pourquoi l'adaptation des arbres aux conditions climatiques dans lesquelles il est cultivé est primordiale. Les besoins en froid et en chaleur et les températures sont aussi considérés comme des caractères héréditaires. Néanmoins leur détermination ne peut pas être mesurée directement car il faut pouvoir déterminer la date de levée dormance. Cela peut se faire par modélisation statistique; plusieurs types de modèles existent et diffèrent par les courbes de réponse aux températures, les températures seuils et l'interaction entre accumulation de froid et accumulation de chaleur (Legave *et al.*, 2013; Darbyshire *et al.*, 2013; Pope *et al.*, 2014). Ces estimations statistiques nécessitent néanmoins de longues séries de notation des dates de floraison, 10 ans minimum. Les besoins en froid et en chaleur ont été déterminés de cette manière pour certains cultivars comme 'Golden Delicious' (Legave *et al.*, 2008), mais jamais sur les individus composant une population en ségrégation. La détermination de la date de levée de dormance peut se faire aussi de manière expérimentale, par prélèvement de rameaux en verger à différentes dates et testés en serre pour leur réaction à des conditions favorables au débourrement (Gibson and Reighard, 2002; Fan *et al.*, 2010), ou alors par le test de Tabuenca (Tabuenca, 1964) qui consiste à prélever les bourgeons à différentes dates, à en extraire les ébauches florales pour les peser afin de suivre leur évolution et d'en déduire leur état d'activité. Ces tests sont efficaces mais nécessitent de forts moyens expérimentaux tels qu'une grande quantité de matériel végétal, de l'espace disponible en serre et de la main d'œuvre pour le prélèvement et le suivi de l'expérimentation. Les événements phénologiques comme la date de débourrement et la date de floraison sont considérés comme des caractères héréditaires, c'est pourquoi la sélection apparaît comme l'une des solutions pour l'adaptation des cultivars aux nouvelles et aux futures conditions climatiques (Labuschagne *et al.*, 2002).

3. Etude des déterminismes génétiques de caractères complexes chez le pommier

L'induction florale et l'influence des températures sur la date de débourrement et de floraison des bourgeons sont des étapes clés du cycle phénologique du pommier. Dans cette thèse, j'ai étudié leur déterminisme génétique. Néanmoins, ces deux caractères sont complexes phénotypiquement et génétiquement.

L'investissement pour le phénotypage de caractères complexes tels que ceux décrits dans les paragraphes précédents peut être lourd en temps, en besoins expérimentaux et en besoins humains. L'utilisation des connaissances et outils apportés par la génétique quantitative permet d'optimiser les processus de sélection par une détermination plus précoce et plus précise du potentiel génétique. De plus, de nouveaux outils de génotypage permettent aujourd'hui d'envisager de nouvelles perspectives pour les approches en génétique quantitative chez le pommier. En effet, le génome du pommier a été récemment séquencé (Velasco *et al.*, 2010), des marqueurs SNP et des puces à ADN ont été développés, comme la puce à 8000 SNP (Chagné *et al.*, 2012) puis celle à 20000 SNP (Bianco *et al.*, 2014) et récemment une puce Affimetrix à 487 000 SNP, dans le cadre du projet FruitBreedomics (Europe, www.fruitbreedomics.com ; Laurens *et al.*, 2010). L'investissement dans divers projets de recherche d'envergure internationale, HiDRAS (Europe, www.hidras.unimi.it), RosBREED (USA, www.rosbreed.org), FruitBreedomics (Europe, www.fruitbreedomics.com ; Laurens *et al.*, 2010) ont permis des avancées majeures sur le développement des populations d'étude, des outils de génotypage et sur l'étude des déterminismes génétiques des caractères principaux tels que la qualité du fruit et la résistance aux maladies. Dans ce qui suit, je retrace brièvement les concepts et les principaux QTLs indentifiés.

3.1. Stabilité et transférabilité des QTLs

La cartographie de QTL est basée sur le concept d'association statistique entre phénotype et génotype. Ces associations sont appelées quantitative trait loci (QTL) (Collard *et al.*, 2005). La cartographie de QTL requiert la construction de carte génétique pour la population d'étude. Le concept de carte génétique repose sur le fait que, dans une population en ségrégation, les gènes et les marqueurs ségrègent via le processus de recombinaison qui a lieu pendant la méiose. Plus les marqueurs sont proches les uns des autres, moins la fréquence de recombinaison est élevée. Ainsi

plus le taux de recombinaison est élevé entre deux marqueurs plus les deux marqueurs sont éloignés l'un de l'autre (Staub *et al.*, 1996; Collard *et al.*, 2005). La distance entre marqueurs est exprimée en centiMorgans, une unité qui est fonction du taux de recombinaison. Le génotypage des individus composant la ou les populations en ségrégation permet d'inférer les événements de recombinaison entre marqueurs et ainsi de construire la carte génétique. La pertinence des QTLs détectés réside en l'hypothèse que les marqueurs moléculaires utilisés sont en fort déséquilibre de liaison avec les gènes responsables de la variabilité du phénotype et ainsi que marqueurs et gènes ont la même ségrégation au sein de la population étudiée. Une grande densité de marquage et de nombreux individus favorisent une cartographie plus précise des QTLs (Collard *et al.*, 2005).

Les études d'un déterminisme génétique par la cartographie de QTLs se sont basées essentiellement sur des populations bi-parentales dont les deux parents sont choisis pour leurs phénotypes contrastés pour le caractère étudié (Blanc *et al.*, 2006; Cavanagh *et al.*, 2008). Ce type d'étude fournit une bonne localisation des régions chromosomiques associées au caractère et peut permettre d'estimer l'effet d'allèles rares si l'un des parents en est porteur (Würschum, 2012). Néanmoins, les QTLs ainsi détectés sont spécifiques de la population étudiée : ce sont uniquement des loci qui ségrègent dans cette population qui peuvent être détectés. De nombreux facteurs peuvent influencer la détection des QTLs : les effets d'épistasie, d'interaction avec l'environnement, le nombre d'individus composant la population, et la densité de marquage (Bernardo, 2008). Le même caractère étudié dans divers fonds génétiques peut donner des résultats différents. Par exemple, l'étude de (King *et al.*, 2000) sur la fermeté du fruit, sur pommier, a révélé des QTLs sur les groupes de Liaison (LG) 1, 2, 6, 8, 10, 12, et 16, alors que l'étude de (Liebhard *et al.*, 2003) sur une population différente a révélé des QTLs sur les LG 3, 6, 10, 12 et 16. Dans ce cas, les QTLs sur les LG6, 10 et 12 colocalisent dans les deux études, alors que les autres QTLs semblent être spécifiques des populations en ségrégation utilisées. De nombreuses études montrent également des effets de l'année sur ce même caractère (Ben Sadok *et al.*, 2015).

3.2. Elargissement de la diversité génétique

L'élargissement de la diversité allélique peut répondre à cette limitation. En plus d'une plus grande stabilité des résultats, les études basées sur une plus large diversité génétique confèrent un

plus grand pouvoir de détection des QTLs, une meilleure estimation de leur effet et position ainsi qu'une meilleure estimation des effets alléliques et de leur distribution (Coles *et al.*, 2010). Il existe plusieurs façons de considérer une diversité allélique plus large.

3.2.1. La génétique d'association

La génétique d'association est une méthode de détection d'association entre génotype et phénotype qui est basée sur une population de diversité composée d'individus dont les apparentements ne sont pas connus ou contrôlés, contrairement à la cartographie de QTL présentée plus haut, et donc la diversité allélique est large (Ingvarsson and Street, 2011; Khan and Korban, 2012). L'avantage de la génétique d'association est que le nombre d'associations détectées n'est pas limité par le nombre de QTLs qui ségrègent dans la population mais plutôt par le nombre réel de QTLs existant pour le trait étudié (Myles *et al.*, 2009). De plus, la résolution des associations détectées est plus fine car la grande diversité génétique de la population permet de prendre en compte tous les événements de recombinaisons qui ont donné naissance à cette population (Ingvarsson et Street, 2011). Dans des populations de diversité, le déséquilibre de liaison (LD) décroît plus rapidement que dans des populations en ségrégation. Par conséquent, la densité de marquage doit être adaptée afin de garantir la pertinence et la robustesse des associations détectées (Khan and Korban, 2012). C'est pourquoi l'application de cette méthode peut être parfois limitée par le nombre marqueurs moléculaires disponibles pour l'espèce en question. De plus, l'historique des événements génétiques qui a mené à la population utilisée, structure, apparentements, sélection, dérive génétique, peut avoir un impact sur la robustesse et la précision des associations détectées (Khan and Korban, 2012). La structure de la population ainsi que les apparentements doivent être pris en compte pour éviter la détection de QTLs faux positifs (Khan and Korban, 2012). La fréquence allélique peut aussi être limitante pour la détection d'association surtout lorsque celle-ci est faible (<5%) car l'effet du locus sur le phénotype sera lui aussi faible (Myles *et al.*, 2009).

La génétique d'association s'avère particulièrement intéressante pour les arbres, et notamment pour les espèces de la famille des *Rosaceae*. En effet, la détection de QTLs est limitée par la taille des populations en ségrégation utilisées, à cause du manque de diversité allélique présent dans une telle population. D'autre part, la possibilité de cartographier finement des QTLs est aussi

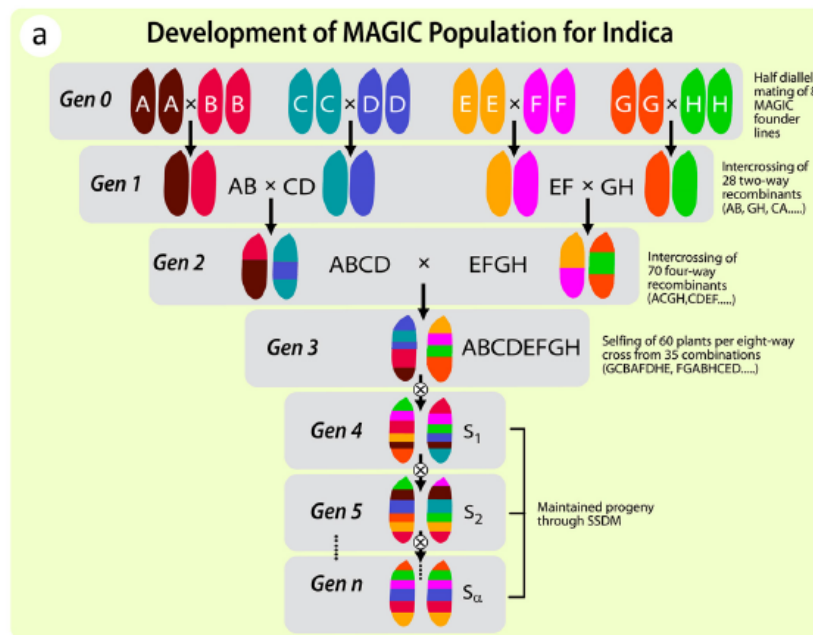


Figure 17 : Développement d'une population MAGIC qui consiste en le croisement de plusieurs lignées, suivi d'une succession d'autofécondations qui aboutissent sur des lignées fortement recombinantes (Bandillo *et al.*, 2013).

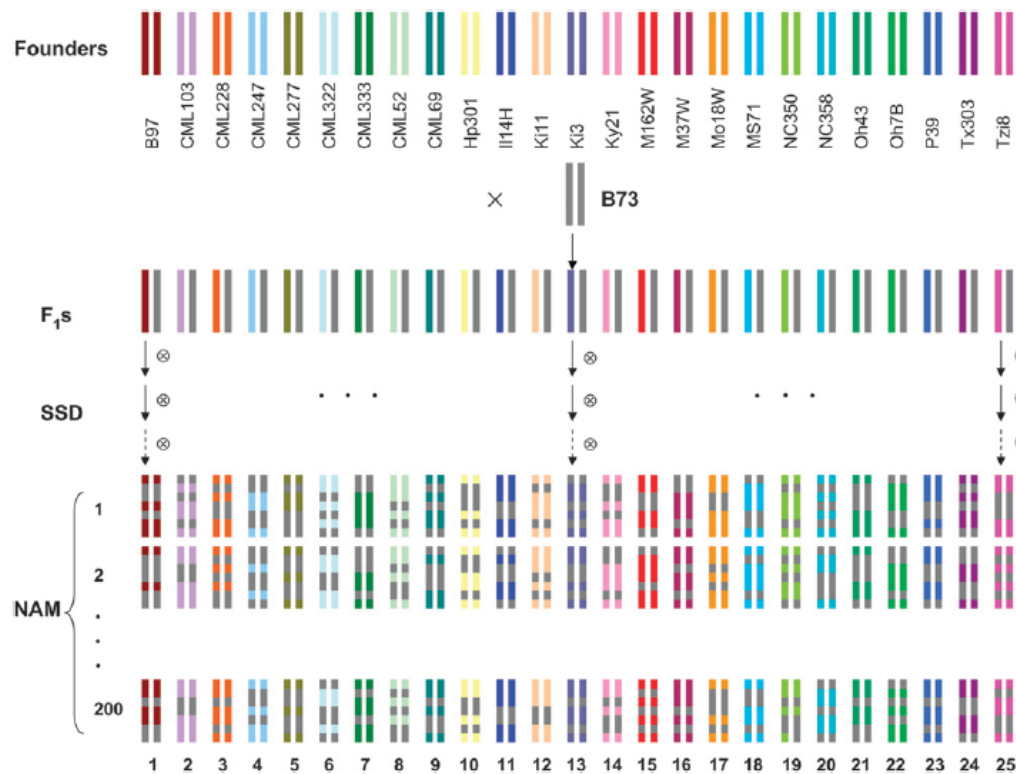


Figure 18 : Dispositif expérimental de type Nested Association Mapping (Yu *et al.*, 2008). La première étape pour la mise en place de ce dispositif est le croisement entre un parent récurrent (ici B73) et plusieurs lignées fondatrices (en couleurs). Puis une série d'autofécondation est réalisée pour obtenir des lignées recombinantes dont les génomes sont composés de morceaux venant du parent récurrent et d morceaux venant des lignées fondatrices.

limitée par le petit nombre d'individus recombinants, jusqu'à quelques centaines. La génétique d'association permet de surmonter ces limitations pour les raisons citées avant (Khan and Korban, 2012). Récemment, une core-collection de pommier a été définie, ce qui ouvre cette voie pour les années prochaines (Lassois *et al.*, 2016).

3.2.2. Méthodes de détection de QTLs multi-parentales

A l'interface entre génétique d'association et cartographie de QTLs sur famille biparentale, se trouvent des méthodes de détection de QTL basées sur plusieurs populations apparentées entre elles. Ces méthodes permettent de cumuler les avantages des deux approches : des fréquences alléliques plus grandes, un plus grand nombre de recombinaisons, une plus grande puissance de détection et une résolution plus fine de leur cartographie. Plusieurs types de population existent dont trois seront présentées dans les paragraphes qui suivent : la Multi-parent Advanced Generation Inter-Cross (MAGIC) développée sur les espèces annuelles, la Nested Association Mapping (NAM) appliquée chez le maïs et d'autres espèces annuelles comme le sorgho, et le Pedigree Based Analysis (PBA) qui s'adapte aux espèces allogames comme le pommier, et qui a été utilisé dans cette thèse.

Les populations MAGIC sont issues d'un croisement entre plusieurs lignées, suivi de plusieurs étapes d'inter-croisements et terminées par une succession d'auto-fécondations qui aboutissent à un grand nombre de lignées recombinantes (Figure 17) (Cavanagh *et al.*, 2008; Bandillo *et al.*, 2013). Le but de ces populations est de provoquer les recombinaisons afin de pouvoir cartographier finement des QTLs.

La méthode NAM repose sur un dispositif expérimental composé d'un grand nombre de lignées qui dérivent de croisements avec un parent commun (Figure 18). Par exemple, dans le cas du maïs, une lignée fondatrice a été croisée avec 25 autres lignées, par auto-fécondation. Une population F2 est développée pour chaque croisement, puis par auto-fécondation des lignées recombinantes sont développées à partir des individus issus de ces populations F2. C'est sur ces lignées recombinantes que des QTLs seront cartographiés. Les objectifs de la NAM sont de capturer la diversité génétique de l'espèce étudiée, d'exploiter les événements de recombinaison ancestraux, d'augmenter le pouvoir de détection des QTL et la précision de leur position (Yu *et al.*, 2008).

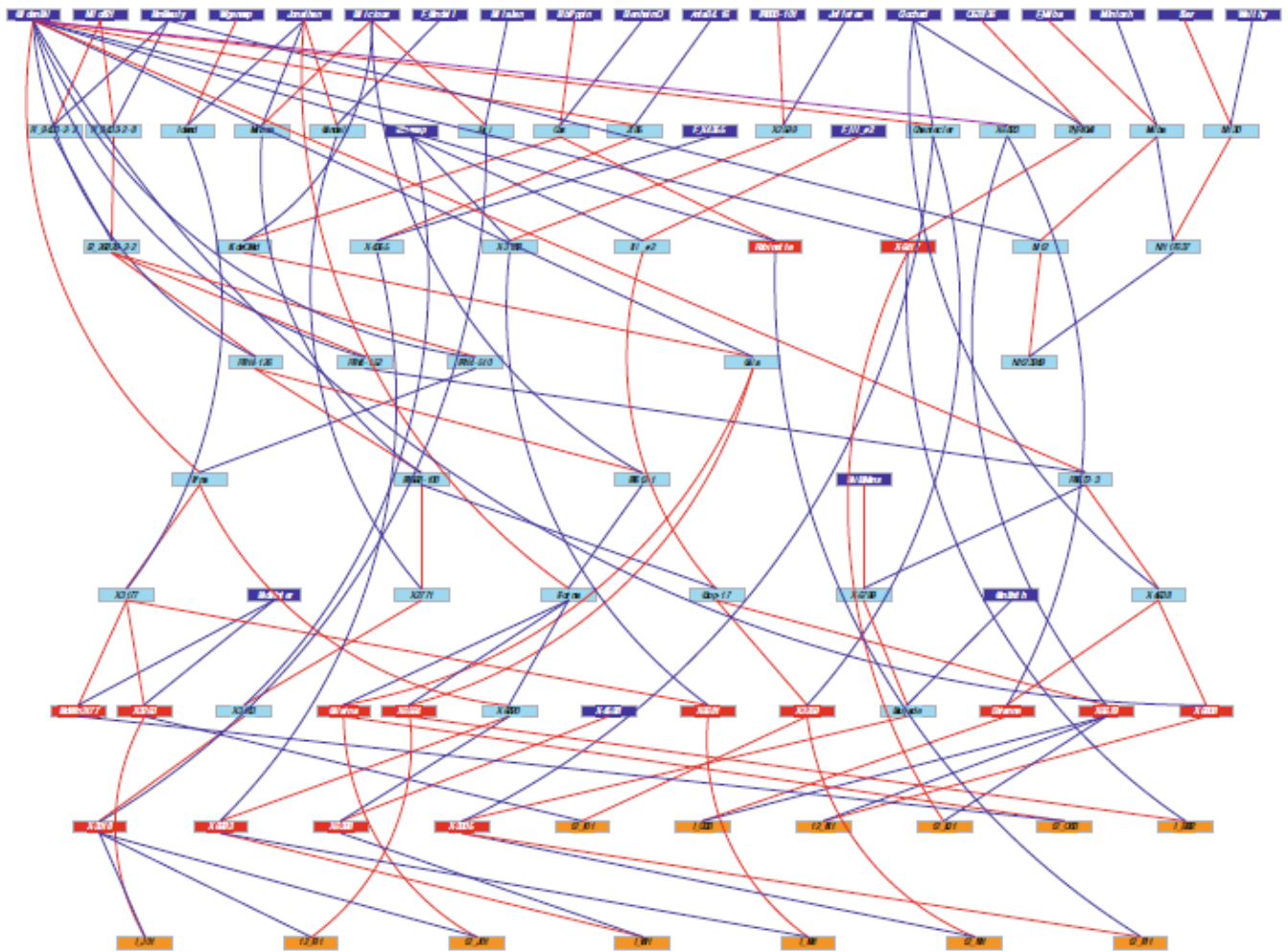


Figure 19 : Structure de pedigree pour une analyse de type PBA du projet FruitBreedomics. En orange, les populations en ségrégation, en rouge leurs parents, en bleu foncés les individus fondateurs et en bleu clair les individus intermédiaires. Les populations en ségrégation sont reliées à leurs parents, en rouge avec le parent femelle et en bleu avec le parent mâle.

Le Pedigree Based Analysis (PBA) (Bink *et al.*, 2002, 2008, 2012, 2014) est une méthode de cartographie de QTL dont les objectifs sont communs à ceux de la NAM. La méthode PBA propose d'utiliser une population d'étude composée de plusieurs descendance interconnectées par leur pedigree (Figure 19, pour un exemple de structure de population). L'augmentation de la diversité allélique apportée par ce type de dispositif permet une exploration plus approfondie des QTLs responsables de variabilité du caractère étudié (Bink *et al.*, 2002). Néanmoins, la prise en compte des apparentements entre individus est nécessaire pour améliorer la précision de la cartographie des QTLs. Dans la méthode PBA, les apparentements sont pris en compte par Identity By Descent (IBD, notion présentée dans la suite de ce document). Pour chaque position du génome, une matrice d'apparentement est estimée (Bink *et al.*, 2012). Pour cela, les ancêtres de chacun des parents sont spécifiés. Les individus fondateurs sont identifiés, ainsi que les individus intermédiaires entre parents et fondateurs, lorsqu'ils sont connus (Bink *et al.*, 2012). La méthode PBA modélise des QTLs bi-alléliques. Ceci peut apparaître comme une simplification de la réalité. Mais cela est plus efficace que d'estimer le nombre d'allèles par QTL, ce qui peut être statistiquement limité lorsque les effets des allèles sont peu contrastés. Les QTLs multi-alléliques sont modélisés par plusieurs QTL bi-alléliques localisés dans des loci adjacents (Bink *et al.*, 2008). C'est cette méthode qui a été choisie dans ce travail.

3.3. Estimation des héritabilités et des effets génétiques

La génétique quantitative peut aussi permettre d'estimer et de prédire le potentiel génétique des individus à partir des mêmes marqueurs moléculaires ayant servis à identifier les QTLs.

Une variable phénotypique P peut être exprimée de manière simple par l'équation suivante (Gallais, 1990):

$$P_{ij} = \mu + G_i + E_j + G_i * E_j + \varepsilon_{ij}$$

où G est le facteur 'Génotype', E le facteur 'Environnement', au sens large du terme, $G * E$ l'interaction entre les facteurs 'Génotype' et 'Environnement' et ε_{ij} la résiduelle du modèle.

Cette modélisation du phénotype est utilisée pour séparer de l'effet génotypique de l'effet de l'environnement, ainsi la variance phénotypique peut être, elle aussi, décomposée de la manière suivante :

$$\sigma_P^2 = \sigma_G^2 + \sigma_E^2 + \sigma_{G \times E}^2 + \sigma_\epsilon^2$$

où σ_P^2 , σ_G^2 , σ_E^2 , $\sigma_{G \times E}^2$ et σ_ϵ^2 sont les variances phénotypiques, génotypiques, environnementales, associées à l'interaction 'Génotype' et 'Environnement' et résiduelle, respectivement. A partir des composantes de la variance phénotypique, l'héritabilité au sens large peut être estimée :

$$H = \frac{\sigma_G^2}{\sigma_G^2 + \sigma_{G \times E}^2/p + \sigma_\epsilon^2/n * p}$$

où n est le nombre de répétition par génotypes et p le nombre d'environnement, d'après (Gallais, 1990).

A partir de ce type de modèle, les effets respectifs de l'environnement et des génotypes peuvent être estimés. Dans de nombreuses études de génétique quantitative, des modèles mixtes sont utilisés dans lesquels les effets génotypiques sont considérés comme des effets aléatoires (van Eeuwijk *et al.*, 2010; Segura *et al.*, 2008; Celton *et al.*, 2011, 2014; Guitton *et al.*, 2012; Ben Sadok *et al.*, 2015; Virlet *et al.*, 2015). Cette démarche permet de décomposer la variance les phénotypes (Celton *et al.*, 2011; Guitton *et al.*, 2012; Ben Sadok *et al.*, 2015) et d'estimer les valeurs génétiques individuelles par les Best Linear Unbiased Predictors (BLUPs) qui représentent une estimation des effets aléatoires des génotypes.

Les marqueurs moléculaires peuvent permettre de disséquer cette valeur génétique globale en effets locaux et d'estimer la valeur génétique grâce à la compilation de leurs effets alléliques. Les marqueurs moléculaires peuvent donc permettre de prédire la valeur génétique avant que l'individu n'ait été phénotypé. (Lande and Thompson, 1990) ont proposé une méthode d'estimation qui consiste à additionner les effets alléliques à chaque QTLs, dans le but d'estimer un score génotypique par individu. Pour des caractères complexes régulés par un grand nombre de loci régulés par un grand nombre de loci à petits effets, peu de QTLs sont en général détectés, et n'expliquent qu'une faible part de variance génétique. Le cumul de leurs effets peut ne pas

suffire à estimer précisément la valeur génétique et entraîner un fort taux d'erreur de prédiction (Francia *et al.*, 2005; Collard and Mackill, 2008; Jannink *et al.*, 2010; Kumar *et al.*, 2012; Cooper *et al.*, 2014). Pour pallier à ce problème, Meuwissen *et al.* (2001) ont proposé d'estimer la valeur génétique en utilisant les effets alléliques de marqueurs densément répartis sur le génome. Ainsi, aucun a priori sur le déterminisme génétique du trait étudié n'est nécessaire car tous les loci sont pris en compte, même ceux dont l'effet est faible. La sélection génomique, basée sur ce concept, consiste en l'estimation d'effets alléliques de marqueurs répartis sur tout le génome à partir d'une population d'entraînement génotypée et phénotypée. Puis ces modèles sont appliqués sur des populations de sélection uniquement génotypées pour lesquelles les valeurs génétiques des individus seront prédites à partir de leurs génotypes et des modèles précédemment calibrés (Kumar *et al.*, 2012; Zhao *et al.*, 2015).

3.4. Apparentements

Les méthodes de détection de QTL multi-parentales, génétique d'association ou d'estimation de la valeur génétique nécessitent de prendre en compte les apparentements entre individus afin d'éviter la détection d'associations génotype-phénotype qui seraient un artefact dû à l'apparentement, et d'augmenter la précision des estimations (Würschum, 2012). De plus, la prise en compte des apparentements permet d'optimiser les modèles d'estimation des effets des marqueurs en diminuant le nombre de paramètres (Blanc *et al.*, 2006), et de prendre en compte les événements de recombinaison dans le pedigree ainsi que dans les descendance (Bink *et al.*, 2012).

L'apparentement est mesuré par la probabilité qu'à un locus donné deux individus partagent le même allèle issu d'un ancêtre commun (Eding and Meuwissen, 2001). Cette probabilité varie suivant le degré d'apparentement entre les deux individus. Un degré d'apparentement moyen peut être mesuré grâce à l'information apportée par la connaissance des parents et ancêtres des individus concernés. L'apparentement peut être aussi calculé à partir de l'information apportée par les marqueurs en se basant sur le principe d'Identity By State (IBS), si à un locus donné deux individus portent le même allèle alors il est probablement hérité d'un ancêtre commun. La première étape est de calculer un score de similarité entre chaque couple d'individu à un locus donné, avec la formule suivante :

$$S_{xy,l} = \frac{1}{4} * (I_{11} + I_{12} + I_{21} + I_{22}) \text{ (Eding and Meuwissen, 2001)}$$

où I_{ij} est une variable prenant la valeur de 1 lorsque au locus l , l'allèle i de l'individu x est identique à l'allèle j de l'individu y , sinon I_{ij} prend la valeur 0. La moyenne des $S_{xy,l}$ pour tous les loci donne une estimation de l'apparentement moyen entre les individus x et y en calculant la proportion d'allèles partagés entre individus (Lynch, 1988). Cette méthode nécessite que l'hypothèse suivante soit vérifiée : les allèles des individus fondateurs doivent être uniques. Cette méthode donne une bonne estimation de l'apparentement avec des marqueurs moléculaires multi-alléliques, comme les marqueurs microsatellites. En revanche, l'estimation est beaucoup moins bonne avec des marqueurs moléculaires bi-alléliques tels que les SNP car il est plus difficile de s'assurer que deux individus partageant un allèle commun l'aient hérité d'un ancêtre commun.

Enfin, un apparentement peut être calculé selon le principe d'Identity By Descent (IBD) qui fait appel à l'information apportée par le pedigree, par les marqueurs, et par la probabilité de recombinaison entre position et marqueur. La valeur d'apparentement par descendance entre deux individus représente la probabilité qu'un allèle tiré au hasard à un locus donné pour le premier individu soit identique par descendance à un allèle tiré au hasard pour le second individu au même locus, autrement dit que cet allèle soit hérité d'un ancêtre commun. L'IBD entre deux individus peut être calculé à toute position du génome. Contrairement à l'IBS, le calcul d'IBD se fait donc toujours suivant un groupe de fondateur identifiés (Thompson, 2013). L'IBD procure une estimation de l'apparentement entre deux individus à l'échelle du locus. Cette information peut être combinée à l'échelle du génome afin de donner une valeur globale de l'apparentement entre deux individus, mais peut aussi être utilisée à l'échelle du locus afin de modéliser plus précisément l'effet du locus sur le phénotype.

L'IBD mesure l'apparentement par rapport à un pedigree connu. Par conséquent, des relations de parenté inconnues ne peuvent pas être modélisées de cette manière (Luan *et al.*, 2012). L'IBS n'émet aucune hypothèse concernant d'éventuels apparentement connus, par conséquent de cette manière peuvent être pris en compte des apparentements inconnus ou plus anciens que ceux déjà connus (Luan *et al.*, 2012). Au-delà d'un certain nombre de générations, quatre selon Luan *et al.* (2012), le calcul de l'apparentement par IBS n'est pas plus informatif que celui par IBD. De plus, le calcul d'apparentement par IBD est moins influencé par la densité de marquage. L'utilisation

de l'une ou l'autre des techniques dépend du dispositif expérimental, des outils génétiques disponibles et du but de l'étude.

3.5. Les statistiques Bayésiennes au secours de la détection de QTLs

Pour les caractères quantitatifs, le nombre de loci impliqués dans la variation phénotypique peut être considérable. Il faut donc pouvoir bâtir des modèles pour tester l'éventuelle présence de QTL sur un grand nombre de loci afin d'en estimer leur effet respectifs. Les généticiens ont fait appel aux statistiques bayésiennes pour répondre à ce besoin.

Alors que les statistiques classiques se basent sur des tests d'hypothèse pour estimer la valeur d'un paramètre d'intérêt, les statistiques bayésiennes, permettent d'inférer une valeur au paramètre d'intérêt par l'utilisation de la distribution de probabilité *a posteriori* de ce paramètre (Shoemaker *et al.*, 1999). Le calcul de la distribution de probabilité se fait suivant le théorème de Bayes :

$$p(\theta|X) = \frac{p(\theta) * p(X|\theta)}{p(X)}$$

où θ représente le paramètre sur lequel porte l'inférence, X les données disponibles. $p(\theta|X)$ représente la distribution *a posteriori* soit la distribution du paramètre θ sachant les données X , $p(\theta)$ la distribution *a priori* du paramètre, $p(X|\theta)$ la vraisemblance du modèle soit la distribution des données X sachant le paramètre θ et $p(X)$ la distribution observées des données (Albert *et al.*, 2012). Ainsi, le théorème de Bayes permet de calculer une distribution de probabilité *a posteriori* en prenant en compte la connaissance détenue *a priori* sur le paramètre, la vraisemblance du modèle par rapport aux données et les données observées. Lorsque le nombre de paramètres est grand, la distribution *a posteriori* est complexe et nécessite une capacité de calcul importante et des algorithmes de calculs efficaces. Le développement d'algorithmes tels que les Chaînes de Markov Monte Carlo (MCMC) rendent cela possible (Shoemaker *et al.*, 1999). A chaque itération de MCMC, la distribution *a posteriori* est calculée à partir des inférences sur les paramètres des itérations précédentes, un nombre suffisant d'itération permet de converger vers des inférences stables pour les paramètres (Yi and Shriner, 2007).

En génétique, les statistiques bayésiennes sont utilisées pour estimer simultanément l'effet de nombreux loci. Des méthodes bayésiennes sont implémentées dans de nombreux algorithmes de détection de QTL ainsi que dans des modèles de sélection génomique. Elles offrent de nombreux avantages par rapport aux statistiques classiques : l'inférence en statistique bayésienne est plus robuste que l'estimation des paramètres en statistique classique, une capacité à estimer de nombreux paramètres, à résoudre des équations pour lesquelles le nombre de données observées est beaucoup plus faible que le nombre de paramètres à inférer, la possibilité d'utiliser les connaissances a priori conjointement à l'information apportée par les données, et la quantification de l'incertitude sur les paramètres par la distribution de probabilité a posteriori (Shoemaker *et al.*, 1999; Yi and Shriner, 2007; Gianola *et al.*, 2009).

4. Intégration de plusieurs caractères

Bien que l'étude des déterminismes génétiques se fasse en général sur un caractère à la fois, la sélection porte sur plusieurs caractères. Par exemple, pour le pommier, les caractères les plus importants sont la qualité du fruit, qui est elle-même composée d'une multitude de sous-caractères, la résistance aux maladies, et les caractères liés au rendement. D'autres caractères pourraient aussi faire l'objet de sélection : la régularité de production, qui est aujourd'hui prise en compte dans les schémas de sélection mais de manière empirique, ou la phénologie des bourgeons en réponse aux températures automno-hivernales. Au total, tous ces caractères sont influencés par un grand nombre de gènes dont certains peuvent être communs à plusieurs caractères, et dont les effets peuvent être soit synergiques soit antagonistes. Ces hypothèses concernant la corrélation génétique des caractères sont importantes à prendre en compte en sélection, et peuvent être envisagées de plusieurs manières suivant le niveau de corrélation entre les caractères et l'objectif visé.

4.1. Analyses conjointes pour des traits corrélés

L'analyse multi-caractères peut répondre à des questions biologiquement très intéressantes (Jiang and Zeng, 1995) comme la nature d'une corrélation génétique entre deux caractères, l'existence éventuelle d'effet de pléiotropie, l'étude du déterminisme génétique d'un caractère à faible héritabilité due à une grande influence de l'environnement sur le phénotype (Alimi *et al.*, 2013).

Le modèle basique pour la modélisation multi-trait est le suivant :

$$y_{ij} = \mu + T_j + G_i + GT_{ij} + \varepsilon_{ij}$$

Avec y_{ij} la valeur du génotype i pour le trait j , T_j l'effet du trait j , G_i l'effet du génotype i , GT_{ij} l'effet d'interaction entre trait j et génotype i et ε_{ij} l'erreur (Sukhwinder-Singh *et al.*, 2012; Alimi *et al.*, 2013). Ce type de modèle peut aussi permettre de prendre en compte les effets additifs des marqueurs, le modèle est alors le suivant :

$$y_{ij} = \mu + T_j + \sum_{f \in F} x_{if}^{add} c_{if}^{add} + x_i^{add} \alpha_j^{add} + GT_{ij} + \varepsilon_{ij}$$

Avec F les cofacteurs choisis (si des cofacteurs sont inclus dans le modèle), x_{if}^{add} et x_i^{add} sont les prédictors additifs du génotype i au cofacteur f et à la position testée, respectivement ; et c_{if}^{add} et α_j^{add} sont les effets respectifs associés au cofacteur et à la position testée (Sukhwinder-Singh *et al.*, 2012; Alimi *et al.*, 2013). Ces modèles sont souvent utilisés sous la forme de modèles mixtes, avec les génotypes considérés comme facteurs aléatoires et les autres facteurs comme facteurs fixes.

Parmi les avantages apportés par la détection conjointe de QTLs pour plusieurs caractères, il y a une plus grande puissance statistique de détection des QTLs, et une plus grande précision de l'estimation des paramètres des modèles utilisés. De plus, ce type de modèle peut permettre de détecter des QTLs pour un caractère peu héritable corrélé à un caractère très héritable (Jiang and Zeng, 1995; Knott and Haley, 2000; Caliński *et al.*, 2000; Sukhwinder-Singh *et al.*, 2012; Alimi *et al.*, 2013). Mais l'un des désavantages de ce type d'analyse est la complexité des modèles bâtis. Ces outils requièrent la capacité à estimer un grand nombre de paramètres, et donc des outils statistiques puissants et adaptés (Alimi *et al.*, 2013). Cette démarche est envisageable pour des caractères qui sont corrélés génétiquement les uns aux autres. Par exemple, chez le pommier, la texture du fruit est un important paramètre de qualité. Il peut être mesuré par plusieurs paramètres, sensoriels ou instrumentaux qui révèlent différents aspects du caractère (Ben Sadok *et al.*, 2015), ce qui rend leur analyse conjointe pertinente.

4.2. Idéotypes et indices de sélection

Les exigences de la sélection portant sur des caractères potentiellement de natures très différentes, il est nécessaire d'envisager leur intégration afin de pouvoir répondre aux objectifs complexes de la sélection.

L'utilisation d'un idéotype est une méthode très efficace pour identifier des objectifs de sélection précis. D'après (Donald, 1968), l'idéotype est un modèle qui maximise la quantité ou la qualité du rendement d'une espèce cultivée. Pour les espèces pérennes, tel que le pommier, un idéotype comporte des caractères additionnels comme les résistances aux maladies et ravageurs, l'architecture de l'arbre, la régularité de production, l'adaptabilité au climat et sa résistance à des stress abiotiques par exemple (Costes *et al.*, 2009). Un idéotype est intimement lié aux connaissances scientifiques des caractères qui le composent. Plus les connaissances sont fines plus l'idéotype est précis. C'est un modèle qui évolue avec l'approfondissement des connaissances scientifiques mais peut aussi varier d'un pays à l'autre, ou en fonction de paramètres tel que le climat par exemple. L'idéotype rassemble les objectifs de sélection, et peut servir de base à l'élaboration d'une stratégie de sélection par indices de sélection.

L'intégration des caractères peut également se faire par l'utilisation d'indices de sélection. Les indices de sélection ont été définis comme la performance phénotypique générale d'un individu sur plusieurs caractères (Hazel, 1943). L'utilisation d'indices de sélection peut permettre d'augmenter l'efficacité de la sélection, par l'amélioration simultanée sur plusieurs caractères. Dans ce travail, j'ai utilisé les concepts d'idéotype et d'indice de sélection pour définir les valeurs génétiques multi-caractères des individus de deux familles en ségrégation qui seront détaillés plus avant dans le dernier chapitre.

Références

- Albert I, Ancelet S, David O, Denis JB, Makowski D, Parent E, Soubeyrand S.** 2012. *Méthodes statistiques bayésiennes- Bases théoriques et application en alimentation, environnement et génétique.*
- Alimi NA, Bink MCAM, Dieleman JA, Magán JJ, Wubs AM, Palloix A, van Eeuwijk FA.** 2013. Multi-trait and multi-environment QTL analyses of yield and a set of physiological traits in pepper. *Theoretical and Applied Genetics* **126**, 2597-2625.
- Anderson JL, Richardson EA.** 1986. Validation of chill unit and flower bud phenology models for 'Montmorency' sour cherry. *Acta Horticulturae* **184**.
- Atkinson CJ, Brennan RM, Jones HG.** 2013. Declining chilling and its impact on temperate perennial crops. *Environmental and Experimental Botany* **91**, 48-62.
- Bandillo N, Raghavan C, Muyco PA, et al.** 2013. Multi-parent advanced generation inter-cross (MAGIC) populations in rice: progress and potential for genetics research and breeding. *Rice* **6**, 11.
- Bangerth F.** 2006. Flower induction in perennial fruit trees: still an enigma? *Acta Horticulturae* **727**, 177-195.
- Bangerth KF.** 2009. Floral induction in mature, perennial angiosperm fruit trees: Similarities and discrepancies with annual/biennial plants and the involvement of plant hormones. *Scientia Horticulturae* **122**, 153-163.
- Ben Sadok I, Tiecher A, Galvez-Lopez D, Bruneau M, Hanteville S, Robic R, Cournol R, Lahaye M, Lasserre-Zuber P, Laurens F.** 2015. Apple fruit texture QTLs: year and cold storage effects on sensory and instrumental traits. *Tree Genetics & Genomes* **accepted**.
- Bernardo R.** 2008. Molecular Markers and Selection for Complex Traits in Plants: Learning from the Last 20 Years. *Crop Science* **48**, 1649.
- Bianco L, Cestaro A, Sargent DJ, et al.** 2014. Development and Validation of a 20K Single Nucleotide Polymorphism (SNP) Whole Genome Genotyping Array for Apple (*Malus × domestica* Borkh). *PLoS ONE* **9**, e110377.
- Bink MCAM, Boer MP, Braak CJF, Jansen J, Voorrips RE, Van de Weg E.** 2008. Bayesian analysis of complex traits in pedigreed plant population. *Euphytica* **161**, 85-96.
- Bink MCAM, Jansen J, Madduri M, et al.** 2014. Bayesian QTL analyses using pedigreed families of an outcrossing species, with application to fruit firmness in apple. *Theoretical and Applied Genetics* **127**, 1073-1090.
- Bink MCAM, Radu Totir L, ter Braak CJF, Winkler CR, Boer MP, Smith OS.** 2012. QTL linkage analysis of connected populations using ancestral marker and pedigree information. *Theoretical and Applied Genetics* **124**, 1097-1113.

Bink MCAM, Uimari P, Sillanpää MJ, Janss LLG, Jansen RC. 2002. Multiple QTL mapping in related plant populations via pedigree-analysis approach. *Theoretical and Applied Genetics* **104**, 751-762.

Blanc G, Charcosset A, Mangin B, Gallais A, Moreau L. 2006. Connected populations for detecting quantitative trait loci and testing for epistasis: an application in maize. *Theoretical and Applied Genetics* **113**, 206-224.

Böhlenius H, Huang T, Charbonnel-Campaa L, Brunner AM, Jansson S, Strauss SH, Nilsson O. 2006. CO/FT Regulatory Module Controls Timing of Flowering and Seasonal Growth Cessation in Trees. *Science* **312**, 1040-1043.

Boss PK, Bastow RM, Mylne JS, Dean C. 2004. Multiple Pathways in the Decision to Flower: Enabling, Promoting, and Resetting. *The Plant Cell* **16**, S18-S31.

Breen KC, Tustin DS, Palmer JW, Close DC. 2015. Method of manipulating floral bud density affects fruit set responses in apple. *Scientia Horticulturae* **197**, 244-253.

Bushakra JM, Sargent DJ, Cabrera A, et al. 2012. Rosaceae conserved orthologous set (RosCOS) markers as a tool to assess genome synteny between *Malus* and *Fragaria*. *Tree Genetics & Genomes* **8**, 643-658.

Butault JP, Delame N, Jacquet F, et al. 2010. L'utilisation des pesticides en France: Etat de lieux et perspectives de réduction.

Caliński T, Kaczmarek Z, Krajewski P, Frova C, Sari-Gorla M. 2000. A multivariate approach to the problem of QTL localization. *Heredity* **84**, 303-310.

Cavanagh C, Morell M, Mackay I, Powell W. 2008. From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. *Current Opinion in Plant Biology* **11**, 215-221.

Celton J-M, Kelner J-J, Martinez S, Bechti A, Touhami AK, James MJ, Durel C-E, Laurens F, Costes E. 2014. Fruit Self-Thinning: A Trait to Consider for Genetic Improvement of Apple Tree. *PLoS ONE* **9**.

Celton JM, Martinez S, Jammes MJ, Bechti A, Salvi S, Legave JM, Costes E. 2011. Deciphering the genetic determinism of bud phenology in apple progenies: a new insight into chilling and heat requirement effects on flowering dates and positional candidate genes. *New Phytologist* **192**, 378-392.

Chagné D, Crowhurst RN, Troggio M, et al. 2012. Genome-Wide SNP Detection, Validation, and Development of an 8K SNP Array for Apple. *PLoS ONE* **7**, e31745.

Chan B, Cain J. 1967. The effect of seed formation on subsequent flowering in apple. *Journal of the American Society for Horticultural Science*, 63-67.

Cilas C, Montagnon C, Bar-Hen A. 2011. Yield stability in clones of *Coffea canephora* in the short and medium term: longitudinal data analyses and measures of stability over time. *Tree Genetics and Genomes* **7**, 421-429.

Coles ND, McMullen MD, Balint-Kurti PJ, Pratt RC, Holland JB. 2010. Genetic Control of Photoperiod Sensitivity in Maize Revealed by Joint Multiple Population Analysis. *Genetics* **184**, 799-812.

Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK. 2005. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica* **142**, 169-196.

Collard BCY, Mackill DJ. 2008. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philosophical Transactions of the Royal Society B: Biological Sciences* **363**, 557-572.

Cooke JEK, Eriksson ME, Junttila O. 2012. The dynamic nature of bud dormancy in trees: environmental control and molecular mechanisms. *Plant, Cell & Environment* **35**, 1707-1728.

Cooper M, Messina CD, Podlich D, Totir LR, Baumgarten A, Hausmann NJ, Wright D, Graham G. 2014. Predicting the future of plant breeding: complementing empirical evaluation with genetic prediction. *Crop and Pasture Science* **65**, 311-336.

Cornille A, Giraud T, Smulders MJM, Roldán-Ruiz I, Gladieux P. 2014. The domestication and evolutionary ecology of apples. *Trends in Genetics* **30**, 57-65.

Costes E, Lauri P-E, Laurens F, Durel C-E, Parisi L. 2009. Construire un idéotype de variétés de pommiers pour des vergers agronomiquement performants et à faibles intrants. *Innovations Agronomiques* **7**, 49-64.

Costes E, Lauri PÉ, Regnard JL. 2006. Analyzing Fruit Tree Architecture: Implications for Tree Management and Fruit Production. In: Janick J, ed. *Horticultural Reviews*. John Wiley & Sons, Inc., 1-61.

Costes E, Sinoquet H, Kelner JJ, Godin C. 2003. Exploring Within-tree Architectural Development of Two Apple Tree Cultivars Over 6 Years. *Annals of Botany* **91**, 91-104.

Darbyshire R, Webb L, Goodwin I, Barlow EWR. 2013. Impact of future warming on winter chilling in Australia. *International Journal of Biometeorology* **57**, 355-366.

Dennis FG. 2000. The history of fruit thinning. *Plant Growth Regulation* **31**, 1-16.

Dirlewanger E, Graziano E, Joobeur T, Garriga-Calderé F, Cosson P, Howad W, Arús P. 2004. Comparative mapping and marker-assisted selection in Rosaceae fruit crops. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 9891-9896.

Donald CM. 1968. The breeding of crop ideotypes. *Euphytica* **17**, 385-403.

Durand J-B, Guitton B, Peyhardi J, Holtz Y, Guédon Y, Trottier C, Costes E. 2013. New insights for estimating the genetic value of segregating apple progenies for irregular bearing during the first years of tree production. *Journal of Experimental Botany* **64**, 5099-5113.

Dyk MM van, Soeker MK, Labuschagne IF, Rees DJG. 2010. Identification of a major QTL for time of initial vegetative budbreak in apple (*Malus x domestica* Borkh.). *Tree Genetics & Genomes* **6**, 489-502.

Eding H, Meuwissen THE. 2001. Marker-based estimates of between and within population kinships for the conservation of genetic diversity. *Journal of Animal Breeding and Genetics* **118**, 141-159.

van Eeuwijk FA, Boer M, Totir LR, et al. 2010. Mixed model approaches for the identification of QTLs within a maize hybrid breeding program. *Theoretical and Applied Genetics* **120**, 429-440.

Falavigna VDS, Porto DD, Buffon V, Margis-Pinheiro M, Pasquali G, Revers LF. 2013. Differential Transcriptional Profiles of Dormancy-Related Genes in Apple Buds. *Plant Molecular Biology Reporter* **32**, 796-813.

Fan S, Bielenberg DG, Zhebentyayeva TN, Reighard GL, Okie WR, Holland D, Abbott AG. 2010. Mapping quantitative trait loci associated with chilling requirement, heat requirement and bloom date in peach (*Prunus persica*). *New Phytologist* **185**, 917-930.

Flachowsky H, Szankowski I, Waidmann S, Peil A, Tränkner C, Hanke M-V. 2012. The MdTFL1 gene of apple (*Malus x domestica* Borkh.) reduces vegetative growth and generation time. *Tree Physiology* **32**, 1288-1301.

Foster T, Johnston R, Seleznyova A. 2003. A Morphological and Quantitative Characterization of Early Floral Development in Apple (*Malus x domestica* Borkh.). *Annals of Botany* **92**, 199-206.

Francia E, Tacconi G, Crosatti C, Barabaschi D, Bulgarelli D, Dall'Aglio E, Valè G. 2005. Marker assisted selection in crop plants. *Plant Cell, Tissue and Organ Culture* **82**, 317-342.

Gallais A. 1990. *Théorie de la sélection en amélioration des plantes*.

Gianola D, Campos G de los, Hill WG, Manfredi E, Fernando R. 2009. Additive Genetic Variability and the Bayesian Alphabet. *Genetics* **183**, 347-363.

Gibson PG, Reighard GL. 2002. Chilling Requirement and Postrest Heat Accumulation in Peach Trees Inoculated with Peach Latent Mosaic Viroid. *Journal of the American Society for Horticultural Science* **127**, 333-336.

Guitton B, Kelner J-J, Velasco R, Gardiner SE, Chagné D, Costes E. 2012. Genetic control of biennial bearing in apple. *Journal of Experimental Botany* **63**, 131-149.

Hanke MV, Flachowsky H, Peil A, Hättasch C. 2007. No Flower no Fruit - Genetic Potentials to Trigger Flowering in Fruit Trees. *Genes, Genomes, and Genomics* **1**, 1-20.

Harris SA, Robinson JP, Juniper BE. 2002. Genetic clues to the origin of the apple. *Trends in Genetics* **18**, 426-430.

Hazel LN. 1943. The Genetic Basis for Constructing Selection Index. *Genetics* **28**.

Heide OM, Prestrud AK. 2005. Low temperature, but not photoperiod, controls growth cessation and dormancy induction and release in apple and pear. *Tree Physiology* **25**, 109-114.

Hoblyn TN, Grubb NH, Painter AC, Wates BL. 1936. Studies in biennial bearing. *International Journal of Pomology and Horticultural Science* **14**: 39-76.

Hsu C-Y, Adams JP, Kim H, et al. 2011. FLOWERING LOCUS T duplication coordinates reproductive and vegetative growth in perennial poplar. *Proceedings of the National Academy of Sciences* **108**, 10756-10761.

Ingvarsson PK, Street NR. 2011. Association genetics of complex traits in plants. *New Phytologist* **189**, 909-922.

Jannink J-L, Lorenz AJ, Iwata H. 2010. Genomic selection in plant breeding: from theory to practice. *Briefings in Functional Genomics* **9**, 166-177.

Jiang C, Zeng Z-B. 1995. Multiple Trait Analysis of Genetic Mapping for Quantitative Trait Loci. *Genetics* **140**, 1111-1127.

Jiménez S, Reighard GL, Bielenberg DG. 2010. Gene expression of DAM5 and DAM6 is suppressed by chilling temperatures and inversely correlated with bud break rate. *Plant Molecular Biology* **73**, 157-167.

Jones K, Bound S, Koen T, Oakford M. 1992. Effect of timing of hand thinning on the cropping potential of Red Fuji apple trees. *Australian Journal of Experimental Agriculture* **32**, 417-420.

Jonkers H. 1979. Biennial bearing in apple and pear: a literature survey. *Scientia Horticulturae* **11**, 303-317.

Juniper B. 2007. The Mysterious Origin of the Sweet Apple. *American Scientist* **95**, 44.

Khan MA, Korban SS. 2012. Association mapping in forest trees and fruit crops. *Journal of Experimental Botany* **63**, 4045-4060.

King GJ, Maliepaard C, Lynn JR, et al. 2000. Quantitative genetic analysis and comparison of physical and sensory descriptors relating to fruit flesh firmness in apple (*Malus pumila* Mill.). *Theoretical and Applied Genetics* **100**, 1074-1084.

Knott SA, Haley CS. 2000. Multitrait Least Squares for Quantitative Trait Loci Detection. *Genetics* **156**, 899-911.

Kotoda N, Hayashi H, Suzuki M, *et al.* 2010. Molecular Characterization of FLOWERING LOCUS T-Like Genes of Apple (*Malus × domestica* Borkh.). *Plant and Cell Physiology* **51**, 561-575.

Kotoda N, Wada M. 2005. MdTFL1, a TFL1-like gene of apple, retards the transition from the vegetative to reproductive phase in transgenic Arabidopsis. *Plant Science* **168**, 95-104.

Kumar S, Bink MCAM, Volz RK, Bus VGM, Chagné D. 2012. Towards genomic selection in apple (*Malus x domestica* Borkh.) breeding programmes: Prospects, challenges and strategies. *Tree Genetics and Genomes* **8**, 1-14.

Labuschagne IF, Louw JH, Schmidt K, Sadie A. 2002. Genetic variation in chilling requirement in apple progeny. *Journal of the American Society for Horticultural Science* **127**, 663-672.

Lander ES, Botstein D. 1989. Mapping Mendelian Factors Underlying Quantitative Traits Using RFLP Linkage Maps. *Genetics* **121**, 185-199.

Lande R, Thompson R. 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* **124**, 743-756.

Lang GA, Early JD, Darnell RL. 1987. Endo-, Para-, and Ecodormancy: Physiological Terminology and Classification for Dormancy Research. *HortScience* **22**.

Lassois L, Denancé C, Ravon E, Guyader A, Guisnel R, Hibrand-Saint-Oyant L, Poncet C, Lasserre-Zuber P, Feugey L, Durel C-E. 2016. Genetic Diversity, Population Structure, Parentage Analysis, and Construction of Core Collections in the French Apple Germplasm Based on SSR Markers. *Plant Molecular Biology Reporter*.

Laurens F, Audergon J-M, Claverie J, Duval H, Germain E, Kervella J, LeLezec M, Lauri P-E, Lespinasse J-M. 2000. Integration of architectural types in French programmes of ligneous fruit species genetic improvement. *Fruits* **55**, 141-152.

Laurens F, Durel CE, Patocchi A, Peil A, Salvi S, Tartarini S, Velasco R, van de Weg E. 2010. Review on apple genetics and breeding programmes and presentation of a new European initiative to increase fruit breeding efficiency. *Journal of Fruit Science* **27**, 102-107.

Lauri PÉ. 2002. From Tree Architecture to Tree Training - An Overview of Recent Concepts Developed in Apple in France. *Korean Journal of Horticultural Science and Technology* **43**, 782-788.

Lauri PE, Laurens F. 2005. Architectural types in apples (*Malus X Domestica* Borkh.). *Crops: growth, quality and biotechnology*. Helsinki, Finland: WFL Publisher, 1300-1313.

Lauri PE, Lespinasse JM. 2000. The Vertical Axes and Solaxe systems in France. *Acta Horticulturae* **513**, 287-296.

Lauri P-E, Terouanne E, Lespinasse JM. 1997. Relationship between the early development of apple fruiting branches and the regularity of bearing—An approach to the strategies of various cultivars. *Journal of Horticultural Science* **72**, 519-530.

Lauri PE, T  rouanne E, Lespinasse JM, Regnard JL, Kelner JJ. 1995. Genotypic differences in the axillary bud growth and fruiting pattern of apple fruiting branches over several years—an approach to regulation of fruit bearing. *Scientia Horticulturae* **64**, 265-281.

Lebon G, Wojnarowiez G, Holzapfel B, Fontaine F, Vaillant-Gaveau N, Cl  ment C. 2008. Sugars and flowering in the grapevine (*Vitis vinifera* L.). *Journal of Experimental Botany* **59**, 2565-2578.

Legave J-M. 2009. Comment faire face aux changements climatiques en arboriculture fruiti  re? *Innovations Agronomiques* **7**, 165-177.

Legave JM, Blanke M, Christen D, Giovannini D, Mathieu V, Oger R. 2013. A comprehensive overview of the spatial and temporal variability of apple bud dormancy release and blooming phenology in Western Europe. *International Journal of Biometeorology* **57**, 317-331.

Legave JM, Farrera I, Almeras T, Calleja M. 2008. Selecting models of apple flowering time and understanding how global warming has had an impact on this trait. *Journal of Horticultural Science & Biotechnology* **83**, 76-84.

Liebh  rd R, Kellerh  ls M, Pfammatter W, Jertmini M, Gessler C. 2003. Mapping quantitative physiological traits in apple (*Malus × domestica* Borkh.). *Plant Molecular Biology* **52**, 511-526.

Link H. 2000. Significance of flower and fruit thinning on fruit quality. *Plant Growth Regulation* **31**, 17-26.

Luan T, Woolliams JA,   deg  rd J, Dolezal M, Roman-Ponce SI, Bagnato A, Meuwissen THE. 2012. The importance of identity-by-state information for the accuracy of genomic selection. *Genetics Selection Evolution* **44**, 28.

Lynch M. 1988. Estimation of Relatedness by Dna Fingerprinting. *Molecular Biology and Evolution* **5**, 584-599.

Malagi G, Marcos RS, Citadin I, Flavio Gilberto H, Bonhomme M, Regnard JL, Legave JM. 2015. The comparison of dormancy dynamics in apple trees grown under temperate and mild winter climates imposes a renewal of classical approaches. *Trees*.

Meuwissen THE, Hayes BJ, Goddard ME. 2001. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics* **157**, 1819-1829.

- Mimida N, Komori S, Suzuki A, Wada M.** 2013. Functions of the apple TFL1/FT orthologs in phase transition. *Scientia Horticulturae* **156**, 106-112.
- Monselise SP, Goldschmidt EE.** 1982. Alternate bearing in fruit trees. *Horticultural Reviews* **4**, 128-173.
- Mouradov A, Cremer F, Coupland G.** 2002. Control of Flowering Time. *The Plant Cell* **14**, s111-s130.
- Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z, Costich DE, Buckler ES.** 2009. Association Mapping: Critical Considerations Shift from Genotyping to Experimental Design. *The Plant Cell* **21**, 2194-2202.
- Neilsen JC, Dennis FG.** 2000. Effects of seed number, fruit removal, bourse shoot length and crop density on flowering in 'Spencer Seedless' apple. *Acta Horticulturae* **527**, 137-146.
- Noiton DAM, Alspach PA.** 1996. Founding Clones, Inbreeding, Coancestry, and Status Number of Modern Apple Cultivars. *Journal of the American Society for Horticultural Science* **121**, 773-782.
- Noiton D, Shelbourne CJA.** 1992. Quantitative genetics in an apple breeding strategy. *Euphytica* **60**, 213-219.
- Pope KS, Da Silva D, Brown PH, DeJong TM.** 2014. A biologically based approach to modeling spring phenology in temperate deciduous trees. *Agricultural and Forest Meteorology* **198-199**, 15-23.
- Randoux M, Davière J-M, Jeauffre J, et al.** 2014. RoKSN, a floral repressor, forms protein complexes with RoFD and RoFT to regulate vegetative and reproductive development in rose. *New Phytologist* **202**, 161-173.
- Remay A, Lalanne D, Thouroude T, Couviour FL, Oyant LH-S, Foucher F.** 2009. A survey of flowering genes reveals the role of gibberellins in floral control in rose. *Theoretical and Applied Genetics* **119**, 767-781.
- Rolland F, Baena-Gonzalez E, Sheen J.** 2006. SUGAR SENSING AND SIGNALING IN PLANTS: Conserved and Novel Mechanisms. *Annual Review of Plant Biology* **57**, 675-709.
- Rosenstock TS, Rosa UA, Plant RE, Brown PH.** 2010. A reevaluation of alternate bearing in pistachio. *Scientia Horticulturae* **124**, 149-152.
- Saito T, Bai S, Ito A, Sakamoto D, Saito T, Ubi BE, Imai T, Moriguchi T.** 2013. Expression and genomic structure of the dormancy-associated MADS box genes MADS13 in Japanese pears (*Pyrus pyrifolia* Nakai) that differ in their chilling requirement for endodormancy release. *Tree Physiology* **33**, 654-667.
- Sargent DJ, Fernández-Fernández F, Ruiz-Roja JJ, Sutherland BG, Passey A, Whitehouse AB, Simpson DW.** 2009. A genetic linkage map of the cultivated strawberry (*Fragaria* ×

ananassa) and its comparison to the diploid *Fragaria* reference map. *Molecular Breeding* **24**, 293-303.

Sauphanor B, Dirwimmer C, Boutin S, et al. 2009. *Analyse Comparative de Différents Systèmes en Arboriculture Fruitière*. INRA.

Schoot C van der, Paul LK, Rinne PLH. 2013. The embryonic shoot: a lifeline through winter. *Journal of Experimental Botany* **65**, 1699-1712.

Segura V, Cilas C, Costes E. 2008. Dissecting apple tree architecture into genetic, ontogenetic and environmental effects: mixed linear modelling of repeated spatial and temporal measures. *New Phytologist* **178**, 302-314.

Seleznova AN, Tustin DS, Thorp TG. 2008. Apple Dwarfing Rootstocks and Interstocks Affect the Type of Growth Units Produced during the Annual Growth Cycle: Precocious Transition to Flowering Affects the Composition and Vigour of Annual Shoots. *Annals of Botany* **101**, 679-687.

Shoemaker JS, Painter IS, Weir BS. 1999. Bayesian statistics in genetics: a guide for the uninitiated. *Trends in Genetics* **15**, 354-358.

Staub JE, Serquen FC, Gupta M. 1996. Genetic Markers, Map Construction, and Their Application in Plant Breeding. *HortScience* **31**, 729-741.

Sukhwinder-Singh, Hernandez MV, Crossa J, Singh PK, Bains NS, Singh K, Sharma I. 2012. Multi-Trait and Multi-Environment QTL Analyses for Resistance to Wheat Diseases. *PLoS ONE* **7**, e38008.

Sung ZR, Chen L, Moon Y-H, Lertpiriyapong K. 2003. Mechanisms of floral repression in *Arabidopsis*. *Current Opinion in Plant Biology* **6**, 29-35.

Tabuenca MC. 1964. Necesidades de frío invernal de variedades de albaricoquero, melocotonero y peral. *Annals of Aula Dei* **7**, 113-132.

Thompson EA. 2013. Identity by Descent: Variation in Meiosis, Across Genomes, and in Populations. *Genetics* **194**, 301-326.

Tränkner C, Lehmann S, Hoenicka H, et al. 2010. Over-expression of an FT-homologous gene of apple induces early flowering in annual and perennial plants. *Planta* **232**, 1309-1324.

Trillot M, Masseron A, Mathieu V, Bergougnoux F, Hutin C, Lespinasse Y. 2002. *Le pommier*. CTIFL.

Velasco R, Zharkikh A, Affourtit J, Dhingra A. 2010. The genome of the domesticated apple (*Malus x domestica* Borkh.). *Nature Genetics* **42**, 833-842.

