



**HAL**  
open science

# Déterminisme génétique de la régularité de production chez le pommier et l'olivier

Agathe Maillard

► **To cite this version:**

Agathe Maillard. Déterminisme génétique de la régularité de production chez le pommier et l'olivier. Sciences agricoles. 2015. hal-02796612

**HAL Id: hal-02796612**

**<https://hal.inrae.fr/hal-02796612>**

Submitted on 5 Jun 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# DETERMINISME GENETIQUE DE LA REGULARITE DE PRODUCTION CHEZ LE POMMIER ET L'OLIVIER

Agathe MAILLARD



## RESUME

L'alternance de production est un caractère phénotypique au déterminisme génétique complexe. Exprimé par le pommier (*Malus x domestica* Borkh) et l'olivier (*Olea europea*), sa maîtrise présente un grand intérêt agronomique. Pour caractériser les aspects génétiques de ces deux traits, la régularité de production et le rendement, cinq valeurs distinctes sont calculées statistiquement et servent de bases à une détection de QTLs (zones génomiques d'intérêts) sur des cartes génétiques élaborées à cette fin pour les populations en croisement « X3263 x Belrène » pour le pommier et « Olivière x Arbequina » pour l'olivier. Ces cartes sont construites avec des marqueurs moléculaires SNP pour apporter de la précision et des marqueurs SSR qui servent de liens avec d'autres cartes pour les mêmes espèces. Suite aux analyses de données, trois types de comportements ont été isolés au sein des populations étudiées : alternant, régulier et irrégulier. Les QTLs repérés ont ensuite pu être comparés avec d'autres analyses plus anciennes et certaines similitudes ont été mises en évidence, validant les positions et les fonctions de ces zones. De plus, une première détection des gènes candidats pour ces caractères a été réalisée, les grandes catégories des gènes situés sous les QTLs du pommier ont été listées tandis que les annotations associées aux marqueurs sous les QTLs de l'olivier ont été extraites.

Mots clés : Pommier, *Malus x domestica* Borkh, olivier, *Olea europea*, alternance, rendement, QTLs, cartes génétiques, gènes candidats

## ABSTRACT

The Biennial Bearing is a phenotypic trait which has a complex genetic determinism. Expressed by the apple tree (*Malus domestica* Borkh) as well as the olive tree (*Olea europea*), controlling it is representing a major agronomic point. In order to characterize the genetic aspects of these two traits, the yield and the regularity of production, five different values have been statistically calculated and are the basis for QTL detection on the genetic maps elaborated for this purpose for the apple tree population crossed-pollinated "X3263 x Belrène" and for the olive tree population "Olivière x Arbequina". These maps are tools that were elaborated with molecular markers named SNP in order to bring acuity and microsatellites markers SSR to link these new tools with other ones. Following data analysis, three types of behavior have been found among the studied populations: biennial bearer, irregular and regular bearer. QTLs detected have been compared with ancient analysis and some similar results appeared, validating the locations and the functions of these zones. Furthermore, a first candidate genes detection for these traits have been realized. Large categories of genes situated under the apple tree's QTLs have been listed while the annotations associated to the markers under the olive tree's QTLs have been extracted.

Keywords : Apple tree, *Malus x domestica* Borkh, olive tree , *Olea europea*, biennial bearing, yield, QTLs, genetic maps, candidate genes.



## SOMMAIRE

Résumé	3
Sommaire	5
Remerciements	7
Table des sigles et abréviations	9
Introduction	11
Partie I : Contexte	13
Partie II : Problématiques	27
Partie III : Matériel et Méthodes	31
Partie III : Résultats	51
Partie IV : Discussion	71
Conclusion	83
Bibliographie	85
Table des figures	95
Table des tableaux	96
Table des annexes	97
Table des matières	xvi



## REMERCIEMENTS

En premier lieu, je voudrais remercier mon maître de stage, Bouchaïb Khadari pour m'avoir encadrée lors de mon stage de fin d'étude ainsi que pour ses suggestions quant à l'écriture de ce mémoire. Merci également à Jean-Jacques Kelner pour son aide et ses conseils autour des gènes candidats et sur la discussion des résultats obtenus.

J'adresse mes remerciements à tous les membres de l'équipe AFEF pour leur accueil chaleureux, leur bonne humeur et leur bienveillance tout au long de ce stage. Particulièrement, un grand merci à Evelyne Costes pour sa disponibilité, son suivi général et son appui en statistiques, à Alix Allard pour sa gentillesse, ses conseils en cartographie et en détection de QTLs et à Sébastien Martinez pour ses explications sur le phénotypage et les expérimentations. Merci également à Isabelle Farrera pour son aide avant ma présentation au GIS Fruits, à Benoît Pallas pour son aide en statistiques et à Françoise Dosba pour ses conseils pour l'oral au GIS Fruits. Enfin merci à Véronique, Pierre-Eric-Lauri, Jean-Luc Regnard, Jean-Michel Legave, Lucille, Lamia et Gerardo pour leurs conseils et leur gentillesse. Un grand merci à Agnès Doligez et Timothée Flutre de l'équipe DAAV pour le temps qu'ils m'ont accordé et leurs conseils d'utilisation du logiciel R. Un grand merci aussi à Sylvain Santoni pour ses suggestions et sa bienveillance.

Je voudrais également remercier Hajar Chouiki pour m'avoir accordé du temps et expliqué son travail sur les marqueurs SNP de l'olivier ainsi que Gauthier Sarah pour ses explications sur les annotations des marqueurs.

J'adresse toute ma reconnaissance aux membres du GIS Fruits pour m'avoir permis de réaliser ce stage et donné l'opportunité de présenter mon travail au cours de l'une des réunions.

Merci à Vassilia Theodorou pour m'avoir suivie lors de l'écriture de ce mémoire et pour ses suggestions.

Pour finir, je remercie ma famille ainsi que mes colocataires Mathilde et Kim pour leur soutien dans l'écriture de ce mémoire de fin d'étude. Merci aussi à tous mes amis purpanais pour avoir rendu inoubliables ces cinq années d'études.





## TABLE DES SIGLES ET ABREVIATIONS

ADN : Acide DésoxyriboNucléique  
A<sub>f</sub> : Effet allélique femelle  
AFEF : Architecture et Fonctionnement des Espèces Fruitières  
AFLP : Amplified Fragment-Length Polymorphism  
AGAP : Amélioration Génétique et Adaptation des Plantes Méditerranéennes et Tropicales  
AIC : Akaike Information Criterion  
A<sub>m</sub> : Effet allélique mâle  
ANOVA : Analysis Of Variances  
ANR : Agence Nationale de la Recherche  
ARCAD : Agropolis Resource Centre for Crop Conservation, Adaptation and Diversity  
ARN : Acide RiboNucléique  
BBI : Biennial Bearing Index (Indice de Régularité de Production)  
BIC : Bayesian Information Criterion  
BLAST : Basic Local Alignment Search Tool  
BLUPs : Best Linear Unbiased Prediction (Bilan Linéaire Universel