Vilanova S, Sargent DJ, Arús P, Monfort A. 2008. Synteny conservation between two distantly-related Rosaceae genomes: *Prunus* (the stone fruits) and *Fragaria* (the strawberry). *BMC Plant Biology* **8**, 67.

Virlet N, Costes E, Martinez S, Kelner J-J, Regnard J-L. 2015. Multispectral airborne imagery in the field reveals genetic determinisms of morphological and transpiration traits of an apple tree hybrid population in response to water deficit. *Journal of Experimental Botany* **66**, 5453-5465.

Waldner W, Knoll M. 1998. The influence of fruit load on biennial bearing of Fuji apple. *Compact Fruit Tree* **31**, 1-11.

Westwood MN. 1993. *Temperate Zone Pomology: Physiology and culture*.

Wilcox J. 1944. Some factors affecting apple yields in the Okanagan Valley: tree size, tree vigor, biennial bearing, and distance of planting. *Scientific Agriculture* **25**: 189.

Würschum T. 2012. Mapping QTL for agronomic traits in breeding populations. *Theoretical and Applied Genetics* **125**, 201-210.

Yao J-L, Xu J, Cornille A, et al. 2015. A microRNA allele that emerged prior to apple domestication may underlie fruit size evolution. *The Plant Journal* **84**, 417-427.

Yi N, Shriner D. 2007. Advances in Bayesian multiple quantitative trait loci mapping in experimental crosses. *Heredity* **100**, 240-252.

Yu J, Holland JB, McMullen MD, Buckler ES. 2008. Genetic Design and Statistical Power of Nested Association Mapping in Maize. *Genetics* **178**, 539-551.

Zhao Y, Mette MF, Reif JC. 2015. Genomic selection in hybrid breeding. *Plant Breeding* **134**, 1-10.

Result Chapter 1

New insights into genetic determinisms of budbreak and flowering time in apple through a multi-family QTL discovery and identification of putative candidate genes linked to temperature requirements

Paper submitted to *Journal of Experimental Botany*, the 16th October 2015, accepted with revision- Revised version submitted the 12th of December 2015.

Investigated questions

Do budbreak and flowering dates represent the genetic variability on chill and heat requirement?

How are they impacted by geographical site and climate?

What candidate genes, possibly involved in chill and heat perception, colocalize with major QTLs?

New insights into genetic determinisms of budbreak and flowering time in apple through a multi-family QTL discovery and identification of putative candidate genes linked to temperature requirements

A. Allard^{1,2}, M.C.A.M. Bink³, S. Martinez¹, J.J. Kelner², J.M. Legave¹, M. di Guardo^{4,5}, E.A. Di Pierro⁶, F. Laurens⁷, W.E. van de Weg⁵, E. Costes¹

1-Institut National de la Recherche Agronomique (INRA), UMR 1334, AGAP CIRAD-INRA-Montpellier SupAgro, F-34398 Montpellier, France

2- Montpellier SupAgro, UMR 1334, AGAP CIRAD-INRA-Montpellier SupAgro, F-34398 Montpellier, France

3- Biometris, Wageningen University and Research centre, Droevendaalsesteeg 1, P.O. Box 16, 6700AA, Wageningen, The Netherlands.

4- Research and Innovation Centre, Fondazione Edmund Mach, San Michele all'Adige, Trento, Italy

5- Wageningen UR Plant Breeding, Wageningen University and Research Centre, Droevendaalsesteeg 1, P.O. Box 16, 6700AA, Wageningen, The Netherlands.

6 - Department of Biosciences, University of Milan, Via Celoria 26, 20133 Milan, Italy

7- INRA, UMR1345, Institut de Recherche en Horticulture et Semences IRHS, INRA, Agrocampus-Ouest, Université d'Angers, SFR 4207 QUASAV, F-49071 Beaucouzé, France

A. Allard : alix.allard@supagro.inra.fr

M.C.A.M Bink: marco.bink@wur.nl

S. Martinez: martines@supagro.inra.fr

J.J. Kelner : jean-jacques.kelner@supagro.inra.fr

J.M. Legave: legave@supagro.inra.fr

M. Di Guardo: mario.diguardo@fmach.it

E.A. Di Pierro: erica.dipierro@unimi.it

F. Laurens : francois.laurens@angers.inra.fr

W.E.van de Weg : eric.vandeweg@wur.nl

E. Costes, corresponding author : costes@supagro.inra.fr;

telephone: +33 (0)4 67 61 75 08; fax: +33 (0)4 67 61 55 96

Date of submission: 16/10/2015

Total word count: 10101

Figure number: 5

Color figures: 2 (Figures 1 and 5)

Table number: 4

Supplementary data: Figures: 7; Tables: 3

Highlight: QTLs and candidates genes for the regulation of budbreak and flowering time and new hypotheses on importance of chilling temperature perception in growth resumption at spring time.

Abstract

In temperate trees, growth resumption at spring time results from chilling and heat requirements and is an adaptive trait under global warming. This study aimed at deciphering the genetic determinism of budbreak and flowering time in apple using five related full-sib families. Both traits were observed over three years and two sites and expressed in calendar and degree-days. Mixed linear models were used to extract the best linear unbiased predictors (BLUPs) for the genotype factor and genotype x year interaction. QTL mapping for these BLUPs was done with an integrated genetic map containing 6849 SNPs, from which haplotypes were computed, and with a Bayesian approach using pedigree-based analysis. Four major genomic regions were identified on LG7, LG10, LG12 and LG9, the latter appearing to be the most stable across progenies, sites and years. Homologs of *AGL24* and *FT* were predicted under QTLs on LG9 and LG12, respectively, and *DORMANCY ASSOCIATED MADS box* genes were found under additional QTLs on LG8 and LG15. This study suggests that chilling and heat requirements share, in part, a common genetic determinism and that chilling perception mechanisms could be common to perennial and annual plants. It highlighted progenitors with favorable alleles for new breeding strategies for apple adaptation to climate change.

Keywords: budbreak, *DAM* gene, dormancy, *Malus x domestica* (Borkh), pedigree-based analysis, QTL

Introduction

Global warming has an impact on tree phenology, and rising temperatures in late autumn and winter can lead to problematic conditions regarding bud dormancy release. Dormancy is a mechanism developed by perennial plants growing in temperate climates to overcome periods of cold temperatures and to protect buds from frost. Dormancy has been described as being composed of three states: (i) paradormancy, which is regulated by physiological factors outside the bud, such as apical dominance; (ii) endodormancy, which is regulated by physiological factors inside the bud and during which the perception of chilling temperatures is active; and (iii) ecodormancy, which is regulated by environmental factors, especially temperature (Lang *et al.*, 1987). Dormancy release is mainly driven by exposure to chilling temperatures (Heide and Prestrud, 2005; Schoot *et al.*, 2013), whereas growth rate is driven by warm temperatures (Wigge, 2013). The chilling requirement (CR) corresponds to the amount of exposure time to cold temperatures required for dormancy release and is usually expressed in chilling hours (Anderson and Richardson, 1986). The heat requirement (HR) represents the amount of exposure time to warm temperatures required to reach a particular physiological stage, generally budbreak or flowering time. Both stages are considered to result from the cumulated time necessary to satisfy CR and HR.

CR and HR have been extensively studied in cultivated trees (Citadin *et al.*, 2001; Egea *et al.*, 2003; Ruiz *et al.*, 2007; Albuquerque *et al.*, 2008; Guo *et al.*, 2014). A lack of synchronization between flowering and temperature causes bud frost, extended flowering time, poor pollination and fruit setting (Atkinson *et al.*, 2013). In the Rosaceae family, it has been established that at the species level and in a fixed environment, CR has more influence on flowering date than HR for *Prunus* and apple (Egea *et al.*, 2003; Ruiz *et al.*, 2007; Albuquerque *et al.*, 2008; Fan *et al.*, 2010; Celton *et al.*, 2011). Moreover, some inter-dependency between CR and HR has been suggested: a lack of chilling causes an extended need for warm temperatures, probably due to a residual effect of dormancy, whereas extended exposure to chilling temperatures leads to a reduced HR (Ruiz *et al.*, 2007). Models with an overlap between CR and HR fulfilment have been proposed for phenological stages dates prediction in order to account for a possible influence of cold temperature after dormancy release (Pope *et al.*, 2014). For a given cultivar, the number of days required to fulfill CR and HR depend on the climatic region, in particular, in sub-

tropical and Mediterranean regions where dormancy tends to be released later (Legave *et al.*, 2012). The impact of increasing temperatures on phenology has been extensively studied in different climates: temperate (Howe *et al.*, 2000; Wolfe *et al.*, 2005; Legave *et al.*, 2009; Fujisawa and Kobayashi, 2010), Mediterranean and sub-tropical (Gordo and Sanz, 2009; Legave *et al.*, 2009; Grab and Craparo, 2011; Darbyshire *et al.*, 2013; Ghrab *et al.*, 2014). In the context of global warming, the trend tends towards an earlier growth resumption, which would be due to a reduced time for HR fulfilment (Wolfe *et al.*, 2005; Gordo and Sanz, 2009; Fujisawa and Kobayashi, 2010; Grab and Craparo, 2011; Legave *et al.*, 2012). In the coming decades, climate models predict a reduction in chill accumulation that could be an advantage for countries at high latitudes such as Norway (Campoy *et al.*, 2011), since it would still allow chill accumulation while avoiding spring frost (Luedeling, 2012). However, in warmer regions, a major loss of chill accumulation is expected (Campoy *et al.*, 2011).

The concept of common physiological mechanisms between flowering promoted by vernalization and dormancy has been reinforced since Chouard's work in 1960 (Horvath, 2009; Rios *et al.*, 2014). In fruit trees, the comprehension of dormancy molecular control has been improved by the study of an evergrowing mutant (*evg*) in peach in which six *Dormancy Associated MADs-box (DAM)* genes have been identified (Bielenberg *et al.*, 2008; Jimenez *et al.*, 2009). These *DAM* genes (especially *ppDAM5* and *ppDAM6*) have been shown to belong to the *SHORT VEGETATIVE PHASE (SVP)/AGAMOUS-LIKE 24 (AGL24)* gene family in *Arabidopsis thaliana* (Yamane *et al.*, 2011; Wells *et al.*, 2015), whose members control flowering and are up-regulated by vernalization (Michaels *et al.*, 2003). Their orthologs in peach and Japanese pear have been shown to be involved in dormancy establishment and release, regulation of chilling and heat perception, and could also be part of a down-regulation system for bud development (Jiménez *et al.*, 2010; Saito *et al.*, 2013). In apple, four *MdDAM* genes have been identified (Mimida *et al.*, 2015). *MdDAMa* (MDP00003222567) expression peaked at dormancy establishment, and gradually decreased under the influence of chilling temperatures until dormancy release (Falavigna *et al.*, 2013). Different chilling treatments have revealed other differentially expressed genes with a putative role in dormancy regulation, especially a FLOWERING LOCUS C-like gene (*FLC*-like) located at the top of LG9 (Porto *et al.*, 2015). Other regulatory genes belong to the *MADS-box* such as *FLOWERING LOCUS T (FT)* homologs would be involved in the transition between reproductive and vegetative growth, growth

cessation in response to short photoperiod and winter temperatures, dormancy requirement, and promotion of early flowering in trees (Böhlenius *et al.*, 2006; Kotoda *et al.*, 2010; Tränkner *et al.*, 2010; Hsu *et al.*, 2011; Srinivasan *et al.*, 2012). Since budbreak and flowering time have been shown to be heritable traits (Labuschagne *et al.*, 2002), genetic breeding could be a way to adapt fruit trees to a changing environment. Several QTLs have been detected with bi-parental families on *Prunus* and *Malus* species. On *Prunus*, a co-localization between flowering date and CR QTL has confirmed the hypothesis of a common genetic determinism and the high complexity of their genetic control (Dirlewanger *et al.*, 2012). In apple, the top of linkage group 9 (LG9) has been identified as an important QTL region in two progenies (Dyk *et al.*, 2010; Celton *et al.*, 2011), and complex genetic control has been revealed with several QTLs on different LG, depending on the progeny (Liebhard *et al.*, 2003; Segura *et al.*, 2007; Celton *et al.*, 2011). However, single bi-parental analyses did not provide results transferable to other progenies, except for LG9 in apple. This problem that can be addressed by QTL mapping on multiparental populations which extend the genetic diversity and allow the comparison of allele performance in different genetic backgrounds (Pauly *et al.*, 2012). This strategy has been shown to increase QTL detection power, accuracy of QTL positions, and robustness of QTL effects estimation (Bink *et al.*, 2002; Blanc *et al.*, 2006; Liu *et al.*, 2012). However, false positive QTLs can be detected with such populations due to relationships between individuals. Pedigree-based analysis (PBA) overcame this issue by taking into account relationships between individuals using the concept of Identity By Descent (IBD), which combines information from both pedigree and markers (Luan *et al.*, 2012). In addition, considering marker data for common ancestors makes it possible to trace the source of favorable alleles.

In this study, our first aim was to map QTLs linked to bud phenology and to distinguish those linked to CR or HR through a multi-family and pedigree-based analysis. A second aim was to identify underlying candidate genes, paying special attention to related *DAM* genes. The third aim of the study was to identify founders, parents or individuals with interesting genotypes and to trace their transmission along the pedigree. For these purposes, several variables that represented variation due to CR or HR or both were studied. Five families and their genetic relationships were considered in a pedigree-based analysis. This study revealed four important loci on LG7, LG9, LG10 and LG12, and two minor loci on LG8 and LG15. For the first time in apple, *Dormancy Associated MADS-box* genes, homologs to *AGL24*, *FT* and *FLC* in *Arabidopsis*

thaliana, were mapped under four QTLs. Moreover, progenitors with favorable alleles were identified that could open new perspectives for breeding.

Materials and Methods

Plant Material

Five related full-sib families were considered. Two families were grown at INRA's Diascope experimental unit in Montpellier (coordinates: 43°36'35''N; 3°58'50''E) and three at INRA's Angers experimental station (coordinates: 47°29'7.656 N; 0°36'47.646). The Montpellier families were derived from a cross between a color mutant of 'Delicious', 'Starkrimson' and 'Granny Smith', and between X-3263 and 'Belrene' (Fig. 1). They were designated SG and XB, and were composed of 115 and 58 progenies, respectively. SG progenies were repeated twice. For both families, trees were grafted on Pajam I apple rootstock; SG was planted in 2007 and XB in 2005. Trees were grown with minimal training and pruning. The Angers families were derived from crosses between X-3263 and X-3259, X-3259 and X-3305, and X-3305 and 'Rubinette', and were designated HIVW, N, and P, respectively (Fig. 1). They were composed of 171, 42 and 45 individuals, respectively, each with a single replicate per individual. The trees were trained in vertical axis with an annual manual thinning with 1 fruit per inflorescence. At both sites, pest and disease management was performed consistently with professional practices.

The families were interconnected through a complex pedigree (Fig. 1). X-3263 is the parent of both XB and HIVW, X-3259 is the parent of both N and HIVW, X-3305 is the parent of both N and P, and 'Chantecler' is the mother of both X-3305 and X-3259. Moreover, 'Golden Delicious' is the founder of several parents, 'Chantecler', 'PRI672-3', 'PRI14-126' and 'Rubinette'. 'Granny Smith' is the parent of 'Baujade' and the SG family. Compared to the low chilling cultivar 'Anna' and the high chilling cultivar 'Golden Delicious' (requiring 300 and 1000 chilling hours, respectively) (Labuschagne *et al.*, 2002), 'Granny Smith' and 'Delicious' are considered as relatively low chilling cultivars since they require 600 and 700 chilling hours, respectively (<http://www.orangeppin.com>). Parents X-3305, X-3259 and X-3263 were selected by INRA for their high fruit quality and scab resistance.

Phenotypic assessment

Two phenological stages were observed: budbreak (BB), and beginning of flowering (BF) (Fleckinger, 1964). Each stage was evaluated at the tree scale, and the stage was considered to be reached when 50% of the buds reached it. To avoid missing a stage, the assessment was made three times a week from mid-March to May. Progenies were phenotyped in three consecutive years, 2012, 2013 and 2014.

Year characterization

For each year and site, dormancy release date and flowering date were predicted, for the reference high CR cultivar ‘Golden Delicious’, for which CR and HR are known, using the model of Legave *et al.* (2012). This model uses a sequential approach where (i) dormancy release is predicted when CR is fulfilled, (ii) heat perception starts once CR is fulfilled, and (iii) flowering date is predicted when HR is fulfilled. The input data of this model are the mean daily temperatures of the corresponding sites and years.

Phenological stage modeling

Since BB and BF dates result from fulfillment of both CR and HR, we intended to differentiate their effect by using two different units: calendar days (CD) and Growing Degree Hours (GDH). Budbreak in calendar days (BB_CD) was calculated from the first of January, since most models consider a fixed date to start the calculation of chilling unit accumulation whatever the variety (Bidabé, 1967). This variable was considered as representative of variations in time to fulfill both CR and HR. Budbreak date in growing degree hours (BB_GDH) was calculated between the date of dormancy release estimated for Golden Delicious by the model of Legave *et al.* (2012) and the observed budbreak date for each genotype. Growing degree hours accounted for temperatures above a certain threshold and in a certain range, following the classical method used in apple to quantify the amount of warm temperature cumulated (Anderson and Richardson, 1986):

$$\text{If } T_H < T_U, GDH1 = \frac{F \cdot (T_B - T_U)}{2} * \left[1 + \cos \left(\pi + \frac{\pi \cdot (T_H - T_B)}{T_U - T_B} \right) \right]$$

$$\text{If } T_H > T_U, GDH2 = (T_B - T_U) * \left[1 + \cos \left(\frac{\pi}{2} + \frac{\pi}{\frac{2(T_H - T_U)}{T_C - T_U}} \right) \right]$$

where

GDH is the accumulation of growing degree hours during one hour

TH is the hourly temperature

TB is the base temperature (4°C)

TU is the optimum temperature (25°C)

TC is the critical temperature (36°C)

The threshold temperatures also correspond to values generally used for fruit trees, according to Anderson and Richardson (1986). Finally, the heat accumulation was also quantified between the dates of budbreak and the beginning of flowering (Delta_GDH; for abbreviations, see Table 1). BF, which was highly correlated with BB_CD, was not considered in later analyses.

Phenotype modeling

Linear models were used to assess the effect of factors, Year, Site, Genotype, and the genotype x year interaction for each trait, BB_CD, BB_GDH and Delta_GDH:

$$y = X_1\beta_1 + X_2\beta_2 + Z_1u_1 + Z_2u_2 + e$$

where y is the variable. X_1 is a matrix of dimension ‘number of individual trees’ x 2, associated with the site fixed effect. X_2 is a matrix of dimension ‘number of trees’ x 3, associated with the year fixed effect. β_1 and β_2 are vectors of length 2 and 3, respectively, associated with site and year fixed effects, respectively. Z_1 is a matrix of dimension ‘number of trees’ x ‘number of genotypes’, associated with the genotype random effect. Z_2 is a matrix of dimension ‘number of trees’ x (‘number of genotypes’*3), associated with the genotype x year interaction random effect. u_1 and u_2 are vectors of length, ‘number of genotypes’ and ‘number of genotypes’*3, respectively, associated with genotype and interaction year x genotype random effects, respectively. Finally, e is the vector of residual variance.

Preliminary analysis revealed that variances differed significantly between years, and correlations between years were positive. The heterogeneous variances per year were modeled in

the residual term and the interaction factor, and correlations between years were accounted for by fitting an unstructured variance-covariance matrix on the residual term in the linear mixed models. All these models were solved with ASReml-R software (Gilmour *et al.*, 1995).

The mixed model yielded the best unbiased linear predictors (BLUPs) for genotype and genotype x year interaction random factors for BB_CD, BB_GDH, and Delta_GDH. These BLUP values were used for QTL mapping. For each trait, a QTL analysis was performed for the genotypic BLUPs (G_j) and for the interaction BLUPs (G_j*Y_k), resulting in a total of 12 variables. The mixed models also yielded estimates for the variance components and the mean broad sense heritability was calculated as:

$$h^2 = \frac{\sigma_G^2}{\sigma_G^2 + \frac{\sigma_{GxY}^2}{k} + \frac{\sigma_\varepsilon^2}{n}}$$

where σ_G^2 is the variance of genotype effect, σ_{GxY}^2 is the variance of genotype x year interaction effect, σ_ε^2 is the variance of the residual term, k is the number of years, and n is the total number of observations.

It was not possible to assess the interaction between genotypes and sites because each family was present in one site only. Consequently, in order to assess the stability of QTLs across sites, each trait was analyzed with three datasets: the full set containing the two sites (multisite analysis); the two complementary subsets of two Montpellier families (Mtp analysis); and three Angers families (Ang analysis; for abbreviations, see Table 1).

SNP marker data

The five full-sib families and their progenitors (when available; see Fig. 1) were genotyped with the Infinium® 20K SNP array (Bianco *et al.*, 2014) at the Fondazione Edmund Mach according to the procedures described by Chagné *et al.* (2012) and Antanaviciute *et al.* (2012). A total of 6849 of the SNPs were used in this study, after (i) having passed the Excel-based forerunner of the ASSIsT pipeline (Di Guardo *et al.*, 2015), for 27 full-sib families, (ii) having shown robust performance over the pedigrees of these families (Van de Weg *et al.*, unpublished), and (iii) having shown robust performance on the current germplasm. For the latter, genotype score consistency was assessed between individuals and their progenies and parents. When errors

systematically arose for a marker, it was removed from the dataset. The recombination pattern was checked in order to identify spurious double recombination events and problematic markers were removed. The genetic positions of these SNP were taken from a pre-runner of the consensus genetic map based on 21 of the above 27 families. The marker order of this map was validated through co-segregation patterns across all 27 families and their pedigrees. (Van de Weg, personal communication). Next, sets of SNPs were integrated into haploblocks. The genetic map was then divided into successive 1 cM segments and the SNPs of each segment were assigned to a single haploblock. Haplotypes were composed using FlexQTLTM and PediHaplotyper software (Voorrips *et al.*, 2016). At the marker level, double recombination occurred in 5.2% of the LGs, whereas at the haplotype level, they occurred in 0.5% of the LGs. Since they were spread over the genome and progenies, they were not considered as problematic. Haploblocks along the genome were highly informative, considering the proportion of informative meioses (Fig. S1).

Bayesian QTL mapping methodology

The 12 variables, i.e., the genotypic BLUP and three genotype x year interaction BLUPs for BB_CD, BB_GDH, and Delta_GDH were analyzed using a linear model that comprised an intercept μ , the regressions on the QTL covariates $\mathbf{a}$, and a model residual $\mathbf{e}$, as:

$$\mathbf{y} = 1\mu + \mathbf{W}\mathbf{a} + \mathbf{e} \quad (1)$$

where $\mathbf{W}$ is the design matrix for the QTL effects. The Bayesian modeling including the number of QTLs is a random variable and, consequently, the number of columns in $\mathbf{W}$ is not fixed (Bink *et al.*, 2008). A bi-allelic model is assigned to a QTL with alleles denoted by Q and q , with only additive effects modeled, and the covariate values of $[QQ, Qq, qq]$ are equal to $[1, 0, -1]$. The frequency of the allele Q among founder individuals is denoted by f_a and the linkage map positions of the QTL are given by vector λ . The QTL genotypes of individuals are *a priori* unknown and modeling is based on the independent assignment of alleles to founders and segregation indicators to trace transmission from parents to offspring (Bink *et al.*, 2002; Thompson *et al.*, 2008). Our Bayesian modeling assigns uniform priors to the variables μ and λ , while assigning normal priors to the vectors $\mathbf{a}$ and $\mathbf{e}$ in (1), i.e., $\mathbf{a} \sim N(\mathbf{0}, I\sigma_a^2)$ and $\mathbf{e} \sim N(\mathbf{0}, I\sigma_e^2)$. The variables σ_a^2 and σ_e^2 are the per-QTL explained variance and the residual variance, with priors being inverse Gamma distributions (Bink *et al.*, 2008). The number of QTLs was assigned

a Poisson prior where different values, i.e., 5, 10, were used to assess the sensitivity of posterior inference to the prior assumptions. Only results for a prior mean of 5 are reported since the other values yield similar results and inferences.

Markov chain Monte Carlo (MCMC) simulation, as implemented in FlexQTL™ software (Bink *et al.*, 2014), was applied to obtain samples from the joint posterior distribution of the model parameters:

$$f(\mu, \mathbf{a}, \lambda, \sigma_a^2, \sigma_e^2 | \mathbf{y}) \quad (2)$$

The MCMC algorithms have been previously described (Bink *et al.*, 2008) and are omitted here. The Monte Carlo accuracy was monitored and the length of the simulation chains was required to be equivalent to at least 100 effective chain samples. The required lengths of the Markov chain simulations varied among traits but never exceeded 500,000 iterations. 1000 samples were stored for each simulation and are thus available for statistical inference. The inference on the number of QTLs was based on a pairwise comparison of models differing from each other by one QTL, by taking twice the natural log of Bayes factors (Kass and Raftery, 1995), denoted as 2*lnBF. Values for 2*lnBF that are greater than 2, 5 and 10 indicate positive, strong, and decisive evidence, respectively, favoring the largest QTL model. Similar to Bink *et al.* (2014), the inferences on QTL positions are based on posterior QTL intensities, and the inference on QTL contributions are based on the posterior mean estimates of the QTL effect sizes. Posterior probabilities of QTL genotypes were also estimated and the same thresholds as in Bink *et al.* (2014) was used.

Identification of genes underlying QTLs

For each QTL region, flanking markers were identified and localized on the first and the third version of the apple genome (kindly provided by the Fondazione Edmund Mach, now publicly available at <https://www.rosaceae.org/>) because gene prediction was performed on the first version only. The list of genes under the QTL interval was screened on the first version in order to identify genes related to flowering time, *Arabidopsis thaliana* flowering pathway, temperature response, dormancy and vernalization. Genes involved in the regulation of cell cycles or associated with plant hormones that have been previously identified in Celton *et al.* (2011) were

not considered in the present study. The presence of candidate genes in the same interval and the corresponding scaffold was checked on the third version.

Results

Year and site characterization

The predicted dormancy release dates and the observed flowering dates for ‘Golden Delicious’ in each year and at both sites gave insights into differences in climatic conditions (Fig. 2, and Figure S2 to S4 for daily temperatures). At both sites, flowering time happened later in 2013 and earlier in 2014 than in other years.. For Angers, the dormancy release date predicted for ‘Golden Delicious’ was almost stable, with a maximum of a three-day difference between years, whereas a larger variance was observed in the number of days to fulfill the HR, which varied between 77 and 95 days. Therefore, in Angers, CR was easily met and the variation in flowering date would mainly be due to the variation in time to fulfill HR. In Montpellier, the dormancy release date varied between years, up to 11 days between the earliest (2013) and the latest (2012) date. The number of days to fulfill HR also varied by 15 days between 2012 (lowest) and 2013 (highest value). Therefore, in Montpellier, the variation in flowering date was due to the variation in time to fulfill both CR and HR.

Phenotypes and heritabilities

Phenotype distributions differed between sites: the families from Montpellier (XB and SG) had a wider distribution than the families from Angers, where distributions of BB_GDH were particularly narrow for P, N and HIVW (Fig. 3). Heritabilities were 0.87, 0.38 and 0.41 for BB_CD, BB_GDH and Delta_GDH, respectively. Correlation tests revealed a significant and positive correlation between BB_CD and BB_GDH ($r^2=0.89$) (Fig. S2), while Delta_GDH was not correlated to other variables.

QTL discovery for genotypic BLUPs

Four regions considered as major were identified on LG7, LG9, LG10 and LG12 because of their high $2*\ln BF$ values, their stability across sites and variables (Fig. 4 and Table 2) and, in some cases, because of the presence of a candidate gene in the QTL interval. The most stable of these regions was located at the top of LG9 (Figs. 4A to 4I and Table 2). It gave a QTL signal for

BB_CD, BB_GDH and Delta_GDH, and for the multisite as well as for site-specific analyses, even though the signal was not significant in Angers. This QTL explained from 11.5 to 18.2% of the trait variance (Table 2). In this region, MDP0000126259, a homolog of *FLOWERING LOCUS C* (*FLC*-like), was predicted in both the first and third versions of the apple genome (Table 3). Moreover, MDP0000143531, a homolog of *AGL24* in *Arabidopsis thaliana*, was predicted below the lower end of the QTL.

The second region of major interest is on LG12 (Figs. 4A to 4F), where QTLs were found for BB_CD and BB_GDH, which explained from 16.7 to 27.6% of the trait variance (Table 2). According to the Bayes factor, this LG carried one QTL for BB_CD in Mtp analysis and BB_GDH for multisite and Mtp analyses (Table 2). However, the QTL signal is spread over a large zone covering three possible peaks, until 67 cM. The trace plot (Fig. 4, left panel) indicates a switching signal between the last two peaks. There was evidence for the presence of MDP0000132050, also designated as *MdTFL1*, just below the third peak of the QTL (Table 3). On LG7, there was a QTL mapped for BB_CD and BB_GDH (Figs. 4A to 4F) and which explains from 10.4 to 21.3% of the variance (Table 2). The QTL mapped on LG10 for BB_CD and BB_GDH (Figs. 4a to 4f), explained from 5.6 to 9.5% of the variance (Table 2). A weak but non-significant signal was found at that position for Delta_GDH (Figs. 4G to 4I).

Two additional regions were considered as important because *DAM* genes were predicted in the QTL interval. At the bottom of LG15, a QTL was detected for BB_CD and BB_GDH (Figs. 4 and Table 2). In the third version of the genome, MDP0000322567, designated as *MdDAMa*, and a homolog to *AGL24* in *Arabidopsis thaliana* was predicted close to the beginning of the QTL (Table 3). This prediction is in contradiction with the prediction of the first version of the genome in which this gene was predicted on LG16 (Table 3). On LG8, a QTL was detected for BB_CD for multisite and Mtp analyses (Figs. 4A, 4B and Table 2). In this region, MDP0000259294 and MDP0000527190, designated as *MdDAMc* and *MdDAMd*, respectively, were predicted in the first version of the genome (Table 3).

Other QTLs were mapped in this study but are not detailed because of low $2 \times \ln \text{BF}$ values or lack of stability across sites and families. In general, the QTLs mapped in the multisite analyses were also mapped in Mtp analyses, whereas there were fewer common QTLs with Angers analyses. Moreover, the $2 \times \ln \text{BF}$ values were lower in Angers than in other analyses (Fig. 4).

Year x Genotype Interaction-QTL mapping

Similarly to those observed for genotypic BLUPs, most QTLs for GxY interaction BLUPs co-localized between multisite (Table 4) and Mtp analyses (Table S1). Very few QTLs were detected for Ang analyses. QTLs mentioned in the previous section were also detected for interaction BLUPs in 2012 and 2014, especially on LG7, LG9, LG10, LG12 and LG15. However, new QTLs were detected on LG4 and LG11 for BB_CD in 2012 (Table 4). The QTL on LG4 was also detected for Delta_GDH in 2012. On LG8, a new QTL was detected for BB_CD in 2012. On LG15, three QTL regions were detected for Delta_GDH in 2012.

Estimated QTL genotypes

The estimated genotypes at each QTL can be used for three purposes: (i) to determine families that contributed to QTL discovery by identifying the heterozygous parents; (ii) to identify the haplotype alleles linked in coupling phase to the desired QTL alleles; and (iii) to evaluate the transmission of favorable alleles across the pedigree.

For BB_CD, SG and XB contributed the most to QTL detection. Indeed, the parents ‘Granny Smith’, ‘Delicious’ and ‘Belrene’ were estimated as heterozygous at most QTLs. The families HIVW, N and P also contributed to some QTL segregation since the parents X-3263 and X-3305 were estimated as heterozygous for QTLs on LG10 and LG15, respectively (Fig. 5A). However, the four progenitors of Angers families, X-3305, X-3263, X-3259 and ‘Rubinette’, were estimated as homozygous for most QTLs.

For Delta_GDH, HIVW, N and SG contributed the most to QTL mapping since X-3259 and ‘Granny Smith’ were estimated as heterozygous at most QTLs. XB also contributed since X3263 was estimated as heterozygous at some QTLs (Fig. 5B), whereas X-3305 and ‘Rubinette’ were estimated as homozygous for all QTLs.

If we consider that favorable alleles are those that confer low values to BB_CD, most parents and founders were homozygous at the QTL on LG9 with the favorable allele, except ‘Granny Smith’ and ‘Belrene’. On LG15, parents or founders exhibiting favorable alleles were X-3305, ‘Granny Smith’, ‘Delicious’ and ‘Belrene’. X-3305 probably inherited the favorable allele from its paternal grandfather ‘Granny Smith’, while for ‘Delicious’, it was probably inherited from its

parent ‘Winesap’ (Figs. S6 and S7). Haplotype information is available in the Table S3 in the ‘Annexe’ Section of this documents (p.190). On LG7, X-3259 was homozygous with the favorable allele, and ‘Delicious’ and ‘Belrene’ were heterozygous. On LG10, only ‘Red Winter’ was identified as the homozygous parent with the favorable allele among parents and founders. On LG12, the uncertainty of the QTL position did not allow the identification of a favorable individual. A similar approach can be taken to identify favorable alleles for Delta_GDH (data not shown).

Discussion

Genetic determinism of budbreak and flowering time and QTL interpretation regarding CR and HR

BB_CD was assumed to account for both HR and CR. In contrast, BB_GDH represented the variation of heat accumulation between the date of dormancy release estimated for ‘Golden Delicious’ and budbreak. Finally, Delta_GDH represented the heat accumulation between budbreak and flowering. Heritability of budbreak in calendar days (0.87) was much higher than in Growing Degree Hours (0.38) and between flowering and budbreak (0.41). This suggests that the genotypic variation in CR may have a greater effect on genotype budbreak time than HR, consistently with previous studies (Egea *et al.*, 2003; Ruiz *et al.*, 2007; Alburquerque *et al.*, 2008; Fan *et al.*, 2010; Celton *et al.*, 2011). The lower heritability of BB_GDH could also be due to a loss of variability in the computation. Indeed, the dormancy release date was estimated for ‘Golden Delicious’ for each site and year and did not account for a possible variability of dormancy release date among genotypes. For deciduous fruit trees, two methods have been employed to determine CR fulfillment. One is to expose cuttings harvested at different stages to a controlled warm condition for a period of time, with scoring of floral bud break (Gibson and Reighard, 2002). Another is to measure the weight of floral buds before and after exposition to warm condition for a period of time (Tabuenca, 1964). Alternatively, the estimation of dormancy release date from budbreak date could be performed by modelling but would require long series of observations to determine the CR and HR of a genotype (Legave *et al.*, 2012).

Budbreak date in either CD or GDH shared four major QTLs on LG7, LG9, LG10 and LG12. This is consistent with the positive correlation between BLUPs of BB_CD and BB_GDH and

suggests that the genetic control of these traits is partially shared. However, other chromosomal regions were detected, suggesting a multi-genic control.

Four detected regions could result from chilling perception since they colocalized with candidate genes linked to this mechanism. Under the major QTL at the top of LG9, MDP0000126259, an *FLC*-like gene, was predicted in the first version of the genome. This gene has been shown to be differentially expressed in a CR study, supporting a role in repression of bud growth during ecodormancy (Porto *et al.*, 2015). Close to the same region in both the first and third versions of the apple genome, MDP0000143531, which is a homolog to *AGL24*, a *MADS-box* transcription factor belonging to the same gene family as *DAM* genes, was also predicted (Yamane *et al.*, 2011; Wells *et al.*, 2015). Since *AGL24* is a floral promoter regulated by an *FLC*-independent vernalization pathway (Michaels *et al.*, 2003), this suggests complementary roles of these genes in chilling perception and endodormancy regulation. Two other QTLs, LG15 and LG8, co-localized with three of the four *DAM* genes annotated in the apple genome which constitute a cluster of genes unique to *Rosaceae* that play a role in the establishment of endodormancy and its maintenance (Mimida *et al.*, 2015). These *DAM* genes were predicted on LG8, *MdDAMc* and *MdDAMd*, and on LG15, *MdDAMa*. The prediction of *MdDAMa* on LG15 in the third version of the apple genome, instead of LG16 in the first version, is more consistent with both LG8 and LG15 homology (Velasco *et al.*, 2010) and the present QTL detection. Finally, at the bottom of LG12, an *FT* homolog was mapped, such as in Guitton *et al.* (2012). In perennial species such as *Populus*, *FT* homologs have been shown to be linked to growth cessation in response to winter temperature, photoperiod and dormancy establishment (Böhlenius *et al.*, 2006; Kotoda *et al.*, 2010; Tränkner *et al.*, 2010; Hsu *et al.*, 2011; Srinivasan *et al.*, 2012). This constitutes a new and complementary assumption for dormancy regulation in apple.

Regarding the perception of warm temperatures after budbreak, represented by Delta_GDH, a QTL was also detected at the top of LG9 despite a weak signal. Since growth resumption after CR fulfilment and under favorable temperatures results from the capacity of cells to divide and elongate (Rohde and Bhalerao, 2007), genes involved in the regulation of cell cycles or associated with plant hormones have been identified under QTLs associated with budbreak, particularly at the top of LG9 (Celton *et al.*, 2011). In this study, we focused on candidate genes

related to chill perception, but the high stability of the QTL at the top of LG9 across studies (Dyk *et al.*, 2010; Celton *et al.*, 2011) and the candidate genes predicted in the same region suggest that it could contain a cluster of key genes involved in chill perception and cell cycle reactivation. This is consistent with the conclusion of Wisniewski *et al.* (2015) that dormancy, growth regulation and budbreak are likely to be interdependent mechanisms.

Contribution of QTLs to specific phenological stages or climatic conditions

In this study, differences in temperatures between years and sites were taken into account in the computation of GDH traits and fixed effects in mixed linear models. Regions detected for genotype x year interaction BLUPs were generally similar to those detected for genotypic BLUPs, suggesting that specific temperature reinforced the effect of one or several QTLs. However, new regions were specifically detected in 2012 and 2014 on LG4, LG8, LG11 and LG15. Since 2012 was characterized by an extended time required to release dormancy for ‘Golden Delicious’ in Montpellier (Fig. 2), warm winter conditions could enhance the effect of known regions such as the top of LG9 and trigger other regions. Finally, as suggested by Celton *et al.* (2011), minor QTLs could contribute to additional genetic effects in specific years and climatic conditions for specific stages.

The wider variability of phenotypes in Montpellier compared to Angers can be due to differences between parents, operators or climatic conditions. Consequently, the interaction between site and genotype could not be estimated. Even though CR and HR were unknown for most parents, it is likely that they have different requirements. Also, despite the same protocol used between the two sites, phenotyping was carried out by different operators, possibly leading to specific effects that could not be distinguished from the site effect. Finally, at Mediterranean sites, dormancy release dates are more impacted by changing temperatures than in cooler climates (Legave *et al.*, 2012). However, the high CR cultivar ‘Golden Delicious’ was used in these estimations and lower CR cultivars could be less impacted by climate differences. Since most QTLs detected for multisite analysis were common with Mtp analysis, the climatic conditions in Montpellier could be more suitable to reveal differences in CR among progenies because of a larger variability in the time to CR fulfilment than in Angers.

Comparison of QTL analysis between single family and multi-family studies

Compared to previous studies (Dyk *et al.*, 2010; Celton *et al.*, 2011), more QTLs were detected, among which the QTL at the top of LG9 was common whereas QTLs on LG7, LG10 and LG12 were new revealed regions. This confirms the expected advantages of multi-population studies (Bink, 2002; Blanc *et al.*, 2006; Liu *et al.*, 2012), despite the small size of several families which was compensated by relatedness due to common parent. However, the percentage of variance explained by most QTLs was lower than in previous ones whereas the total variance explained by all the QTLs was higher for most traits. This could be due to genetic variability spread over more genomic regions. Finally, the size of QTL intervals was similar between studies, suggesting that the QTL position, e.g., on LG9, could be precisely estimated in single family studies, mainly due to the number of individuals considered (Collard *et al.*, 2005).

Contribution of families and breeding perspectives

The families that segregate at major QTLs for BB_CD and BB_GDH are SG, XB and HIVW, whereas SG, XB, N and P segregate for Delta_GDH. Breeding for low CR will certainly be more crucial for Mediterranean and sub-tropical than for temperate climates since low chill cultivars could be one solution to avoid dormancy breaking problems (Campoy *et al.*, 2011). Parents ‘Granny Smith’, ‘Belrène’, X-3305 and ‘Delicious’ displayed favorable alleles, therefore their descendants are expected to provide interesting progenitors possibly bearing and cumulating the favorable alleles. Since low HR may lead to early budbreak and bud damage due to spring frost in either warm or cold climates, the recommendation would therefore be to avoid low HR varieties whatever the climate. In addition, high HR could be difficult to fulfill in cold climates and should be avoided as well. In any case, HR could be used to adjust budbreak timing to other constraints such as the phenology of pollinator varieties or insects.

In spite of major QTLs detected for budbreak timing, we demonstrated the highly polygenic control of budbreak. In a breeding context, many small effect QTLs could contribute to an increase in the total variance explained and the prediction robustness. In this perspective, genome-wide selection models will be a complementary approach to QTL analysis in order to evaluate the genetic value of individuals by summing up allelic effects along the genome, provided that the marker density is high enough (Jannink *et al.*, 2010).

Supplementary Data

Figure S1: Proportion of informative meiosis for haploblock markers along the genome. On the X-axis, the beginning and the end of the chromosomes are represented by vertical dashed lines.

Figure S2: Daily temperature during chilling and heat fulfilment in winter 2011-2012 and spring 2012.

Figure S3: Daily temperature during chilling and heat fulfilment in winter 2012-2013 and spring 2013.

Figure S4: Daily temperature during chilling and heat fulfilment in winter 2013-2014 and spring 2014.

Figure S5: Correlation plot and r value between genotypic BLUP for BB_CD and BB_GDH.

Figure S6: Estimated genotypes and corresponding haplotypes for a QTL located on LG15 between 105 and 111 cM of the SG family and its progenitors. The name of the individual and the alleles for four successive haploblocks located at 107, 108, 109 and 110 cM, respectively, on LG15 are given in each box. The color of the boxes indicates the estimated genotypes at the QTL. Red and blue boxes correspond to homozygous individuals with alleles for high (QQ) and low (qq) genotypic values, respectively, green boxes correspond to heterozygous individuals, and gray boxes to individuals for which no distinction could be made between heterozygosity and homozygosity.

Figure S7: Estimated genotypes and corresponding haplotypes for a QTL located on LG15 between 105 and 111 cM of the P family and its progenitors. For the legend, see Fig. S3.

Table S1: QTL parameters for interaction BLUPs for the Montpellier analysis

Table S2: Haplotype composition for the QTL at the bottom of LG15

Table S3: Genes underlying major QTLs first and third versions of apple genomes

563

564 **Acknowledgments**

565 For early access to the integrated map, we acknowledge Herma Koehorst-van Putten, Yolanda
566 Noordijk, Sara Longhi, Johannes Kruisselbrink, Fabrizio Costa, Giulia Pagliarini, Thomas
567 Letschka, Lidia Lozano-Luis, Stefano Tartarini and Luca Gianfranceschi; for genotyping for the
568 20,000 SNP array, Michela Troggio; for field phenotyping, Sylvain Hanteville and Roland Robic;
569 for phenotype modeling, Marie Denis and David Cros; for access to the cc2 cluster and data
570 management, Bertrand Pittolat and Enrique Ortega-Abboud; for access to the third version of the
571 apple genome, Ricardo Velasco and Alessandro Cestaro.

572 This work was co-funded by the Seventh Framework Program of the EU under the
573 FruitBreedomics project N° 265582: Integrated Approach for increasing breeding efficiency in
574 fruit tree crops (www.FruitBreedomics.com). The views expressed in this work are the sole
575 responsibility of the authors and do not necessary reflect the views of the European Commission.

576

References

Alburquerque N, García-Montiel F, Carrillo A, Burgos L. 2008. Chilling and heat requirements of sweet cherry cultivars and the relationship between altitude and the probability of satisfying the chill requirements. *Environmental and Experimental Botany* **64**, 162-170.

Anderson JL, Richardson EA. 1986. Validation of chill unit and flower bud phenology models for 'Montmorency' sour cherry. *Acta Horticulturae* **184**.

Antanaviciute L, Fernández-Fernández F, Jansen J, Banchi E, Evans KM, Viola R, Velasco R, Dunwell JM, Troggio M, Sargent DJ. 2012. Development of a dense SNP-based linkage map of an apple rootstock progeny using the *Malus* Infinium whole genome genotyping array. *BMC Genomics* **13**.

Atkinson CJ, Brennan RM, Jones HG. 2013. Declining chilling and its impact on temperate perennial crops. *Environmental and Experimental Botany* **91**, 48-62.

Bianco L, Cestaro A, Sargent DJ, et al. 2014. Development and Validation of a 20K Single Nucleotide Polymorphism (SNP) Whole Genome Genotyping Array for Apple (*Malus × domestica* Borkh). *PLoS ONE* **9**, e110377.

Bidabé B. 1967. Action de la température sur l'évolution des bourgeons de pommier et comparaison de méthodes de contrôle de l'époque de floraison. *Annales de physiologie végétale* **9**, 65-68.

Bink MC a. M. 2002. On flexible finite polygenic models for multiple-trait evaluation. *Genetics Research* **80**, 245-256.

Bink MCAM, Boer MP, Braak CJE, Jansen J, Voorrips RE, Van de Weg E. 2008. Bayesian analysis of complex traits in pedigreed plant population. *Euphytica* **161**, 85-96.

Bink MCAM, Jansen J, Madduri M, et al. 2014. Bayesian QTL analyses using pedigreed families of an outcrossing species, with application to fruit firmness in apple. *Theoretical and Applied Genetics* **127**, 1073-1090.

Bink MCAM, Uimari P, Sillanpää MJ, Janss LLG, Jansen RC. 2002. Multiple QTL mapping in related plant populations via pedigree-analysis approach. *Theoretical and Applied Genetics* **104**, 751-762.

Blanc G, Charcosset A, Mangin B, Gallais A, Moreau L. 2006. Connected populations for detecting quantitative trait loci and testing for epistasis: an application in maize. *Theoretical and Applied Genetics* **113**, 206-224.

Böhlenius H, Huang T, Charbonnel-Campaa L, Brunner AM, Jansson S, Strauss SH, Nilsson O. 2006. CO/FT Regulatory Module Controls Timing of Flowering and Seasonal Growth Cessation in Trees. *Science* **312**, 1040-1043.

Campoy JA, Ruiz D, Egea J. 2011. Dormancy in temperate fruit trees in a global warming context: A review. *Scientia Horticulturae* **130**, 357-372.

Celton JM, Martinez S, Jammes MJ, Bechti A, Salvi S, Legave JM, Costes E. 2011*a*. Deciphering the genetic determinism of bud phenology in apple progenies: a new insight into chilling and heat requirement effects on flowering dates and positional candidate genes. *New Phytologist* **192**, 378-392.

Celton J-M, Martinez S, Jammes M-J, Bechti A, Salvi S, Legave J-M, Costes E. 2011*b*. Deciphering the genetic determinism of bud phenology in apple progenies: a new insight into chilling and heat requirement effects on flowering dates and positional candidate genes. *New Phytologist* **192**, 378-392.

Chagné D, Crowhurst RN, Troggio M, et al. 2012. Genome-Wide SNP Detection, Validation, and Development of an 8K SNP Array for Apple. *PLoS ONE* **7**, e31745.

Citadin I, Raseira MCB, Herter FG, da Silva JB. 2001. Heat requirement for blooming and leafing in peach. *Hortscience* **36**, 305-307.

Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK. 2005. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica* **142**, 169-196.

Darbyshire R, Webb L, Goodwin I, Barlow EWR. 2013. Impact of future warming on winter chilling in Australia. *International Journal of Biometeorology* **57**, 355-366.

Di Guardo M, Micheletti D, Bianco L, et al. 2015. ASSIsT: An Automatic SNP ScorIng Tool for in- and outbreeding species. *Bioinformatics*.

Dirlewanger E, Quero-García J, Le Dantec L, et al. 2012. Comparison of the genetic determinism of two key phenological traits, flowering and maturity dates, in three *Prunus* species: peach, apricot and sweet cherry. *Heredity* **109**, 280-292.

Dyk MM van, Soeker MK, Labuschagne IF, Rees DJG. 2010. Identification of a major QTL for time of initial vegetative budbreak in apple (*Malus x domestica* Borkh.). *Tree Genetics & Genomes* **6**, 489-502.

Egea J, Ortega E, Martínez-Gómez P, Dicenta F. 2003. Chilling and heat requirements of almond cultivars for flowering. *Environmental and Experimental Botany* **50**, 79-85.

Falavigna V da S, Porto DD, Buffon V, Margis-Pinheiro M, Pasquali G, Revers LF. 2013. Differential Transcriptional Profiles of Dormancy-Related Genes in Apple Buds. *Plant Molecular Biology Reporter* **32**, 796-813.

Fan S, Bielenberg DG, Zhebentyayeva TN, Reighard GL, Okie WR, Holland D, Abbott AG. 2010. Mapping quantitative trait loci associated with chilling requirement, heat requirement and bloom date in peach (*Prunus persica*). *New Phytologist* **185**, 917-930.

Fleckinger J. 1964. *Le bon jardinier*. Paris, France: Grisvard P. & Chaudun V.

Fujisawa M, Kobayashi K. 2010. Apple (*Malus pumila* var. *domestica*) phenology is advancing due to rising air temperature in northern Japan. *Global Change Biology* **16**, 2651-2660.

Ghrab M, Ben Mimoun M, Masmoudi MM, Ben Mechlia N. 2014. Chilling trends in a warm production area and their impact on flowering and fruiting of peach trees. *Scientia Horticulturae* **178**, 87-94.

Gibson PG, Reighard GL. 2002. Chilling Requirement and Postrest Heat Accumulation in Peach Trees Inoculated with Peach Latent Mosaic Viroid. *Journal of the American Society for Horticultural Science* **127**, 333-336.

Gilmour A, Thompson R, Cullis B. 1995. Average Information REML: an efficient Algorithm for Variance Parameter Estimation in Linear Mixed Models. *Biometrics* **51**, 1440-1450.

Gordo O, Sanz JJ. 2009. Long-term temporal changes of plant phenology in the Western Mediterranean. *Global Change Biology* **15**, 1930-1948.

Grab S, Craparo A. 2011. Advance of apple and pear tree full bloom dates in response to climate change in the southwestern Cape, South Africa: 1973-2009. *Agricultural and Forest Meteorology* **151**, 406-413.

Guitton B, Kelner J-J, Velasco R, Gardiner SE, Chagné D, Costes E. 2012. Genetic control of biennial bearing in apple. *Journal of Experimental Botany* **63**, 131-149.

Guo L, Dai J, Ranjitkar S, Yu H, Xu J, Luedeling E. 2014. Chilling and heat requirements for flowering in temperate fruit trees. *International Journal of Biometeorology* **58**, 1195-1206.

Heide OM, Prestrud AK. 2005. Low temperature, but not photoperiod, controls growth cessation and dormancy induction and release in apple and pear. *Tree Physiology* **25**, 109-114.

Howe GT, Saruul P, Davis J, Chen THH. 2000. Quantitative genetics of bud phenology, frost damage, and winter survival in an F2 family of hybrid poplars. *Theoretical and Applied Genetics* **101**, 632-642.

Hsu C-Y, Adams JP, Kim H, et al. 2011. FLOWERING LOCUS T duplication coordinates reproductive and vegetative growth in perennial poplar. *Proceedings of the National Academy of Sciences* **108**, 10756-10761.

Jannink J-L, Lorenz AJ, Iwata H. 2010. Genomic selection in plant breeding: from theory to practice. *Briefings in Functional Genomics* **9**, 166-177.

Jiménez S, Reighard GL, Bielenberg DG. 2010. Gene expression of DAM5 and DAM6 is suppressed by chilling temperatures and inversely correlated with bud break rate. *Plant Molecular Biology* **73**, 157-167.

Kass RE, Raftery AE. 1995. Bayes Factors. *Journal of the American Statistical Association* **90**, 773-795.

Kotoda N, Hayashi H, Suzuki M, *et al.* 2010. Molecular Characterization of FLOWERING LOCUS T-Like Genes of Apple (*Malus × domestica* Borkh.). *Plant and Cell Physiology* **51**, 561-575.

Labuschagne IF, Louw JH, Schmidt K, Sadie A. 2002. Genetic variation in chilling requirement in apple progeny. *Journal of the American Society for Horticultural Science* **127**, 663-672.

Lang GA, Early JD, Darnell RL. 1987. Endo-, Para-, and Ecodormancy: Physiological Terminology and Classification for Dormancy Research. *HortScience* **22**.

Legave JM, Blanke M, Christen D, Giovannini D, Mathieu V, Oger R. 2012. A comprehensive overview of the spatial and temporal variability of apple bud dormancy release and blooming phenology in Western Europe. *International Journal of Biometeorology* **57**, 317-331.

Legave JM, Christen D, Giovannini D, Oger R. 2009. Global Warming in Europe and Its Impacts on Floral Bud Phenology in Fruit Species. *Acta Horticulturae* **838**, 21-26.

Liebhards R, Kellerhals M, Pfammatter W, Jertmini M, Gessler C. 2003. Mapping quantitative physiological traits in apple (*Malus × domestica* Borkh.). *Plant Molecular Biology* **52**, 511-526.

Liu W, Reif JC, Ranc N, Porta GD, Würschum T. 2012. Comparison of biometrical approaches for QTL detection in multiple segregating families. *Theoretical and Applied Genetics* **125**, 987-998.

Luan T, Woolliams JA, Ødegård J, Dolezal M, Roman-Ponce SI, Bagnato A, Meuwissen THE. 2012. The importance of identity-by-state information for the accuracy of genomic selection. *Genetics Selection Evolution* **44**, 28.

Luedeling E. 2012. Climate change impacts on winter chill for temperate fruit and nut production: A review. *Scientia Horticulturae* **144**, 218-229.

Michaels SD, Ditta G, Gustafson-Brown C, Pelaz S, Yanofsky M, Amasino RM. 2003. AGL24 acts as a promoter of flowering in *Arabidopsis* and is positively regulated by vernalization. *The Plant Journal* **33**, 867-874.

Mimida N, Saito T, Moriguchi T, Suzuki A, Komori S, Wada M. 2015. Expression of DORMANCY-ASSOCIATED MADS-BOX (DAM)-like genes in apple. *Biologia Plantarum* **59**, 237-244.

Pope KS, Da Silva D, Brown PH, DeJong TM. 2014. A biologically based approach to modeling spring phenology in temperate deciduous trees. *Agricultural and Forest Meteorology* **198-199**, 15-23.

Porto DD, Bruneau M, Perini P, Anzanello R, Renou J-P, Santos HP dos, Fialho FB, Revers LF. 2015. Transcription profiling of the chilling requirement for bud break in apples: a putative role for FLC-like genes. *Journal of Experimental Botany* **66**, 2659-2672.

Rohde A, Bhalerao RP. 2007. Plant dormancy in the perennial context. *Trends in Plant Science* **12**, 217-223.

Ruiz D, Campoy JA, Egea J. 2007. Chilling and heat requirements of apricot cultivars for flowering. *Environmental and Experimental Botany* **61**, 254-263.

Saito T, Bai S, Ito A, Sakamoto D, Saito T, Ubi BE, Imai T, Moriguchi T. 2013. Expression and genomic structure of the dormancy-associated MADS box genes MADS13 in Japanese pears (*Pyrus pyrifolia* Nakai) that differ in their chilling requirement for endodormancy release. *Tree Physiology* **33**, 654-667.

Schoot C van der, Paul LK, Rinne PLH. 2013. The embryonic shoot: a lifeline through winter. *Journal of Experimental Botany* **65**, 1699-1712.

Segura V, Denancé C, Durel CE, Costes E. 2007. Wide range QTL analysis for complex architectural traits in a 1-year-old apple progeny. *Genome* **50**, 159-171.

Srinivasan C, Dardick C, Callahan A, Scorza R. 2012. Plum (*Prunus domestica*) Trees Transformed with Poplar FT1 Result in Altered Architecture, Dormancy Requirement, and Continuous Flowering. *PLoS ONE* **7**.

Tabuenca MC. 1964. Necesidades de frio invernal de variedades de albaricoquero, melocotonero y peral. *Annals of Aula Dei* **7**, 113-132.

Thompson E, Balding DJ, Bishop M, Cannings C. 2008. Linkage analysis. *Handbook of statistical genetics*. Chichester, 1141-1167.

Tränkner C, Lehmann S, Hoenicka H, *et al.* 2010. Over-expression of an FT-homologous gene of apple induces early flowering in annual and perennial plants. *Planta* **232**, 1309-1324.

Velasco R, Zharkikh A, Affourtit J, Dhingra A. 2010. The genome of the domesticated apple (*Malus x domestica* Borkh.). *Nature Genetics* **42**, 833-842.

Voorrips RE, Bink MCAM, Kruisselbrink JW, Koehorst-Van Putten, Van de Weg WE. 2016. PediHaplotyper: software for consistent assignment of SNP haplotypes in pedigrees. *Molecular Breeding*.

Wells CE, Vendramin E, Tarodo SJ, Verde I, Bielenberg DG. 2015. A genome-wide analysis of MADS-box genes in peach [*Prunus persica* (L.) Batsch]. *BMC Plant Biology* **15**, 41.

Wigge PA. 2013. Ambient temperature signalling in plants. *Current Opinion in Plant Biology* **16**, 661-666.

Wisniewski M, Norelli J, Artlip T. 2015. Overexpression of a peach CBF gene in apple: a model for understanding the integration of growth, dormancy, and cold hardiness in woody plants. *Plant Physiology* **6**, 85.

Wolfe DW, Schwartz MD, Lakso AN, Otsuki Y, Pool RM, Shaulis NJ. 2005. Climate change and shifts in spring phenology of three horticultural woody perennials in northeastern USA. *International Journal of Biometeorology* **49**, 303-309.

Yamane H, Ooka T, Jotatsu H, Hosaka Y, Sasaki R, Tao R. 2011. Expressional regulation of PpDAM5 and PpDAM6, peach (*Prunus persica*) dormancy-associated MADS-box genes, by low temperature and dormancy-breaking reagent treatment. *Journal of Experimental Botany* **62**, 3481-3488.

Tables and Figures

Table 1: Abbreviations used in this paper and their meaning

Abbreviation	Meaning
XB	Full-sib family derived from the cross between X-3263 and ‘Belrene’
SG	Full-sib family derived from the cross between ‘Starkrimson’ and ‘Granny Smith’
N	Full-sib family derived from the cross between X-3305 and X-3259
P	Full-sib family derived from the cross between ‘Rubinette’ and X-3305
HIVW	Full-sib family derived from the cross between X-3263 and X-3259
CR	Chilling Requirement
HR	Heat Requirement
BB	Budbreak date
BF	Beginning of flowering date
delta_GDH	Time between BF and BB expressed in GDH
CD	Calendar Days
GDH	Growing Degree Hours
BB_CD	Budbreak date expressed in CD
Multisite	Analysis with full dataset containing the five progenies
Mtp	Analysis with only progenies grown in Montpellier
Ang	Analysis with only progenies grown in Angers
BV	Breeding Value
GBV	Genome-wide Breeding Value

Table 2: Parameters associated with the genotypic BLUP QTLs. The first column indicates the type of analysis, the second one the trait concerned, and the following columns indicate the LG where the QTL is located, the 2*lnBF value at the LG scale for a 1 QTL over a 0 QTL model, the 2*lnBF value at the local scale, the position of the QTL region in cM, the additive QTL effect, the frequency of the positive allele, its variance and its percentage of variance explained. The 2*lnBF values for multi-QTL models were not presented because none of them passed the significance threshold. The QTLs with a 2*lnBF at the LG scale higher than 5 are in bold type.

Analysis	trait	LG	2*lnBF_LG	max 2*lnBF_loc	Pos (cM)	add	fq	var	%var
Multisite	BB_CD	3	4.8	7.0	1-21	0.77	0.40	0.28	4.6
		7	9.4	7.5	54-88	1.17	0.51	0.69	11.1
		8	3.8	4.8	11-47	0.98	0.71	0.39	6.3
		9	29.9	8.8	0-11	1.76	0.21	1.03	16.6
		10	10.9	8.6	22-35	0.92	0.65	0.35	5.6
		12	29.9	8.6	37-43	2.05	0.80	1.32	21.3
		15	9.1	9.5	105-111	0.99	0.70	0.42	6.7
	BB_GDH	7	29.4	6.3	56-84	189.8	0.24	13070	21.3
		9	30.1	8.8	4-10	193.0	0.82	11139	18.2
		10	9.4	9.2	28-34	106.4	0.53	5640	9.2
		12	29.6	7.3	19-43	182.0	0.67	14726	24.0
	delta_GDH	8	3.5	5.5	1-9	59.4	0.64	1628	5.2
		9	4.4	6.9	4-14	98.9	0.24	3572	11.5
Mtp	BB_CD	7	4.0	5.9	56-84	1.45	0.45	1.0	10.4
		8	5.1	5.0	9-31	1.15	0.51	0.7	6.6
		9	4.8	7.0	2-14	1.63	0.51	1.3	13.3
		10	3.0	5.1	26-40	1.14	0.60	0.6	6.2
		12	6.1	6.6	19-67	1.88	0.62	1.7	16.7
	BB_GDH	7	7.2	6.8	54-86	153.3	0.47	11723	12.0
		9	10.2	8.6	2-12	183.7	0.49	16858	17.2
		10	3.2	6.3	26-40	108.9	0.59	5722	5.8
		12	10.4	6.7	19-43	189.0	0.56	17609	18.0
	delta_GDH	9	3.6	5.5	2-14	107.4	0.40	5538	15.9
		14	3.4	4.9	21-37	89.7	0.46	3993	11.5
An	BB_CD	6	4.5	8.0	1-15	0.55	0.59	0.1	6.6
		10	4.3	6.4	16-34	0.65	0.50	0.2	9.5
		15	4.0	4.4	77-109	0.70	0.65	0.2	10.0
	BB_GDH	no QTL							
	delta_GDH	2	2.5	4.0	13-35	60.9	0.52	1851	6.9
		3	2.7	5.2	59-79	58.8	0.48	1724	6.4

Result Chapter 1

Table 3: Candidate gene information underlying QTLs. The first five columns contain information about the QTL: LG, position in cM, flanking markers and their physical position on the first and third versions of the apple genome. Columns 6 to 9 contain information about candidate genes underlying the QTLs, or that are located nearby: their name and annotation on the apple genome and/or their homolog in *Arabidopsis thaliana*, their position on the first version of the genome, the related literature reference, and their position on the third version when available. The information related to the first version of the genome is available at the website <http://www.rosaceae.org/>, whereas that of the third version was provided by the Fondazione Edmund Mach. Physical positions are in Mega base pair (Mbp). All genes underlying major QTLs are available in supplementary Table S3.

LG	QTL position (cM)	QTL flanking markers	Flanking marker position		Candidate gene	Candidate gene position	Reference	Candidate gene position Apple genome v3
			Apple genome v1	Apple genome v3		Apple genome v1		
9	Start: 0	SNP_FB_0771340	Chr9:0.8	Chr9: 1.50	MDP0000143531 (<i>AGL24</i>)	Start: chr9: 4.52	Porto <i>et al.</i> (2015)	MDC021673.64
	End: 10	SNP_FB_0774491	Chr9:2.4	Chr9: 3.07		End: chr9: 4.54		Scaffold300219 Start: chr9:5.21 End: chr9:5.22
12	Start: 43	SNP_FB_0141443	Chr12:23.5	Chr12: 28.2	MDP0000126259 (<i>FLC-like</i>)	Start: chr9: 0.695 End: chr9: 0.697	Guitton <i>et al.</i> (2012)	MDC021142.191
	End: 67	SNP_FB_0155354	Chr12:31.2	Chr12: 37.4	MDP0000132050 (<i>MdFT1</i>)	Start: chr12: 31.58 End: chr12: 31.58		Scaffold300114 Start: chr12: 37.86 End: chr12: 37.87
15	Start: 105	SNP_FB_0914634	Chr15: 39.1	Chr15: 48.5	MDP0000322567 (<i>MdDAMa</i> , <i>AGL24</i>)	MDC020688.360 Start: chr16: 17.76 End: chr16: 17.77	Mimida <i>et al.</i> (2015) Saito <i>et al.</i> (2013)	MDC020688.360
	End: 111	SNP_FB_0327408	Chr15: 43.0	Chr15:53.1				Scaffold317868 Start: chr15: 46.66 End: chr15: 46.67
8	Start: 11	R_8450890_L8_36_1	Chr8: 7.50	Chr8: 8.2	MDP0000527190 (<i>MdDAMd</i>)	MCD008471.150 Start: chr8: 23.93 End: chr8: 23.95	Mimida <i>et al.</i> (2015) Saito <i>et al.</i> (2013)	
	End: 47	* SNP_FB_0759171	Chr8: 23.30	Chr8: 27.0		MDC020948.189 Start: chr8: 23.86 End: chr8: 23.87		
					MDP0000259294 (<i>MdDAMc</i>)		Mimida <i>et al.</i> (2015) Saito <i>et al.</i> (2013)	

Result Chapter 1

Table 4: Parameters associated with the interaction BLUP QTLs. The first column indicates the type of analysis, the second one the trait concerned, and the third one the interaction BLUPs concerned. The following columns have the same meaning as in Table 2.

* 2*lnBF values for the comparison of a model with three QTLs to a model with two QTLs

Analysis	Trait	BLUP	LG	2lnBF_LG	Max 2lnBF_loc	pos	add	fq	var	%var
Multisite	BB_CD	Int 12	4	10.1	7.4	27-37	0.45	0.46	0.101	14.5
			8	13.8	9.1	47-55	0.44	0.53	0.096	13.7
			11	5.4	8.1	7-17	0.45	0.41	0.096	13.7
			15	5.5	8.1	101-115	0.32	0.60	0.049	7.0
		Int 14	4	2.9	4.6	9-25	0.18	0.34	0.015	5.4
			7	2.1	3.6	18-54	0.15	0.54	0.012	4.2
			9	3.9	7.5	2-10	0.22	0.41	0.024	8.7
	BB_GDH	Int 14	3	3.9	6.2	67-79	27.6	0.22	266	5.8
			5	2.7	5.9	14-24	29.8	0.33	393	8.6
			7	12.7	8.3	54-60	38.6	0.27	585	12.7
			9	11.2	8.0	2-10	47.1	0.34	996	21.6
			10	3.8	6.5	26-36	27.1	0.40	351	7.6
			12	13.6	6.2	17-29	27.2	0.34	333	7.2
		Int 12	2	3.8	5.9	21-31	332.7	0.88	23653	7.0
			4	29.9	7.0	13-29	328.5	0.44	53264	15.8
			15	10.6*	14.1	1-3	684.8	0.10	83240	24.6
			15	10.6*	7.0	9-25	319.5	0.56	50297	14.9
			15	10.6*	8.4	59-67	383.2	0.89	29065	8.6
			17	10.2	9.0	37-43	309.8	0.09	15541	4.6
	delta_GDH	Int 14	8	6.8	6.3	9-31	92.7	0.61	4096	14.0
			12	5	5.1	51-67	86.7	0.57	3684	12.6
			15	3.2	6.8	7-17	100.2	0.67	4426	15.1
		Int 12	4	3.8	5.0	29-59	0.41	0.41	0.08	26.1
			7	3.7	4.5	54-86	0.22	0.59	0.022	7.2
An	BB_CD	Int 13	no QTL							
		Int 14	14	3.9	6.5	1-21	0.33	0.56	0.054	18.8
		Int 12	no QTL							
	BB_GDH	Int 13	no QTL							
		Int 14	no QTL							
		Int 12	no QTL							
	delta_GDH	Int 13	3	6.9	6.3	65-87	126.5	0.41	7730	7.7
		Int 14	no QTL							
		Int 12	no QTL							
		Int 13	no QTL							

Result Chapter 1

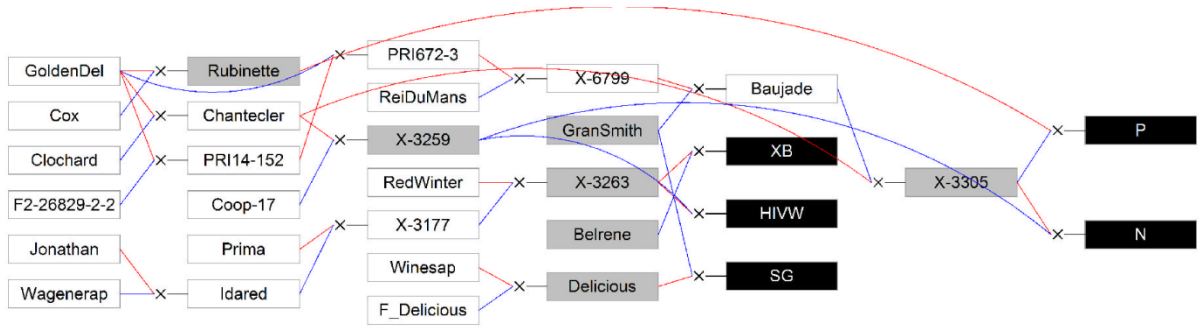


Figure 1: Genetic relationships between the five full-sib families. The full-sib families are represented by black boxes; the parents by gray boxes; the founders and other members of the pedigree by white boxes. Blue and red lines link the father and mother, respectively, to its progeny. GoldenDel = Golden Delicious, ReiDuMans = Reinette du Mans, Wagenerap = Wagenerapfel; for XB, HIVW, SG, N and P, see Table 1 for abbreviations and their meaning.

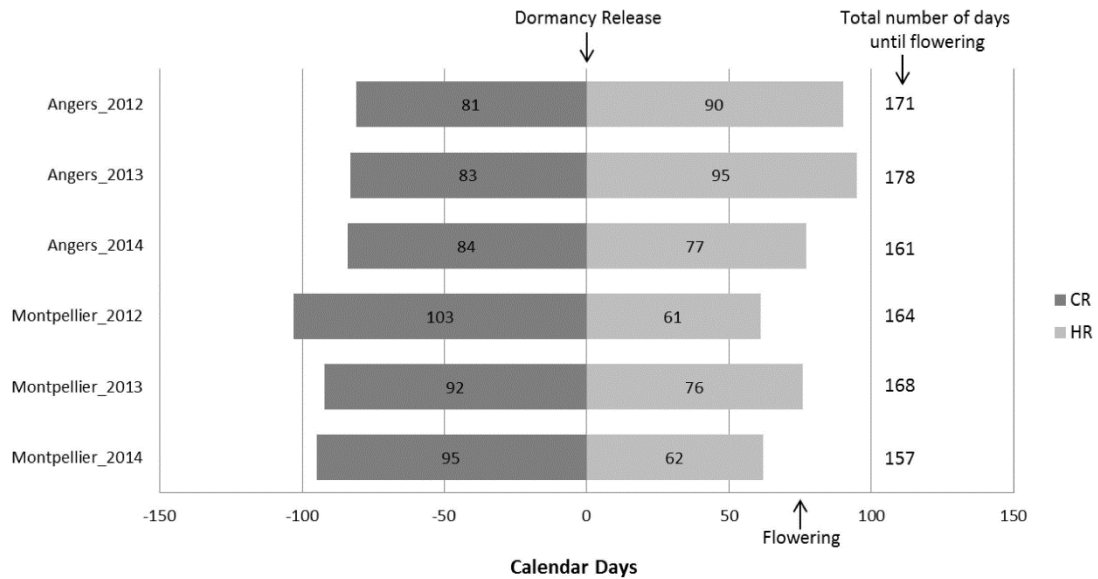


Figure 2: Characterization of climatic years (2012, 2013, and 2014) and sites (Angers and Montpellier) by observed flowering dates and estimation of dormancy release date of cultivar 'Golden Delicious' with the model of Legave *et al.* (2012). The dark gray blocks represent the estimated number of days before dormancy release, and the light gray blocks the number of days between dormancy release and observed flowering dates.

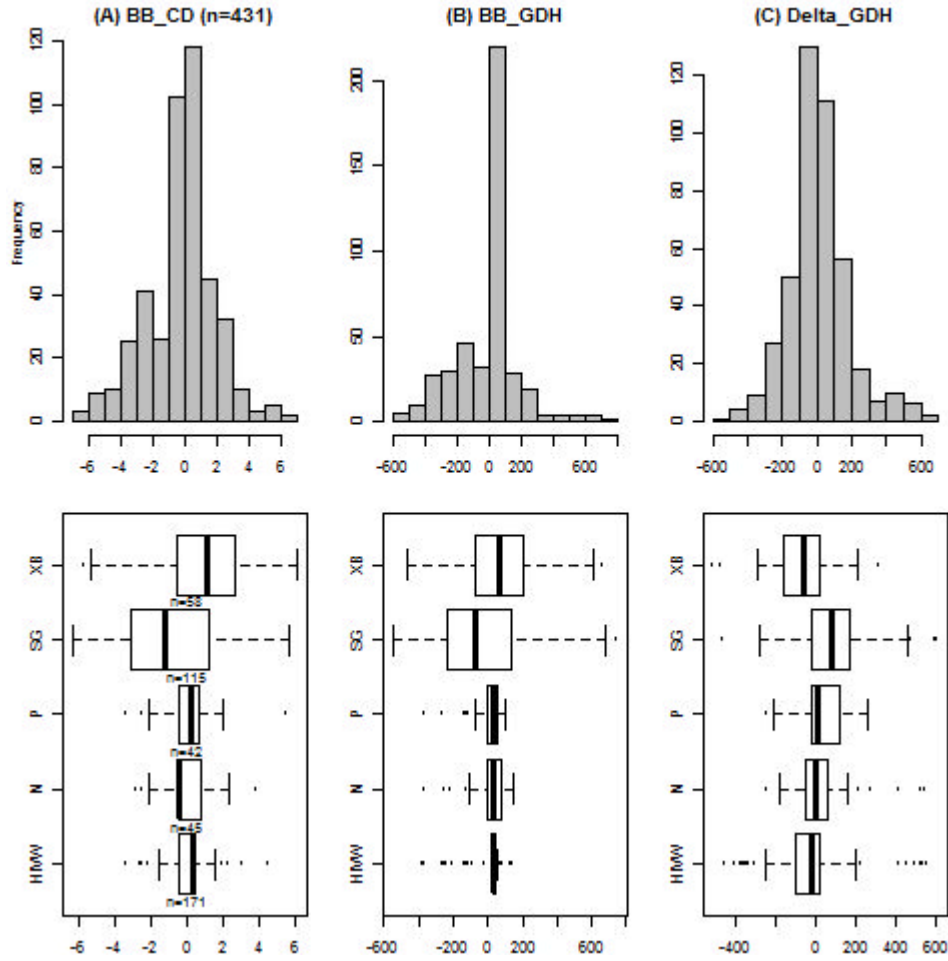


Figure 3: Upper part: histograms of genotypic Best Linear Unbiased Predictor (BLUP) for BB_CD (A), in BB_GDH (B) and Delta_GDH (C). Lower part: boxplots for each full-sib family (XB, SG, P, N, HIVW) and for BB_CD (left), BB_GDH (middle) and Delta_GDH (right). For each boxplot, the bold line represents the median, the extremities of the box represent the first and third quartile, from left to right, respectively, and the whiskers represent extreme values. See Table 1 for family abbreviations.

Result Chapter 1

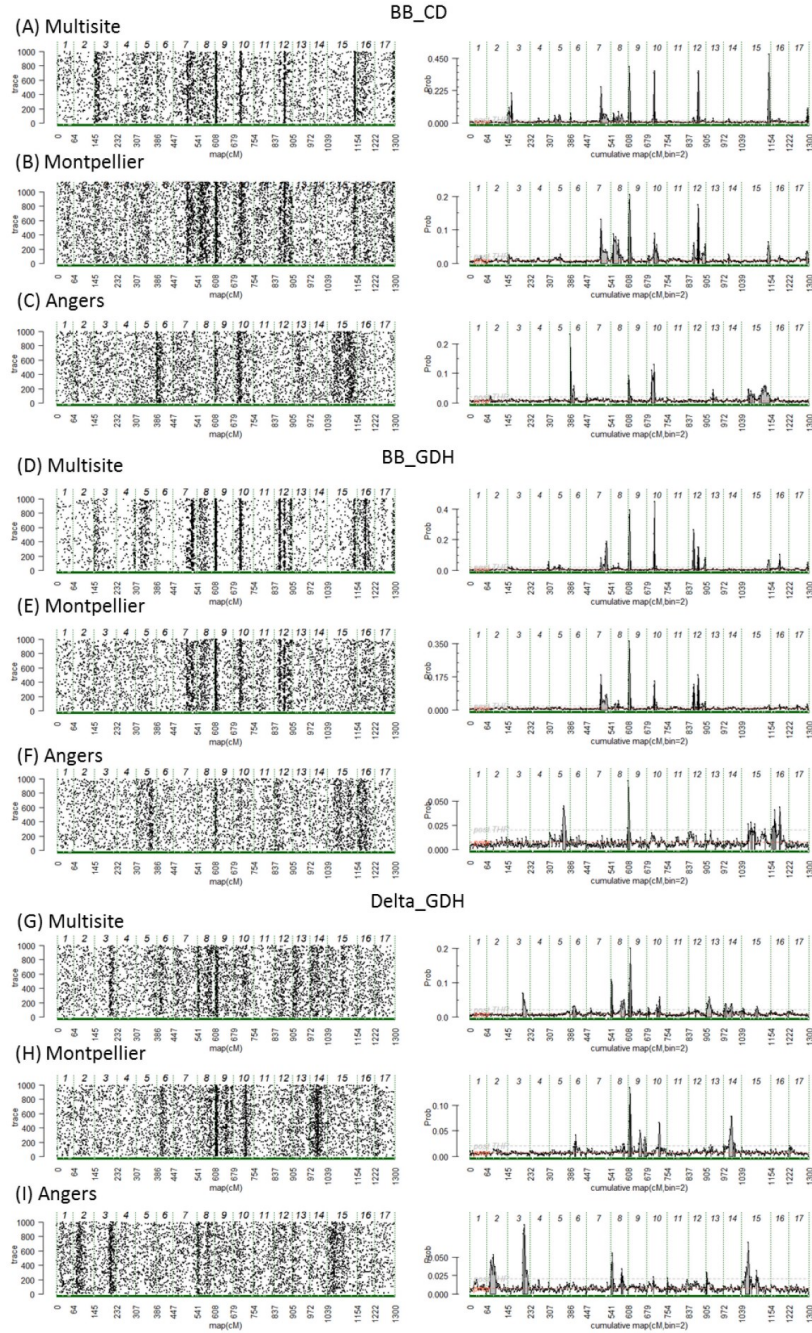


Figure 4: Trace plots (left) and posterior probability of QTL positions (right) along the genome for genotypic BLUPs, for trait BB_CD in the (a) multisite, (b) Montpellier, and (c) Angers analyses, for trait BB_GDH in the (d) multisite, (e) Montpellier, and (f) Angers analyses, and for trait Delta_GDH in the (g) multisite, (h) Montpellier, and (i) Angers analyses. The beginning and the end of the chromosomes are represented by vertical dashed lines. The solid gray areas on the right side correspond to regions with positive evidence ($2 \cdot \ln BF_{10} > 2$) for the presence of QTLs.

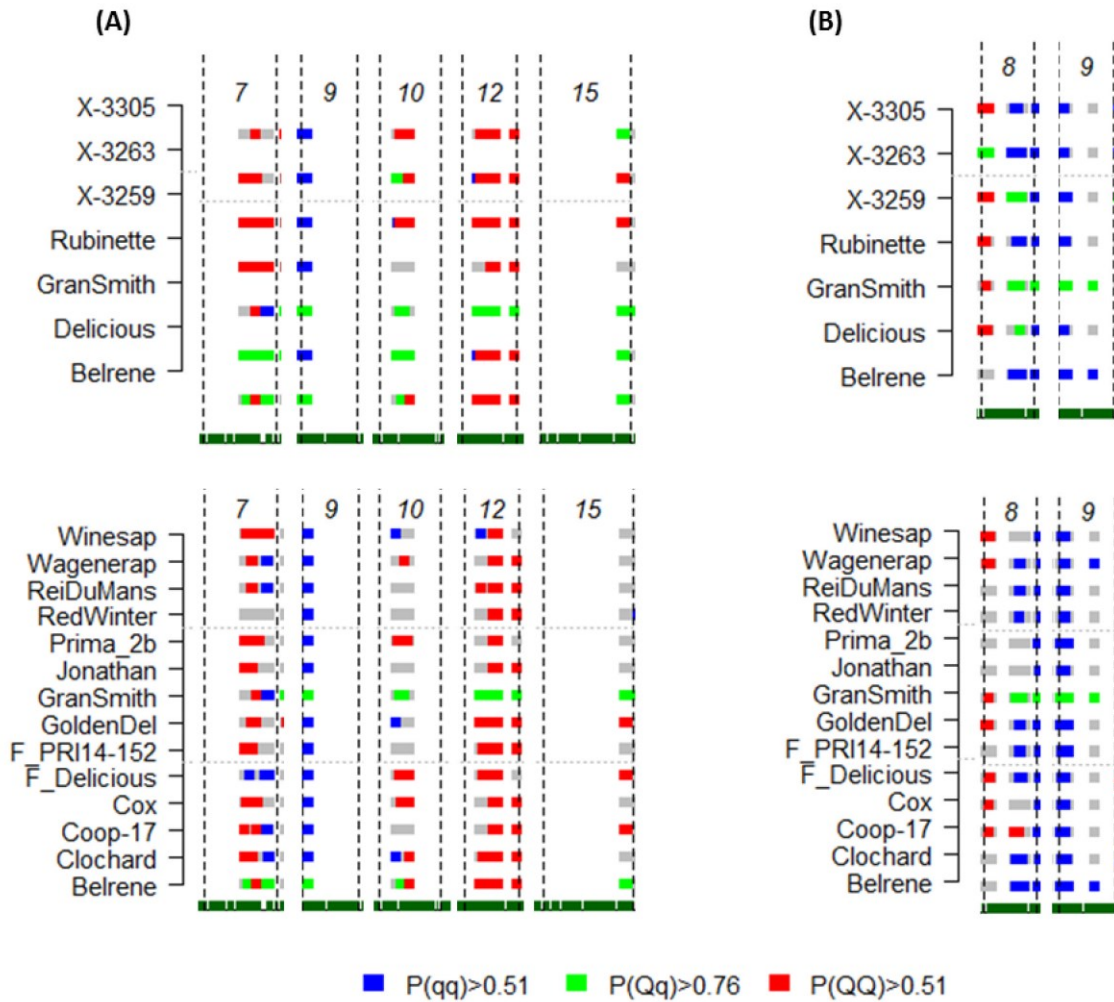


Figure 5: Posterior estimates of QTL genotype probabilities. Estimates are plotted for main regions for genotypic BLUPs for (a) BB_CD and (b) Delta_GDH. The beginning and the end of the chromosomes are represented by vertical dashed lines. Blue, green and red colors indicate positive evidence for QTL genotypes qq, qQ and QQ, respectively; q and Q refer to low and high phenotypic values, respectively. Gray colors indicate ignorance for a given genotype. Estimated genotypes of the parents and founders are presented, respectively.

Supplementary Material

New insights into genetic determinisms for chilling and heat requirements in apple and putative candidate genes through a multi-family QTL discovery

A. Allard, J.M. Legave, S. Martinez, J.J. Kelner, M.C.A.M. Bink, M. di Guardo, E.A. Di Pierro, F. Laurens, W.E. van de Weg, E. Costes

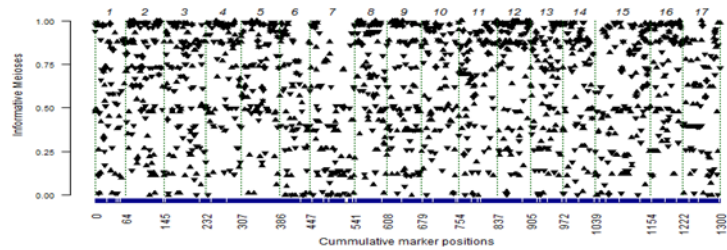


Figure S1: Proportion of informative meioses for haploblock markers along the genome. On the X-axis, the beginning and the end of the chromosomes are represented by vertical dashed lines.

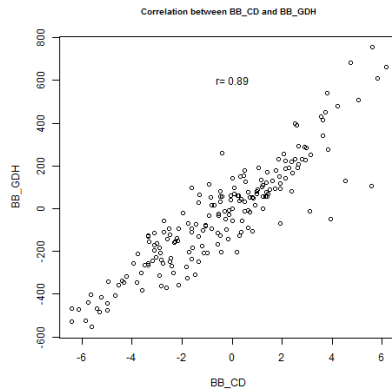


Figure S2: Correlation plot and r value between genotypic BLUP for BB_CD and BB_GDH.

Result Chapter 1

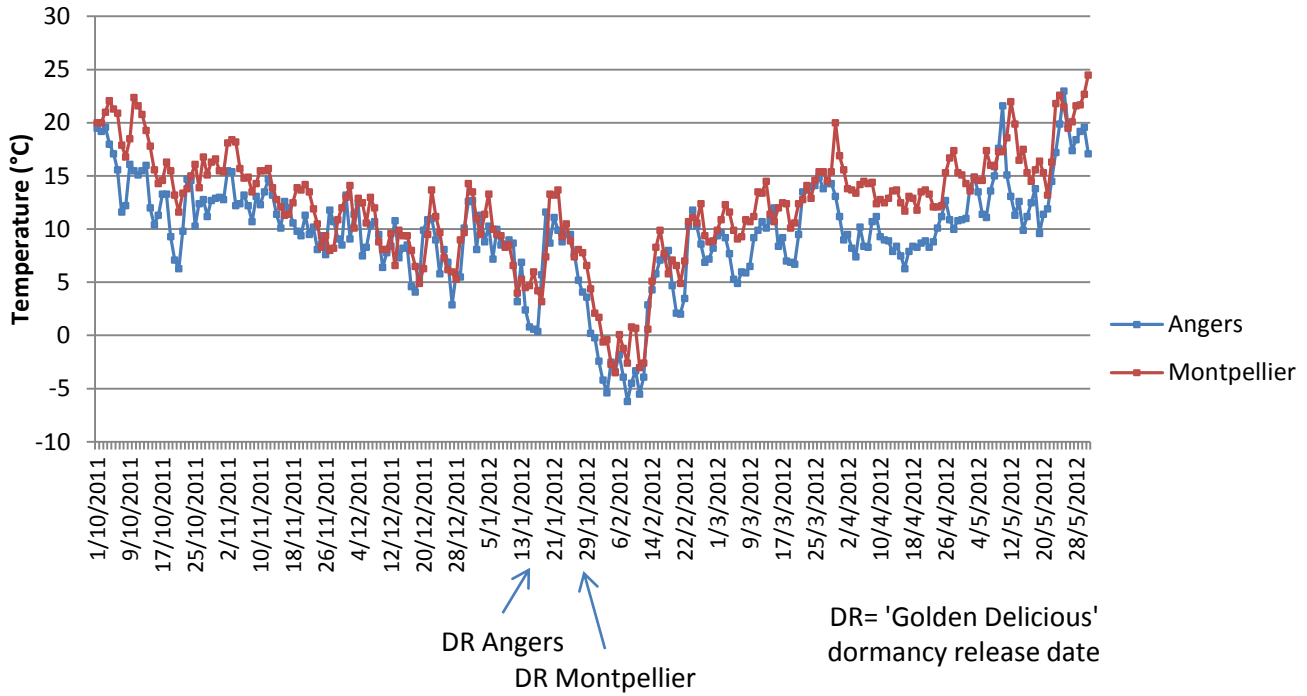


Figure S3: Daily temperature during chilling and heat fulfilment in winter 2011-2012 and spring 2012.

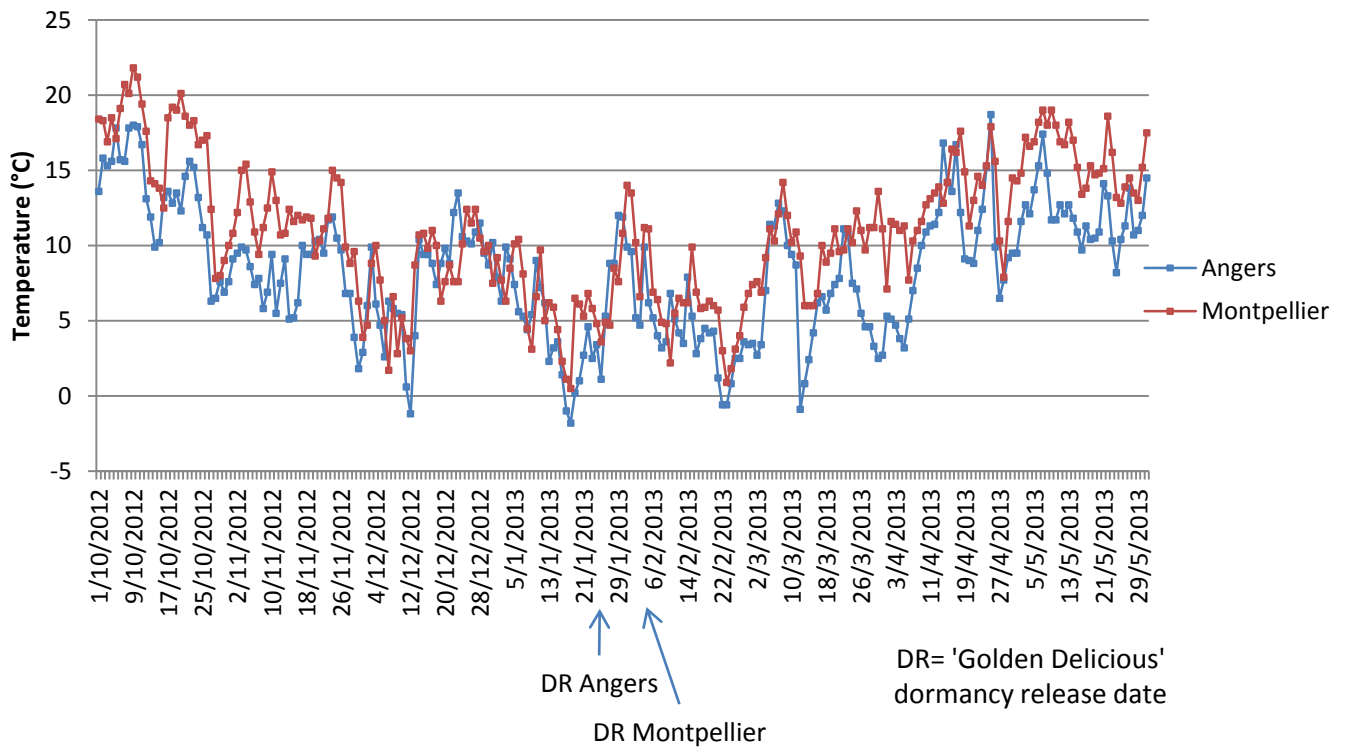


Figure S4: Daily temperature during chilling and heat fulfilment in winter 2013-2014 and spring 2014.

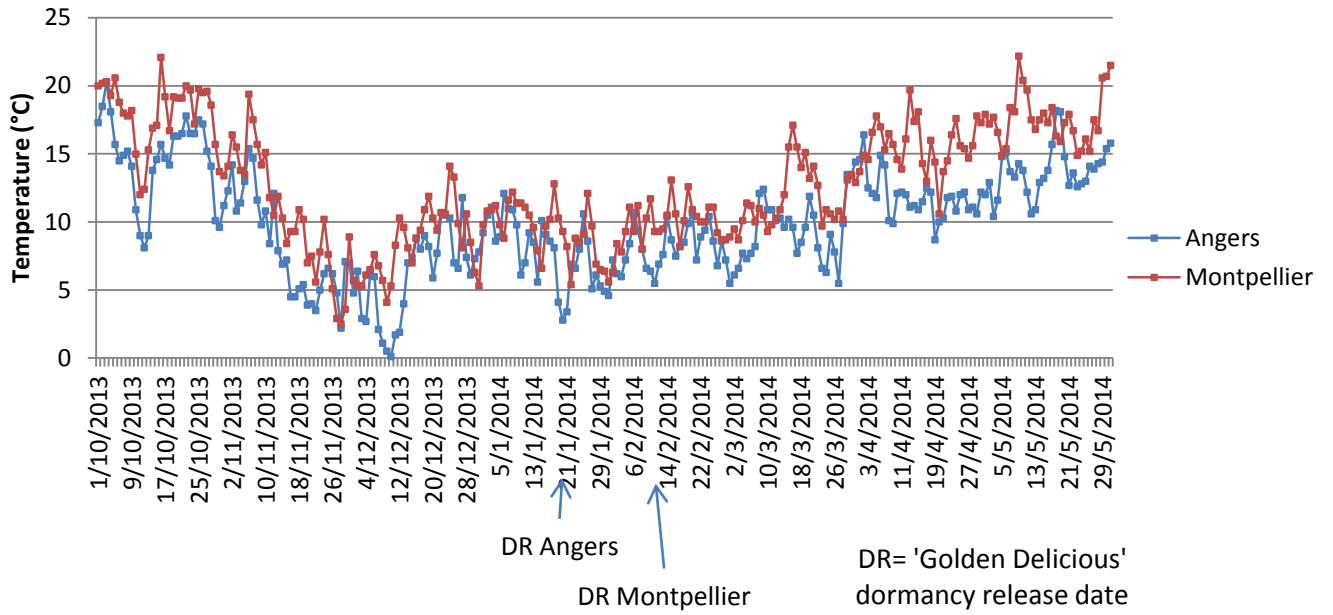


Figure S5: Daily temperature during chilling and heat fulfilment in winter 2013-2014 and spring 2013.

Table S1: Parameters associated with the interaction BLUP QTLs for the Montpellier analysis. The first column indicates the type of analysis, the second one the trait concerned, and the following columns indicate the LG where the QTL is located, the 2*lnBF value at the LG scale for a 1 QTL over a 0 QTL model, the 2*lnBF value at the local scale, the position of the QTL region in cM, the additive QTL effect, the frequency of the positive allele, its variance and its percentage of variance explained. The 2*lnBF values for multi-QTL models were not presented because none of them passed the significance threshold. The QTLs with a 2*lnBF at the LG scale higher than 5 are in bold type.

Analysis	Trait	LG	2lnBF_LG	Max 2lnBF_loc	Pos (cM)	add	fq	var	%var
BB_CD	int 12	4	4.3	5.5	11-35	0.52	0.52	0.13	13.8
		8	6	5.8	39-65	0.43	0.54	0.09	9.2
		12	3.1	3.9	39-61	0.37	0.49	0.07	7
	int 13	15	2.3	5.3	101-115	0.42	0.62	0.08	8.3
		4	2.8	4.7	1-15	0.76	0.44	0.28	8.6
		14	4.3	6.8	45-65	1	0.46	0.49	15
	int 14	3	2.5	4.6	59-79	0.27	0.47	0.04	10.3
		7	3.2	4	14-40	0.26	0.55	0.03	9.2
		9	3.1	5.5	2-14	0.27	0.5	0.04	10.4
BB_GDH	int 12	8	4.8	5.3	1-37	338.2	0.47	56921	8.3
		12	4.4	5	33-67	360.1	0.48	64681	9.4
		15	3.3	5.7	97-109	349.9	0.6	58620	8.5
	int 13	7	6.8	5.7	58-86	204.5	0.44	20573	9.2
		8	2.9	4.7	15-45	162	0.49	13127	5.9
		9	10.3	8.1	2-12	256.4	0.53	32756	14.7
		10	5.9	8.4	26-34	182.4	0.62	15753	7.1
		12	11.3	6.2	19-45	245.3	0.53	29951	13.4
		14	5.3	6.7	43-55	278.1	0.33	34392	15.4
	int 14	3	2.4	4.3	59-85	28.3	0.46	398	0.2
		7	5.3	6.4	52-64	36.9	0.55	675	12.6
		9	4.3	6.5	2-12	39.2	0.5	767	14.4
		12	5.3	6.2	11-43	32	0.54	507	9.5
delta_GDH	int 12	15	3.8	5.1	53-77	247.4	0.46	30392	11.8
		8	4.7	5.9	11-31	102.5	0.45	5197	13.4
	int 14	12	3.3	5.4	53-67	90.4	0.51	4089	10.5
		15	2.6	4.9	7-23	101.4	0.54	5105	13.1

Result Chapter 1

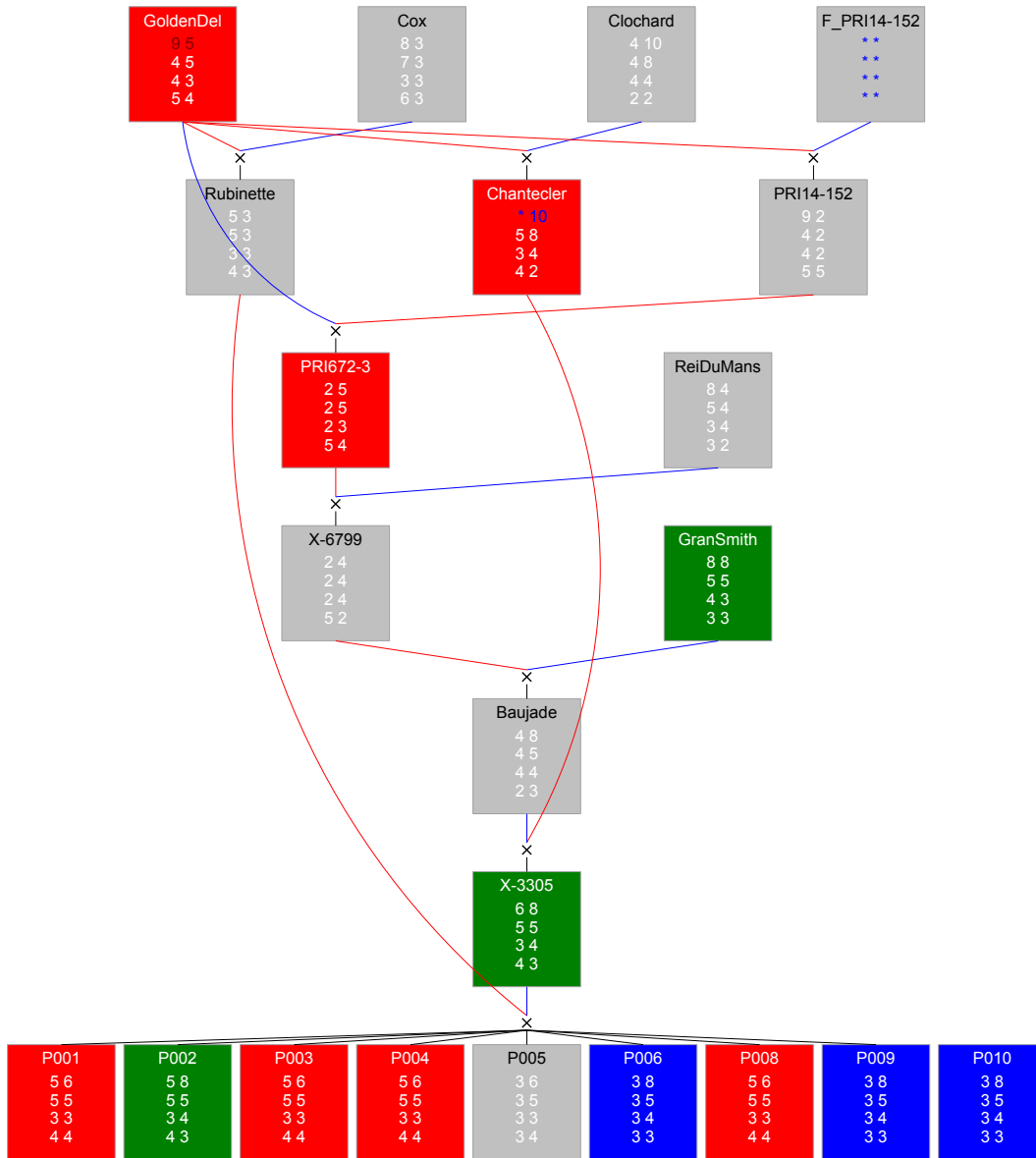


Figure S6: Estimated genotypes and corresponding haplotypes for a QTL located on LG15 between 105 and 111 cM of the P family and its progenitors. The name of the individual and the alleles for four successive haploblocks located at 107, 108, 109 and 110 cM, respectively, on LG15 are given in each box. The color of the boxes indicates the estimated genotypes at the QTL. Red and blue boxes correspond to homozygous individuals with alleles for high (QQ) and low (qq) genotypic values, respectively, green boxes correspond to heterozygous individuals, and gray boxes to individuals for which no distinction could be made between heterozygosity and homozygosity.

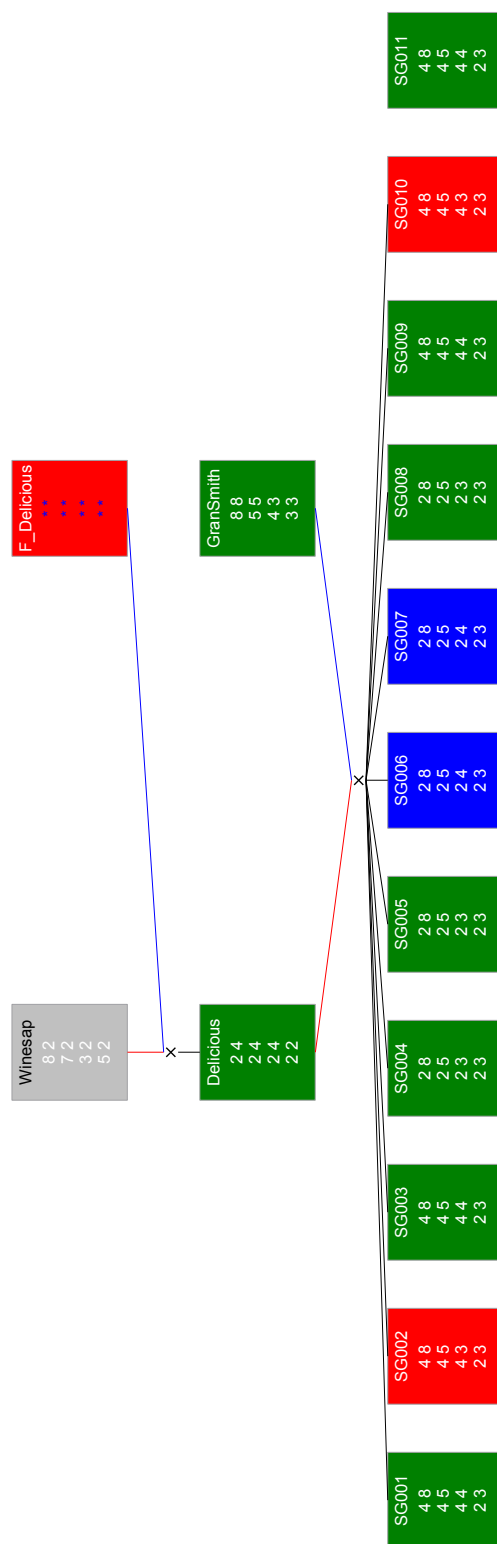


Figure S7: Estimated genotypes and corresponding haplotypes for a QTL located on LG15 between 105 and 111 cM of the SG family and its progenitors. For the legend, see Fig. S6.

Result Chapter 1

Table S2: Indication of haplotype composition for the QTL located on LG15 between 105 and 107 cM. The first column indicates the haplotype name, the second its position on LG15 and the number of SNP markers composing it, the third the haplotype allele ID, the fourth the frequency of the allele, and the following columns the composition of the haplotype allele at each marker. The name of the markers are given in the last column. Missing values are indicated by a ‘-’ symbol.

Haplotype	Position (cM)	Marker_nb	Haplo_ID	freq	M1	M2	M3	M4	M5	M6	M7	M8		
fp15_107	107	8	1	9	-	-	-	-	-	-	-	-	M1:	F_0323470_L15_PA
		8	2	89	1	2	1	1	2	2	2	1	M2:	F_0323476_L15_PA
		8	3	58	1	1	2	1	2	2	1	2	M3:	F_0323478_L15_PA
		8	4	185	1	2	1	2	1	1	2	2	M4:	F_0323485_L15_PA
		8	5	26	1	2	2	1	2	2	1	2	M5:	F_0323486_L15_PA
		8	6	154	2	2	1	1	1	2	1	2	M6:	F_0324449_L15_PA
		8	7	122	1	2	2	1	2	2	2	2	M7:	F_0324471_L15_PA
		8	8	285	1	2	1	1	2	2	1	2	M8:	F_0324474_L15_PA
		8	9	2	2	2	1	1	1	2	2	1		
		8	10	2	2	2	1	1	1	1	2	2		
		8	11	0	1	-	1	1	2	2	1	2		
		8	12	0	1	2	1	2	1	2	1	2		
		8	13	0	1	-	1	2	1	1	2	2		
		8	14	0	1	-	2	1	2	2	1	2		
		8	15	0	1	2	1	1	2	2	-	2		
		8	16	0	2	2	1	1	1	2	-	2		
		8	17	0	1	2	1	2	1	1	-	2		
		8	18	0	1	2	2	1	2	2	-	2		
		8	19	0	1	2	1	2	-	1	-	2		
		8	20	0	2	2	1	1	-	2	-	2		
		8	21	0	1	2	1	2	2	1	-	2		
		8	22	0	1	2	1	2	1	1	2	1		
fp15_108	108	5	1	5	-	-	-	-	-	-	-	-	M1:	F_0323965_L15_PA
		5	2	88	2	2	1	1	1	-	-	-	M2:	F_0323976_L15_PA
		5	3	59	2	1	2	2	2	2	-	-	M3:	F_0323978_L15_PA
		5	4	190	1	2	2	1	1	-	-	-	M4:	F_0323993_L15_PA
		5	5	344	2	1	1	1	2	-	-	-	M5:	F_0323994_L15_PA
		5	6	2	2	2	2	1	2	-	-	-		
		5	7	242	2	1	2	1	2	-	-	-		
		5	8	2	2	1	2	1	1	-	-	-		
fp15_109	109	2	1	6	-	-	-	-	-	-	-	-	M1:	F_0042270_L10_PA
		2	2	90	2	2	-	-	-	-	-	-	M2:	F_0042295_L10_PA
		2	3	547	1	1	-	-	-	-	-	-		
		2	4	289	1	2	-	-	-	-	-	-		
		2	5	0	1	-	-	-	-	-	-	-		
		2	6	0	2	-	-	-	-	-	-	-		
fp15_110	110	4	1	5	-	-	-	-	-	-	-	-	M1:	F_0915567_L15_76_1
		4	2	275	1	1	2	1	-	-	-	-	M2:	F_0915575_L15_76_1
		4	3	226	2	2	2	1	-	-	-	-	M3:	F_0915601_L15_76_1
		4	4	179	2	2	1	2	-	-	-	-	M4:	F_0326218_L15_PA
		4	5	125	1	2	2	1	-	-	-	-		
		4	6	1	2	2	2	2	-	-	-	-		
		4	7	121	2	2	1	1	-	-	-	-		

Result Chapter 2

Genetic determinisms of flowering regularity in apple tree: a multi-family QTL detection based on statistical indices extracted from branch sequence analyses

Paper project, currently being reviewed by co-authors before revision to *Journal of Experimental Botany*

Questions investigated

How to measure and characterize precisely the flowering habit of a tree over years?

Does partial phenotyping of tree, by sampling branches and annual shoot, allows estimating the whole tree habit?

How flowering synchronisation at the branch scale is related to the regular, irregular or biennial habit at the tree scale?

How does years, site, and thinning practices influence tree habit?

Genetic determinisms of flowering regularity in apple tree: a multi-family QTL detection based on statistical indices extracted from branch sequence analyses

Jean-Baptiste Durand^{1,2(*)}, **Alix Allard**³, **Baptiste Guitton**³, **Jean Peyhardi**^{2,4}, **Yann Guédon**², **Catherine Trottier**⁴, **E. van de Weg**⁵, **Marco Bink**⁶, **Evelyne Costes**³

¹Grenoble University, Laboratoire Jean Kuntzmann, BP53, F-38041 Grenoble Cedex 9, France

²CIRAD/Inria, Virtual Plants Team, UMR AGAP, F-34398 Montpellier, France

³INRA, UMR AGAP, Equipe Architecture et Fonctionnement des Espèces Fruitières, F-34398 Montpellier, France

⁴Institut de Mathématiques et de Modélisation de Montpellier, Université Montpellier 2, F-34095 Montpellier Cedex 9, France

⁵Wageningen UR Plant Breeding, Wageningen University and Research Centre, Droevendaalsesteeg 1, P.O. Box 16, 6700AA, Wageningen, The Netherlands.

⁶Biometris, Wageningen University and Research centre, Droevendaalsesteeg 1, P.O. Box 16, 6700AA, Wageningen, The Netherlands.

Corresponding author: Jean-Baptiste Durand (+33 4 76 63 57 09 / +33 4 76 63 12 63), jean-baptiste.durand@imag.fr

alix.allard@supagro.inra.fr, Baptiste.Guitton@cirad.fr, Jean.Peyhardi@math.univ-montp2.fr, Yann.Guedon@cirad.fr, Catherine.Trottier@math.univ-montp2.fr, eric.vandeweg@wur.nl, marco.bink@wur.nl, costes@supagro.inra.fr

Running title: Flowering regularity in apple

Submitted: 00-00-0000

Word counts: ~ 10 820 words

6 Figures (colour online-only),

4 Tables

Supporting information: 1 000 words, 2 Figures, 2 Tables

Key words: biennial bearing, *Malus x domestica*, Markov models, entropy, pedigree based analysis, Bayes factor

Short Abstract

In fruit tree species, fluctuations of flowering over years are of great economic importance. As breeding process is slow and fruiting behaviour express only during mature stage, the prediction of biennial behaviour as early as possible is highly desirable. This study aimed to extend previous methods and indices to perform such predictions and enlarge the comprehension of genetic basis of biennial bearing in apple to a multi-parental population.

Successions of vegetative vs. floral annual shoots (AS) were observed along different axis types in apple trees belonging to five related full-sib families. Sequence analyses were based on mixed linear models, including year and site effects, and high-order Markov models for deriving Biennial Bearing Index-like (BBI) and other indices such as auto-regressive coefficient (γ_g), entropy and memories. Tree behaviour was predicted from indexes defined at axis scale, using neural network (NN). The best linear unbiased predictors (BLUPs) were extracted for the genotype factor, and its interactions with year or memory. QTL mapping was performed with an integrated genetic map containing about 7K SNPs, with a Bayesian approach using pedigree-based analysis.

BBI-like and γ_g descriptors were efficient to classify the genotypes and those descriptors, estimated at AS scale, allowed regular and biennial behaviours to be predicted with high confidence despite miss-classifications for irregular genotypes. Five major QTL zones were highlighted for BBI indices and γ_g (LG4, 5, 8 and 10) and entropy (LG9). Allelic effects and segregation patterns showed the complex genetic architecture and highly polygenic control of the considered traits.

Our results showed that sampling *a posteriori* axes within the tree structure is relevant, considering time-saving for phenotyping. Biennial bearing at tree scale appeared to result from synchronism in flowering between AS in a given year and alternation at AS scale between consecutive years, whereas regularity resulted from either asynchronous alternating flowering. The analysis of QTL co-localisations suggest that both tree development and fruiting capacity may contribute to the genetic variation of biennial bearing in apple tree.

Introduction

In fruit tree species, such as the apple (*Malus x domestica* Borkh.), flowering is of great economic importance by its influence on fruit yield and quality. Depending on the years, flowering can be strongly inhibited by concurrent fruiting leading to biennial bearing defined as the irregular production of a tree over consecutive years (Monselise and Goldschmidt 1982). Biennial or irregular fruiting behaviours are characterised by small-sized fruit the years of high production ('ON' years) and low, or even no production during the 'OFF' years. Although the direct cause of biennial bearing is still largely unknown, usually flowering in each year is assumed to depend on both external factors (i.e. photoperiod, temperature) and internal factors such as the carbohydrates availability and hormones, as well as interactions between other organs (leaves, shoot growth and fruits) (Hanke *et al.* 2007; Bangerth, 2009). Moreover, it has been clearly shown that excessive production of a given year negatively affects flower formation for the subsequent year, suggesting that fruit or seed development produces substances inhibiting floral induction for the next year in the nearest meristem (Singh 1948; Jonkers 1979; Monselise and Goldschmidt, 1982; Dennis and Neilsen, 1999). Economically and environmentally sustainable techniques are therefore required for the management of biennial bearing in apple production. For this purpose, models have been proposed to predict the adequate amount of flowers to be left within a tree, accounting for thinning effect on flower production in the next year (Pellerin *et al.*, 2011) or to determine whether an apical meristem flowers or not, based on critical ratios of promoting or inhibiting hormones (Pellerin *et al.*, 2012). Even though this later model was able to reproduce biennial bearing establishment, the spatial arrangement of buds along branches or within trees remained critical.

An alternative and sustainable strategy to thinning practices would be to develop cultivars combining high fruit quality, long term resistance to pest and disease and tree architecture adapted to modern training systems resulting in regular production and decreased manpower cost. Breeding processes in fruit trees are usually slower than most crops species because of the juvenile phase length that increases time between successive breeding generations and the long time required to assess the agronomic performances of pre-selected trees (Laurens *et al.*, 2010). In addition, biennial bearing is only expressed during the mature phase and trees have to be observed for many years before estimating the bearing tendency of a genotype. It is thus of high

interest to diagnostic the bearing habit of a genotype as soon as possible from the beginning of tree production. The development of new tools is thus highly desirable to efficiently select regular apple cultivars as from the first breeding steps. This strategy has been reinforced by the demonstration of genetic determinants for the control of biennial bearing in a family derived from a cross between biennial and regular bearers (“Starkrimson” × “Granny Smith”; SG) (Guitton *et al.*, 2012).

To characterise successive yields and biennial behaviour, different parameters and indices have been proposed (see Durand *et al.*, 2013 for a review). In Durand *et al.* (2013), a new methodology was introduced to characterize the bearing habit of apple tree progenies as early as the first years of production. A trend model was proposed to capture the gradual increase in the number of flowers and a Biennial Bearing Index-like (BBI-like) was derived to measure the intensity of deviations of production around the trend, while an auto-regressive coefficient (denoted by γ) was used to discriminate between alternating and irregular genotypes. An approximation of these indices based on within-tree sampling AS successions along axes was considered. Moreover, an entropy criterion was proposed to assess synchronicity of flowering in a given year. Then, a multiscale analysis of alternation was performed through the analysis of correlations between indices, i.e. relating tree-scale to axis-scale alternation indices. However in this approach, the ability of axis-scale indices to predict genotype habits at tree scale was investigated on one single family (SG), and the sequences were merely used to approximate the total number of flowering AS within trees. In a parallel work, successions between two consecutive annual shoots (AS), as proposed by Lauri *et al.* (1997), have been correlated to short lateral death (‘extinction’) in the perspective of providing early diagnostic of fruiting behaviour in a range of genotypes (Lauri *et al.*, 2014). In the present study, we propose to extend the previous investigations by exploring new methods and indices based on sequence analysis, and by performing a multi-family QTL detection to enlarge the genetic basis of biennial bearing variation. Regarding the characterization of alternation from AS sequences by finer statistical analysis, we assumed that the analysis of entire sequences of successive annual shoots (AS) combined to yield synchronicity in each year would provide new insights on fruiting behaviours of different genotypes. We thus proposed (i) to use not only the total number of flowering AS but also the vegetative AS and the successions of both kinds of AS; (ii) to derive new indices from this sequence analysis for predicting the tree bearing habit at no additional measurement cost;

(iii) to re-examine previous assumptions on the relation between alternation and regularity at tree and axis scales. Regarding genetic, we considered a multiparental population in order to allow the comparison of alleles' performance in different genetic backgrounds (Pauly *et al.*, 2012). We assumed that this approach will increase the number of segregating QTLs, QTL detection power, accuracy of QTL positions, and give more robust estimation of QTL effects (M. C. A. M. Bink 2002) (Blanc *et al.*, 2006; Liu *et al.*, 2012). Moreover, we performed a Pedigree Based Analysis (PBA) to take into account relationships between individuals using the concept of Identity By Descent (IBD) which uses both information from the pedigree and markers (Luan *et al.*, 2012). Our aim was to confirm previously found QTLs on the SG family and find new ones, using new indices and four new progenies, tracing the original source of the favourable alleles and deepening our comprehension of the genetic determinisms of biennial bearing in apple tree.

Materials and methods

Plant material

Five F1 segregating families with known and related pedigree were used (Fig. 1). The first family (SG), as described by (Segura *et al.*, 2006), is derived from a cross between a female parent ('Starkrimson') having strong tendency to biennial bearing and a male parent ('Granny Smith') being prone to regular bearing. This population is composed of 123 genotypes that were each replicated twice in the same environment, among which 115 individuals were genotyped and phenotyped. The second family, described by Celton *et al.* (2011), is issued from a cross between the hybrid X3263 (regular bearer) and the cultivar 'Belrène' (biennial bearer). This family (XB) comprises 324 genotypes of which 50 seedlings were randomly selected for replication resulting in 374 trees in total that were all phenotyped among which 58 were also genotyped. For both families, seedlings were grafted on a semi-dwarfing Pajam I rootstock and planted in 2004 and 2005, respectively, in a random experimental design at the Diascope INRA Montpellier experimental unit (co-ordinates: 43°36'35''N; 3°58'50''E). All trees were grown under irrigated conditions with minimal training, i.e. trees were neither pruned nor the fruit thinned in order not to interfere with flowering and tree architecture traits.

The three other families, called HIVW, N, and P, respectively, were derived from crosses between ‘X-3263’ and ‘X-3259’, ‘X-3259’ and ‘X-3305’ and between ‘X-3305’ and ‘Rubinette’. These families are composed of 171, 42 and 45 individuals, respectively, each with a single replicate per genotype. They were planted in 1992 at the INRA Angers experimental station (coordinates: 47°29’7.656 N; 0°36’47.646). The trees were trained in vertical axis with an annual manual thinning with one fruit per inflorescence. At both sites, pest and disease management was performed consistently with professional practices.

Phenotyping

As described by Durand *et al.* (2013), on the SG family, successions of vegetative vs. floral annual shoots were observed over consecutive years along different types of axes: trunk, long and short axillary shoots (LAS and SAS respectively). Shoots were classified depending on their length and within LAS, proleptic and sylleptic branches were distinguished (Long Proleptic Axillary Shoot, LPAS and Long Sylleptic Axillary Shoot, LSAS; see Segura *et al.*, 2006). Vegetative and floral AS were also distinguished depending on the presence/absence of an inflorescence and floral AS were decomposed into a “bourse” (leafy growth unit ending with an inflorescence) and a “bourse-shoot” (vegetative growth units originating from a lateral meristem of the bourse).

For SG, flowering occurrence was observed along the trunk of each tree, as well as along one LSAS and one LPAS, both sampled on the first trunk’s AS (AS 2004). On each LAS, two short axillary shoots (SAS) per annual AS were sampled and phenotyped the same way. As a consequence, 10 SAS of 5 to 1 years were recorded on LSAS, and 8 SAS of 4 to 1 years were recorded on LPAS. The flowering pattern was described by recording the presence/absence of flowering event on AS (6 possible flowering occurrences on the trunk and LSAS, 5 on LPAS). The data thus consisted in vegetative vs. floral AS in 6 to 1 year sequences. In total, 2,716 sequences were phenotyped, with mean length 3.0.

For the XB, HIVW, N and P families, flowering occurrence was observed along three LPAS only. For each, four and three short axillary shoots (SAS) were sampled on the annual growth unit 2006 and 2007, respectively and were phenotyped as previously described. In the XB family, 7,757 sequences were phenotyped, with mean length 5.3. In HIVW, 1,511 sequences were

phenotyped, with mean length 6.0. In the P family, 442 sequences were phenotyped, with mean length 6.2. In the N family, 905 sequences were phenotyped, with mean length 6.4 (see distributions in Supplementary Figure S1).

Statistical modelling of AS fate sequences

Our approach for modelling alternation is based on the classical BBI and on indices defined in Durand *et al.* (2013): BBI-derived, auto-regressive coefficient and entropy. They were based on counts of flowering AS at whole tree and axis scales. The description of these indices as well as the statistical methodology for classification of genotype bearing habit from indices can be found in Supplementary Material M1. Compared to the original indices, we added a fixed “site” effect, Montpellier (M) or Angers (A), in the trend and auto-regressive models, whenever possible. These models were used to compute the BBI_res_norm (normalized BBI on residuals) and γ_g (auto-regressive coefficient) indices at tree and axis scales.

At the axis scale, sequences of AS fates are denoted $(F_{g,r,\pi,t,\ell})_{\ell \geq 0}$ with $(F_{g,r,\pi,t,\ell} = 0)$ denoting the absence and $(F_{g,r,\pi,t,\ell} = 1)$ the presence of flower for replication r of genotype g at site π , at year t , and at location (or AS) ℓ in the axis. The indices at axis scale, that are denoted by BBI_ax, BBI_norm_ax, BBI_res_norm_ax, and γ^{ax} , were computed as the BBI, BBI_norm, BBI_res_norm and γ_g , respectively but using the total yearly counts of flowering AS in axes sampled within each tree replicate.

These counts were also used to compute the entropy indices, denoted $\overline{Ent}_g$ (entropy based on frequencies) and $\overline{Ent}_{glmm,g}$ (entropy based on a Generalized Linear Mixed Models - GLMM). However, the latter indices are based on the assumption of independent Bernoulli distributions for the successive AS fates $F_{g,r,\pi,t,\ell}$, thus ignoring dependencies and patterns of alternation v. regularity that could be inferred from sequences of AS fates along axes, which were also collected. Thus, we hypothesized that models and indices taking explicitly into account the succession of AS fates would provide new insights on the relationship between alternation at axis and whole tree scales. Prediction of genotype habit using such models would also be expected to yield lower error rates.

For this purpose, we modelled the sequence of AS fates $(F_{g,r,\pi,t,\ell})_{\ell \geq 0}$ through a high-order Markov dependency model (see Costes and Guédon (2011), for discussion of the shortcoming of using a first-order rather than a high-order Markov model). In such process, the variable at time t depends on the M past variables. The values of these M past variables are referred to as the memory m of the process at time t , used to predict the current value. For example, memory “10” means that flowering occurred at time $t - 2$ but not at time $t - 1$. Markov chains with different orders were estimated on the basis of every non-overlapping sequence extracted from the trees (the maximal sequence length was 6). A model of order 2 was chosen by a model selection procedure based on the Bayesian Information Criterion (BIC, see Kass and Raftery (1995), meaning that knowledge on presence or absence of flower at years $t - 1$ and $t - 2$ are necessary for prediction of flowering at year t . Thus, the set of memories was $\{00, 10, 01, 11\}$.

In the case of alternation partly due to genetic, it is expected that some interactions between year t , memory m and genotype g will have an effect on flowering. To model these interactions in the case of binary observations $F_{g,r,\pi,t,m,\ell}$, approaches based on GLMMs are relevant (Molenberghs and Verbeke, 2006). The following GLMM was considered:

$$\log \frac{P(F_{g,r,\pi,t,m,\ell} = 1)}{P(F_{g,r,\pi,t,m,\ell} = 0)} = \lambda + \rho_{\pi} + \mu_m + \phi_t + \theta_{g,m} + \eta_{g,t} + \zeta_{g,r}, \quad (1)$$

where λ is a fixed intercept, ρ_{π} is the fixed effect of site π (with Angers site as reference $\rho_A = 0$), μ_m is the fixed effect of memory m (with reference $\mu_{00} = 0$), ϕ_t is the fixed effect of year t (with reference $\phi_{2006} = 0$) treated as a qualitative variable, variables $\theta_{g,m}$ are independent random interactions between genotype g and memory m with common variance τ_{θ}^2 , variables $\eta_{g,t}$ are independent random interactions between genotype g and year t with common variance τ_{η}^2 , and variables $\zeta_{g,r}$ are independent replication-specific random effects with common variance τ_{ζ}^2 . All random effects are assumed to be mutually independent and Gaussian. Globally, this model consists of a high-order Markov chain for process $(F_{g,r,\pi,t,m,\ell})_{\ell \geq 0}$ where the transitions are treated as GLMMs, so as to introduce fixed and random effects. Modelling the log ratio of the probabilities allows the probability of a binary variable (comprised between 0 and 1) to be

mapped into $\mathbb{R}$. The model parameters were estimated by maximum likelihood and the BLUPs for the random effects were computed using the functions *glmer* and *ranef* of package *lme4* (Bates *et al.*, 2011). Particularly, the BLUPs of $\theta_{g,m}$ were used to highlight genotypes that had low or high probability of AS bearing flowers at year t given that they bore flowers or not at times $t - 1$ and $t - 2$. Similarly the BLUPs of $\eta_{g,t}$ were used to highlight genotypes that had especially low or high probability of AS bearing flowers at “off” or “on” years t .

Predicting tree fruiting behaviour from axis-scale indices

One goal of this study was the prediction of the tree flowering behaviour (regular, irregular or alternate bearing). As this behaviour is a categorical variable which possible values can be seen as three classes, the prediction of the genotype bearing habit is a classification problem. To perform this classification using indices measured at axis scale, we used BLUPs of $\theta_{g,m}$ random effects together with BBI_res_norm_ax , γ^{ax} , $\overline{Ent}_g$ and $\overline{Ent}_{glmm,g}$ as predictors. Classification was addressed by neural networks (NNs), as described in Supplementary Material M2. Note that BLUPS may not be defined for every parameter and every genotype. If for example memory 11 did not occur at all in the axes of some genotype g , $\theta_{g,11}$ cannot be defined. This resulted into missing predictors. Moreover, some predictors could be highly correlated and thus redundant. To handle both issues of predictor absence and redundancy, a principal component analysis (PCA) for partially missing data was used, with the R package *missMDA* (Josse and Husson 2012). The number of principal components (PCs) to choose and the NN regularisation parameter were determined by out-of-sample validation.

Although classification is a relevant method in itself to assess regularity in flowering, it provides a prediction of the genotype flowering behaviour at tree scale through three classes only, which is not more than a qualitative summary of the flowering behaviour. In some other contexts (as genotype ranking), more quantitative assessment of the genotype behaviour at tree scale may be required. This could be achieved through the tree-scale indices BBI_res_norm and γ_g (if they could be measured), since they provide a more accurate description of the fruiting behaviour. Since both indices were unknown, except for the SG family, prediction of both indices must be

performed from the axis-scale indices $\theta_{g,m}$, BBI_res_norm_ax, γ^{ax} , $\overline{Ent}_g$ and $\overline{Ent}_{glimm,g}$. This is a regression problem. To perform regression, NNs were used as nonlinear regression functions (instead of nonlinear classifiers in the case of classification), also using missing data PCA. The NN parameters were estimated by least squares minimization. The optimal number of PCs to be used in both classification and regression NNs was chosen by out-of-sample validation as explained in Supplementary Material M2 (these numbers may not be the same in classification and regression NNs). The predicted values of BBI_res_norm and γ_g are referred to as BBI_res_norm_pred and γ^{pred} , respectively. Due to the nature of NNs, BBI_res_norm_pred and γ^{pred} are necessarily linearly dependent.

Classification and regression were also used in Durand *et al.* (2013) but the models did not use $\overline{Ent}_{glimm,g}$ nor the $\theta_{g,m}$'s so we assessed the gain of using these new indices as predictors. Since the tree-scale indices were known for the SG family only, the classification and regression errors were assessed using this family only. In the case of regressions, BBI_res_norm_pred and γ^{pred} were used as indices in QTL analysis.

Genetic map and QTL mapping

The five full sib families and their progenitors were genotyped with the Infinium® 20K SNP array (Bianco *et al.*, 2014) at the Fondazione Edmund Mach according to the procedures described by Chagné *et al.* (2012) and Antanaviciute *et al.* (2012). An integrated genetic map was used for QTL mapping, this map was composed of 7100 SNPs and is integrated over 27 full-sib families (Van de Weg *et al.*, unpublished). 6849 SNPs were used in this study after careful check of their robustness (Van de Weg *et al.*, 2013; Guardo *et al.*, 2015), consistency and recombination pattern (Allard *et al.*, under revision). Next, sets of single SNPs were integrated into haploblocks, corresponding to successive 1 cM segments in which the SNPs were assigned. Haplotypes were composed using the software FlexQTLTM and PediHaplotyper (Voorrips *et al.*, submitted).

Indices used in QTL analysis were BBI_ax, BBI_norm_ax, BBI_res_norm_ax, γ^{ax} , BBI_res_norm_pred, γ^{pred} , $\overline{Ent}_g$, $\overline{Ent}_{glimm,g}$, the BLUPs for genotypes x memory interactions (

$\theta_{g,00}, \theta_{g,01}, \theta_{g,10}, \theta_{g,11}$) and genotype x year interactions ($\eta_{g,2006}, \eta_{g,2007}, \eta_{g,2008}, \eta_{g,2009}, \eta_{g,2010}, \eta_{g,2011}, \eta_{g,2012}$). Among all the variables, QTLs were detected by using a linear model that comprised an intercept μ , the regressions on the QTL covariates a , and a model residual e , as:

$$y = \mu + Wa + e \quad (2)$$

where W is the design matrix for the QTL effects. A bi-allelic model is assigned to a QTL with alleles denoted by Q and q, with only additive effects modelled and the covariate values of [QQ, Qq, qq] are equal to [1, 0, -1]. The frequency of the allele Q among founder individuals is denoted by fa and the linkage map positions of the QTL are given by vector λ . The QTL genotypes of individuals are a priori unknown and modelling is based on independent assignment of alleles to founders and segregation indicators to trace transmission from parents to offspring (Bink, 2002). The Bayesian modelling assigns uniform priors to the variables μ and λ , while assigning normal priors to the vectors a and e in (2), i.e. $a \sim \mathcal{N}(0, \sigma_a^2 I)$ and $e \sim \mathcal{N}(0, \sigma_e^2 I)$. The variables σ_a^2 and σ_e^2 are the per-QTL explained variance and the residual variance, with priors being inverse Gamma distributions (Bink *et al.*, 2008). The number of QTL was assigned a Poisson prior with different values (i.e., 5, 10) to assess sensitivity of posterior inference to the prior assumptions. Results for prior mean of 5 are reported only, the other values yielding similar results and inferences.

Markov chain Monte Carlo (MCMC) simulation, as implemented in the software FlexQTLTM (Bink *et al.*, 2014), was applied to obtain samples from the joint posterior distribution of the model parameters

$$f(\mu, a, \lambda, \sigma_a^2, \sigma_e^2 | y) \quad (3)$$

The MCMC algorithms previously described (Bink *et al.*, 2008) are omitted here, and details on the following steps (i.e. monitoring of Monte Carlo accuracy, length of the simulation chains) can be found in Allard *et al.* (under revision). The inference on the number of QTL was based on a pairwise comparison of models differing one QTL from each other, by considering twice the natural log of Bayes Factors (Kass and Raftery, 1995), denoted $2 \cdot \ln BF$, which values greater than 2, 5, and 10 indicate positive, strong, and decisive evidence, respectively. Similar to Bink *et*

al. (2014), QTL positions were based on posterior QTL intensities, and QTL contributions on the posterior mean estimates of the QTL effect sizes. Posterior probabilities of QTL genotypes were also estimated, with values of $2 \cdot \ln \text{BF}$ greater than 2, 5 and 10 indicating positive, strong, and decisive evidence, respectively, as in Bink *et al.* (2014).

Results

Modelling AS sequence

The Markov GLMM (1) proposed in Section 2.3 and estimated by merging the data sets for the five families had a BIC value of 44,300. It was compared with the associated model without any random effect (i.e. with $\tau_\theta^2 = \tau_\eta^2 = \tau_\zeta^2 = 0$, BIC = 52,944), with the model containing “genotype” and “replication” random effects only (no interactions with year or memory, BIC = 50,723), with the model without “replication” random effects (BIC = 44,394) and the model without “genotype” random effects and their interactions (BIC = 50,780). The estimates of the fixed effects and variances are given in Table 1 (with the p-values of the tests of the null hypothesis “ $\nu = 0$ ” against the alternative “ $\nu \neq 0$ ”, for parameters ν associated with fixed effects).

It can be deduced from the BIC values of the different models and from the p-values in Table 1 that memory, year, site and replication effects were significant. The replication effect, although significant, induced less variability in flowering than memory and year effects (the latter having the largest variance). The trees in Montpellier had lower flowering probability than those in Angers. The fact that μ_{10} was higher than μ_{01} and μ_{11} highlights that flowering AS were more frequent after a vegetative AS preceded by a flowering AS, than directly after a flowering AS. This is symptomatic of frequent biennial alternation in flowering at axis scale in this dataset. Moreover, the probability of flowering was the highest in 2008, whereas it was especially low in 2010 and 2012, whatever the site.

Empirical standard deviations were computed for the BLUPs of random effects $\theta_{g,m}$ and $\eta_{g,t}$ to estimate their specific variability for each year t or memory m (Table 2). The genetic variability of $\eta_{g,t}$ random interactions increased with years, with $\eta_{g,2012}$ having the largest variability. The random interactions $\theta_{g,10}$ (between genotypes and memory “flowering AS

followed by a vegetative AS”) had the lowest genetic variability and the random interactions $\theta_{g,01}$ between genotypes and memory “vegetative AS followed by a flowering AS” had noticeably higher variability.

Regression of tree-scale indices using axis-scale indices

In the trend models used at axis-scale from the yearly numbers of flowers to compute BBI_res_norm_ax and γ^{ax} (see Supplementary Material M1, models (A) and (I)), including some site effect in addition to the individual random effects led to model non-identifiability. Therefore, this fixed effect was not included. However, the site effect was included in the autoregressive model, i.e. model (I) used to compute γ^{ax} .

Using NNs, BBI_res_norm and γ_g were regressed using three or four PCs, and the best cross-validated correlations were obtained with three PCs. The optimal cross-validated correlation between BBI_res_norm and its prediction BBI_res_norm_pred was 0.71 when using BBI_res_norm_ax, γ^{ax} and $\overline{Ent}_g$ only as predictors, and 0.72 when using the PCs. The optimal cross-validated correlation between γ_g and its prediction γ^{pred} was 0.60 when using BBI_res_norm_ax, γ^{ax} and $\overline{Ent}_g$ only, and 0.64 when using the PCs. Using $\overline{Ent}_{glimm,g}$ instead of (or in addition to) $\overline{Ent}_g$ did not improve correlations. As a conclusion, using Markov GLMMs did not significantly improve prediction of the tree-scale indices.

Classification of the genotypes with respect to tree-scale bearing habit

Classification of the genotypes of the F1 SG family yielded a 37% cross-validated error rate when using BBI_res_norm_ax, γ^{ax} and $\overline{Ent}_g$ only, and 35% when using the PCs on the same dataset (which may be interpreted as the expected error rate on other families). The best prediction was achieved using three PCs. Using $\overline{Ent}_{glimm,g}$ instead of (or in addition to) $\overline{Ent}_g$ did not improve the error rates. Although the improvement is limited, a t-test showed on 50 random test samples that it was significant at level 0.7% (Student’s t-test).

To predict the (unknown) bearing habits of the genotypes of XB or families in Angers (HIVW, N and P), the axis-scale indices were computed on a single model combining all families except SG. The SG genotypes were included since learning the mapping between local indices and classes required the use of genotypes with known classes. A NN model was estimated on the whole data set using three PCs. The confusion between classes is illustrated by the contingency Table 4. No misclassification occurred between regular and alternate bearing genotypes, which highlights that discrimination between both behaviours is easy, and few misclassifications occurred between regular and irregular genotypes. In contrast, irregular genotypes could hardly be discriminated from alternate bearing genotypes.

Analyses of variance (ANOVAs) were performed on the axis-scale indices to assess their abilities to discriminate between the three classes. The p-values are given in Supplementary Table T2 at JXB online. The lowest p-values were obtained with $1e-8$ which proved a significant separation of both indices with $BBI_res_norm_ax, \gamma^{ax}, \theta_{g,01}$ and $\overline{Ent}_g$, with clearly monotonous values with respect to the order “regular, irregular, biennial bearing”. This highlights that these four indices have contrasted values with respect to the classes, and that they are more likely than the other axis-scale indices to allow discriminating between the three classes of bearing behaviour.

The model yielded the following predictions

- SG: 33 (27%) regular, 36 (30%) alternate bearing, 53 (43%) irregular genotypes - with 21% error rate on the learning sample (33% if using $BBI_res_norm_ax, \gamma^{ax}$ and $\overline{Ent}_g$ only)
- XB: 52 (19%) regular, 75 (27%) alternate bearing, 148 (54%) irregular genotypes
- HIVW: 33 (19%) regular, 24 (14%) alternate bearing, 113 (66%) irregular genotypes
- N: 30 (30%) regular, 11 (11%) alternate bearing, 59 (59%) irregular genotypes
- P: 9 (17%) regular, 8 (15%) alternate bearing, 36 (68%) irregular genotypes.

Consistency of the indices for alternation at tree and AS scales

To establish a relationship between the tree and the axis-scale indices for the SG family, a correlation matrix was computed (Table 3). The correlation matrix is given in Supplementary

Table T1 at JXB online for the five combined populations, excluding the tree-scale indices, which cannot be computed except for the SG family.

The strongest correlations were between tree-scale and axis-scale indices were with BBI_res_norm_ax, γ^{ax} and $\theta_{g,01}$ (which correlation is negative with BBI_res_norm_ax, positive with γ^{ax}). Moreover, $\theta_{g,11}$ and $\overline{Ent}_g$ were moderately correlated with BBI_res_norm (negatively) and with γ^{ax} (positively), while $\overline{Ent}_{glmm,g}$ was poorly correlated and $\theta_{g,10}$ was uncorrelated (at level 0.05) with the tree-scale indices.

QTL mapping

Among studied indices, QTLs were detected for BBI_ax, BBI_norm_ax, BBI_res_norm_ax, γ^{ax} , $\overline{Ent}_g$, $\overline{Ent}_{glmm,g}$, $\theta_{g,00}$, $\theta_{g,01}$, $\theta_{g,11}$, $\eta_{g,2006}$, $\eta_{g,2008}$, BBI_res_norm_pred and γ^{pred} (Table 5). In contrast, no QTL were detected for $\theta_{g,10}$, $\eta_{g,2007}$, $\eta_{g,2009}$, $\eta_{g,2010}$, $\eta_{g,2011}$, and $\eta_{g,2012}$.

Five QTLs were detected for the three BBI indexes and BBI_res_norm_pred, but with different robustness (Figure 2, A, B and C). The QTL on LG5 was always detected with a strong evidence with respect to its $2 \cdot \ln BF$ values and stability, for BBI_ax, BBI_norm_ax, BBI_res_norm_ax, and BBI_res_norm_pred. It explained 6.4, 7.7, 6.9, and 8.3% of the variance of each variable, respectively (Figure 2, Table 5). The second QTL was detected on LG4, with a strong evidence for BBI_norm_ax, BBI_res_norm_ax, and BBI_res_norm_pred, explaining 10.9, 11.5, and 13.3% of the variance, respectively. The QTL detected on LG7 had a strong evidence for BBI_ax only and explained 10.7% of its variance. However, its position remains uncertain (Figure 2A). The QTL detected on LG10 for the four BBI indexes, explained 11.4, 9.1, 7.7, and 10% of the variance, respectively, and had a strong evidence for BBI_ax, BBI_norm_ax, and BBI_res_norm_pred. Finally, a QTL was also detected on LG8 for the four BBI indexes, but had a strong evidence for BBI_res_norm_pred only and explained 11.7% of the variance. In addition, a QTL was detected on LG12 for BBI_res_norm_ax, with positive evidence and explained 5.8% of the variance.

For γ^{ax} , three QTL zones were detected on LG4, LG 5 and LG 10 but none with strong evidence. The QTL on LG4 did not colocalize with that of BBI indexes whereas the QTL on LG5 slightly overlapped and that on LG10 colocalized with that of BBI indexes. Four QTLs were mapped for γ^{pred} on LG4, LG5, LG8 and LG10 which all colocalized with BBI_res_norm_pred. QTLs on LG4 and LG8 explained 12% of the variance, and those on LG5 and LG10 explained 10% of the variance (Figure 2, Table 5).

Several QTLs were detected for the entropy variables, $\overline{Ent}_g$ and $\overline{Ent}_{glimm,g}$ (Figure 3A and 3B). A QTL was detected for $\overline{Ent}_g$ on LG12, which explained 5% of the variance, and had a positive evidence for its presence. For $\overline{Ent}_{glimm,g}$, five QTLs were detected among which only one had a strong evidence on LG9 and explained 20% of the variance. The other QTLs were detected on LG1, LG7, LG15, and LG17, and explained 10% of the variance each (Figure 3, Table 5).

Concerning interactions between genotypes and memories, QTLs were mapped for $\theta_{g,00}$, $\theta_{g,01}$ and $\theta_{g,11}$ (Figure 3C-3F, Table 5). For these three variables, a QTL colocalized on LG10 which explained 10, 11.8 and 10.9% of the variance, respectively, and had a strong evidence for $\theta_{g,00}$ and $\theta_{g,01}$. This QTL colocalized with that detected for BBI indexes. Additional QTLs with positive evidence were detected on LG4 for $\theta_{g,01}$ which explained 5.9% of the variance, and on LG11 for $\theta_{g,00}$ which explained 15% of the variance.

Concerning interaction between years and genotypes two QTLs with positive evidence were mapped on LG12 and LG6, for $\eta_{g,2008}$ and $\eta_{g,2006}$, respectively, which explained 7.2% and 14.4% of the variance, respectively (Figure 4A, 4B, and Table 5).

Genotype estimation on main QTLs

Genotype estimation brought two types of information. Firstly, the allelic classes of parents were used to identify in which family QTLs segregated. For this, we considered that if one of the parents of a family was estimated as heterozygous at a QTL, this family segregated for this QTL.

Secondly, genotype estimation was used to identify parents and founders bearing favourable alleles, specific to the variable considered. Hereafter, only QTLs with strong evidence were commented (Figure 5).

For the BBI indexes, the estimated genotypes were similar for QTLs that colocalized (Figure 5). On LG4, bearing a QTL with strong evidence for BBI_norm_ax, BBI_res_norm_ax, and BBI_res_norm_pred, the parent ‘Delicious’ only was estimated as heterozygous and consequently only the SG family segregated. All the other parents, including ‘Granny Smith’ were estimated as homozygous with the low value allele. As bringing low BBI values can be considered as favourable, ‘Granny Smith’ is likely to transmit favourable alleles to its progenies whereas ‘Delicious’, heterozygous, is likely to transmit the favourable allele to half the progeny. On LG5 region bearing a QTL for all BBI indexes, the parents X-3305, X-3259 and ‘Delicious’ were estimated as heterozygous and consequently, all families except XB segregated. Moreover, ‘Granny Smith’ was estimated as homozygous with the high value allele and ‘Rubinette’ as homozygous with the low value allele. Similarly, for the QTL on LG7 for BBI_ax, all the families except XB segregated with parents X-3259 and ‘Delicious’ being heterozygous. For BBI_res_norm_pred QTL on LG8, parents X-3263 and Delicious were estimated as heterozygous, therefore SG, XB, and HIVW segregated. The other parents were estimate as homozygous with the low value allele. For QTL on LG10, all families segregated since at least one parent per family was estimated as heterozygous according to the estimated genotypes for BBI_res_norm_pred (Figure 5D).

The parents estimated as heterozygous for γ^{pred} were the same than for BBI_res_norm_pred, and the homozygous parents were estimated systematically with the opposite allele between the two variables (Figure 6A).

For the only robust QTL detected for $\overline{Ent}_{glimm,g}$ on LG9, ‘Granny Smith’ was estimated as heterozygous, only. The other parents were estimated as homozygous for the high value allele (Figure 6B).

Concerning variables related to memory, , ‘Granny Smith’, ‘Delicious’, and ‘Belrene’ were estimated as heterozygous for $\theta_{g,00}$ for the QTL on LG10 whereas ‘Granny Smith’ and ‘Belrene’

were heterozygous for $\theta_{g,01}$, at the same QTL (Figure 6C). All other parents and founders were homozygous with the high value allele, for $\theta_{g,00}$ and low value allele for $\theta_{g,01}$.

Discussion

Efficiency of indices derived from whole AS sequence analyses at tree scale

In the present study, we estimated a number of indices assumed to capture part of the fruiting behaviour on a large set of genotypes. As previously underlined (Durand *et al.*, 2013), BBI indices and γ_g were negatively correlated and complemented one another. Indeed, γ_g allows distinguishing, among genotypes with high BBI_res_norm, irregular bearing (characterised by values of γ_g close to 0), from alternate bearing genotypes (characterised by high values of γ_g close to -1). Also the new indices $\theta_{g,m}$ estimated from the Markovian GLMM provide complementary information. Their correlations with γ_g and BBI_res_norm can be interpreted as follows: the regular genotypes (lowest values of BBI_res_norm) are those with AS that have a probability to flower above average at year t after flowering at year $t - 1$ (highest $\theta_{g,01}$ and $\theta_{g,11}$, associated with memories 01 and 11). On the contrary, $\theta_{g,00}$ and $\theta_{g,10}$ are less efficient to distinguish between regular and irregular or biennial bearing behaviours because, in any case, the probability to flower after vegetative events is always high.

The relationship between synchronism in flowering in a given year and the regularity of genotypes is highlighted by the positive correlation of entropy $\overline{Ent}_g$ with γ_g and its negative correlation with BBI_res_norm. This suggests that the genotypes with the highest synchronism (lowest value of entropy) are mostly biennial bearers (high values of BBI_res_norm, low values of γ_g). Also, the analysis of correlations between tree and axis-scale descriptors of alternation and synchronism led us to hypothesise that biennial bearing at tree scale results from the conjunction of two phenomena: synchronism in flowering between AS in a given year and biennial alternation at AS scale between consecutive years. On the contrary, regularity at tree scale results from either asynchronous alternating flowering or regular flowering at AS scale. Irregular genotypes exhibit intermediate values for every descriptor, i.e. of both tree and axis

alternation and synchronism. Thus, it can be hypothesised that these genotypes are characterised by partial biennial alternation at AS scale or strong biennial alternation with partial synchronism. However, more complex within-tree organisation of synchronisms could exist, especially at branch scale, as previously proposed for a range of fruit species (Couranjou 1983; Monselise and Goldschmidt, 1982), and has not been considered in the present study.

More insight on the bearing behaviour is also gained by introducing site and year effects and analysing the genotype by year interactions ($\eta_{g,t}$ indices). Considering site fixed effect, Montpellier had a lower flowering probability than Angers whatever the year. This may result from the absence of any thinning practices on the genotypes of XB family at Montpellier that may have hampered their flowering capacity over years (Dennis, 2000). Moreover, despite the mean values of BLUPs for the probability of flowering of all genotypes per site highlighted an opposition phase in the last three years of the experiment (2010 to 2012, Supplementary Figure S2, left), no clear characterisation of years as being ‘ON’ or ‘OFF’ could be made on the mean values per family (Supplementary Figure S2, right). This means that no climatic year could be considered as ‘ON’ or ‘OFF’. Even though critical climatic conditions such as frost (Nagy *et al.*, 2010) or crop load management can synchronise all trees of a given cultivar in a given year (Girona *et al.*, 2010), our results confirm previous finding that different genotypes can be in opposition phase for flowering in a given year, in apple (Durand *et al.*, 2013) and in olive tree (Sadok *et al.*, 2013).

The new indexes at axis scale were expected to improve prediction of bearing habits at tree scale. In this perspective, the predicted BBI_res_norm and γ_g appeared particularly robust based on the QTL detection. But, using these indices, only a 2% improvement in predictions on test samples was obtained in comparison with Durand *et al.* (2013) and the cross-validated classification error was still of 35%, despite the classification error on the whole SG dataset was improved from 33% to 21%. However the miss-classification mainly concerned the irregular genotypes whereas regular and biennial behaviours could be predicted with very high confidence. Even though new technology based on imagery may accelerate the counting of total number of inflorescences per tree (Aggelopoulou *et al.*, 2010; Gongal *et al.*, 2016), we consider that sampling axes within the tree structure with an *a posteriori* observation remains a relevant

strategy, considering time-saving for phenotyping. This strategy could help rejecting biennial and irregular genotypes during the assessment of agronomic performance of genotypes, in breeding programmes. It could also be used on training populations for implementing genomic selection to accelerate breeding progress and overcome long generation intervals and extensive phenotyping in outbred fruit tree crops that still need to (Kumar *et al.*, 2013; Muranty *et al.*, 2015).

Genetic determinisms of bearing behaviour in a multi-family population

The present study yielded five major QTL zones, four for BBI indices and auto-correlation coefficient (LG4, 5, 8 and 10) and one for entropy (LG9), which were partially common with the two previous studies performed on biennial behaviour in the SG family (Guitton *et al.*, 2011 and Durand *et al.*, 2013). QTL on LG4 colocalized with a QTL detected for BBI in (Guitton *et al.*, 2011). QTL on LG8 colocalized with QTLs detected for BBI at both tree and axis scales in Durand *et al.* (2013) and with QTL for BBI, yield and number of flowers per inflorescence in Guitton *et al.* (2011). This zone, located at 8 to 23 cM on LG8, also partially overlapped with QTLs detected on the SG family for a descriptor of tree vegetation density (Virlet *et al.*, 2015), for traits linked to bud break (Celton *et al.*, 2011; Allard *et al.*, under revision), as well as for traits involved in gas exchange and xylem conductance (Regnard *et al.*, 2007; Segura *et al.*, 2008; Lauri *et al.*, 2011). Moreover, QTL on LG10 located between 55 and 78 cM for BBI indices, γ^{ax} , $\theta_{g,00}$ and $\theta_{g,01}$ co-localized with QTLs detected for BBI, precocity and number of seeds per inflorescence in SG family (Guitton *et al.*, 2011) as well as for the percentage of bourses with one fruit on short axes in XB family (Celton *et al.*, 2014).

The QTLs detected on LG1 and LG14 by Durand *et al.* (2013) were confirmed, but in year specific interaction only, whereas the QTL on LG11 could not be confirmed. Actually, re-analysing the same dataset led us to found an inappropriate account for missing or not flowering GUs had been done that had led to a false QTL detection. This was corrected in the present study. Finally, the QTL on LG9 for entropy appeared as a new zone of importance, as it did not co-localized with any QTL previously detected for other traits.

Compared to previous studies (Guitton *et al.*, 2011; Durand *et al.*, 2013), additional QTL were detected for BBIs on LG5 and LG7. This is consistent with the heterozygous parents found for

these QTL in other families than SG. As previously suggested (Bink, 2002; Blanc *et al.*, 2006; Liu *et al.*, 2012), a higher power of detection was obtained in a multi-family context, which brought higher number of individuals, segregating regions, and alleles. It must be noticed that the QTL on LG5, located from 9 to 24 cM, co-localized with QTLs previously detected for variables linked to the tree fruiting capacity (number of fruits and fruit biomass) under soil water restriction (Virlet *et al.*, 2015). Moreover LG5 is homologous of LG10 (Velasco *et al.*, 2010; Bushakra *et al.*, 2012), also involved in the fruiting capacity of the trees.

Altogether co-localisations between present and previous results suggest that both the tree development (LG8) and fruiting capacity (LG5 and LG10) may contribute to the genetic variation of biennial bearing behaviour in apple tree. They reinforce previous assumptions regarding the combined effects of both competition among organs for nutrients, especially carbohydrates, and hormonal signals, especially GA from seeds, on biennial bearing (e.g. Chan and Cain, 1967; Dennis and Neilsen 1999). The present genetic study suggests that alleles present on the different QTL zones may lead to different combinations and therefore degree of alternation, combining effects resulting from either these effects. Further characterisation of allelic variations will be necessary for analysing in more details their relative contribution on the tree phenotype.

Finally, it must be noticed that no evident co-localization was found with QTLs detected for architectural traits measured in the first years of tree development (Segura *et al.*, 2007 and 2009). Even though qualitative notations of architectural traits collected on young trees and their linear combinations could lead to an early diagnostic on biennial bearing (Lauri, *et al.*, 2014), no evident correlations or co-localisation could be found, at least in SG population, that may suggest a genetic correlations between tree architecture and biennial behaviour. However, in the future, including the type of bourse shoot between successive floral AS could improve the characterization of genetic variations and their relationships with architectural factors. Moreover, the tools presently defined to classify genotypes could be wisely used before investigating their physiological behaviour in order to decipher the putative role of fruits and carbon economy on the inhibition of floral induction.

Potential use in breeding of genitors or founders

The multi-family strategy highlighted parents and founders transmitting favourable alleles. Among them, Delicious and Granny Smith belong to the most common cultivars (Gross *et al.*, 2014) whereas others are less or even not characterised for their agronomic performance. Considering the strong evidence of QTLs for BBI_res_norm_pred, γ^{pred} and entropy, these traits could be targeted to obtain appropriate phenotypes, with low BBI_res_norm_pred values combined to medium or high values for γ^{pred} and high Entropy values. Indeed, trees with low Entropy values would be desirable because regular in each year and at all locations within the trees, but such trees were not observed in the studied segregating populations. As underlined by Samach and Smith (2013), the evolutionary advantage of masting (i.e. synchronicity of flowering at tree and population levels) remains questionable. Our results suggest that flowering synchronicity at the tree level could be associated with regularity probably because such trait combination would lead to over-cropping and may exhibit major drawbacks in an agronomic context. Knowing if flowering desynchronization has been selected during the apple domestication process remains an open question of interest.

For BBI_res_norm_pred, Delicious is expected to transmit favourable alleles at the four QTLs with strong evidence to half their progenies since it was estimated as heterozygous. This parent was also responsible for segregation in SG at most QTL also for γ^{pred} and entropy. Therefore, SG segregation appears consistent with the predicted classification for this family which displayed equivalent classes in terms of individual number (27% regular, 30% alternate and 43% irregular). It also confirms the interest of ‘Delicious’ as genitor, consistently with its recurrent use (Noiton and Alspach 1996). The situation is more complex for the other parents and founders for which both favourable and unfavourable alleles could be transmitted. As a similar reasoning could be performed for γ^{pred} and entropy, for which no clear evidence for a higher value of a given parent or founder compared to the others could be drawn. For instance, ‘Granny Smith’ which is known to be a regular genotype (Lespinnasse, 1977) was shown to transmit favourable alleles for BBI_res_norm_pred on LG4 but unfavourable alleles on LG5, favourable alleles with high values at three QTL over four for γ^{pred} , and favourable allele to half the progeny only, at the LG9 QTL for entropy. These results underline the complex genetic architecture and highly polygenic

control of the considered traits. Therefore, a more integrated view based on the calculation of genome wide genetic values will be necessary in the future. Indeed, the QTL genotype probabilities, together with QTL intensity and QTL effect sizes, could be used to predict the genomic breeding value per individual by summing the positional breeding values along the genome (Meuwissen *et al.*, 2001). Moreover, the QTL mapping strategy allowed the detection of loci with high effects, although many small-effect QTLs could contribute to increase the total variance explained. In this perspective, genomic selection models would be complementary to QTL analyses to evaluate the genetic value of individuals by summing allelic effects at each position of the genome (Jannink *et al.*, 2010; Kumar *et al.*, 2012; Muranty *et al.*, 2015).

References

- Aggelopoulou, A. D., D. Bochtis, S. Fountas, K. C. Swain, T. A. Gemtos, and G. D. Nanos. 2010. "Yield Prediction in Apple Orchards Based on Image Processing." *Precision Agriculture* 12 (3): 448-56. doi:10.1007/s11119-010-9187-0.
- Antanaviciute, Laima, Felicidad Fernández-Fernández, Johannes Jansen, Elisa Banchi, Katherine M. Evans, Roberto Viola, Riccardo Velasco, Jim M. Dunwell, Michela Troglio, and Daniel J. Sargent. 2012. "Development of a Dense SNP-Based Linkage Map of an Apple Rootstock Progeny Using the Malus Infinium Whole Genome Genotyping Array." *BMC Genomics* 13: 203. doi:10.1186/1471-2164-13-203.
- Bangerth, K.F. 2009. "Floral Induction in Mature, Perennial Angiosperm Fruit Trees: Similarities and Discrepancies with Annual/biennial Plants and the Involvement of Plant Hormones." *Scientia Horticulturae* 122 (2): 153-63. doi:10.1016/j.scienta.2009.06.014.
- Bianco, Luca, Alessandro Cestaro, Daniel James Sargent, Elisa Banchi, Sophia Derdak, Mario Di Guardo, Silvio Salvi, *et al.* 2014. "Development and Validation of a 20K Single Nucleotide Polymorphism (SNP) Whole Genome Genotyping Array for Apple (*Malus × Domestica* Borkh)." *PLoS ONE* 9 (10): e110377. doi:10.1371/journal.pone.0110377.
- Bink, Marco C. A. M. 2002. "On Flexible Finite Polygenic Models for Multiple-Trait Evaluation." *Genetical Research* 80 (3): 245-56. doi:10.1017/S0016672302005906.
- Bink, M. C. a. M., M. P. Boer, C. J. F. ter Braak, J. Jansen, R. E. Voorrips, and W. E. van de Weg. 2007. "Bayesian Analysis of Complex Traits in Pedigreed Plant Populations." *Euphytica* 161 (1-2): 85-96. doi:10.1007/s10681-007-9516-1.
- Bink, M. C. a. M., J. Jansen, M. Madduri, R. E. Voorrips, C.-E. Durel, A. B. Kouassi, F. Laurens, *et al.* 2014. "Bayesian QTL Analyses Using Pedigreed Families of an Outcrossing Species, with Application to Fruit Firmness in Apple." *Theoretical and Applied Genetics* 127 (5): 1073-90. doi:10.1007/s00122-014-2281-3.
- Blanc, G., A. Charcosset, B. Mangin, A. Gallais, and L. Moreau. 2006. "Connected Populations for Detecting Quantitative Trait Loci and Testing for Epistasis: An Application in Maize." *Theoretical and Applied Genetics* 113 (2): 206-24. doi:10.1007/s00122-006-0287-1.
- Bushakra, Jill M., Daniel J. Sargent, Antonio Cabrera, Ross Crowhurst, Elena Lopez Girona, Riccardo Velasco, V. Vaughan Symonds, *et al.* 2012. "Rosaceae Conserved Orthologous Set (RosCOS) Markers as a Tool to Assess Genome Synteny between *Malus* and *Fragaria*." *Tree Genetics & Genomes* 8 (4): 643-58. doi:10.1007/s11295-011-0450-y.
- Celton, Jean-Marc, Jean-Jacques Kelner, Sébastien Martinez, Abdel Behti, Amina Khelifi Touhami, Marie José James, Charles-Eric Durel, François Laurens, and Evelyne Costes. 2014. "Fruit Self-Thinning: A Trait to Consider for Genetic Improvement of Apple Tree." *PLoS ONE* 9 (3): e91016. doi:10.1371/journal.pone.0091016.

Celton, J-M., S. Martinez, M-J. Jammes, A. Bechti, S. Salvi, J-M. Legave, and E. Costes. 2011. "Deciphering the Genetic Determinism of Bud Phenology in Apple Progenies: A New Insight into Chilling and Heat Requirement Effects on Flowering Dates and Positional Candidate Genes." *New Phytologist* 192 (2): 378-92. doi:10.1111/j.1469-8137.2011.03823.x.

Chagné, David, Ross N. Crowhurst, Michela Troggio, Mark W. Davey, Barbara Gilmore, Cindy Lawley, Stijn Vanderzande, *et al.* 2012. "Genome-Wide SNP Detection, Validation, and Development of an 8K SNP Array for Apple." *PLoS ONE* 7 (2): e31745. doi:10.1371/journal.pone.0031745.

Costes, Evelyne, and Yann Guédon. 2011. "Deciphering the Ontogeny of a Sympodial Tree." *Trees* 26 (3): 865-79. doi:10.1007/s00468-011-0661-8.

Couranjou, J. 1983. "Facteurs variétaux de l'alternance des arbres fruitiers. Résultats de quinze années de recherches chez le prunier domestique." *Fruits* 38 (10): 705-28.

Dennis, F. G. Jr. 2000. "The History of Fruit Thinning." *Plant Growth Regulation* 31 (1-2): 1-16. doi:10.1023/A:1006330009160.

Dennis, Frank G., and John C. Neilsen. 1999. "Physiological Factors Affecting Biennial Bearing in Tree Fruit: The Role of Seeds in Apple." *HortTechnology* 9 (3): 317-22.

Durand, Jean-Baptiste, Baptiste Guitton, Jean Peyhardi, Yan Holtz, Yann Guédon, Catherine Trottier, and Evelyne Costes. 2013. "New Insights for Estimating the Genetic Value of Segregating Apple Progenies for Irregular Bearing during the First Years of Tree Production." *Journal of Experimental Botany*, October, ert297. doi:10.1093/jxb/ert297.

Girona, J., M.H. Behboudian, M. Mata, J. Del Campo, and J. Marsal. 2010. "Exploring Six Reduced Irrigation Options under Water Shortage for 'Golden Smoothee' Apple: Responses of Yield Components over Three Years." *Agricultural Water Management* 98 (2): 370-75. doi:10.1016/j.agwat.2010.09.011.

Gongal, A., A. Silwal, S. Amatya, M. Karkee, Q. Zhang, and K. Lewis. 2016. "Apple Crop-Load Estimation with over-the-Row Machine Vision System." *Computers and Electronics in Agriculture* 120 (January): 26-35. doi:10.1016/j.compag.2015.10.022.

Gross, Briana L., Adam D. Henk, Christopher M. Richards, Gennaro Fazio, and Gayle M. Volk. 2014. "Genetic Diversity in *Malus ×domestica* (Rosaceae) through Time in Response to Domestication." *American Journal of Botany* 101 (10): 1770-79. doi:10.3732/ajb.1400297.

Guardo, Mario Di, Diego Micheletti, Luca Bianco, Herma J. J. Koehorst-van Putten, Sara Longhi, Fabrizio Costa, Maria J. Aranzana, *et al.* 2015. "ASSIsT: An Automatic SNP Scoring Tool for in- and Outbreeding Species." *Bioinformatics*, August, btv446. doi:10.1093/bioinformatics/btv446.

Guitton, Baptiste, Jean-Jacques Kelner, Riccardo Velasco, Susan E. Gardiner, David Chagné, and Evelyne Costes. 2011. "Genetic Control of Biennial Bearing in Apple." *Journal of Experimental Botany*, September, err261. doi:10.1093/jxb/err261.

- Hanke, Magda-Viola, Henryk Flachowsky, Andreas Peil, and Conny Hättasch. 2007. “No Flower No Fruit—genetic Potentials to Trigger Flowering in Fruit Trees.” *Genes Genomes Genomics* 1 (1): 1-20.
- Jannink, Jean-Luc, Aaron J. Lorenz, and Hiroyoshi Iwata. 2010. “Genomic Selection in Plant Breeding: From Theory to Practice.” *Briefings in Functional Genomics*, February, elq001. doi:10.1093/bfpg/elq001.
- Jonkers, H. 1979. “Biennial Bearing in Apple and Pear: A Literature Survey.” *Scientia Horticulturae* 11 (4): 303-17. doi:10.1016/0304-4238(79)90015-3.
- Josse, Julie, and François Husson. 2012. “Selecting the Number of Components in Principal Component Analysis Using Cross-Validation Approximations.” *Computational Statistics & Data Analysis* 56 (6): 1869-79. doi:10.1016/j.csda.2011.11.012.
- Kass, Robert E., and Adrian E. Raftery. 1995. “Bayes Factors.” *Journal of the American Statistical Association* 90 (430): 773-95. doi:10.1080/01621459.1995.10476572.
- Kumar, Satish, Dorian J. Garrick, Marco CAM Bink, Claire Whitworth, David Chagné, and Richard K. Volz. 2013. “Novel Genomic Approaches Unravel Genetic Architecture of Complex Traits in Apple.” *BMC Genomics* 14: 393. doi:10.1186/1471-2164-14-393.
- Lauri, P. É., F. Combe, and L. Brun. 2014. “Regular Bearing in the Apple - Architectural Basis for an Early Diagnosis on the Young Tree.” *Scientia Horticulturae* 174 (July): 10-16. doi:10.1016/j.scienta.2014.05.001.
- Lauri, Pierre-Éric, Olivier Gorza, Hervé Cochard, Sébastien Martinez, Jean-Marc Celton, Véronique Ripetti, Marc Lartaud, Xavier Bry, Catherine Trottier, and Evelyne Costes. 2011. “Genetic Determinism of Anatomical and Hydraulic Traits within an Apple Progeny.” *Plant, Cell & Environment* 34 (8): 1276-90. doi:10.1111/j.1365-3040.2011.02328.x.
- Liu, Wenxin, Jochen C. Reif, Nicolas Ranc, Giovanni Della Porta, and Tobias Würschum. 2012. “Comparison of Biometrical Approaches for QTL Detection in Multiple Segregating Families.” *Theoretical and Applied Genetics* 125 (5): 987-98. doi:10.1007/s00122-012-1889-4.
- Luan, Tu, John A. Woolliams, Jørgen Ødegård, Marlies Dolezal, Sergio I. Roman-Ponce, Alessandro Bagnato, and Theo HE Meuwissen. 2012. “The Importance of Identity-by-State Information for the Accuracy of Genomic Selection.” *Genetics Selection Evolution* 44 (1): 1-7. doi:10.1186/1297-9686-44-28.
- Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. “Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps.” *Genetics* 157 (4): 1819-29.
- Monselise, S. P., and E. E. Goldschmidt. 1982. “Alternate Bearing in Fruit Trees.” In *Horticultural Reviews*, edited by Jules Janick, 128-73. Palgrave Macmillan UK. http://link.springer.com/chapter/10.1007/978-1-349-06519-6_5.

- Muranty, Hélène, Michela Troggio, Inès Ben Sadok, Mehdi Al Rifai, Annemarie Auwerkerken, Elisa Banchi, Riccardo Velasco, *et al.* 2015. "Accuracy and Responses of Genomic Selection on Key Traits in Apple Breeding." *Horticulture Research* 2 (December): 15060. doi:10.1038/hortres.2015.60.
- Nagy, P. T., Z. Szabo, J. Nyeki, and M. C. Dussi. 2010. "Effects of Frost-Induced Biennial Bearing on Nutrient Availability and Fruit Disorders in an Integrated Apple Orchard." In *Xi International Symposium on Plant Bioregulators in Fruit Production*, edited by G. Costa, 884:753-57. Leuven 1: Int Soc Horticultural Science.
- Noiton, Dominique A. M., and Peter A. Alspach. 1996. "Founding Clones, Inbreeding, Coancestry, and Status Number of Modern Apple Cultivars." *Journal of the American Society for Horticultural Science* 121 (5): 773-82.
- Pauly, Laurence, Sandrine Flajoulot, Jérôme Garon, Bernadette Julier, Vincent Béguier, and Philippe Barre. 2012. "Detection of Favorable Alleles for Plant Height and Crown Rust Tolerance in Three Connected Populations of Perennial Ryegrass (*Lolium Perenne* L.)." *Theoretical and Applied Genetics* 124 (6): 1139-53. doi:10.1007/s00122-011-1775-5.
- Pellerin, Brian P., Deborah Buszard, Alex Georgallas, and Richard J. Nowakowski. 2012. "A Novel Framework to Consider Endogenous Hormonal Control of Apple Tree Flowering." *HortScience* 47 (5): 589-92.
- Pellerin, Brian P., Deborah Buszard, David Iron, Charles G. Embree, Richard P. Marini, Douglas S. Nichols, Gerald H. Neilsen, and Denise Neilsen. 2011. "A Theory of Blossom Thinning to Consider Maximum Annual Flower Bud Numbers on Biennial Apple Trees." *HortScience* 46 (1): 40-42.
- Sadok, Inès Ben, Jean-Marc Celton, Laila Essalouh, Amal Zine El Aabidine, Gilbert Garcia, Sebastien Martinez, Naziha Grati-Kamoun, Ahmed Rebai, Evelyne Costes, and Bouchaib Khadari. 2013. "QTL Mapping of Flowering and Fruiting Traits in Olive." *PLoS ONE* 8 (5): e62831. doi:10.1371/journal.pone.0062831.
- Samach, Alon, and Harley M. Smith. 2013. "Constraints to Obtaining Consistent Annual Yields in Perennials. II: Environment and Fruit Load Affect Induction of Flowering." *Plant Science* 207 (June): 168-76. doi:10.1016/j.plantsci.2013.02.006.
- Segura, V., C. Cilas, F. Laurens, and E. Costes. 2006. "Phenotyping Progenies for Complex Architectural Traits: A Strategy for 1-Year-Old Apple Trees (*Malus X Domestica* Borkh.)." *Tree Genetics & Genomes* 2 (3): 140-51. doi:10.1007/s11295-006-0037-1.
- Singh, L. B. 1948. "Studies in Biennial Bearing: III. Growth Studies in 'ON' and 'Off' Year Trees." *Journal of Horticultural Science* 24 (2): 123-48. doi:10.1080/00221589.1948.11513683.
- Van de Weg, W. Eric, Mario Di Guardo, Herma Koehorst-van Putten, Sara Longhi, Y. Noordijk, Helene Muranty, Elisa Banchi, *et al.* 2013. *A Pipeline for Robust Marker Calling from Infinium SNP Arrays for Diploid Crops*. <https://hal.archives-ouvertes.fr/hal-01209967>.

Velasco, Riccardo, Andrey Zharkikh, Jason Affourtit, Amit Dhingra, Alessandro Cestaro, Ananth Kalyanaraman, Paolo Fontana, *et al.* 2010. "The Genome of the Domesticated Apple (*Malus × Domestica* Borkh.)." *Nature Genetics* 42 (10): 833-39. doi:10.1038/ng.654.

Virlet, Nicolas, Evelyne Costes, Sébastien Martinez, Jean-Jacques Kelner, and Jean-Luc Regnard. 2015. "Multispectral Airborne Imagery in the Field Reveals Genetic Determinisms of Morphological and Transpiration Traits of an Apple Tree Hybrid Population in Response to Water Deficit." *Journal of Experimental Botany*, July, erv355. doi:10.1093/jxb/erv355.

Tables and Figures

Table 1. Estimates of fixed effects and variances (with p -values of the significance tests) of the mixed model estimating the probability of flowering at axis scale in the five F1 families.

Nature	Parameter	Estimate	p-value
Intercept	λ	-0.36	0.19
Fixed effects : site	π_M	-1.05	1e-16
	μ_{10}	1.40	1e-16
Fixed effects : memory	μ_{01}	-0.33	1e-07
	μ_{11}	0.31	1e-06
	τ_θ^2	0.54	-
Variances	τ_η^2	2.33	-
	τ_ζ^2	0.22	-
Fixed effects : year	ϕ_{2007}	1.46	1e-7
	ϕ_{2008}	1.81	1e-11
	ϕ_{2009}	1.42	1e-7
	ϕ_{2010}	0.55	0.04
	ϕ_{2011}	1.36	1e-7
	ϕ_{2012}	0.49	0.07

Table 2. Empirical standard deviations of random effects for each kind of interaction of the mixed model estimating the probability of flowering at axis scale in the five F1 families.

Nature of interaction	Parameter	Std. Dev.
Genotype x memory	$\theta_{g,00}$	0.45
	$\theta_{g,10}$	0.35
	$\theta_{g,01}$	0.42
	$\theta_{g,11}$	0.36
Genotype x year	$\eta_{g,2006}$	0.81
	$\eta_{g,2007}$	0.88
	$\eta_{g,2008}$	1.00
	$\eta_{g,2009}$	1.09
	$\eta_{g,2010}$	1.11
	$\eta_{g,2011}$	1.26
	$\eta_{g,2012}$	1.73

Table 3. Correlation coefficients between indices at whole tree scale, and indices and BLUPs at axis scale, with 95% confidence intervals, in the F1 SG family.

	BBI_res_norm	γ_g
γ_g	-0.66 (-0.75, -0.54)	1
BBI_res_norm_ax	0.72 (0.61, 0.80)	-0.61 (-0.72, -0.49)
γ^{ax}	-0.55 (-0.67, -0.41)	0.51 (0.36, 0.63)
$\overline{Ent}_g$	-0.46 (-0.59, -0.30)	0.34 (0.17, 0.49)
$\overline{Ent}_{glmm,g}$	-0.19 (-0.36, -0.01)	0.09 (-0.09, 0.27)
$\theta_{g,01}$	-0.55 (-0.66, -0.41)	0.50 (0.35, 0.62)
$\theta_{g,11}$	-0.21 (-0.40, -0.01)	0.17 (-0.03, 0.36)
$\theta_{g,00}$	0.23 (0.04, 0.39)	-0.22 (-0.39, -0.04)
$\theta_{g,10}$	0.00 (-0.21, 0.21)	-0.07 (-0.28, 0.14)

Table 4. Contingency table for the number of genotypes of class c1 assigned to class c2 by NN on local indices. Cluster 1 corresponds to regular genotypes, cluster 2 to biennial bearing and cluster 3 to irregular genotypes.

		Predicted class c2		
		1	2	3
True class c1	1	33	0	3
	2	0	22	9
	3	0	14	41

Table 5. Parameters associated with the QTL detected for genotypic BLUPs. The first column indicates the variable concerned, the following columns indicate the LG where the QTL is located, $2\ln(\text{BF})$ value at LG scale, $2\ln(\text{BF})$ value at bin scale, the position of the QTL in cM, the position of the QTL peak, its additive effect, the frequency of positive allele and percentage of variance explained, respectively. Only $2\ln(\text{BF})$ values corresponding to the comparison of a model with 0 QTL to a model with 1 QTL are presented. QTLs that appear in bold are QTL with a strong evidence for presence, i.e. with a $2*\ln(\text{BF})$ value higher than 5.

	LG	$2\ln(\text{BF})_{\text{LG}}$	$\max 2\ln(\text{BF})_{\text{bin}}$	pos (cM)	Peak (cM)	add_ef	fq	%var
BBI_ax	4	2,3	4,3	54-65	60-61	0,13	0,32	5
	5	7,5	7,3	9-24	21-22	0,14	0,56	6,4
	7	6,2	5,4	57-94	75-76	0,19	0,66	10,7
	8	2,8	3,7	8-25	14-15	0,11	0,4	4,3
	10	9	8,2	55-78	75-76	0,19	0,64	11,4
BBI_norm_ax	4	7,4	7,3	30-41	34-35	0,32	0,15	10,9
	5	6,1	6,1	3-22	9-10	0,19	0,47	7,7
	7	2,6	4	61-92	73-74	0,2	0,63	8,2
	8	2,8	5,3	8-23	14-15	0,19	0,41	8,2
	10	6,6	7,6	57-78	75-76	0,21	0,6	9,1
BBI_res_norm_ax	4	10,6	8,94	32-41	36-37	0,39	0,12	11,5
	5	5	6	3-24	9-10	0,19	0,5	6,9
	7	1,9	3,1	65-90	71-72	0,18	0,69	5,4
	8	2,1	4,6	10-23	14-15	0,2	0,39	7,7
	10	4,4	7,4	57-78	75-76	0,21	0,58	7,7
BBI_res_norm_pred	12	2	4,1	34-43	38-39	0,18	0,39	5,8
	4	6,3	6,8	32-47	36-37	0,13	0,35	13,3
	5	5,8	8	3-24	21-22	0,1	0,56	8,3
	8	8,4	9	8-23	14-15	0,12	0,42	11,7
	10	9,3	8	57-78	75-76	0,11	0,52	10
γ^{ax}	4	2,7	4,4	34-45	34-35	0,15	0,59	5,3
	5	2,6	4,7	21-36	21-22	0,18	0,41	5,3
	10	3,1	6,4	59-76	75-76	0,18	0,45	5,3
$\gamma^{ax}_{\text{pred}}$	4	6,7	6,8	32-45	36-37	0,11	0,64	12
	5	6,2	8,3	3-30	21-22	0,1	0,42	10
	8	9,4	8,9	8-23	14-15	0,11	0,59	12
	10	9,5	7,8	57-78	75-76	0,1	0,5	10
Ent _g	12	3,1	5,4	6-23	20-21	0,05	0,43	5
	1	3,7	6,4	44-59	48-49	0,05	0,63	10
	7	2,4	4,3	39-56	51-52	0,05	0,54	10
	9	8	7,4	19-34	25-26	0,07	0,75	20
Ent _{gImm,g}	15	4,5	7,1	46-69	56-57	0,05	0,65	10
	17	3,1	5	0-27	0-1	0,05	0,5	10
$\theta_{g,00}$	10	5,4	7,7	59-76	75-76	0,22	0,66	10
	11	3,2	4,8	36-55	46-47	0,28	0,35	15
$\theta_{g,01}$	4	3,8	6	8-23	16-17	0,16	0,38	5,9
	10	8,1	7,3	49-68	53-54	0,24	0,21	11,8
$\theta_{g,10}$	no QTL							
$\theta_{g,11}$	10	1,9	4,8	65-76	75-76	0,14	0,43	6,9
$\eta_{g,2006}$	12	2,5	3,6	34-47	44-45	0,58	0,52	14,4
$\eta_{g,2007}$	no QTL							
$\eta_{g,2008}$	6	2,3	4,8	48-61	54-55	0,41	0,6	7,2
$\eta_{g,2009}$	no QTL							
$\eta_{g,2010}$	no QTL							
$\eta_{g,2011}$	no QTL							
$\eta_{g,2012}$	no QTL							

Result Chapter 2

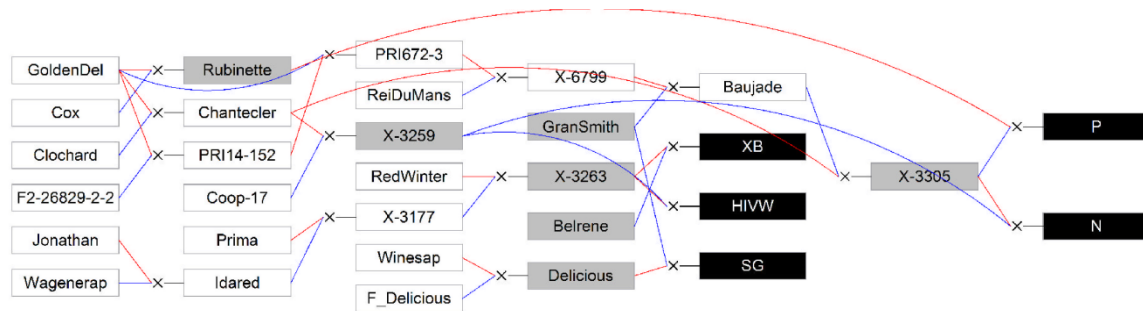
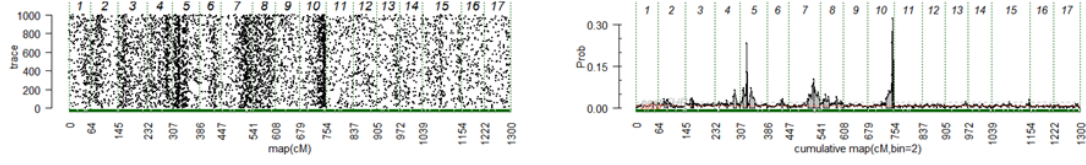
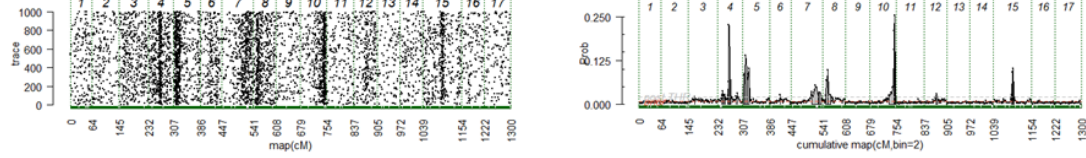


Figure 1. Genetic relationships between the five studied full-sib families (XB, HIVW, SG, N and P; represented by black boxes) and their parents (represented by grey boxes) and founders or other members of the pedigree (represented by white boxes). Blue lines link the father to its progenies and red lines link the mother to its progenies. GoldenDel = Golden Delicious, ReiDuMans = Reinette du Mans, Wagenerap = Wagenerapfel; see text for family abbreviations and supporting information 2 for other abbreviations used in the pedigree.

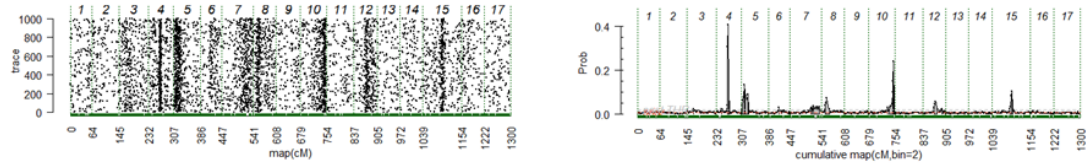
(A) BBI_ax



(B) BBI_norm_ax



(C) BBI_res_norm_ax



(D) γ^{ax}

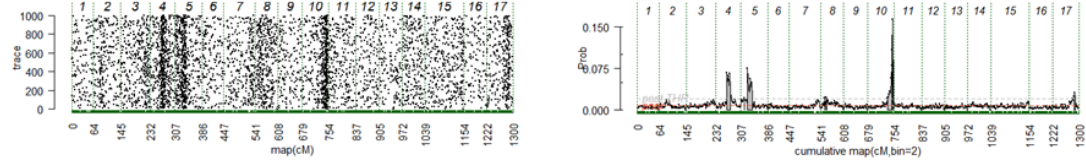


Figure 2. Trace plot of QTL presence along the genome and across iterations (left panel) posterior probability of QTL position along genome (right panel), the beginning and the end of the chromosomes are represented by vertical dashed lines. The variables displayed were (A) BBI_ax , (B) BBI_norm_ax, (C) BBI_res_norm_ax, and (D) Auto-regressive coefficient (γ^{ax}).

Result Chapter 2

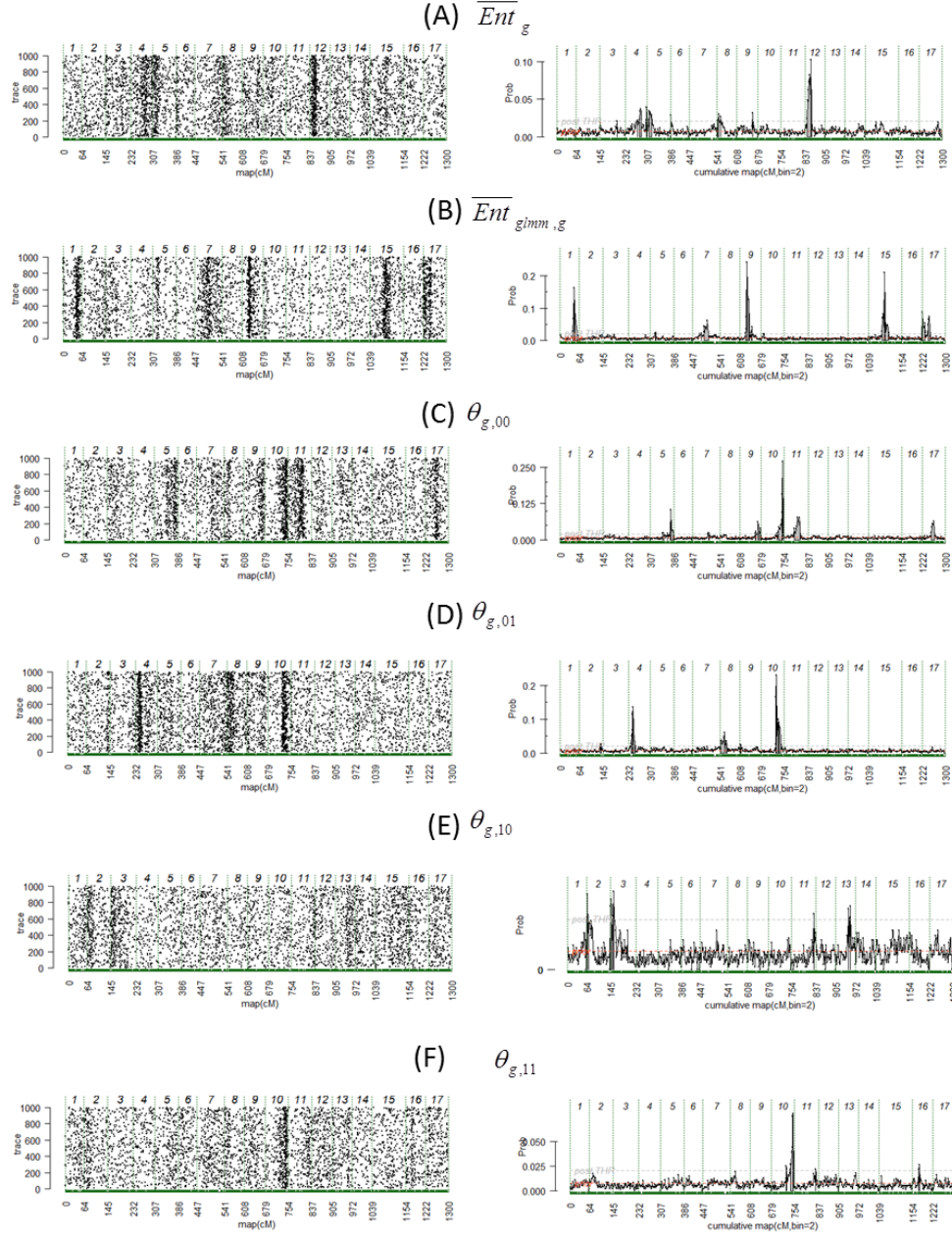


Figure 3. Trace plot of QTL presence along the genome and across iterations (left panel) posterior probability of QTL position along genome (right panel), the beginning and the end of the chromosomes are represented by vertical dashed lines. The variables displayed are entropy: (A) $\overline{Ent}_g$, (B) $\overline{Ent}_{g|mm,g}$ and interactions genotype x memories: (C) $\theta_{g,00}$, (D) $\theta_{g,01}$, and (E) $\theta_{g,10}$ (F) $\theta_{g,11}$.

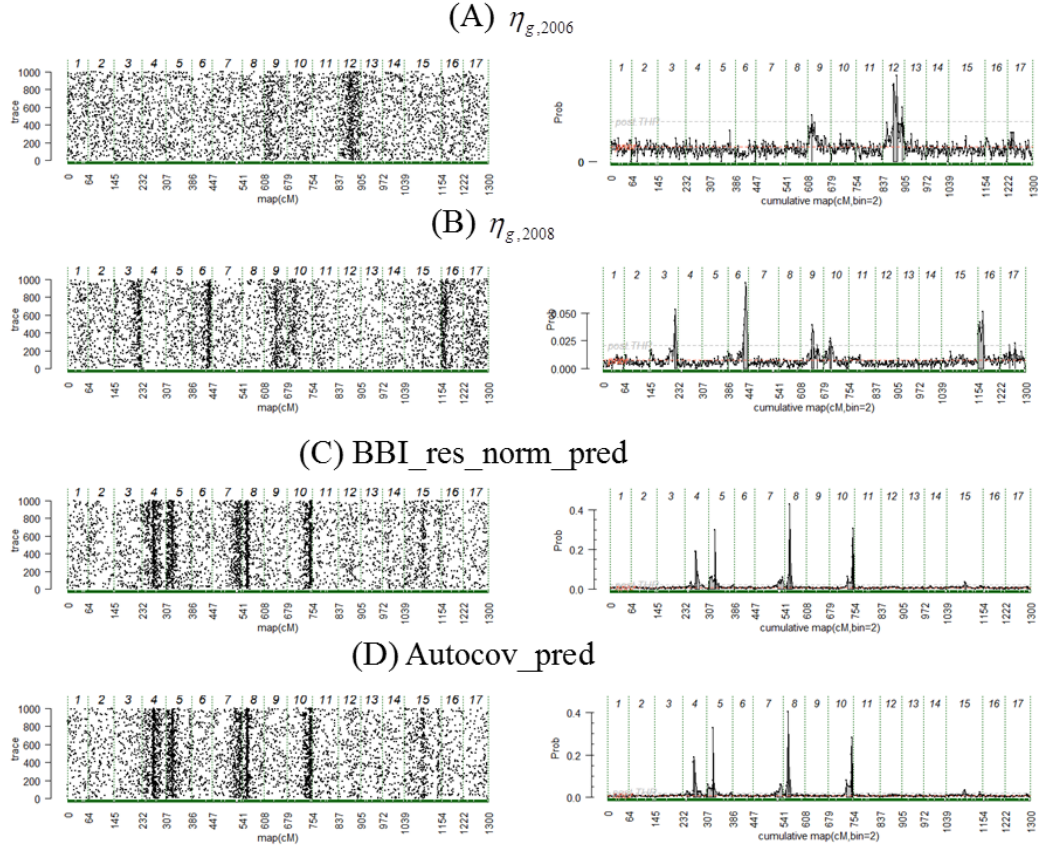


Figure 4. Trace plot of QTL presence along the genome and across iterations (left panel) posterior probability of QTL position along genome (right panel), the beginning and the end of the chromosomes are represented by vertical dashed lines. The variables displayed are the interactions genotype x year: (A) $\eta_{g,2006}$, (B) $\eta_{g,2008}$, and the predicted values of BBI and autocorrelation: (C) BBI_res_norm_pred, and (D) Autocov_Pred.

Result Chapter 2

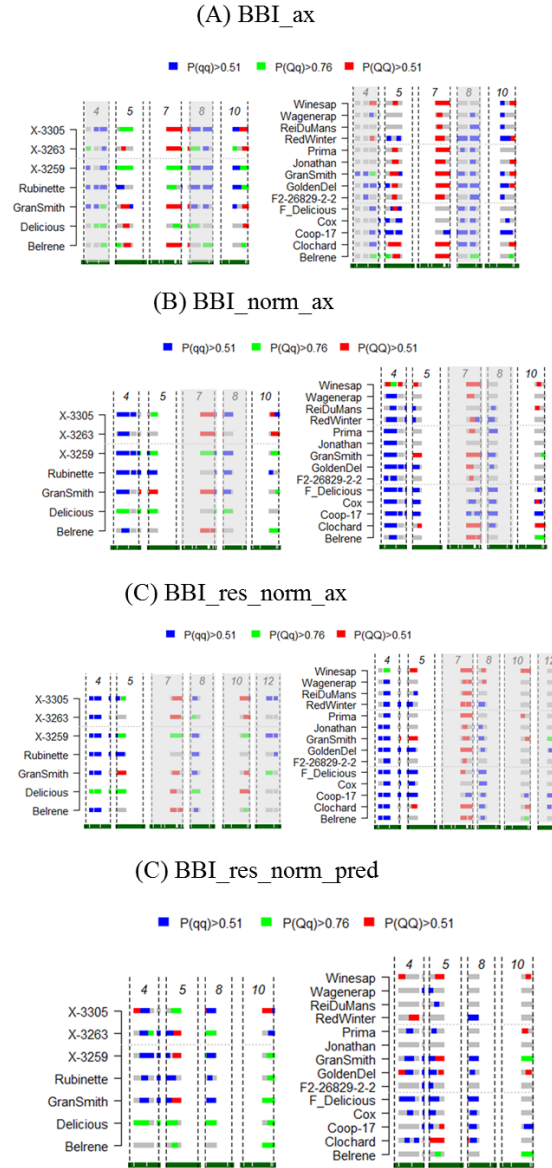


Figure 5. Posterior estimates of QTL genotype probabilities. Estimates are plotted for main QTLs for (A) BBI_ax (B) BBI_norm_ax (C) and BBI_res_norm_ax (D) BBI_res_norm_pred. The beginning and the end of the linkage group (LG) are represented by vertical dashed lines, the number of the LG is given at the top of the figure. For the QTLs with only positive evidence, the LG is highlighted in grey. Blue, green and red colors indicate positive evidence for QTL genotypes qq, qQ and QQ, respectively; q and Q refer to low and high phenotypic values, respectively. Grey colors indicate ignorance for a given genotype. Estimated genotypes of the parents and founders are presented on the left and right side, respectively.

Result Chapter 2

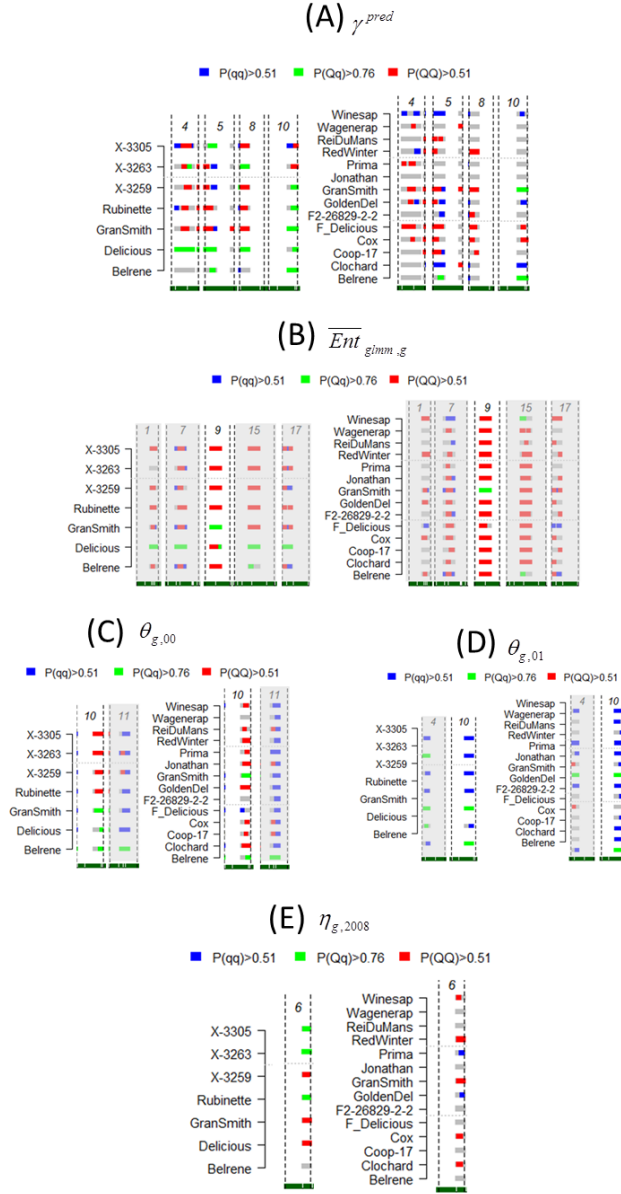


Figure 6. Posterior estimates of QTL genotype probabilities. Estimates are plotted for main regions for (A) γ^{pred} , (B) $\overline{Ent}_{gImm,g}$, (C) $\theta_{g,00}$, (D) $\theta_{g,01}$, (E) $\eta_{g,2008}$. The beginning and the end of the linkage group (LG) are represented by vertical dashed lines, the number of the LG is given at the top of the figure. For the QTLs with only positive evidence, the LG is highlighted in grey. Blue, green and red colors indicate positive evidence for QTL genotypes qq, qQ and QQ, respectively; q and Q refer to low and high phenotypic values, respectively. Grey colors indicate ignorance for a given genotype. Estimated genotypes of the parents and founders are presented on the left and right side, respectively.

Genetic determinisms of flowering regularity in apple tree: a multi-family QTL detection based on statistical indices extracted from branch sequence analyses

Jean-Baptiste Durand^{1,2(*)}, Alix Allard³, Baptiste Guitton³, Jean Peyhardi^{2,4}, Yann Guédon², Catherine Trottier⁴, E. van de Weg⁵, Marco Bink⁶, Evelyne Costes³

Corresponding author: Jean-Baptiste Durand (+33 4 76 63 57 09), jean-baptiste.durand@imag.fr

Supplementary Information

Figures

Figure S1. Histograms of the sequence lengths for the SG (a), XB (b), HIVW (c), P (d) and N (e) families.

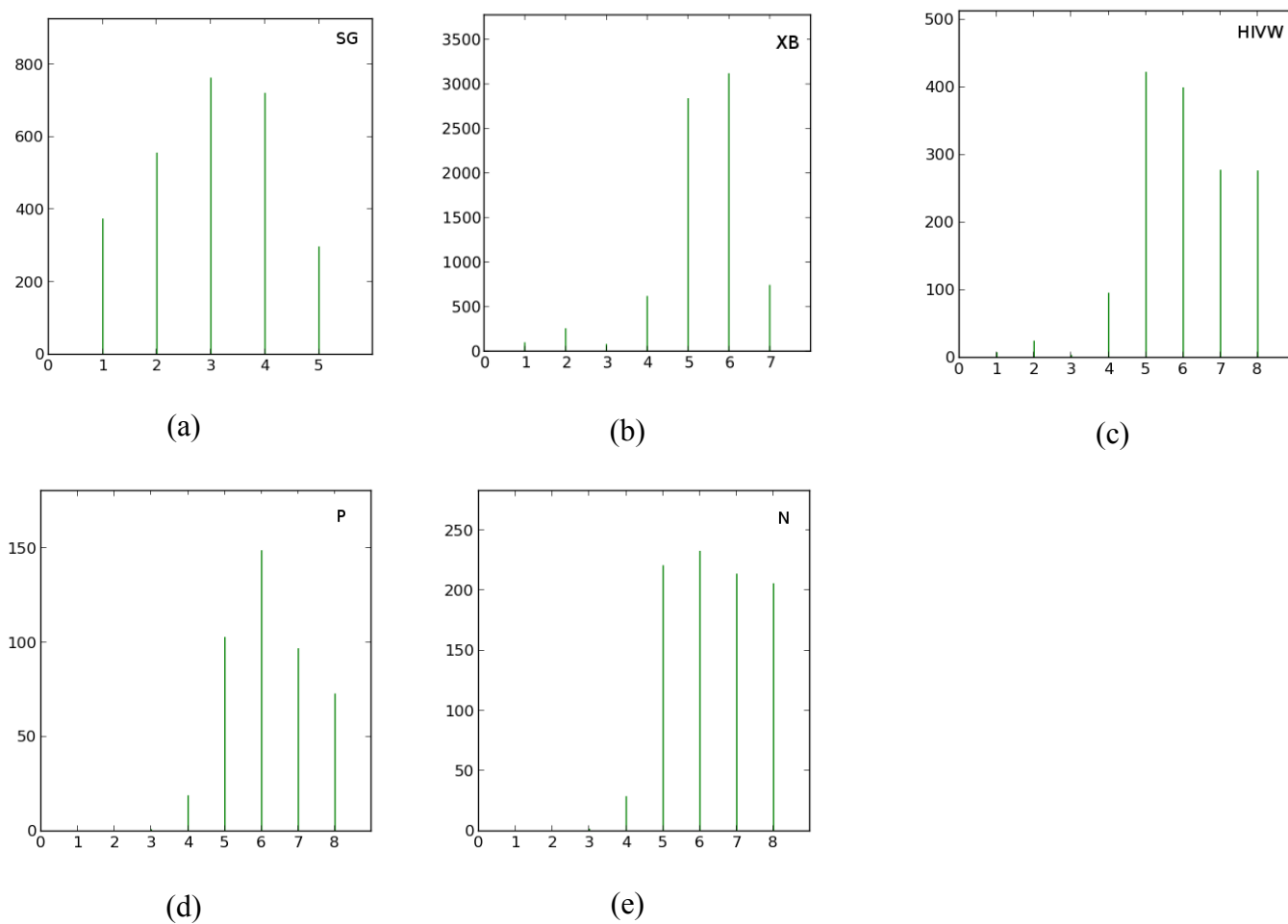
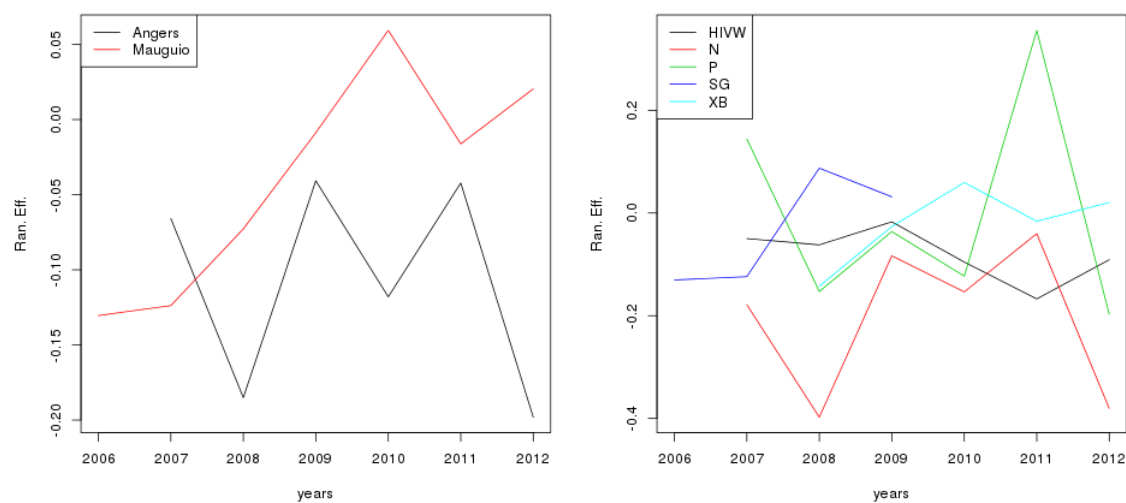


Figure S2. Empirical mean values of BLUPs for genotype and year interactions depending on the site (left) or the family (right).



Tables

Table T1. Correlation coefficients between the $\theta_{g,m}$ BLUP indices and the other indices and BLUPs at axis scale, with 95% confidence intervals, combining in the five apple tree families.

	Genotype x memory 01 $\theta_{g,01}$	Genotype x memory 11 $\theta_{g,11}$	Genotype x memory 00 $\theta_{g,00}$	Genotype x memory 10 $\theta_{g,10}$
BBI_res_norm_ax	-0.45 (-0.51, -0.39)	-0.38 (-0.44, -0.32)	0.07 (-0.01, 0.15)	0.23 (0.16, 0.30)
γ^{ax}	0.44 (0.38, 0.50)	0.37 (0.30, 0.43)	-0.19 (-0.26, -0.11)	-0.31 (-0.38, -0.24)
$\overline{Ent}_g$	0.17 (0.10, 0.24)	-0.16 (-0.24, -0.09)	0.10 (0.02, 0.18)	-0.24 (-0.31, -0.16)
$\overline{Ent}_{glimm,g}$	0.11 (0.03, 0.18)	0.01 (-0.06, 0.09)	-0.09 (-0.17, -0.01)	-0.14 (-0.21, -0.07)
$\theta_{g,01}$	1	0.10 (0.03, 0.18)	-0.37 (-0.44, -0.30)	-0.39 (-0.46, -0.33)
$\theta_{g,11}$		1	-0.39 (-0.46, -0.32)	-0.24 (-0.31, -0.17)
$\theta_{g,00}$			1	-0.12 (-0.20, -0.04)
	γ^{ax}	$\overline{Ent}_g$	$\overline{Ent}_{glimm,g}$	
BBI_res_norm_ax	-0.66 (-0.70, -0.62)	-0.28 (-0.35, -0.21)	-0.42 (-0.48, -0.36)	
γ^{ax}	1	0.12 (0.05, 0.19)	0.23 (0.16, 0.30)	
$\overline{Ent}_g$		1	0.35 (0.28, 0.41)	

Table T2. P-values of the ANOVA tests for the effect of the class of bearing behaviours on axis-scale indices in the F1 SG family. B^{ax} is a shorthand for BBI_res_norm_ax.

Index	$\theta_{g,01}$	$\theta_{g,11}$	$\theta_{g,00}$	$\theta_{g,10}$	$\overline{Ent}_g$	$\overline{Ent}_{glm,g}$	B^{ax}	γ^{ax}
p-value of ANOVA	7.58e-09	0.127	0.0418	0.895	2.06e-08	0.0143	2e-16	2.75e-09

M. Supplementary Description of indices, models and methods

Reminders: indices for characterization of regular, irregular and alternate bearing genotypes.

It was shown in Durand *et al.* (2013) that genotypes can be categorized in three classes of global bearing habit: regular, irregular and alternate bearing. The genotype clustering relied on two indices computed using the production (number of flowers) $Y_{g,r,\pi,t}$ of tree replication r of genotype g at place π and year t , based on trend model

$$Y_{g,r,\pi,t} = \beta + \beta_\pi + \beta_g + (\alpha + \alpha_\pi + \alpha_g + \xi_{g,r})t + \varepsilon_{g,r,\pi,t} \quad (A)$$

where β and α are fixed parameters, β_π and α_π are fixed place effect, β_g and α_g are fixed genotype effects and $\xi_{g,r}$ are independent Gaussian random replication effects, with common variance τ_ξ^2 .

The BBI_res_norm is dedicated to discrimination between regular and other genotypes. This index is defined as

$$\text{BBI_norm} = \frac{\sum_r \sum_{t=2}^{T_{g,r}} |Y_{g,r,\pi,t} - Y_{g,r,\pi,t-1}| / \sum_r (T_{g,r} - 1)}{\sum_r \sum_{t=1}^{T_{g,r}} Y_{g,r,\pi,t} / \sum_r T_{g,r}},$$

$$\text{BBI_res_norm} = \frac{\sum_r \sum_{t=2}^{T_{g,r}} |\hat{\varepsilon}_{g,r,\pi,t} - \hat{\varepsilon}_{g,r,\pi,t-1}| / \sum_r (T_{g,r} - 1)}{\sum_r \sum_{t=1}^{T_{g,r}} Y_{g,r,\pi,t} / \sum_r T_{g,r}}.$$

where $T_{g,r}$ denote the number of measurements for replication r of genotype g , and $\hat{\varepsilon}_{g,r,\pi,t}$ denote the empirical in model (A). BBI_norm is some variant of the usual BBI with some normalisation of the mean absolute differences in production by the mean production.

From a statistical point of view, alternate bearing – as opposed to regular and irregular bearing – can be characterised through negative correlations between successive values of the detrended series of yields. Such correlations can be assessed by an autoregressive model

$$\varepsilon_{g,r,\pi,t} = (\gamma + \gamma_\pi + \gamma_g) \varepsilon_{g,r,\pi,t-1} + u_{g,r,\pi,t} \quad (I)$$

where $\varepsilon_{g,r,t}$ is the same residual as in trend model (A), γ is a fixed parameter, γ_π the fixed deviation from γ for place π , γ_g the fixed deviation from γ for genotype g and $u_{g,r,t}$ the residual of residual $\varepsilon_{g,r,\pi,t}$ of tree replication r of genotype g at time t . It is assumed variables $u_{g,r,\pi,t}$ are independent and Gaussian with mean 0 and variance ρ^2 . The so-called genotype AR coefficient γ_g can be used to discriminate alternate bearing from regular / irregular genotypes. Since

BBI_res_norm and γ_g are computed using global flower counts at tree scale $Y_{g,r,\pi,t}$, they are referred to as tree-scale indices.

Additional entropy indices were used to measure synchronicity in flowering. The classical entropy index is based on sequences of AS fates $(F_{g,r,\pi,t,\ell})_{\ell \geq 0}$, $(F_{g,r,\pi,t,\ell} = 0)$ denoting the absence and $(F_{g,r,\pi,t,\ell} = 1)$ the presence of flower for replication r of genotype g at year t , at place π (either “Montpellier SG”, “Montpellier XB” or “Angers”) at location (or AS) ℓ in the tree. It is defined as

$$\overline{Ent}_{g,\pi} = \frac{1}{R_{g,\pi}} \sum_r \left\{ - \frac{1}{\sum_t n_{g,r,\pi,t}} \sum_t n_{g,r,\pi,t} \left(\hat{p}_{g,r,\pi,t,0} \log \hat{p}_{g,r,\pi,t,0} + \hat{p}_{g,r,\pi,t,1} \log \hat{p}_{g,r,\pi,t,1} \right) \right\} \quad (1)$$

where $R_{g,\pi}$ denotes the number of replications for genotype g , $n_{g,r,\pi,t} = n_{g,r,\pi,t,0} + n_{g,r,\pi,t,1}$ the total number of AS for replication r of genotype g at place π and year t and $n_{g,r,\pi,t,i}$ the number of AS with fate i ($i=0, 1$). For $i=0, 1$, $\hat{p}_{g,r,\pi,t,i} = \frac{n_{g,r,\pi,t,i}}{n_{g,r,\pi,t}}$ is an estimation of the probability of flowering ($i=0$) v. non-flowering.

Note that the variability in the probability of flowering is partly due to genetic variations, but also to climatic effects that are specific to places π and years t . To capture the genetic part of variability only in entropy, we have to eliminate the other effects. Thus, rather than estimating $p_{g,r,\pi,t,i}$ directly using the associated frequency as above, we used some statistical model for $p_{g,r,\pi,t,i}$. Since $F_{g,r,\pi,t,\ell}$ is a binary variable, approaches based on Generalized Linear Mixed Models (GLMMs) are relevant (Molenberghs and Verbeke, 2006). The following GLMM was considered:

$$\log \frac{p_{g,r,\pi,t,1}}{p_{g,r,\pi,t,0}} = \lambda_\pi + \rho_{\pi,g} + \phi_{\pi,t} + \theta_{\pi,g,t}, \quad (2)$$

where λ_π is the fixed effect of place π (with reference $\lambda_{Montpellier} = 0$), $\rho_{\pi,g}$, $\phi_{\pi,t}$ and $\theta_{\pi,g,t}$ are random effects, assumed to be mutually independent and Gaussian, $\rho_{\pi,g}$ being the effect of genotype g at place π , $\phi_{\pi,t}$ the interaction between place π and year t treated as a qualitative variable, and $\theta_{\pi,g,t}$ the interaction between genotype g at place π and year t .

Since $n_{g,r,\pi,t}$ may be very small for some genotypes, maximum likelihood estimation may not converge from a computational point of view. Modelling the log ratio of the probabilities allows the probability of a binary variable (comprised between 0 and 1) to be mapped into $\mathbb{R}$.

The model parameters were estimated by maximum likelihood using the function *glmer* of package *lme4* (Bates *et al.*, 2011). To estimate probabilities $\tilde{p}_{g,r,\pi,t,i}$ corrected from all place-related effects, $\log \frac{\tilde{p}_{g,r,\pi,t,1}}{\tilde{p}_{g,r,\pi,t,0}} = \rho_{\pi,g} + \theta_{\pi,g,t}$ was used in model (2). Then to estimate entropies the $\tilde{p}_{g,r,\pi,t,i}$'s were used in lieu of $p_{g,r,\pi,t,i}$ in equation (1).

Prediction of fruiting behaviour from axis-scale indices

The above indices were used simultaneously to predict genotype habit from subsamples of AS. In our setting, the habit is summed up by a genotype class among $K=3$ possible classes: {regular, alternate, irregular}. The classes obtained for the SG family in Durand *et al.* (2013) by a clustering procedure, using the total number of flowers $Y_{g,r,t}$ for each replication were used as reference since they are the only genotypes with known behaviour at tree scale. The issue of class prediction from indices corresponds to the statistical framework of supervised classification. It was addressed using feed-forward neural networks (NNs in short), which essentially are non-linear multivariate regression models (Bishop, 2006, Chapter 5). These models provide probabilities for each genotype to belong to every possible class. The R-based implementation *nnet* was used (Venables & Ripley, 2002). The NN parameters were estimated by maximum likelihood, except a regularisation parameter that must be chosen by a statistical model selection principle. Assessment of the prediction quality obtained by NNs and selection of the regularisation parameter were achieved by out-of-sample validation. The principle was to use a random partition of the genotypes with known habits between learning (50%) and test (50%) genotypes. The learning genotypes were used to estimate the NN parameters, i.e. the mapping between indices and classes. Thus the classes were considered as known for the learning genotypes. The test genotypes were used to predict their classes, as if these were unknown. Since their classes were actually known, an error rate could be computed on the test set (so-called *test error*). This error rate was likely to vary according to the random partition. Thus, 5 partitions were drawn at random and the error rate was averaged over the 5 test sets. We selected the model (i.e. regularisation parameter) minimizing the test error. This method was adapted to perform non-linear regression to simultaneously predict both tree scale indices BBI_res_norm and γ_g from the axis-scale indices. In this case, the parameters were estimated by minimising the mean square error and the regularisation parameter was chosen by maximising the sum (over each index) of square correlations between the true and the predicted indices. The test correlations are referred to as cross-validated correlations.

Literature Cited in Supplementary Information

Bates D, Maechler M, Dai B. 2011. lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-28. [WWW document] URL <http://lme4.r-forge.r-project.org> [accessed 05 October 2011].

Bishop CM. 2006. Pattern Recognition and Machine Learning. Springer Verlag, 2006.

Molenberghs G., Verbeke G. 2005. *Models for discrete longitudinal data*. New York: Springer.

Venables WN, Ripley BD. 2002. *Modern Applied Statistics with S*, 4th edn. New York: Springer.

Result Chapter 3

Identification of segregating QTLs in two families for ‘pack-out’ traits

Investigated questions

Do the two families, X-3263 x ‘Belrene’ and ‘Starkrimson’ x ‘Granny Smith’, segregate for traits related to sugar content, acidity, firmness, fruit size, yield, and fruit size?

Do the simple ‘pack-out’ traits chosen in this study allow relevant QTL mapping?

Introduction

With fruit, the aim is not to get the highest yield but to reach the best balance between yield and fruit quality, and these two objectives often compete. The concept of ‘pack-out’ corresponds to the proportion of fruit production that is commercially acceptable. Therefore, it is the best overall descriptor of the target of both fruit breeders and the fruit production industry. ‘Pack-out’ therefore includes trait components of both yield and fruit quality (e.g. fruit size, skin colour, firmness, crispness, acidity, and sugar content). Some of these traits were studied in this chapter: yield, fruit size, sugar content, acidity, firmness.

Fruit size and attractiveness are two major criteria for the marketing. Big and nice fruits are put forward by the supermarkets and preferred by the mainstream consumers; as a consequence they are the main targets for the growers. Fruit taste is very complex; it can be decomposed into several main traits such as texture and flavor for example (Costa, 2015), each of them being decomposed into sub-traits. For instance, fruit texture is characterized by firmness, crispness, stiffness, mealiness, juiciness, graininess, and fibrousness; fruit taste is characterized by sweetness, and sourness; and fruit flavor is composed of various volatile compounds such as alcohols, esters and terpenes (Dunemann *et al.*, 2009).

This complexity makes fruit quality very difficult to assess. Besides, most of these traits are very subjective and depend on consumer preferences which vary from one region to the other. Generally, the best accepted varieties in Europe are characterized by a high firmness, high level of sugar and average acidity (Bonany *et al.*, 2013). These preferences can be subdivided in two main sub-groups: on one hand, high sweetness preference independently of acidity and firmness, on the other hand, the acid-firm preference with mid-range sweetness (Bonany *et al.*, 2014). Beside these main trends, preference specificities have been observed in different countries (Bonany *et al.*, 2013). In Poland, Switzerland and Spain people preferred sweet and firm apples with medium to low acidity; in the Netherlands and Germany, people tended to prefer acidic, juicy and crisp apples; and in Spain and Poland, people showed more interest into sugar content than other traits.

Akane	Falstaff	Meran
Arkcharme	Fiesta	Newgold
Arlet	Fuji	Ozarkgold
Belle de	Gala	Pink Lady
Boskoop	Golden	Primgold
Braeburn	Delicious	Red Winter
Cameo	Goldextr	Regali
Corail Pinova	Granny Smith	Reine des
Cox's Orange	Harmonie	Reinettes
Pippin	Honey	Reinette du
Cybele	Idared	Canada
Delbard jubile	Chantecler	RubINETTE
Delbard estivale	Florina	Sunrise
Delbard tardive	Jonagold	Suntan
Delrouge	Melrose	Tentation
Earlygold		
Elstar		

Table 1. List of well-known cultivars used for quality traits comparison. Average sugar content (% Brix), acidity (g ac. Mal./L), and firmness (kg/cm²) were available for these cultivars in CTIFL book (Trillot, 2002).

Sensory analysis is the reference method to assess fruit quality, however it has to be performed with trained panels in order to overcome subjectivity concerns. In addition, instrumental measures are also performed in order to get more objective data: malic acid titration for acidity, refractometry for sugar content and penetrometry for firmness. The complexity and the inter-relations of these traits often lead to low or medium correlations between sensory and related instrumental measurements. For instance, sweet taste is the result of a combination between sweetness and acidity, hence the refractometric index does not provide a reliable estimation of the sensory sweetness. So far, no tool was developed for appropriate assessment of crispness, which is a major criterion of fruit quality for the consumer. On the contrary, firmness assessed instrumentally is relatively well correlated with its sensory evaluation (*HiDRAS*, 2007, Costa *et al.*, 2012). Finally, the long storage ability of apples compared to other fruits, allows year round distribution on the market. As a consequence, the fruit evaluation process has to be performed on storage ability as well.

Among the multitude of traits assessed in apple breeding, four of them will be presented in this work: acidity measured by titration, sugar content measured by refractometry, firmness measured by penetrometry and yield measured by the number of fruits and the cropping weight. Acidity has a strong effect on apple taste. It is mainly driven by malic acid content which is known to have a monogenic determinism with a major locus located at the top of LG16 (Visser *et al.*, 1968; Brown and Harvey, 1971) and called the Ma locus (Maliepaard *et al.*, 1998; Xu *et al.*, 2012). At that locus, it has been shown that there is an incomplete dominance effect of Ma allele over the ma allele, with interaction with two others minor loci (Xu *et al.*, 2012). As a confirmation, two candidate genes, *Mal1* and *Mal2*, were identified as malic acid transporters, and were both located at the top of LG16, and finally *Mal1* expression was shown to be correlated to the acidity of the apples (Khan *et al.*, 2013).

Fructose, glucose, sucrose, and sorbitol are the main components of sugar content in apple (Fuleki *et al.*, 1994). At early stages of fruit development, sugar is accumulated by importation of sucrose and sorbitol from leaves to fruit in order to fuel the energy requirements due to fruit development. The extra imported sugars are stocked as starch which will be metabolized into sugars during fruit ripening (Li *et al.*, 2012). Many genes were identified as coding for enzymes regulating sugar content in fruit. *MdTMT1* and *MdTMT2* were involved in fructose and sucrose

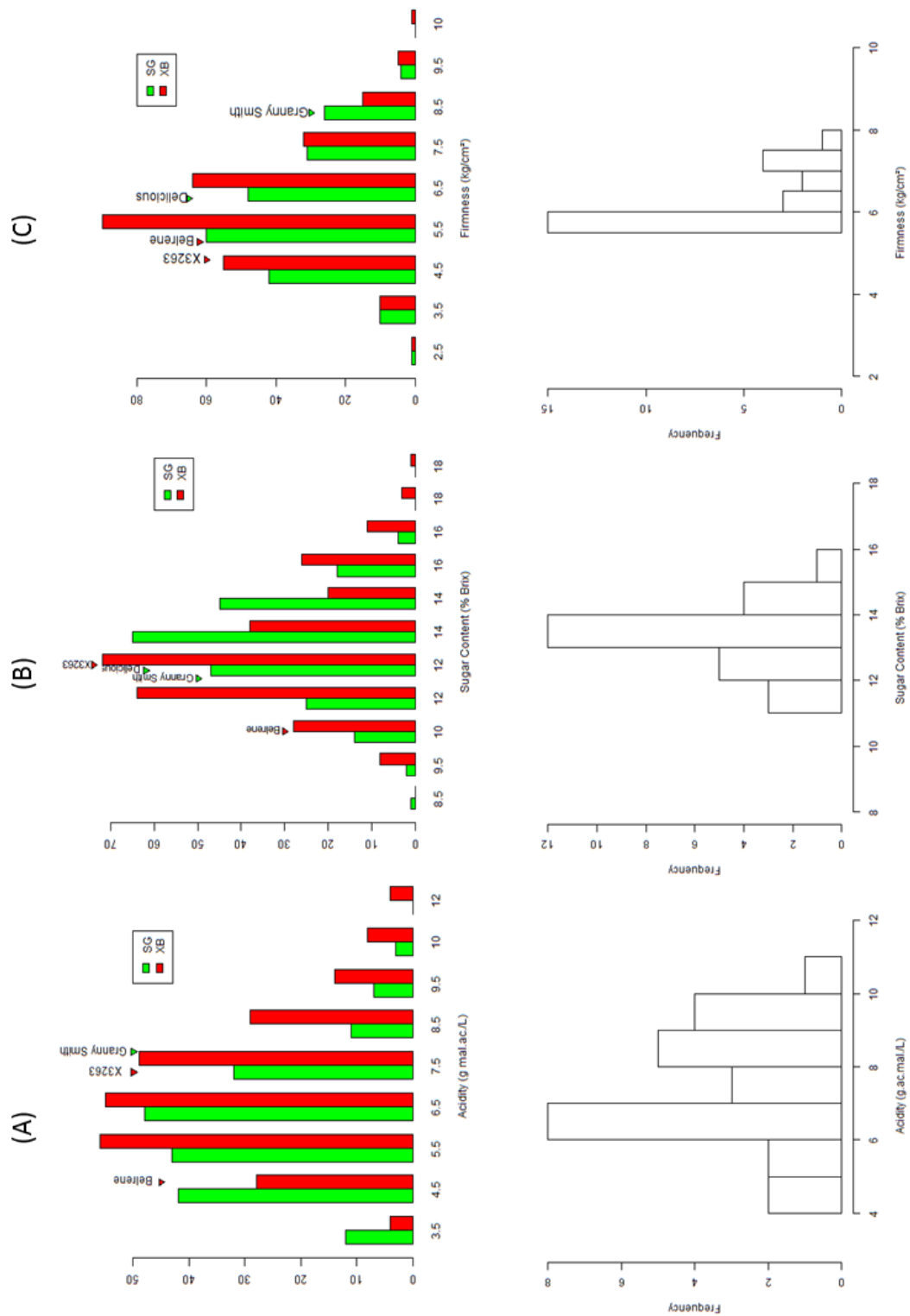


Figure 1. Histogramms of phenotypic distribution of (A) Acidity, (B) Sugar Content, and (C) Firmness. The first line represents the families distributions (parents are represented by arrows) and the second line represents the cultivars distribution.

accumulation in fruit, *MdFK1* was identified as involved in phosphorylation of fructose for sucrose synthesis, *MdFK2* was identified as involved in utilization of fructose in shoots and fruits and fructose accumulation in fruit during cell expansion, *MdNINV3* was identified as controlling sucrose accumulation in cytosol of fruit cells, and *MdSPS5* was identified as involved in sucrose accumulation (Li *et al.*, 2012). QTL mapping studies on sugar content instrumentally measured revealed QTLs located in different regions, on LGs 2, 3, 6, 8, 9, 10, 12, 13, 14 and 15 (Liebhard *et al.*, 2003; Kenis *et al.*, 2008; Guan *et al.*, 2015) and confirmed the hypothesis about the quantitative inheritance of sugar content.

Flesh firmness is determined by cortex tissue architecture, cell turgor, and cell wall composition and structure (Longhi *et al.*, 2012). Mechanical components such as firmness are usually assessed by instrumental measurements using puncture and deformation tests (i.e. penetrometry and compression) (Harker *et al.*, 1997, 2002; Galvez-lopez *et al.*, 2012). Penetrometry is routinely used by growers to estimate the average level of firmness. These measurements are destructive, therefore it is not possible to trace firmness evolution during storage with these techniques. The current challenge is to find an automatic and non-destructive tool to instrumentally assess for firmness. In addition to instrumental measurement, specific sensory methodologies have been set up to assess for texture traits (King *et al.*, 2000; Ben Sadok *et al.*, 2015). Firmness has been shown to be quantitatively inherited, and many QTLs were detected on several LGs depending on the studies and the traits (King *et al.*, 2000; Liebhard *et al.*, 2003; Bink *et al.*, 2014; Costa, 2015). LG10 has been regularly identified as carrying important QTLs for firmness in different studies and different genetic backgrounds. Two main regions of LG10 have been shown to be important. One colocalized with *MdPG1*, a gene encoding for cell wall degrading enzyme (Longhi *et al.*, 2012), located approximately in the middle of the LG; the other one located at the bottom of the linkage group (King *et al.*, 2000; Longhi *et al.*, 2013; Bink *et al.*, 2014; Ben Sadok *et al.*, 2015, *HiDRAS*, 2007). On LG1, the QTL detected by King *et al.* (2000) is in coupling phase with the gene *Rvi6/Vf*, conferring scab resistance. On the top of LG16, there is one QTL which has been shown to colocalize with the major QTL for acidity described in a previous section. There is either pleiotropy between the two traits or the region contain closely located genes (King *et al.*, 2001) and the favorable alleles for both traits are probably in coupling phase (King *et al.*, 2000).

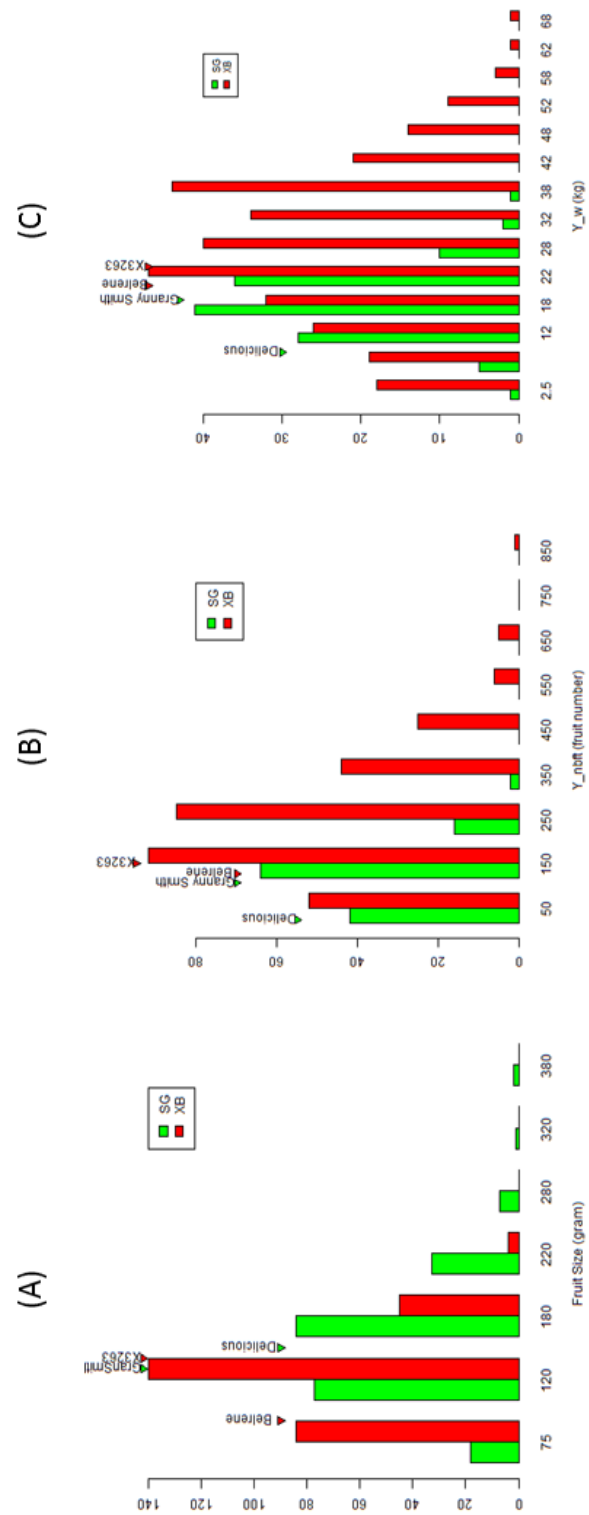


Figure 2. Histogramms of families phenotypic distribution for (A) Fruit Size, (B) Y_nbft, and (C) Y_w. Parents were represented by arrows.

Fruit size was suggested to be linked to cell division and cell expansion during fruit development, and auxins may play a role in regulation of fruit size via cell expansion. The gene *AUXIN Response Factor (ARF1106)*, linked to auxin regulation, was mapped in a strong QTL on LG15 related to fruit size (Devoghalaere *et al.*, 2012). In addition, a recent study on fruit size gene expression revealed an important locus, CrabApple Fruit Size (CAFS) on LG11 (Yao *et al.*, 2015). A metaQTL analysis revealed several QTLs related to fruit size on LG7, 8, 11, 15, and 17 (Costa, 2015).

In this work, we considered five ‘pack-out’ key traits: yield, fruit size, acidity and sugar content and firmness, all assessed by instrumental measurements. The objective is to identify QTLs segregating in two families, derived from different parents and to compare the results to literature. The outputs of this study will be used in the next chapter for a character integration study.

Material and methods

Plant material and sampling

The two progenies used in this study were derived from crosses between ‘Starkrimson’ (called ‘Delicious’ thereafter since ‘Starkrimson is a color mutant of ‘Delicious’) and ‘Granny Smith’, and X-3263 and ‘Belrene’, respectively named SG and XB afterwards. For description of the families, see previous results chapters of this thesis. Fruits were harvested when starch index reached 7 on a 1 to 10 scale (Vaysse, 2002). Harvest time for SG, was spread between the 9th of September and the 3rd of October 2013, and for XB between the 19th of August and the 25th of September 2013. In order to ensure variability within each tree, a representative sample of ten fruits was picked up from each tree. The fruits were sampled in order to represent the fruit variability within the tree. The samples were then brought to a cold chamber in CEHM (Centre Experimental Horticole de Marsillargues) where they were stored for two months at 4-6 °C.

Fruit quality measurements

Four quality traits were measured in an automated way with the ‘Pimprenelle’ tool (www.setop.fr) from the CEHM. Sugar content, firmness and fruit weight were measured for

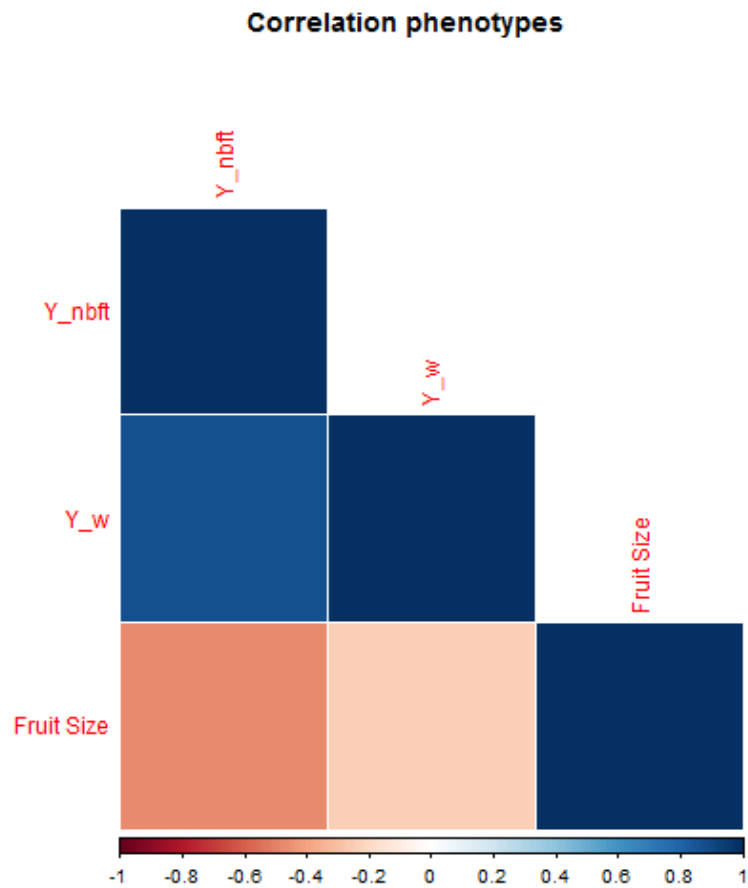


Figure 3. Correlation matrix between Fruit Size, Yield in number of fruit (Y_nbft), and Yield in weight (Y_w). Dark blue indicates high positive correlation, dark red indicates high negative correlation. (Correlations values were obtained with the Spearman method).

each fruit by refractometry (% Brix), electronic penetrometer (kg/cm²) and weighting (g) respectively, and averaged for each sample. For Acidity, juice was extracted from the ten fruits of the sample and mixed together. Malic acid content was measured by titration and expressed in gram of malic acid per liter (g ac.mal./L).

For a comparison purpose, average values for sugar content, acidity, and firmness for well-known cultivars were also presented (Trillot *et al.*, 2002, and Table 1 for cultivar list). These values had to be carefully handled because they represent values averaged over several years and it is not mentioned in which environment and which condition the data were taken. However, they give an idea of the variability of the traits among cultivars and allow a comparison with the progeny values.

Two traits were considered to estimate the level of production: total number of fruit (Y_nbf) per tree, total weight of fruit per tree (Y_w). Y_nbf was assessed by counting fruits on the tree. Y_w was assessed by weighting fruits picked up from the tree and estimating the whole tree production by adding the estimated weight of fallen fruits. Y_nbf and Y_w were measured and averaged over several years, from 2010 to 2014 for SG family and from 2008 to 2014 for XB family, and then averaged over the years.

Heritability and QTL mapping

In this study, heritabilities could not be estimated because no genotype repetitions were available. However, each traits has been previously shown to be heritable in several studies, therefore we considered that they were heritable in this study too. No experimental or environmental bias could be assessed for the same reason and this was taken into account in result interpretation. For genotyping, markers genetic map and QTL mapping methods see previous chapters (1 and 2). Since the lack of repetitions did not allow the estimation of genotypic effects, QTL mapping was realized on raw phenotypes.

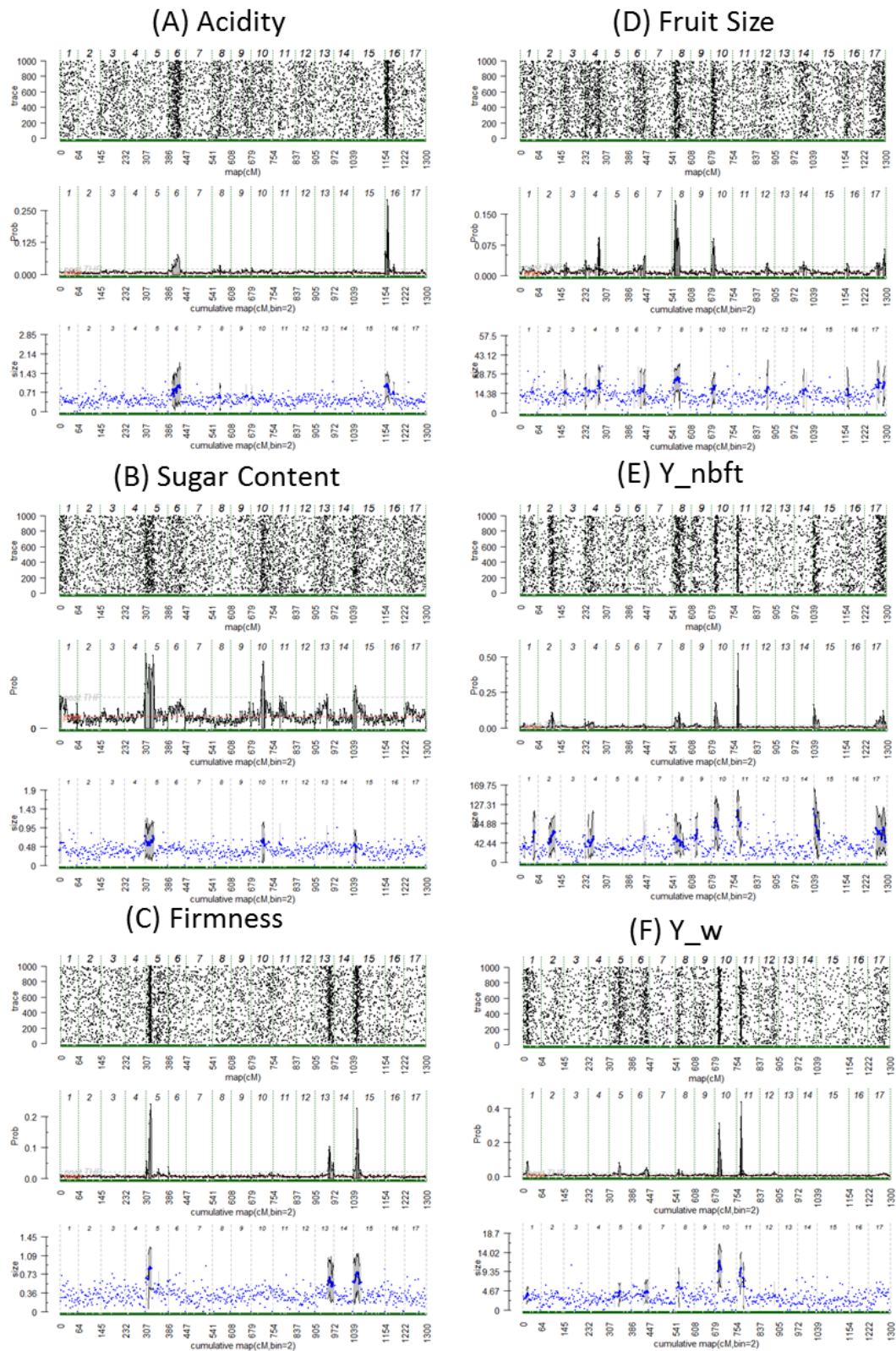


Figure 4. QTL trace plot of QTL presence along the genome and across iterations (top), posterior probability of QTL position along genome (middle), and QTL effects (bottom) for (A) Acidity, (B) Sugar Content, (C) Firmness, (D) Fruit Size, (E) Y_nbft, and (F) Y_w.

Results

Phenotypes

For Acidity, the distribution of XB was slightly shifted towards high values compared to SG distribution. In addition, these two distributions covered the same range than cultivars (Figure 1A). X-3263 and ‘Granny Smith’ had a relatively high level of acidity whereas ‘Belrène’ showed much lower values. Acidity level of ‘Delicious’ could not be measured. For Sugar Content, XB and SG distributions had the same shape, but XB distribution was slightly located towards lower values. The sugar content distribution of the families was wider than that of cultivars (Figure 1B). Concerning the Sugar Content of parents, ‘Belrene’ had a low value, X-3263 had a medium value, such as ‘Granny Smith’ and ‘Delicious’ (Figure 1B). Firmness distributions of families were wider than that of cultivars, and both families had similar distributions (Figure 2A). Concerning, the parents’ firmness values, X-3263 and ‘Belrene’ values were slightly smaller than average, ‘Delicious’ value was slightly above average and ‘Granny Smith’ had a relatively high value (Figure 1C). For fruit size, SG and XB had similar distributions, except that SG had few progenies with big fruits (Figure 2A). Concerning the variables related to the level of production, number of fruits per tree and the total fruit weight revealed that XB family was more productive than SG family since it yielded more fruits and its total fruit weight was higher. Fruit size was negatively correlated to Y_{nbft} and Y_w (Figure 3).

QTL mapping

For Ac, two QTLs were detected (Table 2 and Figure 4A). One major QTL at the top LG16 explained 16.4% of the variance and the other one on LG6 was not very robust regarding its $2*\ln BF$ value (<5) but explained 15% of the variance (Table 2). In addition, the estimated genotypes of the QTL on LG16 revealed that parents X3263 and ‘Delicious’ were probably heterozygous in that position (Figure 5A), which suggested that this QTL segregated in both XB and SG families. Surprisingly, ‘Granny Smith’ was estimated as homozygous for the low value allele (in blue in Figure 5) even though it was the parent with the highest phenotypic value (Figure 1A).

Table 2. Parameters associated with the genotypic BLUPs QTL. The first column indicates the variable concerned, the following columns indicates the LG where the QTL was located, the 2ln(BF) value at the LG scale, the 2ln(BF) value at the bin scale, the position of the QTL in cM, its additive effect, the frequency of favorable allele, its variance and its percentage of variance explained. In this table, the 2lnBF values correspond to the comparison of a model with 0 QTL to a model with 1 QTL. The other 2lnBF values were not presented because they were not relevant.

Variable	LG	2lnBF_LG	Max 2lnBF_bin	pos (cM)	Peak_pos (cM)	add_ef	fq	%var	QTL other studies	R ²
Acidity (g ac.mal./L)	6	3.8	4.69	15-45	34-35	0.94	0.39	15.0	Kenis et al,2008 (expect the Ma gene to map here) Liebhard et al, 2003 Maliéppard et al, 1998	30-35
	16	8.4	5.04	1-15	8-9	0.96	0.48	16.4		42%
Sugar Content (% Brix)	5	2.6	3.79	2-32	15-16	0.6186	0.45	12.7		
	5	9.9	7.40	10-20	17-18	0.8393	0.51	18.9		
	13	3.4	5.39	47-65	52-53	0.6591	0.52	11.7	King et al, 2000	
Firmness (kg/cm ²)									Costa et al, 2015	21.3,
									King et al, 2000	14.3
	15	5	7.27	1-23	14-15	0.7548	0.44	15.1	King et al , 2001 Bink et al, 2014	15

For SC, only one weak QTL was detected at the top of LG5 (Table 2 and Figure 4B), although other non-significant signals appeared on LG10, LG13, and LG15 (Table 2). According to the estimated genotypes at this QTL, only ‘Granny Smith’ was heterozygous, therefore only SG was supposed to segregate at this position. X3263 and ‘Belrene’ were estimated as homozygous for the low value allele (Figure 5B) which was in agreement to the low phenotypic value of ‘Belrene’ and the average value of ‘X3263’ (Figure 1B).

For Firm, three QTLs were mapped, among which, two were robust: on LG5 and LG15 (Table 2 and Figure 4C). They explained 18.9 and 15.1% of the variance, respectively (Table 2). For QTLs on LG5, only parent ‘Granny Smith’ was estimated as heterozygous (Figure 5C), suggesting segregation in SG only, whereas X-3263 and ‘Belrene’ were estimated as homozygous with the low value allele (Figure 5C). However, for QTL on LG15, parents X3263 and ‘Delicious’ were estimated as heterozygous (Figure 6C), therefore both SG and XB segregated for this QTL.

For FS, two QTLs were detected on LG4 and LG8 (Table 3 and Figure 4E). The QTL on LG8 was the strongest since its $2 \times \ln BF$ value was higher than 5 and it explained 13.3% of the variance (Table 3). For this QTL, only parents ‘Granny Smith’ and ‘Delicious’ were estimated as heterozygous, therefore only SG family segregated for this variable. X-3263 and ‘Belrene’ were estimated as homozygous for the low value allele (Figure 6A) and this corresponded to the low phenotypic value of ‘Belrene and the average phenotypic value of X-3263 (Figure 2A).

Several strong QTLs were mapped for Y_nbft on LG8, LG11 and LG17, and positive QTLs were mapped on LG2, LG10, and LG15 (Table 3 and Figure 5B). The QTL on LG8 colocalized with a QTL for fruit size, although the peaks of the two QTLs were not exactly located at the same position. For QTL on LG8, all parents except ‘Belrene’ were estimated as heterozygous, therefore both SG and XB families segregated at that position. For this QTL, ‘Belrene’ was estimated as homozygous for the high value allele, which corresponded to its relatively high phenotypic value (Figure 2B). Finally, for QTLs on LG11, and LG17 only X-3263 was estimated as heterozygous therefore only XB family segregated (Figure 6B).

For Y_w, two QTLs were mapped on LGs 10 and 11, and colocalized with the QTLs mapped on the same LGs for the Y_nbft (Table3 and Figure 5C,). The genotype estimation was similar

Table 3. Parameters associated with the genotypic BLUPs QTL. The first column indicates the variable concerned, the following columns indicates the LG where the QTL was located, the 2ln(BF) value at the LG scale, the 2ln(BF) value at the bin scale, the position of the QTL in cM, its additive effect, the frequency of favorable allele, its variance and its percentage of variance explained. In this table, the 2lnBF values correspond to the comparison of a model with 0 QTL to a model with 1 QTL. The other 2lnBF values were not presented because they were not relevant.

Variable	LG	2lnBF_LG	max_2lnBF_bin	pos (cM)	Pic_pos (cM)	add_ef	fq	%var	QTL other studies	R ²
Fruit size (g)	4	3.1	5.17	49-57	52-53	20.559	0.48	9.5	Costa <i>et al.</i> , 2015 Costa <i>et al.</i> , 2015	18.4
	8	11.1	6.70	9-29	12-13	24.332	0.48	13.3	Devoghalaere <i>et al.</i> , 2012	8.7
									Liebhard <i>et al.</i> , 2003	9%
Total number of fruit per tree (Y_nb_ft)	2	3.5	5.48	41-59	50-51	57.49	0.48	8.0		
	8	5.1	5.48	9-31	22-23	48.17	0.48	5.6		
	10	4.7	6.65	12-28	15-16	94.74	0.45	21.6		
	11	10	9.87	15-29	18-19	113.3	0.41	30.2		
	15	3.5	6.42	0-21	0-1	116.38	0.60	31.5		
	17	5.7	5.74	41-77	66-67	75.82	0.45	13.8		
Total fruit weight per tree (g) (Y_w)	10	10.9	8.1	14-24	17-18	11.97	0.39	40.1		
	11	7.2	9.2	13-27	18-19	1.183	0.5	30.63		

than for Y_nbfr (Figure 6C). The QTLs on LG11 for yield are potentially in the same region than a QTL of fruit size on the same LG revealed by Yao *et al.* (2015).

Discussion

Acidity

In this study, the distribution of acidity among progenies of the two families was continuous and covered the variability of well-known cultivars used for comparison. This finding was slightly different than what described by Xu *et al.* (2012) and Khan *et al.* (2013), where bi-modal distributions were presented. However, the QTL mapping results corresponded to the previous findings about the genetic determinism of acidity: one major locus at the top of LG16 (Xu *et al.*, 2012). However, the percentage of explained variance by that QTL, 16%, was much lower than the same locus detected in Xu *et al.* (2012), which explained more than 40%. As found in other studies (Xu *et al.*, 2012, *HiDRAS*, 2007), another locus was involved at the top of LG6. In this study, the position of the QTL on LG6 could not be precisely defined, indeed it was spread over 30 cM. Besides, there was no clear evidence of colocalization between this QTL and the one found in Xu *et al.* (2012) because of a lack of information about the QTL position in Xu's study. The literature confirmed the results of this study: acidity in apple is mainly driven by one dominant locus on LG16, and minor loci.

Sugar content

For sugar content, the families displayed a wider variability than the cultivars, therefore they may be of interest in breeding since sweetness is a major trait influencing consumer preference (Harker *et al.*, 2008). It has been shown that the inheritance of sugar content was quantitative (Visser *et al.*, 1968; Brown and Harvey, 1971). Sugar content QTL studies have revealed different results depending on the genetic background and the variable: QTLs were detected on LGs 3, 6, 8, 9 and 14 by Liebhard *et al.* (2003), on LGs 2 and 10 for Kenis *et al.*, (2008), on LGs 6, 8, and 12 by Costa (2015), and on LGs 1, 2, 3, 4, 5, 9, 12, 13, 15 and 16 by Guan *et al.* (2015). In this study, there was only one positive evidence for a QTL, several arguments could explain this lack of QTL detection. Sugar content has been shown to be influenced by environmental conditions (Alspach and Oraguzie, 2002; Guan *et al.*, 2015), and since, in this study,

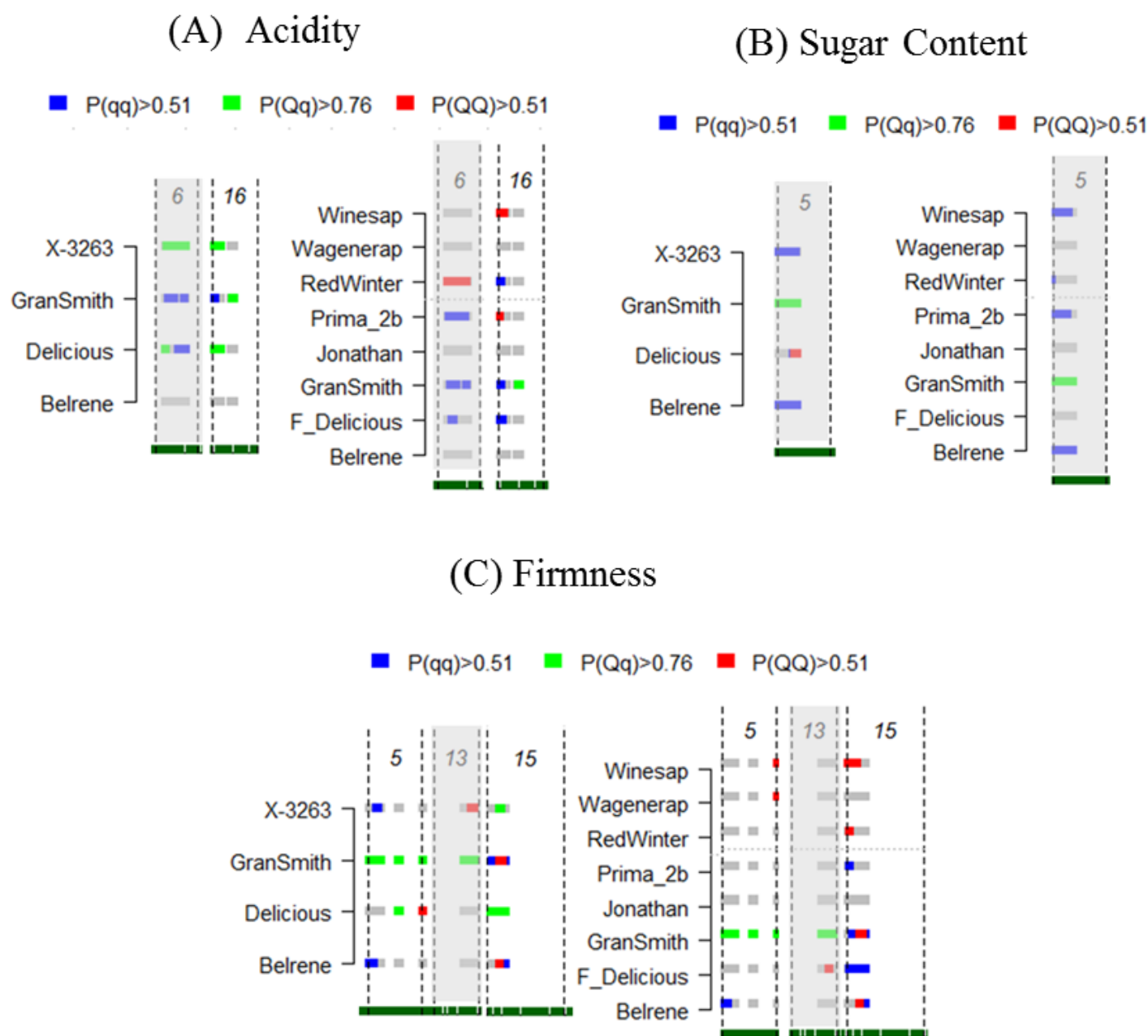


Figure 5. Posterior estimates of QTL genotype probabilities. Estimates are plotted for main QTLs for (A) Acidity, (B) Sugar Content, and (C) Firmness. The beginning and the end of the linkage group (LG) are represented by vertical dashed lines, the number of the LG is given at the top of the plot. For the QTLs with only positive evidence, the LG is highlighted in grey. Blue, green and red colors indicate positive evidence for QTL genotypes qq, qQ and QQ, respectively; q and Q refer to low and high phenotypic values, respectively. Grey colors indicate ignorance for a given genotype. Estimated genotypes of the parents and founders are presented on the left and right side, respectively. This figure was composed from FlexQTL outputs.

environmental effects could not be estimated or taken into account, the QTL detection might have been hampered by the environmental or experimental effects embedded in the trait variability. A lack of contrast between parental allele could be another explanation for the lack of mapped QTL. Finally, highly quantitative traits, such as sugar content, are ruled by many loci with small individual effects, therefore the association detected between phenotypes and genotypes might have not be strong enough to be considered as QTLs in this work.

Fruit Firmness

Progenies firmness was compared to cultivar firmness, and the progenies' distributions were wider than for cultivars. Since firmness is a major trait involved in consumers' preference, it seems logical that cultivars displayed a narrow distribution, which would correspond to the range of firmness acceptance. The wider distribution of the families could be due the genetic effect of the parental admixture. In addition, proper harvest date is difficult to assess for hybrids, and the starch index used was developed for well-known cultivars. This methodological choice may have lead to late harvesting (i.e. close to full maturity), reducing storage abilities of samples.

Fruit firmness has been shown to have a polygenic determinism (King *et al.*, 2001). Many QTLs for flesh firmness were reported in the literature, especially on LG10 where a cluster of QTLs has been identified by several studies (King *et al.*, 2000; Kenis *et al.*, 2008; Bink *et al.*, 2014; Ben Sadok *et al.*, 2015; Costa, 2015), however, in this work, this region was not reported.. The QTL on LG15 mapped in this study has been reported previously (King *et al.*, 2000; Bink *et al.*, 2014; Ben Sadok *et al.*, 2015), and the QTL on LG5 was reported once before by (Ben Sadok *et al.*, 2015) whose study was realized on a family related to XB, derived from X3263 and X-3259. Two QTLs out of three detected in this study were confirmed by the literature. Even though environmental effects were not taken into account, these QTLs appeared relevant. Firmness is a polygenic trait and the results of QTL mapping depends on the genetic background used. According to Ben Sadok *et al.* (2015), harvest and post-harvest fruit texture is important to take into account for a proper understanding of the genetic determinism of fruit texture.

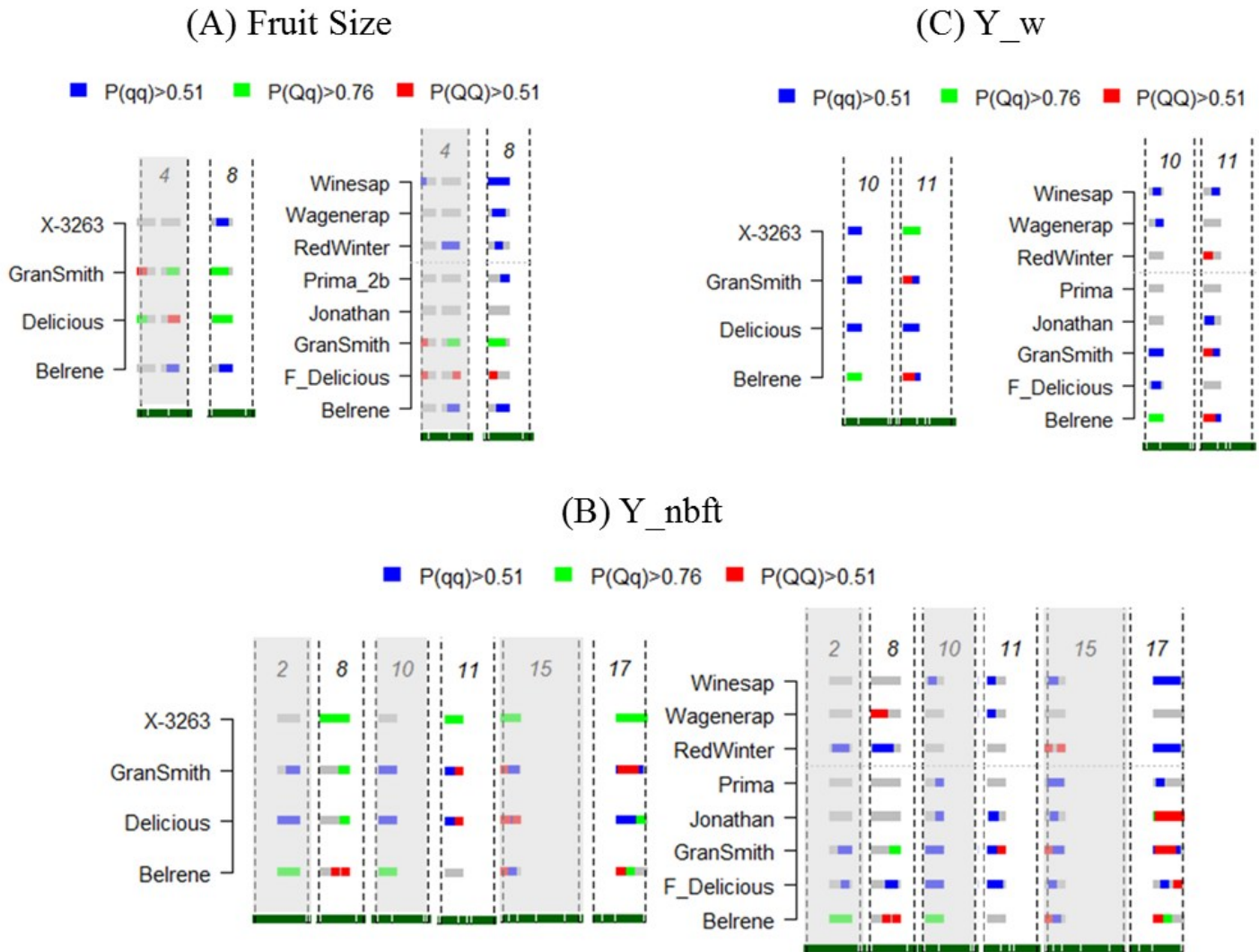


Figure 6. Posterior estimates of QTL genotype probabilities. Estimates are plotted for main QTLs for (A) Fruit Size, (B) Y_{nbft}, and (C) Y_w. The beginning and the end of the linkage group (LG) are represented by vertical dashed lines, the number of the LG is given at the top of the plot. For the QTLs with only positive evidence, the LG is highlighted in grey. Blue, green and red colors indicate positive evidence for QTL genotypes qq, qQ and QQ, respectively; q and Q refer to low and high phenotypic values, respectively. Grey colors indicate ignorance for a given genotype. Estimated genotypes of the parents and founders are presented on the left and right side, respectively. This figure was composed from FlexQTL outputs.

Fruit size and Yield

Fruit size was correlated with fruit yield: the highest the yield was, the smallest the fruits were. The problematic of fruit load has to be considered in combination with the problematic of fruit bearing habit, because heavy fruit loads one year can be responsible for light fruit loads the following year and may be the reason for biennial bearing (Monselise and Goldschmidt, 1982). Big fruits and high yield are favored in apple by marketers and producers, respectively. In this case, a negative phenotypic correlation between the two traits suggested difficulties to satisfy both fruit size and yield requirements. This relationship may be strong in this study because the trees were neither thinned nor pruned. Thinning can be used for fruit load control and may modify the negative correlation mentioned before. This raises the question of how to breed for fruit size and fruit bearing habit. These traits may require two separated breeding steps: one for fruit bearing habit where no pruning and thinning would be realized, and another step for fruit size, where the tree would be conducted in classical growing conditions with pruning and thinning. The phenotypic correlation was partly explained by a genetic correlation illustrated by a QTL on LG8 that colocalized for Y_nbfrt and Fruit Size. In addition, the QTLs mapped on LG11 for yield may colocalized with a QTL mapped for fruit size in Yao *et al.*, (2015). These two regions may harbor either pleiotropic loci or closely linked genes.

Conclusion

This study provided QTLs mapped for 'pack-out' traits for two families derived from two different genetic backgrounds. Even though environmental and experimental effects could not be assessed, several QTLs were confirmed by literature confirming the relevance of the QTLs mapped in this work. For acidity, the QTLs mapped in this study represented the partially monogenic determinism of this traits with the major loci located on LG16. For sugar content and firmness, relatively few QTLs were mapped compared to QTLs mentioned in the literature, and this could be attributed to the genetic background, the small effect loci that may have not been detected as QTLs, or the environmental/experimental conditions. Correlation between fruit size and yield was partly explained by a genetic correlation illustrated by colocalizing QTLs.

Even though, the characters studied were represented by an extremely simplified variable, compared to the assessment of the same characters in breeding, the results were considered as

relevant and allowed the first characterization of the two families, SG and XB, for ‘Pack-out’ traits. In the next chapter, these results will be combined in a multi-character study, with two other important characters, bearing habit and flowering time.

References

- Alspach PA, Oraguzie NC.** 2002. Estimation of genetic parameters of apple (*Malus domestica*) fruit quality from open-pollinated families. *New Zealand Journal of Crop and Horticultural Science* **30**, 219-228.
- Ben Sadok I, Tiecher A, Galvez-Lopez D, Bruneau M, Hanteville S, Robic R, Cournol R, Lahaye M, Lasserre-Zuber P, Laurens F.** 2015. Apple fruit texture QTLs: year and cold storage effects on sensory and instrumental traits. *Tree Genetics & Genomes* **11**, 119.
- Bink MCAM, Jansen J, Madduri M, et al.** 2014. Bayesian QTL analyses using pedigreed families of an outcrossing species, with application to fruit firmness in apple. *Theoretical and Applied Genetics* **127**, 1073-1090.
- Bonany J, Brugger C, Buehler A, et al.** 2014. Preference mapping of apple varieties in Europe. *Food Quality and Preference* **32, Part C**, 317-329.
- Bonany J, Buehler A, Carbó J, et al.** 2013. Consumer eating quality acceptance of new apple varieties in different European countries. *Food Quality and Preference* **30**, 250-259.
- Brown AG, Harvey DM.** 1971. The nature and inheritance of sweetness and acidity in the cultivated apple. *Euphytica* **20**, 68-80.
- Costa F.** 2015. MetaQTL analysis provides a compendium of genomic loci controlling fruit quality traits in apple. *Tree Genetics & Genomes* **11**, 1-11.
- Costa F, Cappellin L, Fontanari M, Longhi S, Guerra W, Magnago P, Gasperi F, Biasioli F.** 2012. Texture dynamics during postharvest cold storage ripening in apple (*Malus × domestica* Borkh.). *Postharvest Biology and Technology* **69**, 54-63.
- Devoghalaere F, Doucen T, Guitton B, et al.** 2012. A genomics approach to understanding the role of auxin in apple (*Malus x domestica*) fruit size control. *BMC Plant Biology* **12**, 7.
- Dunemann F, Ulrich D, Boudichevskaia A, Grafe C, Weber WE.** 2009. QTL mapping of aroma compounds analysed by headspace solid-phase microextraction gas chromatography in the apple progeny 'Discovery' X 'Prima'. *Molecular Breeding* **23**, 501-521.
- Fuleki T, Pelayo E, Palabay RB.** 1994. Sugar Composition of Varietal Juices Produced from Fresh and Stored Apples. *Journal of Agricultural and Food Chemistry* **42**, 1266-1275.
- Guan Y, Peace C, Rudell D, Verma S, Evans K.** 2015. QTLs detected for individual sugars and soluble solids content in apple. *Molecular Breeding* **35**, 1-13.
- Harker FR, Kupferman EM, Marin AB, Gunson FA, Triggs CM.** 2008. Eating quality standards for apples based on consumer preferences. *Postharvest Biology and Technology* **50**, 70-78.

HiDRAS High-Quality Disease Resistant Apples for a Sustainable Agriculture QLK5-CT-2002-01492 PROJECT FINAL REPORT. 2007.

Kenis K, Keulemans J, Davey MW. 2008. Identification and stability of QTLs for fruit quality traits in apple. *Tree Genetics and Genomes* **4**, 647-661.

Khan SA, Beekwilder J, Schaart JG, Mumm R, Soriano JM, Jacobsen E, Schouten HJ. 2013. Differences in acidity of apples are probably mainly caused by a malic acid transporter gene on LG16. *Tree Genetics & Genomes* **9**, 475-487.

King GJ, Lynn JR, Dover CJ, Evans KM, Seymour GB. 2001. Resolution of quantitative trait loci for mechanical measures accounting for genetic variation in fruit texture of apple (*Malus pumila* Mill.). *Theoretical and Applied Genetics* **102**, 1227-1235.

King GJ, Maliepaard C, Lynn JR, et al. 2000. Quantitative genetic analysis and comparison of physical and sensory descriptors relating to fruit flesh firmness in apple (*Malus pumila* Mill.). *Theoretical and Applied Genetics* **100**, 1074-1084.

Liebhart R, Kellerhals M, Pfammatter W, Jertmini M, Gessler C. 2003. Mapping quantitative physiological traits in apple (*Malus × domestica* Borkh.). *Plant Molecular Biology* **52**, 511-526.

Li M, Feng F, Cheng L. 2012. Expression Patterns of Genes Involved in Sugar Metabolism and Accumulation during Apple Fruit Development. *PLoS ONE* **7**, e33055.

Longhi S, Hamblin MT, Trainotti L, Peace CP, Velasco R, Costa F. 2013. A candidate gene based approach validates Md-PG1 as the main responsible for a QTL impacting fruit texture in apple (*Malus x domestica* Borkh). *BMC Plant Biology* **13**, 37.

Longhi S, Moretto M, Viola R, Velasco R, Costa F. 2012. Comprehensive QTL mapping survey dissects the complex fruit texture physiology in apple (*Malus x domestica* Borkh.). *Journal of Experimental Botany* **63**, 1107-1121.

Maliepaard C, Alston FH, Arkel G van, et al. 1998. Aligning male and female linkage maps of apple (*Malus pumila* Mill.) using multi-allelic markers. *Theoretical and Applied Genetics* **97**, 60-73.

Monselise SP, Goldschmidt EE. 1982. Alternate bearing in fruit trees. *Horticultural Reviews* **4**, 128-173.

Trillot M, Masseron A, Mathieu V, Bergougnoux F, Hutin C, Lespinasse Y. 2002. *Le pommier*. CTIFL.

Vaysse P. 2002. Pomme - Code amidon - Aide de décision à la récolte.

Visser T, Schaap AA, Vries DPD. 1968. Acidity and sweetness in apple and pear. *Euphytica* **17**, 153-167.

Xu K, Wang A, Brown S. 2012. Genetic characterization of the Ma locus with pH and titratable acidity in apple. *Molecular Breeding* **30**, 899-912.

Yao J-L, Xu J, Cornille A, *et al.* 2015. A microRNA allele that emerged prior to apple domestication may underlie fruit size evolution. *The Plant Journal* **84**, 417-427.

Result Chapter 4

Integration of phenotypic traits into a pre-defined ideotype:

A study case using genome wide breeding value

and selection index estimation on apple tree

Investigated questions

How to estimate a multi-character genetic potential?

How to deal with synergistic, antagonistic, and non-correlated traits in selection?

How to combine ideotype, selection index concepts and breeding values concept can be combined in a breeding strategy?

Introduction

The phenotype of a plant is composed of a multitude of characters involving different biological processes and different organs, at different plant scales. In quantitative genetics, complex characters are generally studied independently from each other since it is simpler to study a single process without considering the interactions that may exist with other processes or organs. They are generally decomposed into elementary traits, and QTL mapping analyses are run separately on these different traits. For example, in Ben Sadok *et al.* (2015), fruit texture was decomposed in several sensory traits such as firmness, crunchiness, fibrousness, juciness, meltiness, mealiness and graininess in order to assess all the aspects of fruit texture. In apple, many QTLs have been mapped on many different traits: 77 hits for a search with the keys words ‘apple’, ‘Fruit quality’, and ‘QTL’, 76 hits for a search with the keys words ‘apple’, ‘disease resistance’, and ‘QTL’, on Web Of ScienceTM. Some of these studies yielded colocalizing QTLs but, so far, only few attempts have been made toward the integration of the information available per trait into more global view of the phenotype. This is likely to result from the complex nature of the correlation between traits on the one hand and the complex genetic architecture of most traits on the other hand.

Multi-traits analyses are a first step towards trait integration by allowing joint QTL analysis for correlated traits. They have several advantages like increasing the power of QTL detection (Korol *et al.*, 1998; Knott and Haley, 2000; Alimi *et al.*, 2013), allowing QTL detection for low heritability trait by exploiting correlation with a high heritability trait (Jiang and Zeng, 1995), detecting or testing biological hypothesis such as pleiotropy, or pleiotropy vs. close linkage (Jiang and Zeng, 1995; Knott and Haley, 2000; Caliński *et al.*, 2000; Lund *et al.*, 2003; Malosetti *et al.*, 2008; Sukhwinder-Singh *et al.*, 2012; Silva *et al.*, 2012; Korte *et al.*, 2012; Alimi *et al.*, 2013). Breeding schemes on apple tree, and other fruit tree, takes into account a high number of characters such as fruit quality, disease resistance and yield. Among those traits, some are considered as “culling”, i.e. individual that does not perform well for those culling traits are eliminated prior to any harvest (Muranty *et al.*, 2015). Selecting individuals requires a high level of integration, on characters which differ by their biological determinisms but also by the organ concerned, fruits, flowers or whole tree, for example. In particular, breeding on fruit quality is very challenging because it involves a high number of traits, around 30 (F. Laurens, personal

communication), and many of these trait are subjective. Since apple is a fruit available in market all year round, breeding is also focused on the ability to maintain quality during storage. In current breeding programs, the selection of individuals and traits are organized in successive steps (Refer to the breeding scheme in introduction for more details).

Besides fruit quality, other characters have to be considered because, as demonstrated in previous chapters, they are challenging for fruit production stability over years (fruit bearing habit) or for tree adaptation to climatic conditions (chill and heat requirements). Even though breeding goals are already numerous in apple, it is important to think at the next future and find ways to face new challenges concerning the reduction on thinning chemicals and the global climate change. One possible approach to define the trees to be targeted, rely on the concept of ideotype. This concept has been defined as a biological/plant model which behavior can be predicted. For annual crops, an ideotype is a plant model which maximizes the quantity or quality of the yield (Donald, 1968). For perennial species such as trees, in addition to yield, other traits have been proposed as important such as photosynthesis efficiency, sink dynamic, competition, phenology, cold hardiness; and for fruit trees specifically, traits linked to rootstock, growth and development characteristics have been underlined (Dickmann, 1985; Dickmann *et al.*, 1994). Ideotypes are based on the scientific knowledge of traits, the more the knowledge is precise the more the ideotype is precise (Donald, 1968; Dickmann, 1985; Dickmann *et al.*, 1994). They should not be considered as definitive models since they can evolve with the biological knowledge and vary with the environment, the cultural practices (Dickmann *et al.*, 1994), and the marketing demand. Ideotypes have been considered as a comprehensive tool used by breeders to identify breeding goals. More recently, the concept of ideotype in apple has been revisited for underlining its complex multi-traits nature and including the possible changes of targets over time (Costes *et al.*, 2009; Debaeke *et al.*, 2014). The challenge of combining traits as different as disease resistance, fruit quality, regular bearing and other commercial targets, possibly antagonistic between them has also been underlined. In addition to be used as a list of goals to target, ideotype can also be used to build selection indexes. Selection indices have been more used in animal breeding than in plant breeding. Even though it is considered as the most efficient selection strategy for multi-character breeding (Gallais, 1990), few references could be found related to the recent use of selection index in plants (Almeida *et al.*, 2014; Bhering *et al.*, 2012).

Table 1. Recap chart for the character and traits characteristics, and breeding goals.

Character	Trait	Abbreviation	Type of variable for breeding value computation	Reference/ Chapter	Breeding goals
Fruit Quality	Sugar Content	SC	Raw phenotypes	Chapter 3	High sugar content
	Acidity	Ac			High acidity
	Firmness	Firm			High Firmness
	Fruit Size	FS			Big Fruits
Fruit bearing habit	BBI	BBIresnorm	Index	Durand et al, in preparation Chapter 2	Regular :
	Yield autocorrelation	Autocov	Model parameter		Small BBI
		(called γ_g in chapter 2)			High Autocov
Phenology	Budbreak	BBcd	BLUPs	Allard et al, under revision	Adapted to warm climates:
	Delay Budbreak-flowering	Delta		Chapter 1	Early budbreak
					Avoid low heat requirement
Yield		Y_w	Raw phenotypes	Chapter 3	High yield
Harvest Date		HD	Raw phenotypes	Supplementary Information	-

Selection indices have been introduced as the multi-trait phenotypic performance of individuals, and defined as the linear combination of observed phenotypes weighted by coefficients taking into account economic/agronomic/technical value and genetic variance components of each phenotypes (Hazel, 1943; Lin, 1978; Gallais, 1990):

$$I = \sum_{i=1}^n b_i * x_i,$$

where n is the number of traits, b_i the weights for each traits, and x_i the observed phenotypic values (Hazel, 1943). The concept for this type of selection index was to estimate the genetic values from the observed phenotypic values weighted by estimated genetic parameters such as the heritability, and the correlations between traits (Gallais, 1990). For instance, Smith *et al.* (1981) proposed a linear index weighted by the heritability of traits. Another simple form of index is the rank index, $I = \sum_{i=1}^n rank(x_i)$, in which individuals are ranked for each trait, and ranks are added across several traits for each individual. Elston (1963) has proposed a multiplicative index, $I = \prod_{i=1}^n (x_i - k_i)$ where k_i is a threshold required for each trait. Thresholds have been defined as the minimum value accepted and depending on the breeding goals, the quality of the data and the relative importance of each trait. However, selection index theory has been developed before the development of biotechnologies and molecular markers, and at that time models used to estimate genetic values were based on phenotypic observations, only.

Nowadays, the genetic values can be directly estimated with the use of molecular markers, and used in selection indexes. The development of quantitative genetics has allowed a better knowledge of the genetic architecture of traits and the precise identification of loci involved in trait variation. In addition, it permits to estimate the genetic value of individuals more and more precisely based on molecular markers. Marker Assisted Selection (MAS) has been proposed as a genetically assisted breeding strategy for simply inherited traits. MAS has been used in peach breeding to select for low acidity and flat fruit shape. Nowadays, it is currently used in some breeding programs to pyramide scab resistance major genes in apple (Patocchi *et al.*, 2009) and in apricot for sharka resistance (Lambert P. *et al.*, 2009; Patocchi *et al.*, 2009). The use of QTL associated markers scores in the context of MAS has been proposed by Lande and Thompson (1990) under the form of selection indexes for early breeding of traits that were not expressed at

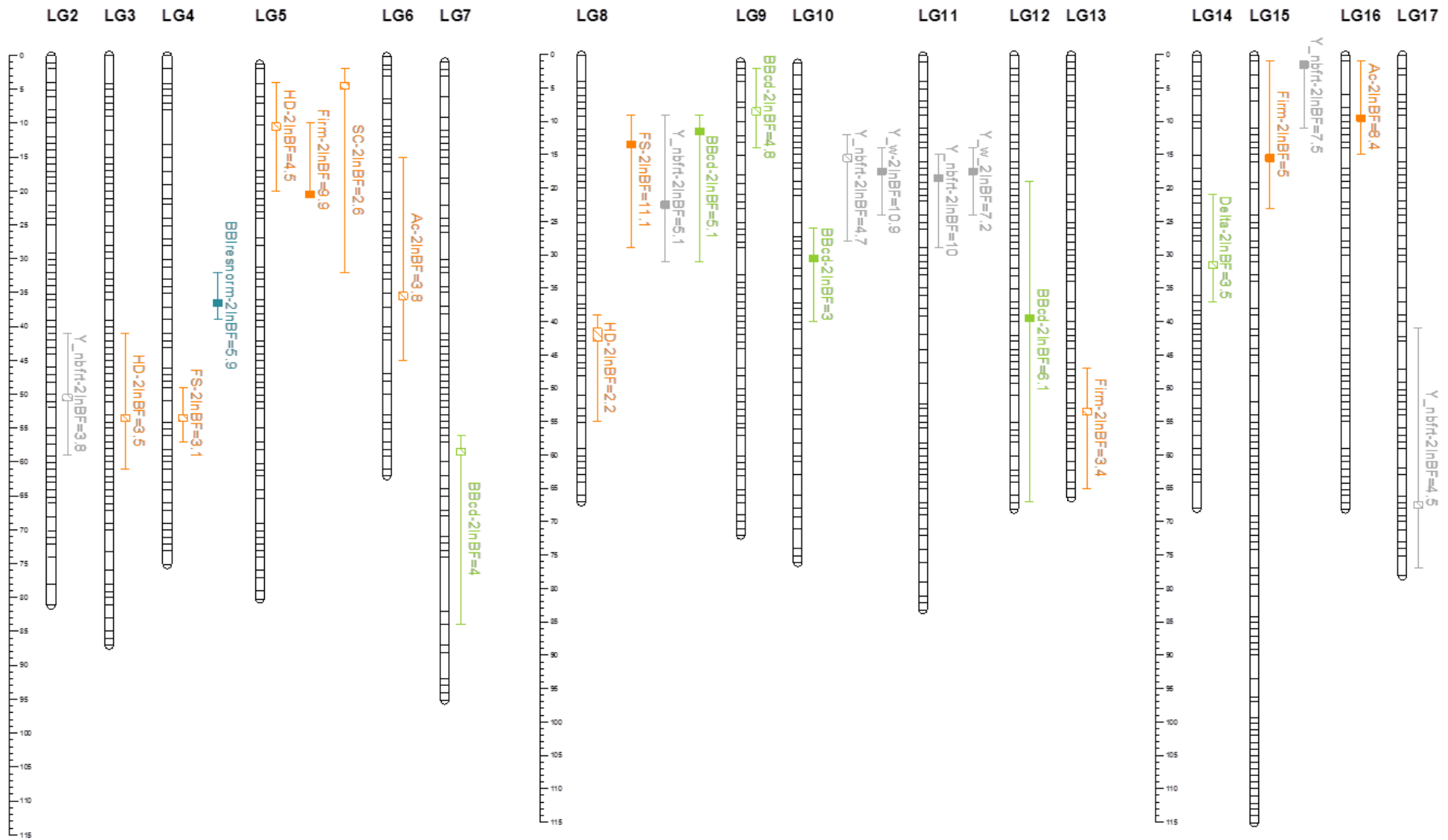


Figure 1. Overview of QTL detected for all the traits studied. QTLs at least positive evidence were shown in this figure. In Orange: QTLs related to fruit quality and fruit maturity. In blue, traits related to bearing habit. In green: traits related to budbreak and flowering phenology. In grey: traits related to yield.

juvenile stages allowing the selection on young plants for which no phenotype was available. Although QTL mapping has allowed progress in the understanding of many quantitatively inherited traits, the use of QTLs in breeding remains limited. Indeed, a review on forest tree breeding has revealed that QTL use was hampered by the over estimation of QTL effects, the lack of colocalization when changing the genetic background, and the low QTL position resolution (Thavamanikumar *et al.*, 2013). Particularly, in *Eucalyptus* molecular markers have been used for management of genetic variation diversity, variation and relationship but the use of QTL was not intensive because of the complexity of the species regarding its polymorphism, the impossibility of developing inbred lines and the few simply inherited traits for (Grattapaglia and Kirst, 2008). Similarly, in fruit trees, few breeding programs presently use the output of the genetics or genomic research (Laurens *et al.*, 2012; Troggio *et al.*, 2012).

In 2001, Meuwissen *et al.* has proposed to estimate breeding values based on genome wide markers effects without previous marker selection. This proposition consisted in a simultaneous estimation of all marker effects evenly and densely spaced over the genome. This implies some technical difficulties linked to the high number of parameters to estimate with the low number of observed data, which can be overcome with new statistical tools such as Bayesian statistics. The genome wide estimation of breeding does not require any prior knowledge on the determinism of the traits which is an advantage compared to MAS, especially for quantitative traits for which it is difficult to identify all the loci involved in the phenotypic variation. Finally, genome wide breeding values has been considered to be the best selection criterion by (Jannink *et al.*, 2010). In apple, Kumar *et al.* (2013) explored genomic selection accuracy for six fruit quality traits, and concluded that genomic selection is a credible alternative to conventional selection for these traits. Also, Muranty *et al.* (2015) have proved that genome wide breeding values could be extended to a larger range of traits and that genomic prediction could accelerate breeding progress in outbred fruit tree crops that still need to overcome long generation intervals and extensive phenotyping costs.

The objectives of the present study was to extend this strategy by combining a breeding value estimation to a selection index approach for combining traits of different natures. For this, we have estimated the breeding values for a range of traits related to fruit quality, yield quantity and stability, and timing of flowering, either genome wide or on QTL zones only, and we have

Table 2. QTLs overview for all the traits studied. First column indicates the variable concerned. Second to fourth columns indicate the linkage group, the $2 \times \ln(\text{Bayes Factor})$ and QTL position in cM. The fifth column indicates the QTL additive effect. The sixth, seventh, and eighth columns indicate the frequency of the positive allele among founders, among parents, and on the overall population, respectively. The last column indicates the percentage of variance explained by QTL.

Variable	LG	2lnBF LG scale	Position (cM)	Additive effect	Favorable allele Frequency among founders	Favorable allele Frequency among parents	Favorable allele Frequency in population	%var
Ac	6	3.8	15-45	0.94	0.45	0.43	0.39	15.0
	16	8.4	1-15	0.96	0.49	0.50	0.48	16.4
Firm	5	9.9	10-20	0.84	0.51	0.50	0.51	18.9
	13	3.4	47-65	0.66	0.47	0.47	0.52	11.7
	15	5	1-23	0.75	0.48	0.49	0.44	15.1
FS	4	3.1	49-57	20.56	0.43	0.44	0.48	9.5
	8	11.1	9-29	24.33	0.44	0.43	0.48	13.3
SC	5	2.6	2-32	0.62	0.43	0.44	0.45	12.7
Yield_nbfrt	2	3.5	41-59	57.49	0.47	0.48	0.48	8.0
	8	5.1	9-31	48.17	0.51	0.50	0.48	5.6
	10	4.7	12-28	96.13	0.40	0.39	0.39	21.4
	11	10	15-29	113.30	0.43	0.42	0.41	30.2
	15	3.5	0-21	116.38	0.61	0.62	0.60	31.5
	17	5.7	41-77	75.82	0.49	0.47	0.45	13.8
Yield_w	10	10.9	14-24	11.97	0.39	0.38	0.39	40.1
	11	7.2	13-27	10.18	0.49	0.50	0.50	30.6
BBIresnorm	4	5.9	32-39	0.38	0.39	0.37	0.38	23.1
Autocov	No QTL							
BBcd	7	4	56-84	1.45	0.52	0.52	0.45	10.4
	8	5.1	9-31	1.15	0.47	0.46	0.51	6.6
	9	4.8	2-14	1.63	0.5	0.51	0.51	13.32
	10	3	26-40	1.14	0.46	0.45	0.60	6.24
	12	6.1	19-67	1.88	0.62	0.63	0.62	16.71
Delta	9	4	2-14	110.00	0.38	0.37	0.39	15.9
	14	3.5	21-37	89.66	0.41	0.41	0.44	11.5
HD	3	3.5	41-61	4.06	0.53	0.53	0.54	13.5
	5	4.5	4-20	3.87	0.53	0.53	0.57	12.1
	8	2.2	39-55	5.67	0.60	0.64	0.70	22.1

compared the results of these two methods . Secondly, we have defined an *a priori* ideotype that would integrate those traits and that was used to weight the traits into a selection index. This ideotype must be considered as a purely speculative study case that would need to be deeply redesigned for a more concrete breeding program. Then, we have explored the correlations among traits in order to identify which traits could be easily associated in the predefined ideotype, based on those correlations. Finally, selection indexes were used to estimate multi-trait breeding values of individual trees. Traits were considered in two steps, separating those considered in the framework of this case study as “culling” and those considered as secondary traits. Two types of calculations were used, additive or multiplicative, for comparison. The overall exercise propose here does not correspond to actual breeding conditions in apple tree, because of lacking of important characters and the arbitrary ideotype used. Rather, it constitutes a case study to elaborate new findings on the possible use of selection indices based on the estimation of genome wide breeding values.

Materials and methods

Plant material and phenotypes

This study gathered all the characters that were studied in the previous chapters. Compared to previous chapters, only two families were used because they were phenotyped for all the traits and implanted in Montpellier. The first one was derived from ‘Starkrimson’ (a mutant of ‘Red Delicious’, hereafter called ‘Delicious’) and ‘Granny Smith’, and called SG; and the second one was derived from X-3263 and ‘Belrene’, and called XB. These two families were described with more details in previous chapters.

Three main characters were considered in this study: fruit quality, fruit bearing habit and phenology that were decomposed into several elementary traits chosen for representing the complexity of the character (Table 1). In addition yield was also considered as it is a culling trait for growers. Fruit quality was represented by sugar content (SC) measured by refractometry in % Brix, acidity (Ac) measured by titration in gram of malic acid per liter (g. ac. mal/L), and firmness (Firm) measured by penetrometry in kg per square centimeter (kg/cm²). Fruit size (FS) was also assessed by estimating the average fruit weight in grams (g). Yield was expressed in number of fruit (Y_nbfrt) and in kg (Y_w) and the harvest date (HD) in calendar days. These

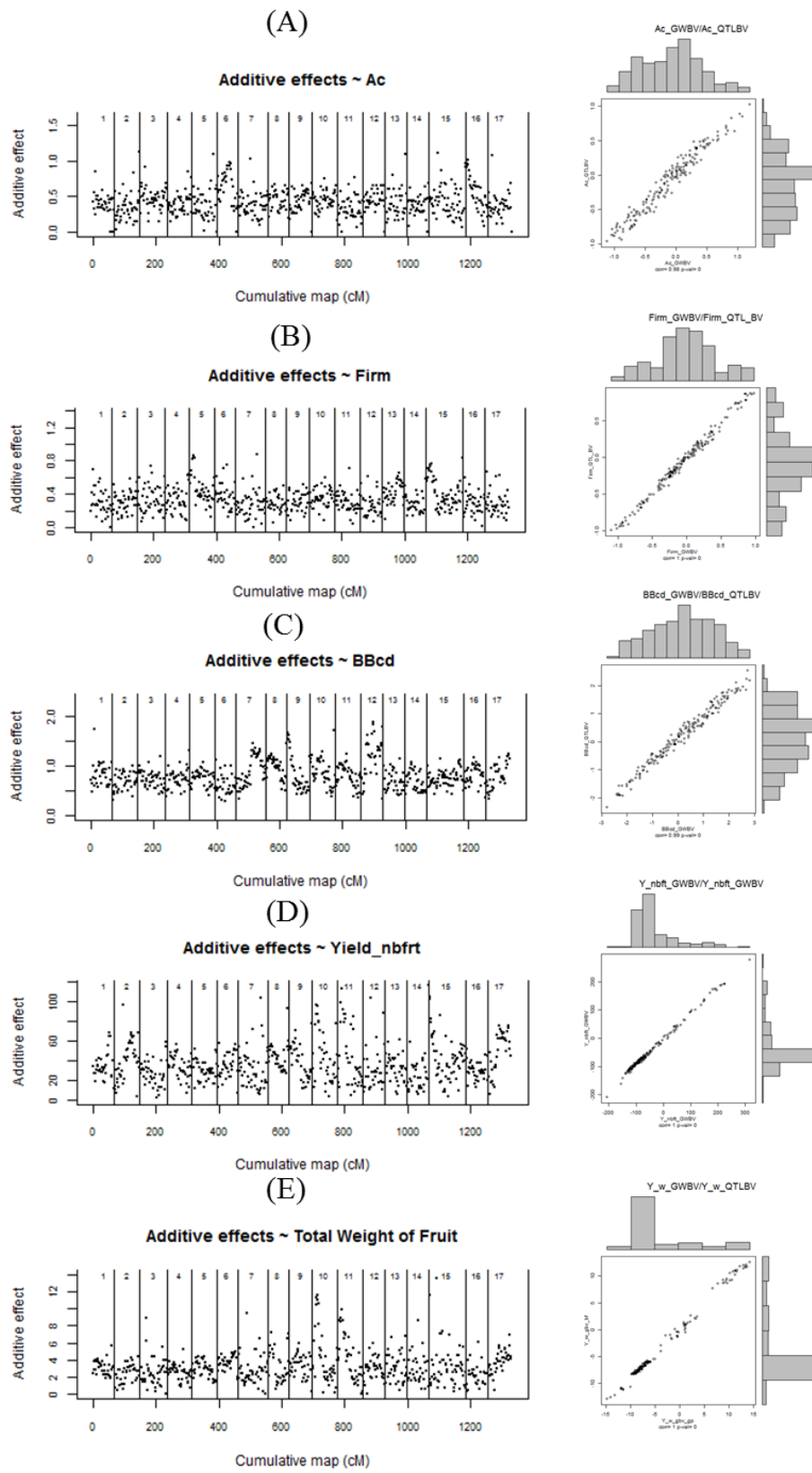


Figure 2. Genetic architecture of traits represented by genetic additive effects along the genome for (A) Ac, (B) Firm, (C) BBcd, (D) Y_nbfrt and (E) Y_w.

characters and elementary traits were quite heterogeneous and of different natures (raw data, BLUPS, model parameters and indices). Raw data were used directly for QTL mapping and breeding value estimation. Bearing habit was assessed by a biennial bearing index (BBIresnorm) and a yield autocorrelation parameter (Autocov) that were used as variables for QTL mapping and breeding value estimation. Finally, concerning phenology, genotypic Best Linear Unbiased Predictors (BLUPs) were extracted from mixed linear models for two traits, budbreak date expressed in calendar days (BBcd) and time between budbreak and flowering expressed in growing degree hours (Delta). These BLUPs were used as variables for QTL mapping and breeding values estimation. All these traits, except yield were shown to be heritable in the previous chapters. See Table 1 for trait abbreviation and recap.

Ideotype definition

To define ideotype, hypothetical breeding goals were determined for each trait. Traits were hierarchized, considering the most important ones as the first targets. This first set of traits contained fruit size, sugar content, yield and regularity. High sugar content and big fruits were considered desirable and thus favored, such as high yields and yield regularity over years. A second set of traits, considered as secondary contained firmness, acidity, and phenology. High firmness was favored since it is generally preferred by Europeans consumers (Bonany *et al.*, 2014). High acidity was favored arbitrarily although it could correspond to preference of some European country such as Germany (Bonany *et al.*, 2014). Concerning traits linked to climate adaptation, we decided to consider warm climatic conditions as target. Therefore individuals with low chilling requirements would be preferable. In complement, in order to avoid too precocious bud break that could be endangered by frost at spring, we favored individuals with high heat requirements. This heating requirement was represented in our data by the number of growing degree hours (GDH) between budbreak and beginning of flowering. See Table 1 for breeding goals recap. All the breeding goals taken together represented the ideotype chosen. This ideotype didn't contain an exhaustive list of traits that may be used in fruit trees breeding (e.g. see Dickmann *et al.*, 1994). Indeed, disease resistance traits, additional fruit quality parameters, growth characteristics could be added in order to have a more complete ideotype.

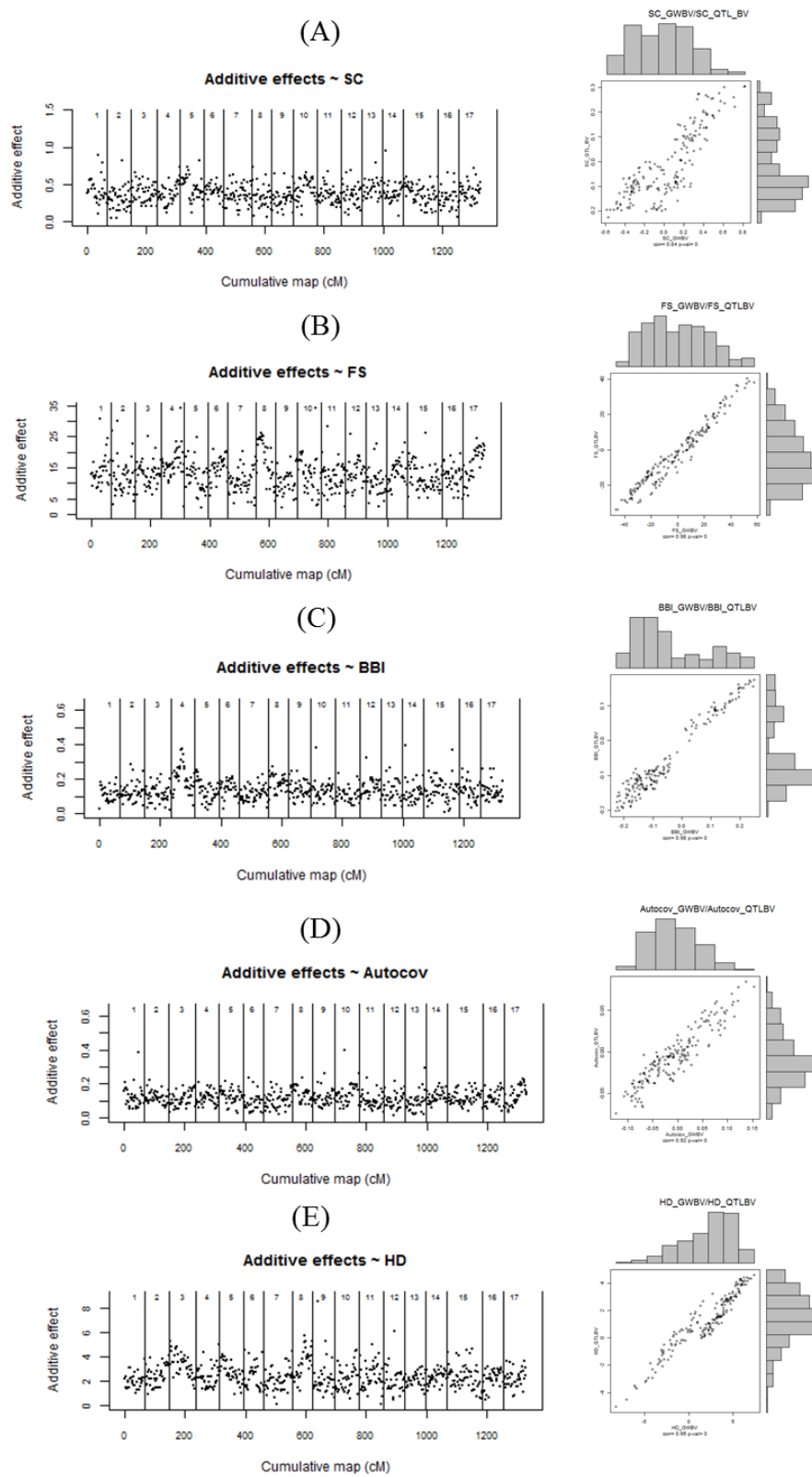


Figure 3. Genetic architecture of traits represented by genetic additive effects along the genome for (A) SC, (B) FS, (C) BBIresnorm, (D) Autocov and (E) HD.

From QTL mapping to estimation of local genetic effect and Breeding Values

The genetic determinism of all traits mentioned in previous paragraphs have been deciphered in a multi-family context by QTL mapping using a Bayesian method, FlexQTL™ (Bink *et al.*, 2014). In the current study, the same Bayesian QTL mapping method was used in order to map QTLs (see previous chapters for QTL mapping priors, parameters and thresholds) and estimate the parameter required for breeding value estimation. Since the QTL mapping consisted in scanning the genetic map for genotype-phenotype associations from fragments of 1 cM, (called bins thereafter), the following parameters were available for each bin: the Bayes Factor, for comparison between models with different number of QTLs, the additive effects, the frequency of positive allele and the genotype estimation for each individual.

The additive effects are parameters from the following model:

$$y \sim N(1\mu + Wa, \sigma_e^2)$$

where y is the phenotypes vector, μ the overall mean, and a the vector of additive effects. X is the design matrix for the environmental effects and W the design matrix for the QTL effects. The prior on the additive effects is: $a \sim N(0, I\sigma_a^2)$ where σ_a^2 is the variance explained per QTL.

Additive effects were used for different purposes. Firstly, they allowed us to compare the distribution and intensity of additive bin effects along the genome and between traits. Secondly, in addition to other parameters such as positive allele frequency and individual genotype estimation, they were used for the computation of two types of breeding values based on the QTL detected or on genome wide parameters estimation (Bink *et al.*, 2014).

The first one cumulated allelic effects of all loci without selection of highest effect loci. It was called Genome Wide Breeding Value (GWBV) thereafter, and resulted from the following formula:

$$Gebv_i = \sum_{bin=1}^{Nbin} \left(\frac{1}{N_{sampl}} * \sum_{t=1}^{N_{sampl}} \sum_{q=1}^{N_{qtl}} I(\lambda_q)^{(t)} w_{i,q}^{(t)} a_q^{(t)} \right)$$

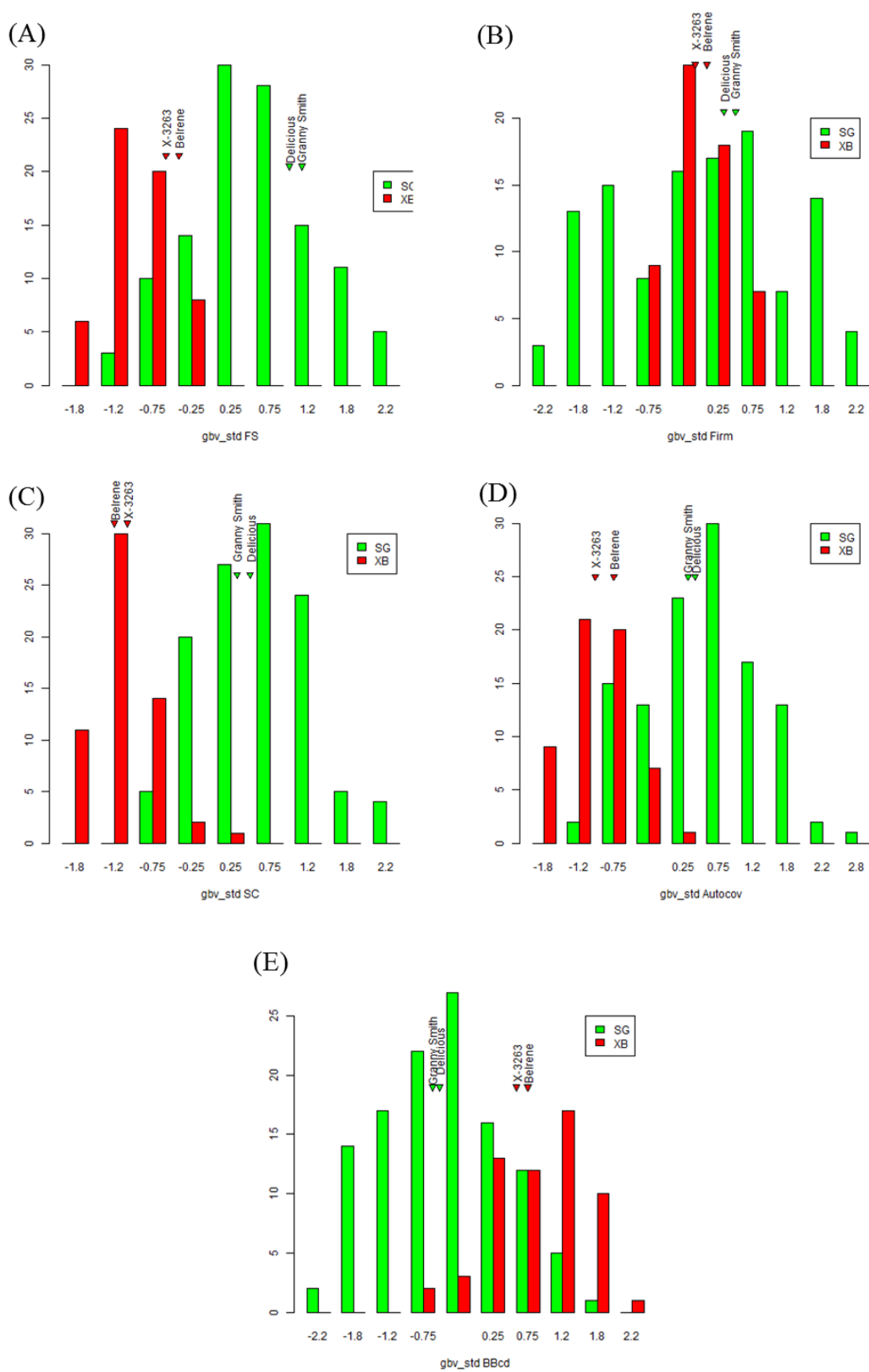


Figure 4. Genome wide breeding value distribution per family for killer traits, SG in green and XB in red, for (A) FS, (B) Firm, (C) SC, (D) Autocov, and (E) BBcd.

where i represents individuals, N_{bin} the number of loci (or bins), t the iteration concerned, N_{sampl} the number of stored samples from the Marcov Chain Monte Carlo, N_{qtl} the number of QTL allowed in the model, $I(\lambda_q)^{(t)}$ an indicator of QTL presence at the iteration, $w_{i,q}^{(t)}$ the estimated genotype and $a_q^{(t)}$ the additive effect.

The second type of breeding value, called QTL Breeding value (QTLBV), had a similar formula but cumulated only the loci with the highest effects. This breeding value could be considered as cumulating QTL effects, only. It resulted from the following formula:

$$G_{ebv_i} = \sum_{bin=1}^{N_{bin}} \left(\frac{1}{N_{sampl}} * \sum_{t=1}^{N_{sampl}} \sum_{q=1}^{N_{qtl}} BF(\lambda_q) I(\lambda_q)^{(t)} w_{i,q}^{(t)} a_q^{(t)} \right)$$

With $BF(\lambda_q)$ an indicator with value 1, if the Bayes Factor of the loci exceeds 2, and 0 otherwise.

GWBV and QTLBV were computed with the same loci parameters estimations, but for GBV all loci were used whereas for QTLBV only loci with a $2 * \ln BF$ value higher than 2 were used. Therefore, GBV represented the genome wide cumulated effects, whereas QTLBV represented the cumulated QTL effects. For this reason, these two breeding values are highly positively correlated especially when QTL could explain a large part of the genetic variability.

In order to be comparable between traits, the breeding value were standardized by subtracting the mean value and dividing by the standard deviation.

Correlations and Principal component analyses

Correlations between GBV and QTLBV were assessed in order to determine which breeding value was the most adequate estimator, depending on the genetic architecture of the traits. Correlations between traits were also assessed using their phenotypes, and their estimated GBV and QTLBV. Correlations between traits were assessed for the two families taken together and for the families separated. All correlations were computed with the R function ‘*rcorr()*’ of the R package ‘Hmisc’ (Harrell *et al.*, 2015). Principal component analyses were used to supplement the correlation analyses between traits, and were realized with the R package FactoMiner (Lê *et al.*, 2008).

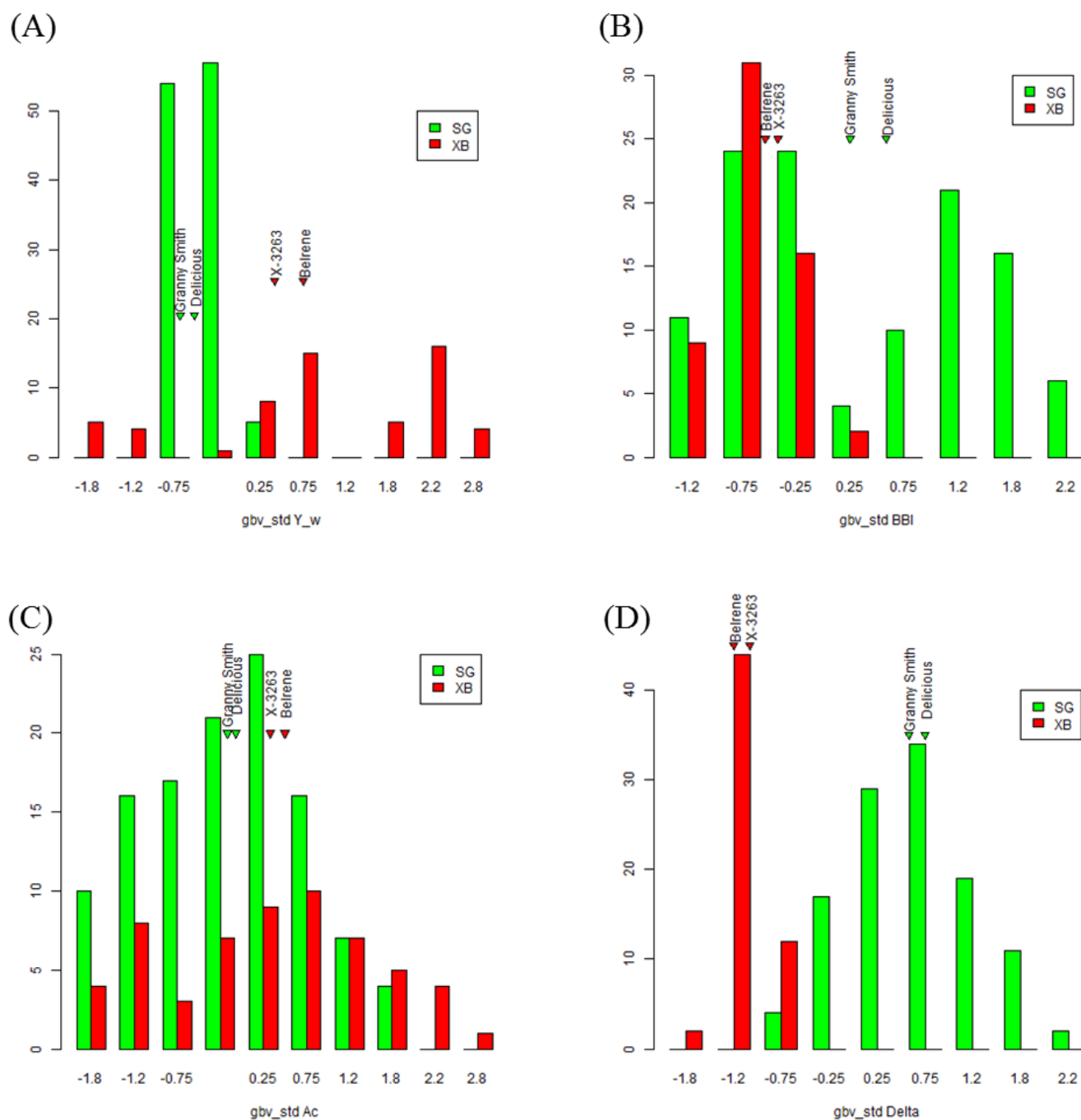


Figure 5. Genome wide breeding value distribution per family for killer traits, SG in green and XB in red, for (A) Y_w, (B) Ac, (C) Delta, and (D) BBI.

Selection Index calculation

Based on the ideotype previously presented, the two sets of traits identified according to their importance and the correlation among traits, we defined a strategy for calculating selection indices. The first set of trait (S1) that was chosen as the most important traits among the traits available, contained fruit size, sugar content, BBiresnorm, Autocov, and yield. The second set of traits (S2) contained acidity, firmness, BBcd and Delta. They were considered as secondary traits, to be combined to the most important one in a second step of selection. For each set of traits, two indexes were chosen, an additive one and a multiplicative one:

$$\begin{aligned} \text{Set 1 Additive Index (S1ad)} = \\ (1) * \text{FruitSize}_{\text{gbv}} + (1) * \text{SugarContent}_{\text{gbv}} + \\ (-1) * \text{BBiresnorm}_{\text{gbv}} + (1) * \text{Autocov}_{\text{gbv}} + \\ (1) * \text{Yield}_{\text{gbv}} \end{aligned}$$

$$\begin{aligned} \text{Set 1 Multiplicative Index (S1m)} = \\ [(1) * \text{FruitSize}_{\text{gbv}}] * [(1) * \text{SugarContent}_{\text{gbv}}] * \\ [(-1) * \text{BBiresnorm}_{\text{gbv}}] * [(1) * \text{Autocov}_{\text{gbv}}] * \\ [(1) * \text{Yield}_{\text{gbv}}] \end{aligned}$$

$$\begin{aligned} \text{Set 2 Additive Index (S2ad)} = \\ (1) * \text{Acidity}_{\text{gbv}} + (1) * \text{Firmness}_{\text{gbv}} + \\ (-1) * \text{BBcd}_{\text{gbv}} + (1) * \text{Delta}_{\text{gbv}} \end{aligned}$$

$$\begin{aligned} \text{Set 2 Multiplicative Index (S2m)} = \\ [(1) * \text{Acidity}_{\text{gbv}}] * [(1) * \text{Firmness}_{\text{gbv}}] * \\ [(-1) * \text{BBcd}_{\text{gbv}}] * [(1) * \text{Delta}_{\text{gbv}}] \end{aligned}$$

All traits were considered with a same weight, but with a difference sign (-1/+1) in order to assign the highest index value to individuals corresponding to the targeted ideotype (Tables 9 and 10). For the multiplicative index, in addition to standardization, the breeding value distributions were shifted toward positive values by adding the absolute value of the lowest value of the distribution. The reason for this second transformation was to avoid negative values that would interfere with interpretability of the multiplicative indexes. For each index, approximately 10 individuals with the highest values were identified as the most desirable individuals.

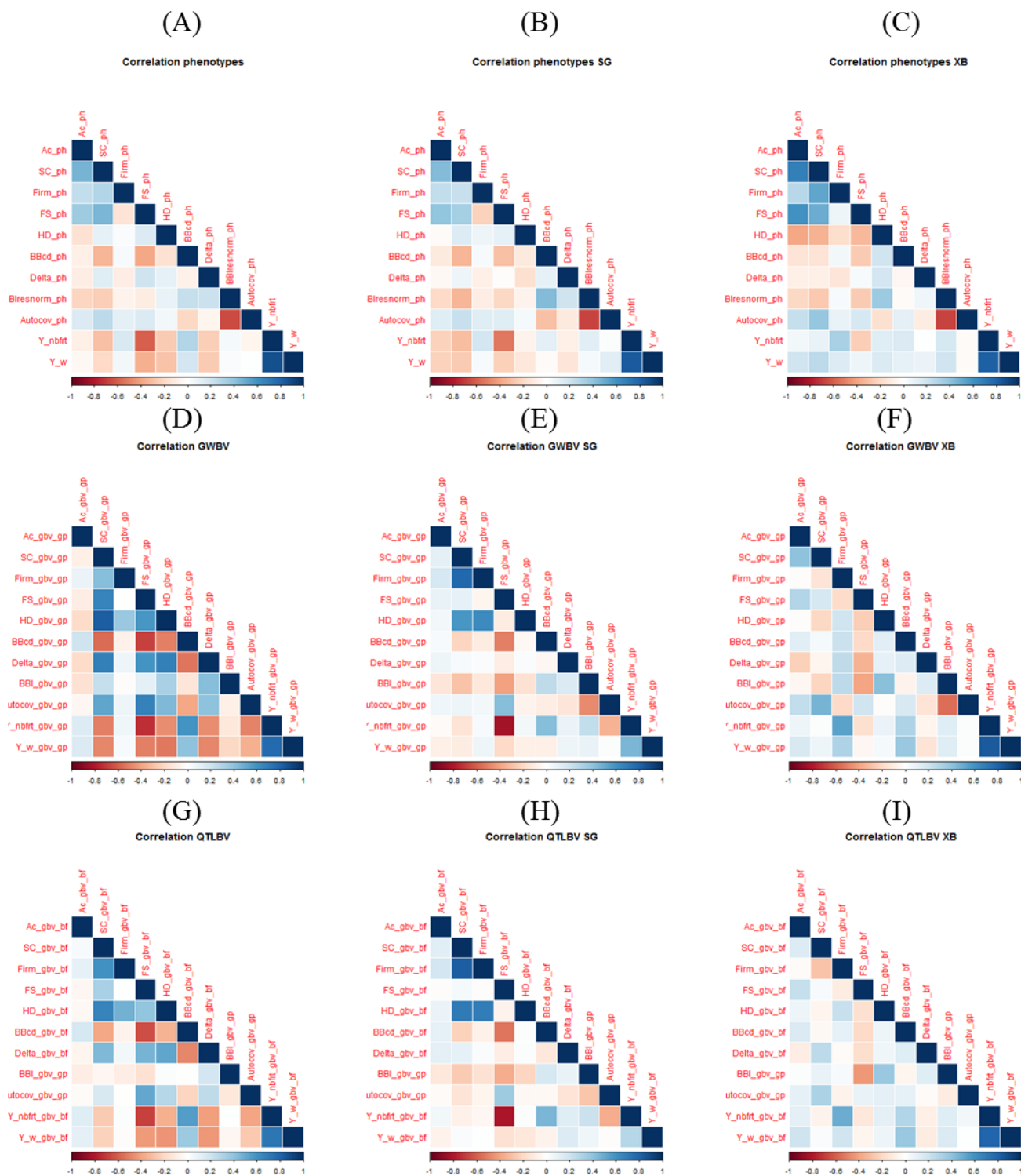


Figure 6. Correlation between phenotypes, (A) for the two families, (B) for SG, (C) for XB, between GWBV (D) for the two families, (E) for SG, (F) for XB, and between QTLBV (G) for the two families, (H) for SG, (I) for XB. Dark blue represents the highly positive correlations, dark red represents the highly negative correlations and pale colors represent the weak correlations.

Results

QTL overview traits genetic architecture and estimation of breeding value

The overview of the QTLs detected for the three characters, fruit quality, phenology, and fruit bearing habit, showed that there were at least one QTL on each linkage group, except LG1 (Figure 1). Few colocalizations between QTLs belonging to different characters were observed. At the top of LG8, QTLs for FS, and BBcd colocalized, and a QTL for Y_nbfrt was mapped in the same region but its peak was lower. At the top of LG15 two QTLs for Y_nbfrt and Firm were localized closeby but their peaks were in different positions (Figure 1). In addition, QTLs explained between 5 and 40% of the variance (Table 2). In total, 20 loci were considered as strong QTLs for one or several traits.

The distribution of QTL effects along the genome gave insights about the genetic architecture of the traits (Figure 2 to 4, left panel). For all traits, many loci with variable effect size were observed in between the peaks which represented QTLs. In a QTL mapping context, small effect loci are ignored whereas in a genome wide context effect of all loci can be taken into account. The correlation between the breeding values calculated either genome wide (GWBV) or on QTLs (QTLBV), gave insights on the impact of the small effect loci (Figure 2 to 4, right panel).

For traits with well-defined QTLs (Ac, Firm, BBcd, Y_nbfrt, and Y_w; Figure 2), GWBV and QTLBV were highly correlated. For these traits, taking into account for small effect loci did not impact the estimation of the breeding value, which could mean that QTLs explained a large part of the variance, and that small effect loci could be considered as insignificant. On the contrary, traits for which no or few weak QTLs were detected (SC, FS, BBIresnorm, Autocov, Delta and HD; Figure 3) had a weak correlation between their respective GWBV and QTLBV. GWBV was chosen over QTLBV for all traits, because it allows accounting for small effect loci which can explain a large part of the genetic variation when cumulated.

Breeding values per family

SG and XB had similar FS and Firm mean values, but the distribution was wider for SG (Figure 4A and 4B). SC and Autocov mean values were higher in SG than in XB, and the

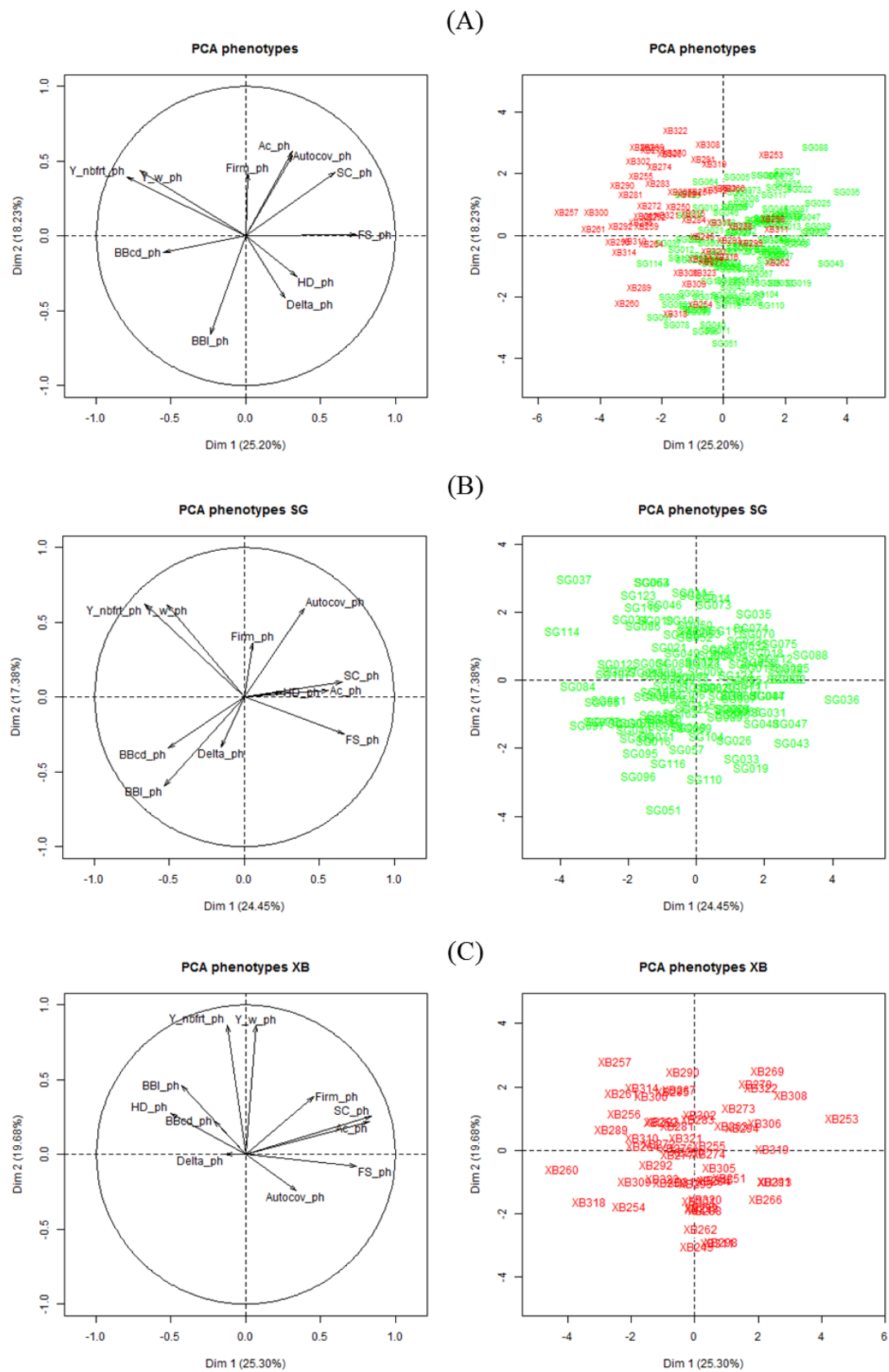


Figure 7. Phenotypes principal component analysis representations, variables in left panel and individuals in right panel for (A) all individuals, (B) SG family and (C) XB family.

distributions were also wider for SG (Figure 4C and 4D). For BBcd, SG distribution was wider but located toward lower value than in XB (Figure 4E). According to the ideotype chosen, SG displayed more desirable individuals than XB for FS, Firm and SC. For Autocov, XB displayed more interesting individuals since it had high Autocov which corresponded to regular individuals. The BBcd distribution included more desirable individuals in SG because the low value corresponded to the targeted precocious individuals.

For the other traits, the distributions included less desirable individuals in the SG family according to the ideotype. By contrast, Y_w displayed a very wide variability in XB, in which the highest yields were found (Figure 5A). Concerning BBiresnorm, XB distribution overlapped SG distribution toward the smallest values, suggesting SG displayed more extreme patterns of fruit bearing irregularity or bienniality than XB (Figure 5B). For Acidity, the distribution of the two families had almost a similar variability with XB having more individuals with very high values (Figure 5C). Concerning Delta, XB distribution was very narrow and with low values (Figure 5D) compared to SG distribution.

Correlation between traits and Principal Component Analyses

Correlations were higher among breeding values (Figure 6D and 6E) than among phenotypes (Figure 6A), and higher among GWBV (Figure 6D) than QTLBV (Figure 6G). This was confirmed by PCAs. PCA on phenotypes revealed that the first two axes explained only 25.23 and 18.16% of the overall variation, respectively (Figure 7A), whereas 5 axes explained almost 75% (Table 3). PCA on GWBV revealed that the first two axes explained 48.68% and 12.63%, respectively (Figure 8A), and four axes were required to explain 75% of the trait variability (Table 6).

Phenotypic values of Acidity were slightly positively correlated with the other fruit quality traits (SC, Firm and FS) (Figure 6A) but breeding values didn't revealed any correlation of Acidity with any other fruit quality traits (Figure 6, and 6G). Firmness was not correlated to other traits, except slightly positively correlated with SC and HD (Figure 6A, 6D, 6G). These two results on Ac and Firm were confirmed by PCA (Figure 7A and 8A).

Table 3. Principal Components Analysis outputs for overall phenotypes.

Variance explained per dimension			Trait correlation with dimensions					
	% variance	cumulative % variance		Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Dim 1	25,2	25,2	BBcd_ph	-0,55	-0,11	0,09	0,21	-0,41
Dim 2	18,2	43,4	Delta_ph	0,26	-0,42	0,22	-0,03	0,03
Dim 3	11,9	55,3	BBI_ph	-0,24	-0,66	0,56	0,03	0,11
Dim 4	9,9	65,2	Autocov_ph	0,31	0,54	-0,54	0,18	0,03
Dim 5	8,5	73,7	Ac_ph	0,31	0,57	0,49	-0,21	-0,19
Dim 6	7,9	81,7	Firm_ph	0,01	0,42	0,40	0,69	-0,16
Dim 7	7,0	88,6	SC_ph	0,59	0,43	0,43	0,06	0,11
Dim 8	4,1	92,7	FS_ph	0,74	0,01	0,16	-0,38	0,18
Dim 9	3,5	96,2	mat_ph	0,34	-0,27	-0,06	0,53	0,60
Dim 10	2,9	99,1	Y_nbfrit_ph	-0,79	0,40	0,13	-0,04	0,35
Dim 11	0,9	100,0	Y_w_ph	-0,71	0,44	0,16	-0,23	0,40

Table 4. Principal Components Analysis outputs for SG phenotypes.

Variance explained per dimension			Trait correlation with dimensions					
	% variance	cumulative % variance		Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Dim 1	24,45	24,45	BBcd_ph	-0,51	-0,34	0,38	-0,03	-0,35
Dim 2	17,38	41,84	Delta_ph	-0,16	-0,34	0,16	0,58	-0,12
Dim 3	11,58	53,42	BBI_ph	-0,54	-0,59	0,12	-0,06	0,38
Dim 4	10,49	63,91	Autocov_ph	0,40	0,59	-0,04	0,06	-0,53
Dim 5	8,99	72,90	Ac_ph	0,56	0,04	0,36	0,31	0,20
Dim 6	8,06	80,95	Firm_ph	0,05	0,36	0,80	-0,19	0,14
Dim 7	6,58	87,53	SC_ph	0,65	0,10	0,31	0,19	0,32
Dim 8	4,37	91,91	FS_ph	0,67	-0,24	-0,37	0,17	0,26
Dim 9	3,99	95,90	mat_ph	0,25	0,03	-0,04	-0,73	0,18
Dim 10	3,00	98,90	Y_nbfrit_ph	-0,67	0,62	0,04	0,11	0,21
Dim 11	1,10	100,00	Y_w_ph	-0,52	0,62	-0,28	0,24	0,35

Table 5. Principal Components Analysis outputs for XB phenotypes.

Variance explained per dimension			Trait correlation with dimensions					
	% variance	cumulative % variance		Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Dim 1	25,30	25,3	BBcd_ph	-0,21	0,22	0,59	-0,31	0,41
Dim 2	19,68	45	Delta_ph	-0,14	0,00	-0,34	0,54	0,44
Dim 3	14,2	59,2	BBI_ph	-0,43	0,46	-0,59	-0,24	0,17
Dim 4	9,87	69,1	Autocov_ph	0,34	-0,24	0,76	0,26	0,04
Dim 5	8,91	78	Ac_ph	0,83	0,22	-0,15	-0,10	0,20
Dim 6	7,87	85,8	Firm_ph	0,45	0,39	0,02	-0,52	-0,17
Dim 7	5,66	91,50	SC_ph	0,84	0,25	-0,02	-0,12	0,21
Dim 8	3,14	94,64	FS_ph	0,74	-0,08	-0,21	0,23	0,26
Dim 9	2,45	97,08	mat_ph	-0,50	0,28	0,24	-0,13	0,56
Dim 10	2,12	99,20	Y_nbfrit_ph	-0,12	0,87	0,20	0,25	-0,28
Dim 11	0,80	100,00	Y_w_ph	0,07	0,86	0,12	0,38	-0,10

A complex pattern of correlation was observed on the other traits (Figure 6A, 7A and 8A). As expected, negative correlation was observed between BBIresnorm and Autocov, even though correlations values were lower with QTLBV (Figure 6A; Figure 7A and 8A). Y_w and FS were negatively correlated in the overall population (Figure 6A, Figure 7A and 8A). SC, FS and HD were only slightly positively correlated to each other. BBcd and Yield were positively correlated to each other in the overall population (Figure 6A, Figure 7A and 8A).

PCAs also revealed that the phenotypes were structured by families according to the first axis which is explained mainly by Y_w and y_nbft, HD, Delta, and FS (Figure 7A, right panel). This structure was stronger for GWBV and the following traits are involved: SC, FS, Autocov, DH, Delta, BBI, and Y_w, Y_nbft, and BBcd (Figure 8A).

The correlations and PCA realized on separated families revealed that the negative correlation between Fruit Size and Yield traits was observed in SG family and not in XB (Figure 6B, 6C, 6E, 6F, 6H and 6I). Similarly, the correlation between BBcd and Y_w was weaker when the two populations were considered separately. Actually, most correlations were lower when the families were separated, as confirmed by PCA on GWBV.

Indexes Comparison

For the first set of traits, the additive and multiplicative indexes (S1ad and S1m respectively) designated the same individuals as having the highest index value (Figure 9). Desirable individuals, as designated by S1ad and S1m, with low breeding values for BBIresnorm and high values for Autocov, were found mostly in SG, even though some XB individuals had low BBI values (Figure 9A). For Yield and Fruit Size, the individuals designated as the most desirable had either high Fruit Size, in SG, or high yield in XB, but it was not possible to associate both traits in one individual (Figure 9B). The most desirable individuals for sugar content, i.e. with high values, were also designated in SG (Figure 9C). Finally, the most desirable individuals appeared highly dependent on the family, as most of the desirable individuals belonged to the SG family, except for yield.

For indexes defined on secondary traits (S2ad and S2m), additive and multiplicative index gave similar result for XB but different results for SG. In SG, the additive index allowed the

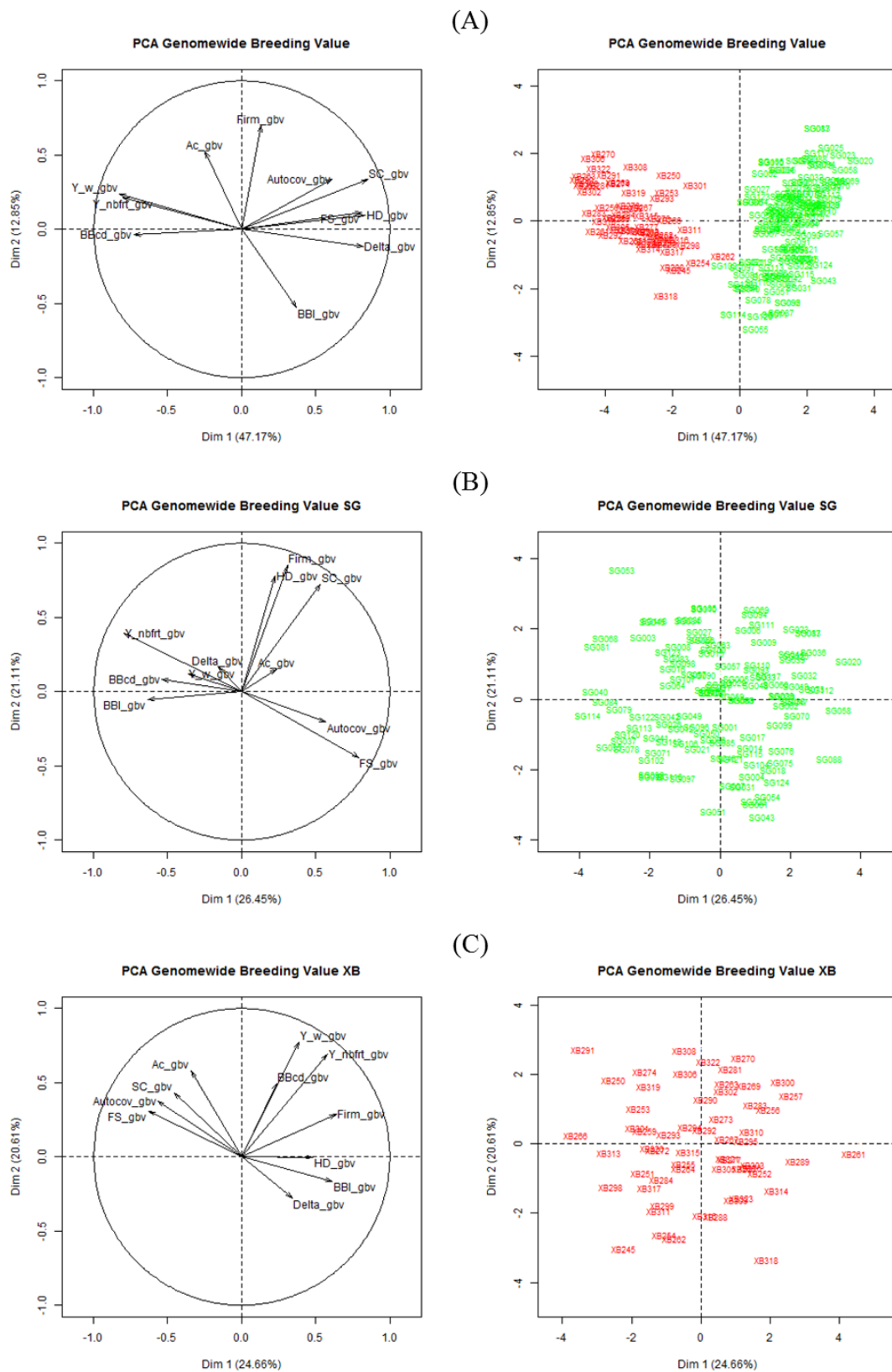


Figure 8. Genome-wide breeding values principal component analysis representations, variables in left panel and individuals in right panel for (A) all individuals, (B) SG family and (C) XB family.

identification of individuals with low BBcd breeding value, whereas the multiplicative index highlighted individuals with high values and few individuals overlapped between the two indexes (Figure 10A). For Delta, both indexes allowed identifying individuals, in the same range of values (Figure 10A). For Firmness, the multiplicative index selected individuals with higher values than the additive index for SG, whereas, for Acidity, both indexes selected different individuals but in the same range of values (Figure 10B). On the contrary, in XB, the individuals with the highest breeding value were almost the same between the additive and multiplicative indexes (Figure 10). Finally, the most desirable individuals designated by S2ad and S2m in each family had low breeding values for BBcd and high breeding values for Delta in SG, but couldn't display high Delta value in XB (Figure 10A). Similarly, the individuals designated as the most desirable had high values for both Firmness and Acidity in SG, but didn't have high firmness values in XB (Figure 10B). See Tables 9 and 10 for a remainder of the ideotypes.

When the indexes calculated on either important or secondary traits were considered jointly, individuals that had the highest value for the two sets of traits and the two kinds of indexes could be identified. In XB, three to four individuals were designated as the most desirable by the four indexes and one to three individuals were designated as the best in SG family (supplementary data).

Discussion

Besides the estimation of breeding values based on molecular markers and the analysis of correlations between traits, this study proposed a method for assessing the multi-trait value of individuals with selection indexes and based on genome wide breeding value estimation. The conditions of the study do not correspond to real conditions in apple breeding. Indeed, important traits were missing such as disease resistance traits and components of fruit quality. Therefore, this method as to be considered as a study case, and would require adaptation if used in real selection conditions (Laurens *et al.*, 2010).

Breeding value estimation

This study considered that depending on the genetic architecture of a trait, QTLs do not have the same efficiency for genetic value estimation. The more quantitative the trait is, the less

Table 6. Principal Components Analysis outputs for overall breeding values.

Variance explained per dimension			Trait correlation with dimensions					
	% variance	cumulative % variance		Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Dim 1	45,80	45,80	Ac_gbv	-0,25	0,52	-0,23	0,75	-0,19
Dim 2	12,56	58,36	SC_gbv	0,85	0,37	0,16	-0,01	-0,03
Dim 3	11,28	69,64	Firm_gbv	0,13	0,78	0,47	-0,21	-0,06
Dim 4	7,77	77,41	FS_gbv	0,80	0,08	-0,37	0,02	0,03
Dim 5	6,66	84,07	Mat_gbv	0,84	0,16	0,25	-0,02	0,17
Dim 6	5,85	89,91	BBcd_gbv	-0,73	-0,03	0,13	0,21	0,28
Dim 7	3,71	93,63	Delta_gbv	0,82	-0,07	0,22	0,15	0,12
Dim 8	2,33	95,96	BBi_gbv	0,39	-0,42	0,56	0,40	0,27
Dim 9	2,06	98,02	Autocov_gbv	0,45	0,22	-0,57	-0,03	0,59
Dim 10	1,24	99,26	Y_nbfrt_gbv	-0,80	0,23	0,20	-0,08	0,31
			Y_w_gbv	-0,83	0,23	0,07	-0,05	0,24

Table 7. Principal Components Analysis outputs for SG breeding values.

Variance explained per dimension			Trait correlation with dimensions					
	% variance	cumulative % variance		Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Dim 1	25,61	25,61	Ac_gbv	0,24	0,15	0,22	-0,23	0,78
Dim 2	21,38	46,99	SC_gbv	0,51	0,74	-0,07	-0,05	-0,06
Dim 3	10,93	57,92	Firm_gbv	0,29	0,85	-0,01	0,10	0,01
Dim 4	9,61	67,53	FS_gbv	0,80	-0,42	-0,18	0,00	0,04
Dim 5	9,33	76,86	Mat_gbv	0,23	0,77	0,14	-0,01	-0,21
Dim 6	7,28	84,14	BBcd_gbv	-0,54	0,05	0,70	-0,09	0,10
Dim 7	6,35	90,49	Delta_gbv	-0,16	0,17	-0,45	-0,54	0,37
Dim 8	4,02	94,51	BBi_gbv	-0,60	-0,07	-0,17	-0,33	-0,16
Dim 9	3,15	97,66	Autocov_gbv	0,46	-0,29	0,32	0,37	0,27
Dim 10	1,42	99,09	Y_nbfrt_gbv	-0,82	0,38	0,06	0,24	0,15
Dim 11	0,91	100,00	Y_w_gbv	-0,40	0,12	-0,51	0,62	0,32

Table 8. Principal Components Analysis outputs for XB breeding values.

Variance explained per dimension			Trait correlation with dimensions					
	% variance	cumulative % variance		Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Dim 1	23,39	23,39	Ac_gbv	-0,23	0,66	0,09	0,28	0,41
Dim 2	20,29	43,69	SC_gbv	-0,37	0,57	0,47	0,01	0,28
Dim 3	13,10	56,79	Firm_gbv	0,69	0,14	-0,01	-0,19	-0,06
Dim 4	10,32	67,10	FS_gbv	-0,57	0,44	-0,34	0,03	0,00
Dim 5	8,03	75,14	Mat_gbv	0,48	-0,08	0,51	0,46	-0,27
Dim 6	7,62	82,75	BBcd_gbv	0,37	0,40	0,27	-0,37	0,12
Dim 7	5,93	88,68	Delta_gbv	0,27	-0,33	0,45	-0,54	0,37
Dim 8	4,68	93,36	BBi_gbv	0,52	-0,24	0,08	0,62	0,40
Dim 9	3,41	96,77	Autocov_gbv	-0,27	0,44	0,63	0,03	-0,47
Dim 10	2,40	99,17	Y_nbfrt_gbv	0,72	0,54	-0,23	-0,04	-0,16
Dim 11	0,83	100,00	Y_w_gbv	0,53	0,66	-0,34	-0,02	0,02

precise the breeding value estimated with QTLs is, because many loci with small effect were not detected as QTL and their cumulated effect could explain a large part of the variance. In this study, two traits can exemplify a poor representation of the breeding value by QTLs: Sugar Content, and Autocov. A single positive QTL was mapped on LG5 for Sugar content which explained a low percentage of the trait variability, even though this trait has been shown to be quantitatively inherited (Visser *et al.*, 1968; Guan *et al.*, 2015). Similarly, no QTL were detected for Autocov even though SG family was known to segregate for fruit bearing habit and QTL have been previously mapped for related traits (Durand *et al.*, 2013). For these traits, the correlation between QTLBV and GWBV was low, suggesting that taking into account small effect loci probably allowed more precise breeding value estimation for quantitatively inherited traits, such as concluded by Meuwissen *et al.* (2001).

In contrast, fruit acidity trait provides an example of trait for which the breeding value defined on QTLs offer an accurate estimation: one major QTL was detected on LG16 probably in linkage disequilibrium with the *Ma* locus responsible for malic acid synthesis, and another minor locus on LG6 (Maliepaard *et al.*, 1998; Xu *et al.*, 2012). In that case, QTLBV and GWBV were highly correlated. A same conclusion could be drawn for traits with more complex genetic architecture but strong QTLs, such as BBcd. In this case, five strong QTLs were detected with meaningful genes mapped in their intervals (Allard *et al.*, in press). As for fruit acidity, a high correlation was found between QTLBV and GWBV. In all cases, the genome wide breeding value estimation, as proposed by Meuwissen *et al.* (2001) can be considered as the ideal selection criterion (Jannink *et al.* 2010), as it is indeed the most precise way to estimate the genetic values, and can be applied on traits with either multi-genic, oligo-genic, or mono-genic genetic architecture.

Effect of intra and inter-family variability and correlations on indices values

For all traits, SG had wider distributions and more desirable values than XB, according to the predefined ideotype. However for yield and regularity, XB may have more interesting individuals than SG. Indeed, SG had a very narrow distribution with low values for yield and a wide variability for fruit size, whereas XB had a narrow distribution with medium values for fruit size and a wider distribution for yield. The overall negative correlation between yield (in both number of fruits and weight) and fruit size may hamper the selection of individuals with the most

Table 9. Breeding goals and index coefficient for the first set of traits.

First set of trait

Trait	Coefficient	Breeding goal
Fruit Size	1	Big fruits are preferred to small fruits
Sugar Content	1	High Sugar Content is necessary for minimum fruit quality
BBIresnorm	-1	Low BBIresnom value are preferred since they represent genotype with low yield variation across years
Autocov	1	High Autocov values are preferred since they discriminate genotypes with biennial fruit bearing habits from genotype with irregular fruiting habit
Yield	1	High Yield are preferred

Table 10. Breeding goals and Index coefficient for second set of traits.

Second set of traits

Trait	Coefficient	Breeding goal
Acidity	1	Medium and high Acidity values are preferred,
Firmness	1	High values are preferred for better texture and better storage ability
BBcd	-1	Low values are preferred in the context of climate change tending toward temperature increase, in order to minimize chilling requirements.
Delta	1	High values are preferred to avoid precocious flowering

desirable values for both traits. This exemplifies trait antagonisms that are common in plant populations either natural or selected (Rieseberg *et al.*, 2003). However, the correlation between yield and fruit size is much stronger in SG than in XB. In addition, even though BB_Iresnorm and Autocov were systematically negatively correlated, more individuals had contrasted bearing behavior in SG than in XB. Consistently with previous results (See Chapter 3), regular individuals with low BB_I-derived indices and high Autocov were found in SG whereas XB contained more irregular individuals. The differences between families for finding appropriate combinations between yield and fruit size or between BB_Iresnorm and Autocov reveal that selection result significantly differ depending on the family. If selection would be applied to both families simultaneously, individuals from XB would be systematically discarded despite the presence of individuals gathering high yield with high fruit size. For all the other traits, the most desirable individuals could be identified by the indexes. These results suggested that the indexes could deal easily with uncorrelated traits and synergic traits whereas only the antagonistic correlations may hamper the selection of the most desirable individuals. This shows that, even though this concept has been initially defined before the use of molecular markers (Hazel, 1943), selection indices can be of interest in modern genetics, as recently shown for animals (e.g. Dekkers 2007; Swan and Brown, 2013; Gonzalez-Recio *et al.*, 2014; Pfeiffer *et al.*, 2015). Indeed, it remains a robust and efficient method for combining traits of different nature (Gallais, 1990).

Grouping characters and breeding steps

A first set of traits, considered as “culling”, were gathered for calculating a first selection index. This may represent a first step of selection scheme during which the individuals with the most undesirable traits are discarded (Laurens *et al.*, 2010). At this step, the two types of index, additive or multiplicative, identified the same individuals as the most desirables. As previously underlined, this may result from both the antagonistic correlation between yield and fruit size combined to the synergistic correlation between BB_Iresnorm and Autocov that may have greatly influenced the values of both indexes towards a few number of individuals.

The second set of trait was composed of slightly or not correlated traits, BB_{cd}, Delta, Firmness and Acidity. Contrarily to what would be performed in a real selection scheme, all

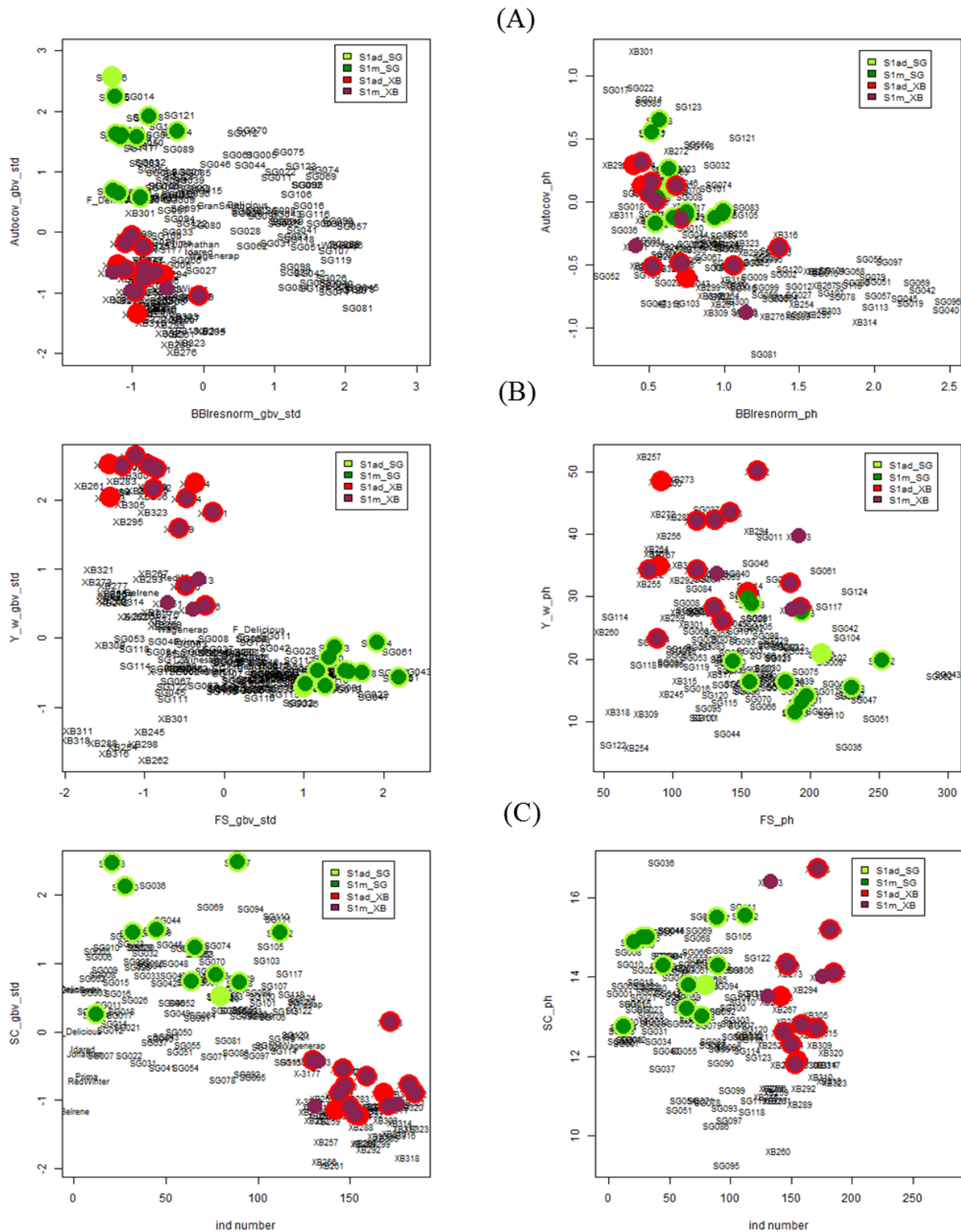


Figure 9. Best individuals repartition through the distribution of breeding values (left panel) and phenotypes (right panel) for (A) BBI plotted against Autocov, (B) Y_w plotted against FS and (C) SC. Individuals highlighted in color are individuals with high index value: for SG family green and dark green for additive and multiplicative index, respectively; for XB family, red and dark red for additive and multiplicative index, respectively.

individuals (not only those selected at the first step) were considered in this second step. Consistently with different distributions of breeding values between families, the two types of indexes, additive and multiplicative, identified behaved differently depending on the family. The same individuals were identified as the most desirables in XB in which low variability was observed whereas different individuals were selected in SG in which a large dispersion was present. For firmness and acidity, both indexes identified similar individuals. On the contrary, the additive index was more efficient than the multiplicative index for BBcd in SG family since it allowed the identification of individuals with lower BBcd values. Even though the reason for a higher efficiency of the additive index remains to be refined, the present analysis revealed the importance of correlations, synergisms or antagonisms, among traits, as well as traits variability and index type in the success of identification of individuals closest to a predefined ideotype.

Accounting for all these parameters may help maximizing the selection efficiency. Even though grouping uncorrelated traits appears pertinent, the more traits considered the more difficult it might be to identify individuals with maximal values in each trait. In this perspective, selection indices may allow finding compromise. Also, grouping synergic traits may help their simultaneous selection, allowing a great gain of time. In contrast, antagonistic traits should be considered separately, and with priorities between them. In this case, hierarchizing traits through the use of successive selection steps is likely to be pertinent. Of course, the traits to gather will highly depend on the pre-defined ideotype(s) which, as previously underlined, can be defined for different context and selection targets (Desclaux *et al.*, 2013; Blum, 2015). However, many efforts are currently deployed for defining crops ideotypes taking benefit from recently developed modelling approaches in annuals (Gu *et al.*, 2014; Martre *et al.*, 2015; Rötter *et al.*, 2015; Van Oijen and Höglind. 2016) and fruit tree (Quilot-Turion *et al.*, 2012; Da Silva *et al.*, 2014).

Conclusion

The purpose of this work consisted in revisiting the concept of selection index, making use of most recent advances in quantitative genetics for estimating genome wide breeding values. For this, we defined an a priori ideotype, combining two sets of traits, either of primordial importance or that could be considered as secondary. As previously shown by Muranty *et al.* (2015), the use of molecular markers to estimate or predict genetic values holds great promises for improving the

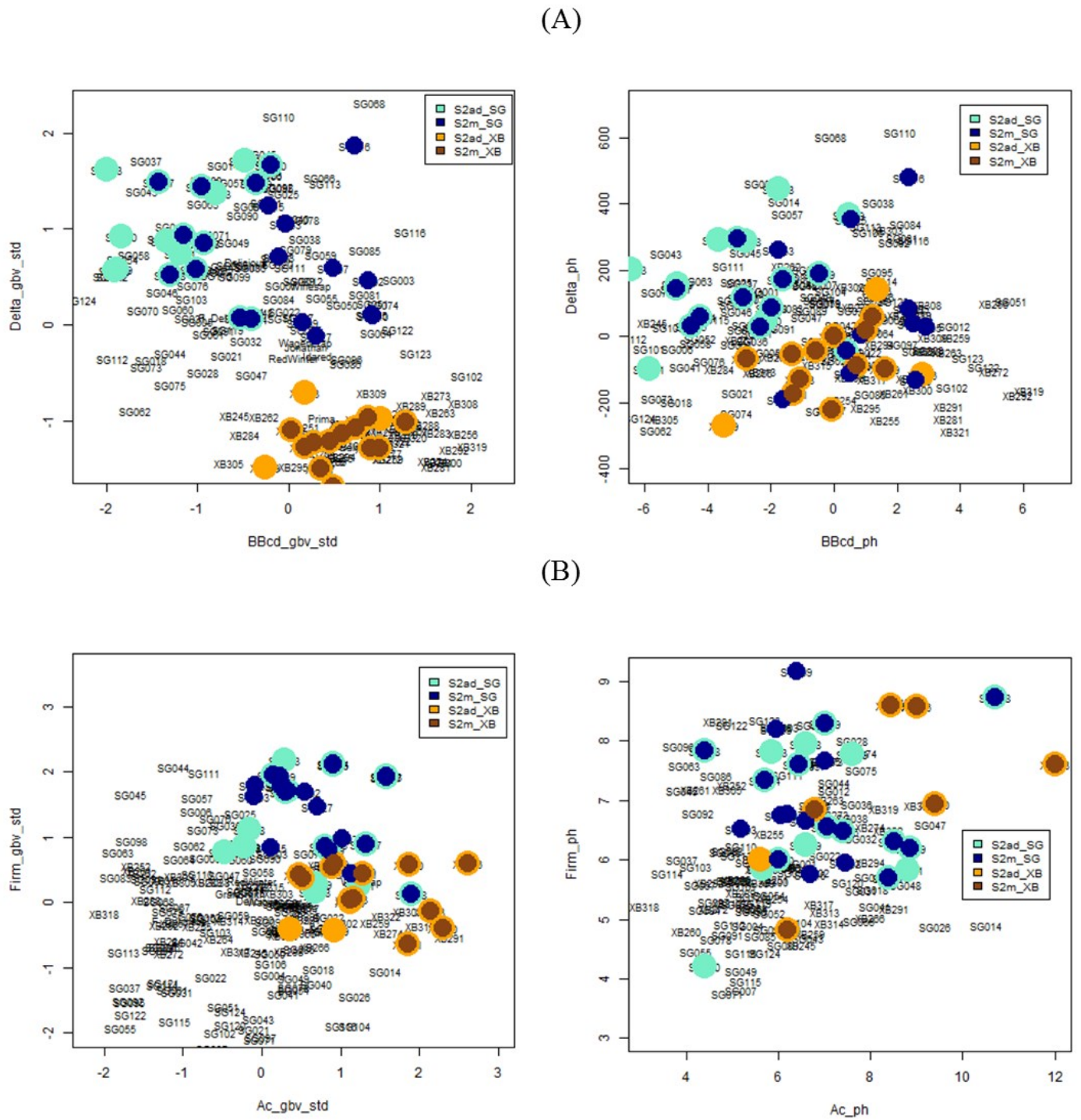


Figure 10. Best individuals repartition through the Distribution of breeding values (left panel) and phenotypes (right panel) for (A) BBcd plotted against Delta, and (B) Firm plotted against Ac. Individuals highlighted in color are individuals with high index value: or SG family, blue and dark blue for additive and multiplicative index, respectively; for XB family, orange and dark orange for additive and multiplicative index, respectively.

efficiency of breeding strategies, especially for tree species for which breeding is a very long process and also for traits with complex phenotypes such as fruit bearing habit, for example. The possibility to estimate individual multi-trait potential at early stages would also be useful for shortening the breeding processes. This study is a proposition for a multi-traits breeding strategy based on ideotype definition, selection index and genome wide breeding value estimation. Such a strategy is reasonably simple to apply to traits with rational breeding goals such as Biennial Bearing Index. It could easily identify the individuals that correspond the most to the breeding goals. However, in other cases, when the breeding goal is less achievable or less clearly defined, this strategy could be used just to eliminate the worst individuals. Finally, the most useful way to use such a strategy would be in the context of genomic selection (Kumar *et al.*, 2012; Muranty *et al.*, 2015) since it would cumulate the advantages of genomic selection, breeding value prediction with phenotyping limited to training populations, and selection indexes, which can be considered as the most efficient method to simultaneously breeding for several characters.

References

- Alimi NA, Bink MCAM, Dieleman JA, Magán JJ, Wubs AM, Palloix A, van Eeuwijk FA.** 2013. Multi-trait and multi-environment QTL analyses of yield and a set of physiological traits in pepper. *Theoretical and Applied Genetics* **126**, 2597-2625.
- Almeida LM, Viana AP, Teixeira do Amaral AJ, Barros Carnero JJ.** 2014. Breeding full-sib families of sugar cane using selection index. *Ciência Rural* **44**, 605-611.
- Ben Sadok I, Tiecher A, Galvez-Lopez D, Bruneau M, Hanteville S, Robic R, Cournol R, Lahaye M, Lasserre-Zuber P, Laurens F.** 2015. Apple fruit texture QTLs: year and cold storage effects on sensory and instrumental traits. *Tree Genetics & Genomes* **11**, 119.
- Bhering LL, Laviola BG, Salgado CC, Sanchez CFB, Rosado TB, Alves AA.** 2012. Genetic gain in physic nut using selection indexes. *Pesquisa Agropecuaria Brasileira* **47**.
- Bink MCAM, Jansen J, Madduri M, et al.** 2014. Bayesian QTL analyses using pedigreed families of an outcrossing species, with application to fruit firmness in apple. *Theoretical and Applied Genetics* **127**, 1073-1090.
- Blum A.** 2015. Towards a conceptual ABA ideotype in plant breeding for water limited environments. *Functional Plant Biology* **42**, 502-513.
- Bonany J, Brugger C, Buehler A, et al.** 2014. Preference mapping of apple varieties in Europe. *Food Quality and Preference* **32, Part C**, 317-329.
- Caliński T, Kaczmarek Z, Krajewski P, Frova C, Sari-Gorla M.** 2000. A multivariate approach to the problem of QTL localization. *Heredity* **84**, 303-310.
- Dekkers JCM.** 2007. Prediction of response to marker-assisted and genomic selection using selection index theory. *Journal of Animal Breeding and Genetics* **124**.
- Desclaux D, Chiffolleau Y, Nolot JM.** 2013. Du concept d'Ideotype à celui de Realttype : gestion dynamique des Innovations Variétales par une approche transdisciplinaire et partenariale. Exemple du blé dur pour l'AB. *Innovations Agronomiques* **32**, 455-466.
- Dickmann DI.** 1985. The ideotype concept applied to forest trees. Pp: 81-101. In: Cannell M. G. R. and Jackson J. E. (eds.) *Attributes of trees as crop plants*. Institute of Terrestrial Ecology, Huntingdon, England, pp.89-101.
- Dickmann DI, Gold MA, Flore JA.** 1994. The Ideotype Concept and the Genetic Improvement of Tree Crops. *Plant Breeding Review* **12**, 163-193.
- Donald CM.** 1968. The breeding of crop ideotypes. *Euphytica* **17**, 385-403.
- Durand J-B, Guitton B, Peyhardi J, Holtz Y, Guédon Y, Trottier C, Costes E.** 2013. New insights for estimating the genetic value of segregating apple progenies for irregular bearing during the first years of tree production. *Journal of Experimental Botany* **64**, 5099-5113.

Elston RC. 1963. A weight free index for the purpose of ranking or selection with respect to several traits at a time. *Biometrics* **19**, 85-97.

Farshadfar E, Rostami-Ahmandvandi H, Shabani A. 2013. Evaluation of drought tolerance in wheat-rye disomic addition lines using agro-physiological indicators and a new integrated selection index (ISI). *Annals of Biological Research* **4**, 70-75.

Gallais A. 1990. *Théorie de la sélection en amélioration des plantes*. Paris, FRA : Masson, 588 p.

Gonzalez-Recio O, Pryce JE, Haile-Mariam M, Hayes BJ. 2014. Incorporating heifer feed efficiency in the Australian selection index using genomic selection. *Journal of Dairy Science* **97**, 3883-3893.

Grattapaglia D, Kirst M. 2008. Eucalyptus applied genomics: from gene sequences to breeding tools. *New Phytologist* **179**, 911-929.

Guan Y, Peace C, Rudell D, Verma S, Evans K. 2015. QTLs detected for individual sugars and soluble solids content in apple. *Molecular Breeding* **35**, 1-13.

Gu J, Yin X, Zhang C, Wang H, Struik PC. 2014. Linking ecophysiological modelling with quantitative genetics to support marker-assisted crop design for improved yields of rice (*Oryza sativa*) under drought stress. *Annals of Botany* **114** (3), 499-511.

Harrell FEJ, Dupont C, and contribution of many others. 2015. Hmisc: Harrell Miscellaneous.

Hathorn A, Chapman SC. 2014. Historical and Prospective Applications of ‘Quantitative Genomics’ in Utilising Germplasm Resources. In: Tuberosa R., In: Graner A, In: Frison E, eds. *Genomics of Plant Genetic Resources*. Springer Netherlands, 93-110.

Hazel LN. 1943. The Genetic Basis for Constructing Selection Index. *Genetics* **28**, 476-490.

Lê S, Josse J & Husson F. 2008. FactoMineR: An R Package for Multivariate Analysis. *Journal of Statistical Software* **25**(1), 1-18.

Jannink J-L, Lorenz AJ, Iwata H. 2010. Genomic selection in plant breeding: from theory to practice. *Briefings in Functional Genomics* **9**, 166-177.

Jiang C, Zeng Z-B. 1995. Multiple Trait Analysis of Genetic Mapping for Quantitative Trait Loci. *Genetics* **140**, 1111-1127.

Knott SA, Haley CS. 2000. Multitrait Least Squares for Quantitative Trait Loci Detection. *Genetics* **156**, 899-911.

Korol AB, Ronin YI, Nevo E, Hayes PM. 1998. Multi-interval mapping of correlated trait complexes. *Heredity* **80**, 273-284.

Korte A, Vilhjalmsón BJ, Segura V, Platt A, Long Q, Nordborg M. 2012. A mixed-model approach for genome-wide association studies of correlated traits in structured populations. *Nature Genetics* **44**, 1066-1074.

Lambert P., Dirlewanger E., Laurens F. 2009. La sélection assistée par marqueurs (SAM) chez les arbres fruitiers: une approche prometteuse au service de l'innovation variétale. *Innovations Agronomiques* **7**, 139-152.

Lande R, Thompson R. 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* **124**, 743-756.

Laurens F, Aranzana MJ, Arús P, et al. 2012. Review on Fruit Genetics and Breeding Programmes and a New European Initiative to Increase Fruit breeding Efficiency. *ISHS Acta Horticulturae* **929**, 95-102.

Lin CY. 1978. Index selection for genetic improvement of quantitative characters. *Theoretical and Applied Genetics* **52**, 49-56.

Lund MS, Sørensen P, Guldbrandtsen B, Sørensen DA. 2003. Multitrait Fine Mapping of Quantitative Trait Loci Using Combined Linkage Disequilibria and Linkage Analysis. *Genetics* **163**, 405-410.

Maliepaard C, Alston FH, Arkel G van, et al. 1998. Aligning male and female linkage maps of apple (*Malus pumila* Mill.) using multi-allelic markers. *Theoretical and Applied Genetics* **97**, 60-73.

Malosetti M, Ribaut JM, Vargas M, Crossa J, van Eeuwijk FA. 2008. A multi-trait multi-environment QTL mixed model with an application to drought and nitrogen stress trials in maize (*Zea mays* L.). *Euphytica* **161**, 241-257.

Martre, P, Quilot-Turion, B, Luquet, D, Memmah, MMOS, Chenu, K, Debaeke, P. 2015. Model-assisted phenotyping and ideotype design. *Crop Physiology* 349-373.

Meuwissen THE, Hayes BJ, Goddard ME. 2001. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics* **157**, 1819-1829.

Patocchi A, Frei A, Frey JE, Kellerhals M. 2009. Towards improvement of marker assisted selection of apple scab resistant cultivars: *Venturia inaequalis* virulence surveys and standardization of molecular marker alleles associated with resistance genes. *Molecular Breeding* **24**, 337-347.

Pfeiffer C, Fuerst-Waltl B, Schwarzenbacher H, Steininger F, Fuerst C. 2015. A comparison of methods to calculate a total merit index using stochastic simulation. *Genetics Selection Evolution* **47** (1), 36-45.

Quilot-Turion B, Ould-Sidi M-M, Kadrani A, Hilgert N, Génard M, Lescourret F. 2012. Optimization of parameters of the 'Virtual Fruit' model to design peach genotype for sustainable production systems. *European Journal of Agronomy* **42**, 34-48.

- Rieseberg LH, Widmer A, Arntz AM, Burke JM.** 2003. The genetic architecture necessary for transgressive segregation is common in both natural and domesticated populations. *Philosophical Transactions of the Royal Society B: Biological Sciences* **358**, 1141-1147.
- Rötter RP, Tao F, Höhn JG, Palosuo T.** 2015. Use of crop simulation modelling to aid ideotype design of future cereal cultivars. *Journal of Experimental Botany* **66**, 3463-3476.
- Silva DD, Han L, Faivre R, Costes E.** 2014. Influence of the variation of geometrical and topological traits on light interception efficiency of apple trees: sensitivity analysis and metamodelling for ideotype definition. *Annals of Botany* **114**, 739-752.
- Silva LDCE, Wang S, Zeng Z-B.** 2012. Multiple trait multiple interval mapping of quantitative trait loci from inbred line crosses. *BMC Genetics* **13**, 67.
- Smith OS, Hallauer AR, Russell WA.** 1981. Use of index selection in recurrent selection programs in maize. *Euphytica* **30**, 611-618.
- Sukhwinder-Singh, Hernandez MV, Crossa J, Singh PK, Bains NS, Singh K, Sharma I.** 2012. Multi-Trait and Multi-Environment QTL Analyses for Resistance to Wheat Diseases. *PLoS ONE* **7**, e38008.
- Swan AA, Brown DJ.** 2013. The impact of measuring adult fleece traits with genomic selection on economic gain in Merino selection indexes. *Proc. Assoc. Advmt. Anim. Breed. Genet.* 233-236.
- Thavamanikumar S, Southerton SG, Bossinger G, Thumma BR.** 2013. Dissection of complex traits in forest trees - opportunities for marker-assisted selection. *Tree Genetics & Genomes* **9**, 627-639.
- Troggio M, Gleave A, Salvi S, Chagné D, Cestaro A, Kumar S, Crowhurst RN, Gardiner SE.** 2012. Apple, from genome to breeding. *Tree Genetics & Genomes* **8**, 509-529.
- Van Oijen M, Höglind M.** 2016. Toward a Bayesian procedure for using process-based models in plant breeding, with application to ideotype design. *Euphytica* **207**, 627-643.
- Visser T, Schaap AA, Vries DPD.** 1968. Acidity and sweetness in apple and pear. *Euphytica* **17**, 153-167.
- Xu K, Wang A, Brown S.** 2012. Genetic characterization of the Ma locus with pH and titratable acidity in apple. *Molecular Breeding* **30**, 899-912.

Discussion Générale

Phénotypes et phénotypage

Une grande partie du travail de cette thèse a porté sur l'étude de deux caractères d'intérêt: la phénologie du débourrement et de la floraison et l'alternance de production. Ces deux caractères ont la particularité d'être complexes à phénotyper car ils nécessitent une évaluation à l'échelle de plusieurs années. Dans ce qui suit, je vais faire un bilan des phénotypages réalisés, de leur limite et des perspectives qui feront suite à ce travail.

Le phénotypage de la phénologie a consisté à noter les dates de débourrement et floraison trois années consécutives, pour les deux sites géographiques et 466 individus. Comme je l'ai déjà souligné précédemment, les perturbations de la floraison dus aux changements climatiques (Legave *et al.*, 2013; Atkinson *et al.*, 2013; Darbyshire *et al.*, 2014) sont liés aux besoins intrinsèques en froid et en chaleur des géotypes (Egea *et al.*, 2003; Ruiz *et al.*, 2007; Albuquerque *et al.*, 2008). Or, les dates de débourrement exprimées en jours calendaires ne suffisent pas à approcher les besoins en froid et en chaleur. Par conséquent, j'ai proposé le calcul de deux variables pour tenter d'approcher les besoins en chaleur entre deux stades consécutifs. L'accumulation de chaleur a été calculée en Growing Degree Hours (GDH) entre la levée de dormance, estimée chez 'Golden Delicious', et le stade de débourrement. De même, les GDH ont été calculés entre le débourrement et le début de la floraison. Cependant, dans cette approche la date de levée de dormance a été considérée constante, puisque estimée pour la variété 'Golden Delicious' pour chaque année et chaque site. Ceci constitue une limite importante de ce travail car il est fortement probable que cette date de levée de dormance soit variable entre géotypes ce qui n'a pas pu être pris en compte.

Pour estimer la levée de dormance chez un géotype trois méthodes sont envisageables. Les deux premières méthodes reposent sur des tests de forçage par exposition de bourgeons ou de rameaux à des températures chaudes (Andreini *et al.*, 2014). Tout d'abord, le test de Tabuenca (Tabuenca, 1964) consiste à prélever en fin d'hiver et à intervalles réguliers des bourgeons floraux sur des arbres implantés en verger, et de suivre l'évolution de leur poids afin d'en déduire une date de reprise d'activité et donc de levée de dormance. Cette méthode longtemps utilisée sur *Prunus* (Legave *et al.*, 2008 ; Andreini *et al.*, 2014) a aussi été testée avec succès sur pommier (Malagi *et al.*, 2015). La deuxième méthode nécessite des prélèvements de rameaux en verger de façon similaire à ceux du test de Tabuenca, mais le suivi de l'état des bourgeons se fait

sur rameaux entiers ou sur boutures (Gibson and Reighard, 2002). L'intensité et le délai moyen de débourrement des bourgeons dans ces conditions de forçage permettent de déduire l'état de satisfaction des besoins en froid et donc la date de levée de dormance (Mauget *et al.*, 1983 ; Champagnat, 1989). Dans une étude sur pêcher, l'identification des besoins en froid et en chaleur de deux populations en ségrégation a été réalisé en effectuant des prélèvements de rameaux en champs à une dizaine de dates successives (Fan *et al.*, 2010).

Ces deux méthodes de phénotypage des besoins en froid par forçage auraient donc pu être envisagées dans ce travail, mais cela n'a pas été le cas pour deux raisons principales. Tout d'abord, les arbres implantés servaient aussi à l'étude d'autres caractères et le prélèvement régulier de rameaux aurait été un traitement trop destructif pour permettre ces phénotypages. La volonté de phénotyper les mêmes arbres pour plusieurs caractères sur des dispositifs d'arbre contenant que peu de réplifications par génotypes est donc peu compatible avec des méthodes destructives. De plus, la mise en place des tests de forçage aurait requis la mise en place de lourds moyens logistiques, et humains dans deux sites et sur un grand nombre de génotypes qui ne pouvait pas être assumées par les équipes encadrantes.

La troisième méthode qui aurait été envisageable est, elle, non destructive car elle s'appuie sur des approches de modélisation existantes (Legave *et al.*, 2009). Actuellement ces modèles sont utilisés pour prédire les dates de floraison de variétés à partir de longues séries d'observation des dates de floraison, des températures observées au cours des mois recouvrant les différentes phases de la dormance, et des besoins en froid et en chaleur connus pour les variétés étudiées (Legave *et al.*, 2008). Il est théoriquement possible d'inverser la prédiction, c'est-à-dire de déterminer les besoins en froid et en chaleur à partir des dates de floraison. Au cours de la thèse, un stage de Master a eu pour objectif d'estimer les besoins en froid et en chaleur des génotypes de la descendance SG à partir de 5 années d'observations (Reig, 2014). En procédant à des échantillonnages de nombre d'années variables dans les séries observées, ce travail a montré une instabilité des prédictions des besoins en froid et en chaleur lorsque le nombre d'années d'observation n'est pas suffisant. Il faut au minimum une dizaine d'années de notation des dates de floraison pour estimer précisément ces besoins en froid et en chaleur. Malgré ces difficultés, cette dernière méthode est probablement la perspective la plus envisageable à l'avenir car de nombreuses années de notation ont déjà été réalisées sur les deux descendance implantées à Montpellier (8 années pour SG et 5 années pour XB).

Discussion Générale

Dans ce travail et dans le but de simplifier le phénotypage, l'alternance de production a été phénotypée rétrospectivement par notations des événements végétatifs et de floraison le long des branches et des rameaux. Ce phénotypage est rendu possible par l'observation des bourses, qui marquent les événements de floraison qui ont lieu lors du développement de l'arbre. Ces bourses sont une particularité des espèces de la sous-famille des *Maloideae*. Néanmoins, la réalisation pratique de ce phénotypage repose sur une expertise avancée de la morphologie du pommier. Dans certains cas la reconnaissance des arrêts de croissance inter et intra-annuels peut s'avérer délicate. Des indices mesurant la variabilité de floraison entre années, la synchronisation des événements de floraison, et la probabilité de floraison d'une pousse annuelle en fonction des événements des années précédentes ont été extraits de ces notations. Des prédictions du comportement global de l'arbre ont été réalisées à partir des descripteurs calculés. L'échantillonnage d'une dizaine d'axes répartis judicieusement dans l'arbre a suffi à estimer le comportement global, avec succès pour les géotypes réguliers et alternants mais avec toutefois une ambiguïté pour les géotypes irréguliers. Une approche basée sur l'existence de corrélations entre paramètres architecturaux et alternance a été envisagée pour prédire de manière précoce le comportement de variétés commerciales de pommier (Lauri *et al.*, 2014). Néanmoins, la cartographie de QTL sur la population 'Starkrimson' x 'Granny Smith' n'a pas permis de révéler de colocalisation entre traits d'architecture et trait d'alternance (voir Chapitre 2, Guittou, communication personnelle). Le phénotypage par imagerie (Aggelopoulou *et al.*, 2010 ; Gongal *et al.*, 2016) est une perspective intéressante d'automatisation et de simplification du phénotypage de l'alternance de production. Il nécessiterait néanmoins le passage dans les vergers pendant plusieurs années consécutives, et à ce titre, la notation rétrospective en un seul passage reste intéressante. La méthode par imagerie aurait cependant l'avantage d'être transférable à d'autres espèces qui n'ont pas de marqueurs morphologiques tels que la bourse, notamment chez les *Prunus* (Costes *et al.*, 2014).

Enfin, la qualité du fruit a été phénotypée sur une plateforme de contrôle de la qualité utilisée dans les vergers commerciaux (Pimprelle, www.setop.fr). L'automatisation partielle du phénotypage bien que coûteuse présente d'avantage de mesurer rapidement plusieurs caractères. Ces mesures sont identiques avec celles pratiquées en sélection: pénétrométrie, réfractomètre, acide titrable... De même des tests sensoriels, qui n'ont pas été considérés dans cette étude sont réalisés dans les vergers par les professionnels, avant commercialisation et en sélection.

Certaines différences sont à noter concernant la texture ou la fermeté lorsque des études plus approfondies sont réalisées notamment pour la détection de QTLs. La pénétrométrie est alors réalisée avec un appareil automatique qui prend en compte la cinétique et mesure donc en plus de la fermeté d'autres caractères comme l'élasticité. Mais la mesure de fermeté issue de cette cinétique est corrélée avec la fermeté « commerciale ». Finalement, le groupe de traits mesurés dans cette étude représente une partie des caractères pris en compte en sélection, mais pas leur totalité et sous-estime donc la complexité de l'évaluation de la qualité des fruits. Cependant l'intention était de pouvoir étudier les combinaisons possibles de caractères dans l'exercice d'évaluation des valeurs génétiques multi-traits, en associant les traits de régularité et adaptatifs à quelques uns plus essentiels liés à la qualité des fruits.

Modélisation des phénotypes : prise en compte des effets années et sites, modélisation multi-trait.

Les caractères étudiés dans ce travail sont de nature très différente, notamment par le fait que la phénologie peut être phénotypée chaque année alors que l'alternance est évaluée sur plusieurs années de floraison conjointement. Par conséquent leur modélisation a aussi été différente. En effet, les traits liés à la phénologie (Chapitre 1) sont mesurables à chaque printemps, les années ont donc été considérées comme des répétitions, dans le but de distinguer les effets des génotypes et les interactions 'génotype' x 'année' des effets des années et sites. Des modèles mixtes comprenant les années et les sites comme effets fixes et les génotypes et interaction 'génotype' x 'année' comme effets aléatoires ont donc été utilisés. De plus, des structures de corrélation ont été ajustées pour modéliser l'hétérogénéité des variances entre années. Par contre, pour l'alternance, les événements de floraison ont été modélisés par analyse de séquences recouvrant plusieurs années consécutives, afin d'estimer les différents descripteurs nécessaires à la caractérisation du comportement des arbres. Ces descripteurs sont de natures différentes : les BBI et les entropies sont des index, γ_g est un paramètre d'autocorrélation, les $\mu_{g,t}$ et les $\theta_{g,m}$ sont des BLUPs, les BBI_res_norm_pred et les γ^{pred} sont des variables prédites. Ce contexte de variables de diverses natures a conduit à la décision de traiter séparément la modélisation des phénotypes et la détection de QTLs.

Dans ce travail, chaque fois que possible l'estimation des interactions Génotype x Environnement a été réalisée dans l'objectif d'identifier des comportements différents selon le

contexte pédo-climatique et ainsi d'appréhender la plasticité des phénotypes considérés (Ben Sadok *et al.*, 2015). De plus, il a été montré que les études multi-environnement sont plus robustes et performantes pour détecter les QTLs que des études séparées par environnement (Alimi *et al.*, 2013). Cependant, les mêmes populations n'étant pas implantées dans les deux sites, l'approche conduite n'a été que partielle. En fait, la famille 'Starkrimson' x Granny Smith' implantée à Montpellier, vient d'être implantée à Angers et à l'inverse la population HIVW a été implantée à Montpellier. Ce nouveau contexte expérimental devrait permettre d'aborder, dans les années à venir, des études multi-environnement et multi-traits. Cependant, les dispositifs ayant été implantés différentes années, des effets âge de l'arbre pourront être considérés mais peut-être pas des effets « année climatique », tout au moins en attendant que les arbres sur le site planté en dernier aient atteint un stade adulte.

Des approches de modélisation et de détection de QTLs multi-trait ont été proposées (Jiang and Zeng, 1995; Knott and Haley, 2000; Lund *et al.*, 2003; Sukhwinder-Singh *et al.*, 2012; Silva *et al.*, 2012). Ces approches sont intéressantes car elles permettent une plus grande précision de détection des QTLs, pour des caractères faiblement héritables ou encore d'identifier des phénomènes de pléiotropie. Néanmoins, cet aspect de modélisation n'a pas été réalisé dans cette thèse car l'aspect multi-caractère a été envisagé pour l'intégration de caractères de nature différente pas forcément corrélés. Plusieurs analyses multi-traits et multi-environnement (Malosetti *et al.*, 2008; Sukhwinder-Singh *et al.*, 2012; Alimi *et al.*, 2013) ont montré que cette approche confère une plus grande puissance de détection de QTLs, en particulier pour les caractères à déterminisme génétique complexe, influencés par de nombreux loci à petits effets. Cette approche pourrait s'appliquer aux caractères abordés dans ce travail à condition d'avoir le dispositif multi-environnement adéquat.

Elargissement de diversité allélique de la population d'étude

Le dispositif expérimental multi-famille a été l'un des points originaux de ce travail et une avancée par rapport aux travaux antérieurs réalisés sur l'alternance et la phénologie. Cinq familles en ségrégation apparentées par leur pedigree ont été utilisées pour la cartographie de QTLs. Certaines familles avaient un parent en commun, d'autres étaient reliées par des ancêtres plus lointains. Au total, sept parents différents ont été utilisés ainsi que 13 individus fondateurs. Parmi les variétés ou hybrides composant le pedigree, quatre font partie des cinq variétés les plus

utilisées en sélection, ‘Golden Delicious’, ‘Red Delicious’, ‘Jonathan’, ‘MacIntosh’ et ‘Cox’s Orange Pippin’ (Noiton and Alspach, 1996). Les parents des populations implantées à Montpellier avaient été initialement choisis pour leur contraste au niveau de leur architecture, et de la qualité de fruit. Les parents des populations implantées à Angers avaient été choisis sur des critères de qualité du fruit. Les cinq populations ont révélé de larges ségrégation sur d’autres caractères, notamment ceux traités dans ce travail, qui s’expliquent en raison de la forte hétérozygotie du pommier.

Grâce à une plus grande diversité allélique et un plus grand nombre d’individus (466) que dans des populations bi-parentales, ce dispositif a permis de détecter plus de QTLs, certains déjà connus et d’autres nouveaux pour les deux principaux caractères étudiés : phénologie du débourrement et alternance de production. De plus, certains gènes candidats pertinents ont pu être localisés dans les zones QTLs, notamment pour la phénologie. Néanmoins, ce dispositif ne représente qu’une partie de la diversité existante chez *Malus x domestica Borkh* car il existe des centaines de variétés dans chacun des plus grands pays producteurs de pommes et de nombreuses espèces sauvages apparentées, *Malus siversii*, *Malus sylvestri* (Cornille *et al.*, 2014). Les loci impliqués dans la variation des différents traits étudiés ne sont donc sans doute pas exhaustivement représentés par les QTLs détectés. Dans le cadre du projet FruitBreedomics un dispositif multi-familles a été développé (Figure 19 de la synthèse bibliographique), il comporte 25 familles reliées par des parents et des ancêtres communs, le pedigree comporte de nombreuses variétés commerciales européennes, et il est implanté dans sept pays européens. Dans ce type de dispositif, la diversité allélique est large grâce au grand nombre de parents utilisés. Les événements de recombinaison sont aussi nombreux grâce au développement de populations en ségrégation. Tous ces critères sont favorables à l’étude des déterminismes génétiques de caractères complexes. Les descendances composant ce dispositif comportent environ 50 individus, alors que le dispositif expérimental de cette étude comporte des familles avec plus de 100 individus notamment SG et HIVW. Les dispositifs multi-descendance chez le pommier ont plusieurs points communs avec d’autres dispositifs multi-parentaux développés chez les espèces annuelles. La récurrence de certains parents, notamment ‘Golden Delicious’, rappelle la présence d’un ou plusieurs parents récurrent dans les populations NAM (Yu *et al.*, 2008), le grand nombre de fondateurs rappelle les populations multi-parentales MAGIC (Cavanagh *et al.*, 2008). Le dispositif FruitBreedomics permet donc une détection plus exhaustive des QTLs responsables de variation de caractères quantitatifs grâce à sa grande diversité allélique. Le dispositif de cette

étude peut permettre la cartographie plus fine de QTLs grâce au grand nombre d'individus par famille. Le dispositif FruitBreedomics avait été mis en place dans le cadre d'un précédent projet Européen (HiDRAS) pour étudier la qualité du fruit (Kouassi *et al.*, 2009; Bink *et al.*, 2014), et il serait intéressant à l'avenir de diversifier ses cibles phénotypiques. Néanmoins, le phénotypage de ce type de populations n'est envisageable que dans le cadre de collaboration entre plusieurs équipes de recherche car il requiert des moyens humains, logistiques, et financiers conséquents. La difficulté consiste cependant à homogénéiser les protocoles d'observation entre partenaires. Les core-collections sont d'autres types de population qui contiennent une diversité allélique maximale et optimisée pour une espèce donnée (Lassois *et al.*, 2016). A l'avenir, il serait intéressant de multiplier des approches de détection de QTLs complémentaires, avec des dispositifs multi-parentaux pour maximiser le nombre de QTL détectés, ainsi que des approches de génétique d'association avec des core-collections, pour avoir une cartographie plus fine des QTLs. L'implantation de la core-collection définie à Angers par Lassois *et al.* (2016) sur le site de Montpellier depuis 2014 ouvre cette perspective.

Densité de la carte génétique et définition d'haplotypes

Le contexte multi-parental de ce travail a requis d'utiliser une carte génétique consensus établie sur le génotypage d'une vingtaine de populations en ségrégation (Van de Weg *et al.*, non publié). Cette carte comporte 7500 SNP issus d'une puce à 20 000 SNP. Au cours de la thèse, j'ai réalisé la carte génétique avec les marqueurs de la puce à 20 000 SNP pour la famille SG, Les populations XB et HIWH ont été réalisées par une stagiaire de master (Maillard, 2015) et par I. Ben Sadok (Sadok *et al.*, 2015), respectivement. La comparaison des cartes mono-population avec la carte consensus a révélé que la grande majorité des marqueurs étaient cartographiés dans le même ordre dans toutes les cartes, mais que les distances génétiques pouvaient varier en fonction des cartes.

Suite à une erreur, pour la population XB composée d'environ 300 individus, seuls 58 ont été génotypés avec la puce 20 000 SNPs (Bianco *et al.*, 2014), le reste de la population a été génotypé avec la puce 8000 SNP (Chagné *et al.*, 2012).

Pour densifier le génotypage des individus restants, j'ai cherché à densifier le génotypage en imputant les génotypes manquants à partir des marqueurs communs entre les deux puces. Mais la quantité trop faible de marqueurs informatifs communs aux deux puces a empêché une

imputation fiable. Ce fut une grosse perte pour le dispositif expérimental car XB était la famille comportant le plus d'individus et elle était apparentée à une autre grande famille présente dans le dispositif, HIVW, par le parent X-3263. L'utilisation de la totalité de la population XB aurait peut-être pu permettre une plus grande précision de la cartographie de QTL par la présence de nombreuses recombinaisons pour le parent X-3263. Un nouvel effort de génotypage sera nécessaire pour cette population.

La définition d'une carte génétique consensus requiert l'utilisation de marqueurs moléculaires informatifs pour une majorité de populations, d'où la réduction drastique du nombre de marqueurs entre la puce et la carte finale. Pour la cartographie de QTLs, une grande densité de marquage n'est pas nécessaire, car la précision de la position des QTLs détectés est limitée par le nombre de recombinaisons, qui correspond au nombre d'individus utilisés. Pour ce travail, l'utilisation d'une carte génétique à 7500 markers aurait peut-être été démesurée comparée à la taille de la population étudiée (466), néanmoins une carte basée sur des haplotypes a été utilisée. A partir des SNP, des haplotypes ont été estimés afin d'avoir des marqueurs multi-alléliques plus informatifs. Un haplotype est composé des SNP présents dans un intervalle de 1cM, et les allèles de l'haplotype sont les différentes combinaisons alléliques des SNP concernés (Voorrips *et al.*, 2016). 926 haplotypes ont été estimés de cette manière et ont servi à la cartographie de QTLs. Ceci a permis d'améliorer la détection des événements de recombinaison et donc des QTLs, même si le nombre de recombinaisons reste limitant. De plus, l'utilisation d'haplotypes a permis de retracer les pedigrees jusqu'aux fondateurs.

Détection de QTLs avec FlexQTL™

Suite à la décision de modéliser les phénotypes en amont de la détection de QTLs, le logiciel ASReml (Butler *et al.*, 2009) a été utilisé pour ajuster les modèles mixtes pour la phénologie. Ce logiciel aurait pu être aussi utilisé pour détecter des QTLs en insérant des effets pour chaque marqueur dans les modèles. Cependant, l'estimation de nombreux effets pour un nombre de données observées limité engendre des problèmes méthodologiques qui peuvent être traités par les statistiques Bayésiennes (Gianola *et al.*, 2009). Des modèles de détection de QTLs bayésiens sont méthodologiquement possibles avec ASReml, mais ce choix méthodologique n'a pas été retenu car il est apparu trop complexe et risqué compte tenu du haut niveau de compétence en modélisation et statistique qu'il aurait requis. Le choix qui a été fait pour répondre à ces besoins

a été d'utiliser le logiciel Bayésien FlexQTL (Bink *et al.*, 2002, 2008, 2012, 2014; Rosyara *et al.*, 2013) dans un cadre collaboratif européen, le projet FruitBreedomics (Laurens *et al.*, 2010).

Dans ce travail, toutes les détections de QTL ont été réalisées avec ce logiciel car il répondait aux besoins de l'étude sur plusieurs points. En effet, il a été conçu pour fonctionner sur des populations out-cross multi-parentales telles que celles utilisées dans ce travail. Il permet aussi de prendre en compte les apparentements entre individus selon le principe de l'Identity By Descent (IBD) qui calcule la probabilité que deux allèles de deux individus à un même locus soient hérités d'un ancêtre commun (Thompson, 2013). C'est une méthode précise et fiable d'estimation des apparentements. Ensuite, FlexQTLTM fonctionne dans un environnement Bayésien qui lui confère la capacité à estimer simultanément les effets additifs de nombreux marqueurs, ainsi qu'un grand nombre de paramètres comme le nombre, la position et la significativité des QTLs, les fréquences et effets alléliques des loci testés. Les sorties du logiciel sont très complètes et tous les paramètres calculés sont disponibles sous format .csv. En contrepartie, la quantité d'information disponible peut-être difficile à gérer. Enfin, plusieurs paramètres doivent être pris en compte pour déterminer la présence d'un QTL, valeur du Bayes Factor à l'échelle du groupe de liaison, probabilité de présence du QTL, stabilité de la détection des QTLs au fil des itérations. Ceci rend la détermination de la présence d'un QTL un peu complexe parfois subjective.

De plus, tous les paramètres estimés dans FlexQTL permettent à la fois d'identifier les loci correspondant à des QTLs et de calculer des 'breeding values' grâce aux effets additifs, aux fréquences alléliques et aux génotypes estimés à chaque loci. FlexQTL permet de prendre en compte des effets fixes environnementaux ou expérimentaux ce qui lui confère un avantage par rapport à d'autres logiciels comme MapQTL (Van Ooijen, 2004). C'est aussi un inconvénient car il n'est pas possible de modéliser finement un phénotype en estimant des effets d'interaction ou en ajustant des structures de corrélation de variance (d'où notre choix de procéder à la modélisation des phénotypes en amont de la détection des QTLs. Par ailleurs, il faut noter que FlexQTL ne considère que deux allèles par locus ce qui peut être considéré comme une limitation compte tenu du fort polymorphisme du pommier.

Estimation de la valeur génétique des individus

Les paramètres estimés par FlexQTLTM ont permis le calcul de deux types de valeurs

génétiques : une prenant en compte tous les loci, GWBV, l'autre prenant en compte uniquement les loci ayant un effet significatif sur la variabilité du caractère, QTLBV. La comparaison des deux types de valeur génétique a confirmé que la GWBV est plus adaptée aux caractères quantitatifs car elle permet de prendre en compte tous les loci à faible effet, qui ne sont pas détectés comme QTLs (Francia *et al.*, 2005). Pour les caractères régulés par un petit nombre de loci, GWBV et QTLBV sont équivalentes, car les QTLs permettent d'expliquer une grande partie de la variabilité génétique. Plusieurs études ont mentionné la limite des QTLs dans l'explication de la variance des caractères très quantitatifs, pour lesquels de nombreux loci à faibles effets qui, cumulés, expliquent une grande partie de la variance phénotypique (Francia *et al.*, 2005; Kumar *et al.*, 2012a). C'est peut-être l'une des raisons pour laquelle la sélection assistée par marqueur a été principalement réalisée sur des caractères simples, notamment la résistance à la tavelure chez le pommier (Kellerhals *et al.*, 2016) ou pour la chair rouge (Chagné *et al.*, 2013).

La méthode d'estimation de valeur génétique sur la base des effets de marqueurs répartis sur le génome, initialement proposée par Meuwissen *et al.* (2001), représentée par GWBV dans ce travail, est une alternative intéressante d'estimation de valeur génétique pour des caractères complexes, car aucun a priori n'est fixé quant à l'architecture génétique du caractère considéré et il n'est donc pas nécessaire de sélectionner les marqueurs suivant un seuil de significativité.

Analyses Multi-traits

Dans ce travail, l'aspect multi-trait a été abordé par la définition d'un idéotype variétal. Pour chacun des caractères étudiés dans la thèse, des objectifs de sélection ont été décidés. Certains choix arbitraires ont dû être faits, comme le choix d'individus adaptés à des conditions climatiques de low chilling ou celui d'individus à forte acidité. D'autres choix auraient pu être faits. En effet, en sélection, il n'existe pas un unique idéotype, en particulier sur les composantes de la qualité du fruit qui reste un caractère très subjectif. En revanche, d'autres objectifs de sélection sont plus rationnels : un arbre « idéal » produirait beaucoup et régulièrement, serait facile à conduire et résistant aux bio-agresseurs. L'originalité de l'idéotype choisi dans ce travail réside dans l'intégration de deux caractères originaux : la phénologie et l'alternance de production. Ces deux caractères ont été combinés à des composants de qualité et ceci a permis d'identifier des synergies et antagonismes entre caractères qui seraient à considérer dans la stratégie de sélection.

Discussion Générale

Les caractères ont aussi été répartis en deux groupes suivant leur importance : d'un côté le rendement, sa régularité, la taille du fruit et le taux de sucre ; et de l'autre la phénologie, l'acidité et la fermeté. L'intérêt de grouper les caractères est de hiérarchiser leur importance. Là aussi, les choix de hiérarchisation ont été en partie arbitraires. En effet l'acidité et la fermeté sont des caractères aussi très importants et auraient pu être inclus dans le premier groupe. L'identification de synergies et d'antagonismes ainsi que la hiérarchisation des caractères sont d'importants critères pour définir les étapes de sélection.

Pour chaque groupe un index de sélection a été construit et pour chaque individu une valeur multi-caractères a été calculée et les meilleurs individus selon l'idéotype ont été identifiés. Le programme de sélection développé en partenariat entre l'INRA et Novadi utilise en routine des index aux différentes étapes du cycle de sélection mais ils sont, pour le moment, basés uniquement sur des données phénotypiques. L'originalité de ce travail de thèse est d'avoir proposé des index basés sur les valeurs génétiques estimées à partir des marqueurs, les GWBV.

Ce travail a porté sur l'estimation de la valeur génétique des individus à partir des marqueurs moléculaires. Il serait envisageable d'entrer dans une démarche de prédiction des valeurs génétiques sur la base des marqueurs, suivant une approche de sélection génomique. La sélection génomique est pertinente pour les caractères complexes à phénotyper car elle permet de réduire le besoin de phénotypage à la population d'entraînement. Chez le pommier, deux études ont abordé cette thématique. La première porte sur plusieurs traits de qualité du fruit (Kumar *et al.*, 2012), et la seconde sur des 'killer' traits relatifs à la productivité des arbres, chute et rendement, et des critères de qualité des fruits notamment liés à l'apparence des fruits (Muranty *et al.*, 2015). Ces deux études se différencient par les relations d'apparentement considérées entre population d'entraînement et population de prédiction. Dans un cas, les deux populations sont des sous-groupes des mêmes populations en ségrégation (Kumar *et al.*, 2012) alors que dans l'autre cas la population de prédiction contient les descendants des individus composant la population d'entraînement (Muranty *et al.*, 2015). Cela illustre bien l'un des enjeux majeurs de la sélection génomique qui est le choix des populations d'entraînement et de leurs apparentements avec les populations de prédiction. Dans ce cadre, un projet français, Fruit SelGen, vise à appliquer la sélection génomique sur deux populations de grandes taille, XB et HIVW, et à organiser les populations d'entraînement et de prédiction de manière similaire à Kumar *et al.* (2012b). Il est aussi envisagé d'utiliser le dispositif FruitBreedomics dans la lignée du travail de Muranty *et al.*,

(2015).

Conclusion

Au final, un travail considérable de phénotypage a été réalisé au cours de cette thèse qui constitue une base de données précieuse pour les années à venir. Il a reposé sur la volonté d'agréger de nombreux traits d'intérêt sur les mêmes populations. A ce titre, les populations étudiées ont fait et font encore l'objet de phénotypage pour de nombreux caractères (Virlet *et al.*, 2015) (Sadok *et al.*, 2015). L'approche initiée par ce travail pourra donc se poursuivre et s'étendre à d'autres caractères à l'avenir. Cela requerra sans doute de poursuivre aussi les efforts de modélisation des phénotypes tels que réalisés dans ce travail, mais aussi dans des contextes plus intégrés. Notamment, l'utilisation de modèles structure-fonction paraît une voie prometteuse pour permettre d'évaluer des traits plus complexes, non directement mesurables, notamment des caractères intégrés à l'échelle de l'arbre entier comme l'interception de la lumière par l'ensemble du feuillage. Cette démarche permettrait également d'accompagner la définition des idéotypes par des expérimentations *in silico* (par simulations), en permettant d'alléger le travail expérimental et d'explorer des combinaisons complexes de caractères.

Au plan génétique, ce travail a démontré l'intérêt d'un élargissement des populations étudiées, initié par les projets HiDRAS et FruitBreedomics, ainsi que celui des sites d'expérimentations. La poursuite de cette démarche est actuellement initiée par la définition de core-collections et par la multiplication de populations partagées dans des environnements climatiquement variés, à l'échelle nationale et internationale. Enfin, cette étude a abouti à la proposition de combiner la définition d'idéotypes multi-caractères et le calcul d'index de sélection avec des approches innovantes de sélection génomique pour aborder l'intégration de plusieurs cibles de sélection, de nature très différentes. L'avenir dira si cette proposition trouvera application plus concrète dans les schémas de sélection des plantes pérennes.

Références

- Aggelopoulou AD, Bochtis D, Fountas S, Swain KC, Gemtos TA, Nanos GD.** 2010. Yield prediction in apple orchards based on image processing. *Precision Agriculture* **12**, 448-456.
- Alburquerque N, García-Montiel F, Carrillo A, Burgos L.** 2008. Chilling and heat requirements of sweet cherry cultivars and the relationship between altitude and the probability of satisfying the chill requirements. *Environmental and Experimental Botany* **64**, 162-170.
- Alimi NA, Bink MCAM, Dieleman JA, Magán JJ, Wubs AM, Palloix A, van Eeuwijk FA.** 2013. Multi-trait and multi-environment QTL analyses of yield and a set of physiological traits in pepper. *Theoretical and Applied Genetics* **126**, 2597-2625.
- Andreini L, de Cortázar-Atauri IG, Chuine I, Viti R, Bartolini S, Ruiz D, Campoy JA, Legave JM, Audergon J-M, Bertuzzi P.** 2014. Understanding dormancy release in apricot flower buds (*Prunus armeniaca* L.) using several process-based phenological models. *Agricultural and Forest Meteorology* **184**, 210-219.
- Atkinson CJ, Brennan RM, Jones HG.** 2013. Declining chilling and its impact on temperate perennial crops. *Environmental and Experimental Botany* **91**, 48-62.
- Ben Sadok I, Martinez S, Moutier N, Garcia G, Leon L, Belaj A, De La Rosa R, Khadari B, Costes E.** 2015. Plasticity in Vegetative Growth over Contrasted Growing Sites of an F1 Olive Tree Progeny during Its Juvenile Phase. *PLoS ONE* **10**, e0127539.
- Bianco L, Cestaro A, Sargent DJ, et al.** 2014. Development and Validation of a 20K Single Nucleotide Polymorphism (SNP) Whole Genome Genotyping Array for Apple (*Malus × domestica* Borkh). *PLoS ONE* **9**, e110377.
- Bink MCAM, Boer MP, Braak CJF, Jansen J, Voorrips RE, Van de Weg E.** 2008. Bayesian analysis of complex traits in pedigreed plant population. *Euphytica* **161**, 85-96.
- Bink MCAM, Jansen J, Madduri M, et al.** 2014. Bayesian QTL analyses using pedigreed families of an outcrossing species, with application to fruit firmness in apple. *Theoretical and Applied Genetics* **127**, 1073-1090.
- Bink MCAM, Radu Totir L, ter Braak CJF, Winkler CR, Boer MP, Smith OS.** 2012. QTL linkage analysis of connected populations using ancestral marker and pedigree information. *Theoretical and Applied Genetics* **124**, 1097-1113.
- Bink MCAM, Uimari P, Sillanpää MJ, Janss LLG, Jansen RC.** 2002. Multiple QTL mapping in related plant populations via pedigree-analysis approach. *Theoretical and Applied Genetics* **104**, 751-762.
- Butler DG, Cullis BR, Gilmour AR, Gogel BJ.** 2009. ASReml-R reference manual.
- Cavanagh C, Morell M, Mackay I, Powell W.** 2008. From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. *Current Opinion in Plant Biology* **11**, 215-221.

- Chagné D, Crowhurst RN, Troggio M, *et al.*** 2012. Genome-Wide SNP Detection, Validation, and Development of an 8K SNP Array for Apple. *PLoS ONE* **7**, e31745.
- Chagné D, Lin-Wang K, Espley RV, *et al.*** 2013. An Ancient Duplication of Apple MYB Transcription Factors Is Responsible for Novel Red Fruit-Flesh Phenotypes. *Plant Physiology* **161**, 225-239.
- Champagnat P.** 1989. Rest and activity in vegetative buds of trees. *Annales des sciences forestières*. EDP Sciences, 9s-26s.
- Cornille A, Giraud T, Smulders MJM, Roldán-Ruiz I, Gladieux P.** 2014. The domestication and evolutionary ecology of apples. *Trends in Genetics* **30**, 57-65.
- Costes E, Crespel L, Denoyes B, Morel P, Demene M-N, Lauri P-E, Wenden B.** 2014. Bud structure, position and fate generate various branching patterns along shoots of closely related Rosaceae species: a review. *Frontiers in Plant Science* **5**, 666.
- Darbyshire R, Webb L, Goodwin I, Barlow EWR.** 2014. Challenges in predicting climate change impacts on pome fruit phenology. *International Journal of Biometeorology* **58**, 1119-1133.
- Egea J, Ortega E, Martínez-Gómez P, Dicenta F.** 2003. Chilling and heat requirements of almond cultivars for flowering. *Environmental and Experimental Botany* **50**, 79-85.
- Fan S, Bielenberg DG, Zhebentyayeva TN, Reighard GL, Okie WR, Holland D, Abbott AG.** 2010. Mapping quantitative trait loci associated with chilling requirement, heat requirement and bloom date in peach (*Prunus persica*). *New Phytologist* **185**, 917-930.
- Francia E, Tacconi G, Crosatti C, Barabaschi D, Bulgarelli D, Dall'Aglio E, Valè G.** 2005. Marker assisted selection in crop plants. *Plant Cell, Tissue and Organ Culture* **82**, 317-342.
- Gianola D, Campos G de los, Hill WG, Manfredi E, Fernando R.** 2009. Additive Genetic Variability and the Bayesian Alphabet. *Genetics* **183**, 347-363.
- Gibson PG, Reighard GL.** 2002. Chilling Requirement and Postrest Heat Accumulation in Peach Trees Inoculated with Peach Latent Mosaic Viroid. *Journal of the American Society for Horticultural Science* **127**, 333-336.
- Gongal A, Silwal A, Amatya S, Karkee M, Zhang Q, Lewis K.** 2016. Apple crop-load estimation with over-the-row machine vision system. *Computers and Electronics in Agriculture* **120**, 26-35.
- Jiang C, Zeng Z-B.** 1995. Multiple Trait Analysis of Genetic Mapping for Quantitative Trait Loci. *Genetics* **140**, 1111-1127.
- Knott SA, Haley CS.** 2000. Multitrait Least Squares for Quantitative Trait Loci Detection. *Genetics* **156**, 899-911.

- Kouassi AB, Durel C-E, Costa F, *et al.*** 2009. Estimation of genetic parameters and prediction of breeding values for apple fruit-quality traits using pedigreed plant material in Europe. *Tree Genetics & Genomes* **5**, 659-672.
- Kumar S, Bink MCAM, Volz RK, Bus VGM, Chagné D.** 2012*a*. Towards genomic selection in apple (*Malus x domestica* Borkh.) breeding programmes: Prospects, challenges and strategies. *Tree Genetics and Genomes* **8**, 1-14.
- Kumar S, Chagné D, Bink MCAM, Volz RK, Whitworth C, Carlisle C.** 2012*b*. Genomic Slection for Fruit Quality Traits in Apple (*Malus x domestica* Borkh.). *Plos One* **7**(5), e36674.
- Lassois L, Denancé C, Ravon E, Guyader A, Guisnel R, Hibrand-Saint-Oyant L, Poncet C, Lasserre-Zuber P, Feugey L, Durel C-E.** 2016. Genetic Diversity, Population Structure, Parentage Analysis, and Construction of Core Collections in the French Apple Germplasm Based on SSR Markers. *Plant Molecular Biology Reporter*, 1-18.
- Laurens F, Durel CE, Patocchi A, Peil A, Salvi S, Tartarini S, Velasco R, van de Weg E.** 2010. Review on apple genetics and breeding programmes and presentation of a new European initiative to increase fruit breeding efficiency. *Journal of Fruit Science* **27**, 102-107.
- Legave JM, Blanke M, Christen D, Giovannini D, Mathieu V, Oger R.** 2013. A comprehensive overview of the spatial and temporal variability of apple bud dormancy release and blooming phenology in Western Europe. *International Journal of Biometeorology* **57**, 317-331.
- Legave JM, Farrera I, Almeras T, Calleja M.** 2008. Selecting models of apple flowering time and understanding how global warming has had an impact on this trait. *Journal of Horticultural Science & Biotechnology* **83**, 76-84.
- Legave JM, Farrera I, Calleja M, Oger R.** 2009. Modelling the Dates of F1 Flowering Stage in Apple Trees, as a Tool to Understanding the Effects of Recent Warming on Completion of the Chilling and Heat Requirements. In: Dixon GR, ed. *I International Symposium on Horticulture in Europe*. Leuven 1: Int Soc Horticultural Science, 153-160.
- Lund MS, Sørensen P, Guldbrandtsen B, Sorensen DA.** 2003. Multitrait Fine Mapping of Quantitative Trait Loci Using Combined Linkage Disequilibria and Linkage Analysis. *Genetics* **163**, 405-410.
- Malosetti M, Ribaut JM, Vargas M, Crossa J, van Eeuwijk FA.** 2008. A multi-trait multi-environment QTL mixed model with an application to drought and nitrogen stress trials in maize (*Zea mays* L.). *Euphytica* **161**, 241-257.
- Mauget J-C, Guinard J, Frizot N.** 1983. Etude de la levée de dormance et du débourrement des bourgeons de noyer (*Juglans regia* L., cv. 'Franquette') soumis à des températures supérieures à 15 C au cours de leur période de repos apparent. *Agronomie* **3**, 745-750.
- Meuwissen THE, Hayes BJ, Goddard ME.** 2001. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics* **157**, 1819-1829.

- Muranty H, Troggio M, Sadok IB, *et al.*** 2015. Accuracy and responses of genomic selection on key traits in apple breeding. *Horticulture Research* **2**, 15060.
- Noiton DAM, Alspach PA.** 1996. Founding Clones, Inbreeding, Coancestry, and Status Number of Modern Apple Cultivars. *Journal of the American Society for Horticultural Science* **121**, 773-782.
- Rosyara UR, Bink MCAM, Weg E van de, Zhang G, Wang D, Sebolt A, Dirlewanger E, Quero-Garcia J, Schuster M, Iezzoni AF.** 2013. Fruit size QTL identification and the prediction of parental QTL genotypes and breeding values in multiple pedigreed populations of sweet cherry. *Molecular Breeding* **32**, 875-887.
- Ruiz D, Campoy JA, Egea J.** 2007. Chilling and heat requirements of apricot cultivars for flowering. *Environmental and Experimental Botany* **61**, 254-263.
- Sadok IB, Tiecher A, Galvez-Lopez D, Lahaye M, Lasserre-Zuber P, Bruneau M, Hanteville S, Robic R, Cournol R, Laurens F.** 2015. Apple fruit texture QTLs: year and cold storage effects on sensory and instrumental traits. *Tree Genetics & Genomes* **11**, 1-20.
- Silva LDCE, Wang S, Zeng Z-B.** 2012. Multiple trait multiple interval mapping of quantitative trait loci from inbred line crosses. *BMC Genetics* **13**, 67.
- Sukhwinder-Singh, Hernandez MV, Crossa J, Singh PK, Bains NS, Singh K, Sharma I.** 2012. Multi-Trait and Multi-Environment QTL Analyses for Resistance to Wheat Diseases. *PLoS ONE* **7**, e38008.
- Tabuenca MC.** 1964. Necesidades de frio invernal de variedades de albaricoquero, melocotonero y peral. *Annals of Aula Dei* **7**, 113-132.
- Thompson EA.** 2013. Identity by Descent: Variation in Meiosis, Across Genomes, and in Populations. *Genetics* **194**, 301-326.
- Van Ooijen JW.** 2004. *MapQTL5: Software for the mapping of quantitative trait loci in experimental populations*. Wageningen, Netherlands: Kyazma B. V.
- Virlet N, Costes E, Martinez S, Kelner J-J, Regnard J-L.** 2015. Multispectral airborne imagery in the field reveals genetic determinisms of morphological and transpiration traits of an apple tree hybrid population in response to water deficit. *Journal of Experimental Botany*, erv355.
- Voorrips RE, Bink MCAM, Kruisselbrink JW, Koehorst-Van Putten,, Van de Weg WE.** 2016. PediHaplotyper: software for consistent assignment of SNP haplotypes in pedigrees. *Molecular Breeding*.
- Yu J, Holland JB, McMullen MD, Buckler ES.** 2008. Genetic Design and Statistical Power of Nested Association Mapping in Maize. *Genetics* **178**, 539-551.

Annexes

Table S3: List of genes underlying major QTLs in first and third versions of the apple genome. For each QTL the list of gene is not exhaustive, only the 40 first genes were included in the tables, and complete version of these tables will be available online with the electronic version of the corresponding paper (see Chapter result 1). Each table contains the list of genes underlying a QTL. In each table, the name of the genes mapped underlying the QTL is given, such as its position, and the chromosome where it is mapped, for first and third version of the apple genome. In addition, the *Arabidopsis thaliana* homologues are also given.

Annexes

Table S3-a: For the QTL on LG7 (54-88cM)

MDP	Apple Genome version 3				Apple Genome Version 1			
	MDC	start_MDC	stop_MDC	CHR	Scaffold ID	Match ID	Organism	Percent Identity
MDP0000319415	MDC053127.000	23903	23937	chr7	chr8	AT2G28880.1	Arabidopsis thaliana	61,69
MDP0000593675	MDC053127.000	14911	15061	chr7	chr8	AT1G48850.3	Arabidopsis thaliana	65,62
MDP0000264634	MDC001065.312	10884	11244	chr7	chr10	AT3G07990.1	Arabidopsis thaliana	78,81
MDP0000429101	MDC016907.372	3206	4510	chr7	chr11	ATMG00860.1	Arabidopsis thaliana	42,73
MDP0000504101	MDC016907.372	1694	1757	chr7	chr17	AT2G24120.1	Arabidopsis thaliana	54,17
MDP0000440507	MDC015392.179	176	579	chr7	chr11	AT3G42170.1	Arabidopsis thaliana	33,96
MDP0000599911	MDC015392.179	166	666	chr7	chr4	AT3G42170.1	Arabidopsis thaliana	35,14
MDP0000377292	MDC006126.328	2143	3390	chr7	chr11	ATMG00750.1	Arabidopsis thaliana	58,93
MDP0000520226	MDC006126.328	310	3390	chr7	chr7	ATMG00750.1	Arabidopsis thaliana	66,07
MDP0000197716	MDC012509.415	3518	3593	chr7	chr7	AT5G38280.1	Arabidopsis thaliana	48,25
MDP0000215259	MDC012509.415	1776	1947	chr7	chr7	AT4G18250.1	Arabidopsis thaliana	43,77
MDP0000147759	MDC005386.449	2287	2345	chr7	chr7	AT4G23160.1	Arabidopsis thaliana	32,61
MDP0000260235	MDC005386.449	4046	4238	chr7	chr16	AT4G23160.1	Arabidopsis thaliana	46,77
MDP0000173609	MDC012509.418	4588	5388	chr7	chr6	AT1G66920.2	Arabidopsis thaliana	52,5
MDP0000191969	MDC012509.418	3233	3920	chr7	chr7	AT1G66920.2	Arabidopsis thaliana	39,49
MDP0000201680	MDC012509.418	3348	3414	chr7	chr2	AT1G66920.1	Arabidopsis thaliana	32,83
MDP0000285161	MDC012509.418	3233	3920	chr7	chr0	AT1G66920.2	Arabidopsis thaliana	43,77
MDP0000386624	MDC012509.418	3233	3920	chr7	chr7	AT1G66920.1	Arabidopsis thaliana	44,47
MDP0000411824	MDC012509.418	3233	3920	chr7	chr2	AT1G66920.2	Arabidopsis thaliana	49,02
MDP0000648787	MDC010408.173	7403	7534	chr7	chr7	ATMG00860.1	Arabidopsis thaliana	40
MDP0000205049	MDC010757.403	6769	6856	chr7	chr7	AT5G36930.1	Arabidopsis thaliana	33,99
MDP0000287797	MDC010757.403	8072	8133	chr7	chr7	AT5G36930.1	Arabidopsis thaliana	42,95
MDP0000291647	MDC010757.403	8072	8133	chr7	chr7	AT5G36930.1	Arabidopsis thaliana	42,76
MDP0000617629	MDC008789.315	13506	13777	chr7	chr7	AT5G48760.2	Arabidopsis thaliana	88,83
MDP0000156587	MDC001601.431	7145	7173	chr7	chr4	AT5G02450.1	Arabidopsis thaliana	86,6
MDP0000327175	MDC001601.431	5640	5696	chr7	chr7	AT5G02450.1	Arabidopsis thaliana	86,6
MDP0000427047	MDC001601.431	168	200	chr7	chr7	AT1G12060.1	Arabidopsis thaliana	48,91
MDP0000571533	MDC001601.431	168	198	chr7	chr10	AT1G12060.1	Arabidopsis thaliana	54,4
MDP0000570445	MDC006005.267	1549	1699	chr7	chr0	ATMG00810.1	Arabidopsis thaliana	34,15
MDP0000626134	MDC006005.267	163	1626	chr7	chr7	AT4G23160.1	Arabidopsis thaliana	44,05
MDP0000626135	MDC006005.267	1846	2865	chr7	chr7	AT4G23160.1	Arabidopsis thaliana	32,02
MDP0000131219	MDC019622.269	1789	2007	chr7	chr7	AT5G51960.1	Arabidopsis thaliana	62,96
MDP0000211188	MDC019622.269	1789	2007	chr7	chr7	AT5G51960.1	Arabidopsis thaliana	62,73
MDP0000011202	MDC018841.202	4437	4505	chr7	chr7	AT2G38905.1	Arabidopsis thaliana	89,8
MDP0000226107	MDC018841.202	4437	4512	chr7	chr1	AT2G38905.1	Arabidopsis thaliana	92,59
MDP0000252251	MDC013775.128	2383	2820	chr7	chr7	AT3G48990.1	Arabidopsis thaliana	72
MDP0000689933	MDC013775.128	2383	2937	chr7	chr12	AT3G48990.1	Arabidopsis thaliana	72,73
MDP0000153032	MDC018211.424	316	4161	chr7	chr7	AT3G09670.2	Arabidopsis thaliana	53,72

Annexes

Table S3-b: For the QTL on LG8 (11-47cM)

MDP	Apple Genome version 3				Apple Genome version 1			
	MDC	start_MDC	stop_MDC	CHR	Scaffold ID	Match ID	Organism	Percent Identity
MDP0000169265	MDC004122.239	2615	3257	chr8	chr8	AT5G10530.1	Arabidopsis thaliana	51,7
MDP0000662096	MDC003563.158	6377	6544	chr8	chr8	AT5G10820.1	Arabidopsis thaliana	56,91
MDP0000311291	MDC003563.158	6377	6642	chr8	chr8	AT4G31870.1	Arabidopsis thaliana	71,74
MDP0000662097	MDC003563.158	8823	8966	chr8	chr8	AT5G57870.1	Arabidopsis thaliana	47,55
MDP0000271659	MDC003563.144	10093	10179	chr8	chr8	AT5G57870.1	Arabidopsis thaliana	58,26
MDP0000662097	MDC003563.144	8499	8642	chr8	chr8	AT5G57870.1	Arabidopsis thaliana	47,55
MDP0000219695	MDC003563.144	8499	8642	chr8	chr8	AT5G57870.1	Arabidopsis thaliana	56,26
MDP0000136645	MDC005353.298	4307	4502	chr8	chr8	AT5G25080.1	Arabidopsis thaliana	57,87
MDP0000252635	MDC005353.298	4307	4373	chr8	chr8	AT5G25080.1	Arabidopsis thaliana	55,17
MDP0000599667	MDC009428.359	6888	6947	chr8	chr8	AT3G27060.1	Arabidopsis thaliana	65,04
MDP0000753874	MDC022595.316	2572	3052	chr8	chr16	ATMG00860.1	Arabidopsis thaliana	39,68
MDP0000835728	MDC051507.000	20544	20908	chr8	chr0	AT2G24170.1	Arabidopsis thaliana	70,09
MDP0000208034	MDC051507.000	19836	20449	chr8	chr0	AT5G10840.1	Arabidopsis thaliana	94
MDP0000136645	MDC051507.000	1059	1254	chr8	chr8	AT5G25080.1	Arabidopsis thaliana	57,87
MDP0000326004	MDC051507.000	10771	10960	chr8	chr8	AT4G31840.1	Arabidopsis thaliana	55,84
MDP0000128730	MDC051507.000	5442	5520	chr8	chr8	AT2G24020.1	Arabidopsis thaliana	74,83
MDP0000759922	MDC051507.000	10771	10960	chr8	chr8	AT4G31840.1	Arabidopsis thaliana	55,84
MDP0000252634	MDC051507.000	5442	5520	chr8	chr8	AT2G24020.1	Arabidopsis thaliana	51,02
MDP0000252635	MDC051507.000	1188	1254	chr8	chr8	AT5G25080.1	Arabidopsis thaliana	55,17
MDP0000542887	MDC051507.000	17939	18241	chr8	chr8	AT4G31830.1	Arabidopsis thaliana	62
MDP0000235045	MDC051507.000	22875	22988	chr8	chr8	AT5G10840.1	Arabidopsis thaliana	86,31
MDP0000154271	MDC051507.000	31785	31898	chr8	chr8	AT5G10840.1	Arabidopsis thaliana	74,6
MDP0000191805	MDC015623.248	305	497	chr8	chr1	AT4G31840.1	Arabidopsis thaliana	59,26
MDP0000326004	MDC015623.248	305	497	chr8	chr8	AT4G31840.1	Arabidopsis thaliana	55,84
MDP0000128730	MDC015623.248	4872	4950	chr8	chr8	AT2G24020.1	Arabidopsis thaliana	74,83
MDP0000759922	MDC015623.248	305	497	chr8	chr8	AT4G31840.1	Arabidopsis thaliana	55,84
MDP0000252634	MDC015623.248	4872	4950	chr8	chr8	AT2G24020.1	Arabidopsis thaliana	51,02
MDP0000214797	MDC017161.153	2018	2028	chr8	chr8	AT4G08900.1	Arabidopsis thaliana	90,74
MDP0000225569	MDC017161.153	2018	2028	chr8	chr12	AT4G08900.1	Arabidopsis thaliana	90,57
MDP0000927700	MDC017161.153	2008	2028	chr8	chr13	AT4G08900.1	Arabidopsis thaliana	92,45
MDP0000531300	MDC006103.280	3091	3679	chr8	chr0	ATMG00860.1	Arabidopsis thaliana	47,3
MDP0000714148	MDC006103.280	2799	3899	chr8	chr10	ATMG00860.1	Arabidopsis thaliana	41,73
MDP0000835728	MDC012768.336	11815	12179	chr8	chr0	AT2G24170.1	Arabidopsis thaliana	70,09
MDP0000208034	MDC012768.336	11151	11713	chr8	chr0	AT5G10840.1	Arabidopsis thaliana	94
MDP0000259807	MDC012768.336	52	590	chr8	chr8	AT3G14460.1	Arabidopsis thaliana	36,33
MDP0000259808	MDC012768.336	5650	5802	chr8	chr8	AT2G24520.1	Arabidopsis thaliana	84,68
MDP0000259809	MDC012768.336	15380	15688	chr8	chr8	AT5G10840.1	Arabidopsis thaliana	76,76
MDP0000675585	MDC012768.336	2410	2963	chr8	chr8	AT3G14460.1	Arabidopsis thaliana	37,1

Annexes

Table S3-c: For the QTL on LG9 (2-10cM)

MDP	Apple Genome Version 3				Apple Genome Version 1			
	MDC	start_MDC	stop_MDC	CHR	Scaffold ID	Match ID	Organism	Percent Identity
MDP0000315893	MDC054300.000	28920	28998	chr9	chr9	AT3G01910.1	Arabidopsis thaliana	82,32
MDP0000317368	MDC054300.000	8504	10058	chr9	chr9	AT5G14170.1	Arabidopsis thaliana	70,41
MDP0000399538	MDC054300.000	2174	3475	chr9	chr9	AT3G27150.1	Arabidopsis thaliana	47,2
MDP0000541368	MDC054300.000	6477	6514	chr9	chr0	AT1G67430.1	Arabidopsis thaliana	85,23
MDP0000802641	MDC054300.000	17012	17509	chr9	chr9	AT3G48520.1	Arabidopsis thaliana	70,42
MDP0000497473	MDC020366.147	4451	4531	chr9	chr9	AT3G07880.1	Arabidopsis thaliana	55,15
MDP0000661029	MDC020366.147	4415	4531	chr9	chr9	AT3G07880.1	Arabidopsis thaliana	52,86
MDP0000723395	MDC005055.979	1163	2599	chr9	chr9	AT1G19260.1	Arabidopsis thaliana	45,44
MDP0000489254	MDC018604.394	2870	3139	chr9	chr15	AT4G23160.1	Arabidopsis thaliana	35,77
MDP0000583525	MDC018604.394	2606	3037	chr9	chr9	AT4G23160.1	Arabidopsis thaliana	40,86
MDP0000583526	MDC018604.394	3215	4084	chr9	chr9	AT4G23160.1	Arabidopsis thaliana	34,77
MDP0000170863	MDC005034.332	5178	5215	chr9	chr9	AT1G27400.1	Arabidopsis thaliana	85,63
MDP0000170864	MDC005034.332	6055	6072	chr9	chr9	AT5G14170.1	Arabidopsis thaliana	71,88
MDP0000170865	MDC005034.332	9222	9523	chr9	chr9	AT3G27120.1	Arabidopsis thaliana	67,71
MDP0000178948	MDC005034.332	5178	5215	chr9	chr17	AT1G27400.1	Arabidopsis thaliana	85,63
MDP0000178949	MDC005034.332	6055	6072	chr9	chr17	AT5G14170.1	Arabidopsis thaliana	72,61
MDP0000200967	MDC005034.332	7136	7453	chr9	chr0	AT5G14170.1	Arabidopsis thaliana	83,96
MDP0000221672	MDC005034.332	5178	5215	chr9	chr9	AT1G67430.1	Arabidopsis thaliana	85,23
MDP0000227849	MDC005034.332	5305	5374	chr9	chr9	AT1G67430.1	Arabidopsis thaliana	87,58
MDP0000312773	MDC005034.332	6176	6838	chr9	chr9	AT3G27120.1	Arabidopsis thaliana	60,78
MDP0000315892	MDC005034.332	6176	7734	chr9	chr9	AT5G14170.1	Arabidopsis thaliana	71,13
MDP0000317368	MDC005034.332	6176	7734	chr9	chr9	AT5G14170.1	Arabidopsis thaliana	70,41
MDP0000417953	MDC005034.332	1061	2332	chr9	chr9	AT3G27150.1	Arabidopsis thaliana	48,8
MDP0000892979	MDC005034.332	1061	2332	chr9	chr17	AT3G27150.1	Arabidopsis thaliana	48,8
MDP0000257116	MDC019947.414	4863	5224	chr9	chr17	AT3G24068.1	Arabidopsis thaliana	26,56
MDP0000399538	MDC021140.166	11931	13223	chr9	chr9	AT3G27150.1	Arabidopsis thaliana	47,2
MDP0000407639	MDC021140.166	11014	11054	chr9	chr9	AT3G27160.1	Arabidopsis thaliana	80,26
MDP0000895750	MDC021140.166	11931	13223	chr9	chr9	AT3G27150.1	Arabidopsis thaliana	49,87
MDP0000153608	MDC007434.177	526	669	chr9	chr0	AT5G14180.1	Arabidopsis thaliana	60,66
MDP0000231401	MDC007434.177	2641	2806	chr9	chr9	AT3G27200.1	Arabidopsis thaliana	67,92
MDP0000254167	MDC007434.177	7833	7929	chr9	chr17	AT5G40870.1	Arabidopsis thaliana	93,27
MDP0000284556	MDC007434.177	2641	2806	chr9	chr17	AT3G27200.1	Arabidopsis thaliana	67,92
MDP0000319946	MDC007434.177	22988	23181	chr9	chr9	AT3G27170.1	Arabidopsis thaliana	79,66
MDP0000430643	MDC007434.177	577	664	chr9	chr9	AT5G14180.1	Arabidopsis thaliana	52,63
MDP0000593400	MDC007434.177	763	823	chr9	chr9	AT5G14180.1	Arabidopsis thaliana	66,9
MDP0000430643	MDC013151.188	1648	1744	chr9	chr9	AT5G14180.1	Arabidopsis thaliana	52,63
MDP0000593400	MDC013151.188	2099	2159	chr9	chr9	AT5G14180.1	Arabidopsis thaliana	66,9
MDP0000269268	MDC000732.217	3199	3381	chr9	chr9	AT5G14180.1	Arabidopsis thaliana	60,98

Annexes

Table S3-d: For QTL on LG10 (22-34cM)

MDP_V3	Apple Genome Version 3				Apple Genome Version 1			
	MDC	start_MDC	stop_MDC	CHR	Scaffold ID	Match ID	Organism	Percent Identity
MDP0000259621	MDC012186.152	5826	6093	chr10	chr13	AT4G33920.1	Arabidopsis thaliana	74,93
MDP0000144523	MDC021505.238	3378	3434	chr10	chr3	AT4G33840.1	Arabidopsis thaliana	57,63
MDP0000426713	MDC021505.238	24381	24734	chr10	chr2	ATMG00860.1	Arabidopsis thaliana	29,37
MDP0000493499	MDC021505.238	25649	25736	chr10	chr3	ATMG00860.1	Arabidopsis thaliana	29,37
MDP0000342190	MDC021652.194	3693	4476	chr10	chr10	AT3G42170.1	Arabidopsis thaliana	26,95
MDP0000441259	MDC021652.194	4876	5340	chr10	chr3	AT3G42170.1	Arabidopsis thaliana	33,17
MDP0000859983	MDC021652.194	1268	1586	chr10	chr3	AT3G54130.1	Arabidopsis thaliana	62,79
MDP0000234284	MDC020136.403	13814	13945	chr10	chr3	AT4G26650.1	Arabidopsis thaliana	75,29
MDP0000276873	MDC020136.403	32796	33083	chr10	chr13	AT4G30010.1	Arabidopsis thaliana	66,2
MDP0000322095	MDC020136.403	7947	8512	chr10	chr3	AT5G20400.1	Arabidopsis thaliana	30,7
MDP0000571392	MDC020136.403	35543	35977	chr10	chr1	AT4G23160.1	Arabidopsis thaliana	35,46
MDP0000718123	MDC020136.403	35596	36217	chr10	chr4	AT4G23160.1	Arabidopsis thaliana	34,8
MDP0000835978	MDC020136.403	21280	21729	chr10	chr3	AT4G33800.1	Arabidopsis thaliana	38,2
MDP0000214891	MDC018551.257	3274	3503	chr10	chr3	AT1G72840.2	Arabidopsis thaliana	35,16
MDP0000214895	MDC018551.257	43473	43709	chr10	chr3	AT1G49960.1	Arabidopsis thaliana	59,17
MDP0000571531	MDC018551.257	50960	51379	chr10	chr3	AT5G04800.4	Arabidopsis thaliana	75,89
MDP0000680996	MDC018551.257	50960	51372	chr10	chr3	AT5G04800.4	Arabidopsis thaliana	81,3
MDP0000298812	MDC012316.130	4376	4976	chr10	chr3	AT4G23160.1	Arabidopsis thaliana	34,08
MDP0000507279	MDC012316.130	3707	4532	chr10	chr2	AT4G23160.1	Arabidopsis thaliana	52,31
MDP0000308301	MDC021685.130	55792	56107	chr10	chr2	AT5G08630.1	Arabidopsis thaliana	63,44
MDP0000638907	MDC021685.130	56192	56252	chr10	chr0	AT5G08630.1	Arabidopsis thaliana	63,44
MDP0000743997	MDC021685.130	55888	56119	chr10	chr11	AT5G08630.1	Arabidopsis thaliana	57,83
MDP0000779011	MDC011163.557	7427	7484	chr10	chr3	AT4G33840.1	Arabidopsis thaliana	57,76
MDP0000738712	MDC005921.376	2957	3457	chr10	chr15	AT1G21280.1	Arabidopsis thaliana	24,24
MDP0000145431	MDC002004.466	13146	14661	chr10	chr3	AT4G23160.1	Arabidopsis thaliana	51,3
MDP0000161955	MDC002004.466	8908	9162	chr10	chr10	AT4G37760.1	Arabidopsis thaliana	74,74
MDP0000161956	MDC002004.466	14146	14442	chr10	chr10	AT4G23160.1	Arabidopsis thaliana	44,93
MDP0000191389	MDC002004.466	8908	9162	chr10	chr5	AT1G58440.1	Arabidopsis thaliana	75,56
MDP0000288317	MDC002004.466	14104	14661	chr10	chr3	AT4G23160.1	Arabidopsis thaliana	47,4
MDP0000525059	MDC002004.466	10376	14661	chr10	chr3	AT4G23160.1	Arabidopsis thaliana	48,08
MDP0000638870	MDC002004.466	8908	9162	chr10	chr1	AT4G37760.1	Arabidopsis thaliana	64,63
MDP0000710449	MDC002004.466	26356	26931	chr10	chr10	AT4G23160.1	Arabidopsis thaliana	32,84
MDP0000168914	MDC004263.238	3343	4056	chr10	chr0	AT5G60440.1	Arabidopsis thaliana	55,62
MDP0000434703	MDC004263.238	7395	8078	chr10	chr0	AT5G60440.1	Arabidopsis thaliana	59,69
MDP0000434709	MDC004263.238	15368	15866	chr10	chr0	AT4G23160.1	Arabidopsis thaliana	49,32
MDP0000453922	MDC004263.238	15554	15990	chr10	chr12	AT4G23160.1	Arabidopsis thaliana	40,14
MDP0000453923	MDC004263.238	15048	15413	chr10	chr12	AT4G23160.1	Arabidopsis thaliana	49,57
MDP0000657806	MDC004263.238	15048	15413	chr10	chr12	AT4G23160.1	Arabidopsis thaliana	50,43

Annexes

Table S3-e: For QTL on LG12 (43-67cM)

MDP_V3	MDC	start_MDC	stop_MDC	CHR	Scaffold ID	Match ID	Organism	Percent Identity
MDP0000604562	MDC007008.338	1823	4522	chr12	chr0	AT4G23160.1	Arabidopsis thaliana	33,88
MDP0000148515	MDC016945.267	1755	1921	chr12	chr0	AT4G08900.1	Arabidopsis thaliana	84,91
MDP0000272122	MDC003783.513	17832	18600	chr12	chr0	AT4G23160.1	Arabidopsis thaliana	33,33
MDP0000272122	MDC004474.234	8906	9639	chr12	chr0	AT4G23160.1	Arabidopsis thaliana	33,33
MDP0000442581	MDC009298.262	298	1638	chr12	chr0	AT4G29090.1	Arabidopsis thaliana	26,62
MDP0000611595	MDC019537.442	1375	1585	chr12	chr0	AT4G23160.1	Arabidopsis thaliana	53,52
MDP0000471066	MDC010910.199	4488	5440	chr12	chr0	AT4G23160.1	Arabidopsis thaliana	39,86
MDP0000492911	MDC005987.217	1858	5359	chr12	chr0	AT3G01410.2	Arabidopsis thaliana	33,02
MDP0000664678	MDC013562.120	283	1524	chr12	chr0	AT1G19260.1	Arabidopsis thaliana	43,69
MDP0000627852	MDC022566.486	2787	3208	chr12	chr0	AT4G23160.1	Arabidopsis thaliana	55,24
MDP0000511140	MDC022370.623	5467	7067	chr12	chr0	AT4G23160.1	Arabidopsis thaliana	39,02
MDP0000597468	MDC010910.199	4631	5382	chr12	chr0	ATMG00300.1	Arabidopsis thaliana	45,38
MDP0000715905	MDC011919.329	9303	13423	chr12	chr0	AT4G23160.1	Arabidopsis thaliana	37,45
MDP0000306945	MDC011919.329	11536	12332	chr12	chr0	ATMG00810.1	Arabidopsis thaliana	30,23
MDP0000505998	MDC000308.142	35	772	chr12	chr0	ATMG00860.1	Arabidopsis thaliana	43,31
MDP0000163803	MDC022544.277	2331	3343	chr12	chr0	AT3G14810.1	Arabidopsis thaliana	36,84
MDP0000200771	MDC000962.711	787	1617	chr12	chr0	AT1G13570.1	Arabidopsis thaliana	39,13
MDP0000211187	MDC000962.711	787	1578	chr12	chr0	AT1G13570.1	Arabidopsis thaliana	38,55
MDP0000232513	MDC000962.711	787	1617	chr12	chr0	AT1G13570.1	Arabidopsis thaliana	38,46
MDP0000301246	MDC017774.149	758	826	chr12	chr0	AT2G43400.1	Arabidopsis thaliana	77,08
MDP0000932247	MDC017774.149	7552	7902	chr12	chr0	AT1G36280.1	Arabidopsis thaliana	76,79
MDP0000932247	MDC001677.177	5640	5990	chr12	chr0	AT1G36280.1	Arabidopsis thaliana	76,79
MDP0000744045	MDC016739.182	1857	2304	chr12	chr0	AT4G23160.1	Arabidopsis thaliana	43,62
MDP0000396099	MDC009298.262	590	1732	chr12	chr0	AT4G29090.1	Arabidopsis thaliana	26,93
MDP0000133209	MDC005063.466	1414	2587	chr12	chr0	AT3G17490.1	Arabidopsis thaliana	24,31
MDP0000733836	MDC011919.329	12924	13423	chr12	chr0	AT4G23160.1	Arabidopsis thaliana	36,33
MDP0000622807	MDC000816.265	3066	3327	chr12	chr0	AT4G23160.1	Arabidopsis thaliana	37,93
MDP0000206211	MDC010598.348	24429	25430	chr12	chr0	AT4G23160.1	Arabidopsis thaliana	45,19
MDP0000695488	MDC001677.177	3575	4527	chr12	chr0	AT3G42170.1	Arabidopsis thaliana	30
MDP0000408030	MDC009298.262	298	1638	chr12	chr0	AT4G29090.1	Arabidopsis thaliana	26,68
MDP0000713992	MDC019537.442	2482	2853	chr12	chr0	AT4G23160.1	Arabidopsis thaliana	42,31
MDP0000166990	MDC015758.276	1733	3439	chr12	chr0	AT1G45616.1	Arabidopsis thaliana	33,9
MDP0000305045	MDC007092.711	1583	1665	chr12	chr0	AT1G56140.1	Arabidopsis thaliana	37,59
MDP0000494200	MDC054532.000	4806	5526	chr12	chr0	ATMG00860.1	Arabidopsis thaliana	42,52
MDP0000399239	MDC008705.117	3770	3983	chr12	chr0	AT4G23160.1	Arabidopsis thaliana	53,73
MDP0000678942	MDC009298.262	4064	5898	chr12	chr0	AT1G43760.1	Arabidopsis thaliana	26,18
MDP0000212693	MDC053832.000	993	2991	chr12	chr0	AT4G23160.1	Arabidopsis thaliana	47,44
MDP0000314752	MDC013154.303	1693	2040	chr12	chr0	AT5G15500.2	Arabidopsis thaliana	36,97

Annexes

Table S3-f: For the QTL on LG15 (105-111cM)

MDP	Apple Genome Version 3				Apple Genome Version 1			
	MDC	start_MDC	stop_MDC	CHR	Scaffold ID	Match ID	Organism	Percent Identity
MDP0000608190	MDC010969.379	46094	46280	chr15	chr15	AT4G37870.1	Arabidopsis thaliana	84,77
MDP0000248937	MDC008226.279	12217	12249	chr15	chr15	AT5G56500.2	Arabidopsis thaliana	72,09
MDP0000248939	MDC008226.279	17089	17372	chr15	chr15	AT4G37870.1	Arabidopsis thaliana	70,3
MDP0000376941	MDC006771.191	2306	2929	chr15	chr10	ATMG00860.1	Arabidopsis thaliana	32,03
MDP0000410690	MDC006771.191	3650	3829	chr15	chr15	ATMG00860.1	Arabidopsis thaliana	29,17
MDP0000415614	MDC006771.191	1871	3829	chr15	chr0	ATMG00860.1	Arabidopsis thaliana	29,37
MDP0000445759	MDC006771.191	2306	2929	chr15	chr0	ATMG00860.1	Arabidopsis thaliana	32,03
MDP0000474486	MDC006771.191	1310	2971	chr15	chr12	ATMG00860.1	Arabidopsis thaliana	31,25
MDP0000543263	MDC006771.191	2348	2860	chr15	chr17	ATMG00860.1	Arabidopsis thaliana	28,57
MDP0000586410	MDC006771.191	2227	3043	chr15	chr4	ATMG00860.1	Arabidopsis thaliana	29,17
MDP0000615549	MDC006771.191	2263	2860	chr15	chr0	ATMG00860.1	Arabidopsis thaliana	28,23
MDP0000739933	MDC006771.191	2263	2860	chr15	chr11	ATMG00860.1	Arabidopsis thaliana	28,23
MDP0000172827	MDC004361.543	8859	9123	chr15	chr15	AT1G58290.1	Arabidopsis thaliana	85,65
MDP0000234947	MDC004361.543	8859	9123	chr15	chr8	AT1G58290.1	Arabidopsis thaliana	82,67
MDP0000151238	MDC012322.235	2679	3011	chr15	chr8	AT5G49665.1	Arabidopsis thaliana	57,05
MDP0000657752	MDC012322.235	665	1092	chr15	chr8	AT4G37890.1	Arabidopsis thaliana	45,89
MDP0000937404	MDC012322.235	2679	3011	chr15	chr15	AT5G49665.1	Arabidopsis thaliana	55,31
MDP0000090281	MDC000033.254	13939	14235	chr15	chr15	AT2G22670.4	Arabidopsis thaliana	57,01
MDP0000158493	MDC000033.254	19509	19604	chr15	chr15	AT1G34070.1	Arabidopsis thaliana	27,1
MDP0000200927	MDC000033.254	14077	14346	chr15	chr0	AT5G65670.1	Arabidopsis thaliana	55
MDP0000289559	MDC000033.254	19509	19604	chr15	chr15	AT1G34070.1	Arabidopsis thaliana	26,79
MDP0000158493	MDC011236.664	14173	14268	chr15	chr15	AT1G34070.1	Arabidopsis thaliana	27,1
MDP0000289559	MDC011236.664	14173	14268	chr15	chr15	AT1G34070.1	Arabidopsis thaliana	26,79
MDP0000618087	MDC003263.398	22111	22326	chr15	chr15	AT5G65660.1	Arabidopsis thaliana	46,25
MDP0000175282	MDC001175.360	14337	15008	chr15	chr8	AT5G65640.1	Arabidopsis thaliana	67,63
MDP0000178782	MDC001175.360	14337	15008	chr15	chr8	AT5G65640.1	Arabidopsis thaliana	67,63
MDP0000644807	MDC001175.360	14337	15008	chr15	chr15	AT5G65640.1	Arabidopsis thaliana	65,7
MDP0000204592	MDC013787.224	7967	8286	chr15	chr15	AT3G42170.1	Arabidopsis thaliana	23,4
MDP0000298374	MDC013787.224	7543	7560	chr15	chr2	AT3G42170.1	Arabidopsis thaliana	25,91
MDP0000412867	MDC013787.224	5986	6370	chr15	chr13	AT3G42170.1	Arabidopsis thaliana	37,5
MDP0000390616	MDC008396.418	1706	2743	chr15	chr7	ATMG00860.1	Arabidopsis thaliana	34,16
MDP0000391997	MDC008396.418	1528	2149	chr15	chr0	ATMG00860.1	Arabidopsis thaliana	39,58
MDP0000431271	MDC008396.418	101	1861	chr15	chr15	ATMG00750.1	Arabidopsis thaliana	57,14
MDP0000521027	MDC008396.418	2624	2743	chr15	chr14	ATMG00860.1	Arabidopsis thaliana	40
MDP0000581852	MDC008396.418	101	1166	chr15	chr6	ATMG00750.1	Arabidopsis thaliana	53,57
MDP0000585943	MDC008396.418	1901	2743	chr15	chr15	ATMG00860.1	Arabidopsis thaliana	36,46
MDP0000588806	MDC008396.418	695	1828	chr15	chr15	ATMG00750.1	Arabidopsis thaliana	51,79
MDP0000601210	MDC008396.418	516	1161	chr15	chr8	ATMG00750.1	Arabidopsis thaliana	50,91

Résumé

Chez le pommier, la sélection porte principalement sur des traits majeurs tels que la qualité du fruit, la résistance aux maladies, et le rendement. D'autres caractères secondaires prennent aujourd'hui de l'importance car ils sont liés à des risques agronomiques et adaptatifs. Ainsi, la régularité de production est gérée par l'éclaircissage manuel ou chimique, mais la volonté de réduire l'utilisation de produits phytosanitaires requiert de nouvelles méthodes de gestion. La sélection de variétés régulières est une solution envisagée. Par ailleurs, il a été montré que le changement climatique entraîne des perturbations de la floraison chez le pommier dans les climats doux qui pourraient s'étendre à l'avenir à d'autres zones géographiques. Ce risque pourrait être limité par la sélection de variétés adaptées. Les objectifs de sélection chez le pommier sont donc multiples et portent sur des caractères de nature différente. Le premier objectif de la thèse a été d'étudier les déterminismes génétiques de la régularité et de la date de floraison dans un dispositif multi-population. Le deuxième objectif a été d'intégrer plusieurs caractères majeurs et secondaires dans un idéotype prédéfini et d'estimer des valeurs génétiques multi-caractères avec des indices de sélection.

Le dispositif expérimental était composé de cinq familles en ségrégation dont les parents sont reliés par leur pedigree. Des cartographies de QTL ont été réalisées grâce à un modèle bayésien permettant de prendre en compte les apparentements, le Pedigree Based Analysis (PBA). Ces analyses ont été réalisées avec d'une carte génétique intégrée composée d'environ 7000 SNPs issus d'une puce à 20 000 SNPs. Le modèle Bayésien de détection de QTLs a aussi permis de calculer des valeurs génétiques basées soit sur l'intégralité du génome ou seulement sur les QTLs détectés. Plusieurs caractères majeurs et secondaires ont été intégrés dans un idéotype prédéfini, et leurs corrélations analysées. Des valeurs génétiques multi-caractères ont été estimées avec des indices de sélection sur la base de l'idéotype afin d'identifier les meilleurs individus.

Les analyses QTL sur la phénologie de la floraison ont permis l'identification de plusieurs QTLs robustes sur les groupes de liaison (LG) 7, 10, 12, et 9. Plusieurs gènes candidats ont été cartographiés sous ces QTLs : un homologue du gène *FT* sous le QTL du LG12, un homologue d'*AGL24* sous le QTL du LG9 et deux gènes *DAM* sous deux QTLs mineurs sur les LG8 et LG15. Pour la régularité de production, plusieurs indices mesurés à l'échelle de la pousse annuelle ont été élaborés et utilisés comme prédicteurs du comportement global de l'arbre. De plus, les analyses QTL ont révélé plusieurs QTLs déjà identifiés dans des études précédentes sur les LG4, 8 et 10, mais aussi de nouveaux QTL sur les LGs 5 et 7. Les corrélations entre traits ont permis de révéler des synergies, entre régularité de production et taille du fruit, et des antagonismes, entre rendement et taille/régularité de production, par apport à l'idéotype fixé. De plus, les valeurs d'index de sélection ont permis de révéler un potentiel différent pour les deux familles utilisées. Finalement, l'estimation de valeurs génétiques multi-caractères des individus sur la base des marqueurs moléculaires vise à ouvrir la perspective d'une stratégie de sélection génomique simultanée sur plusieurs caractères.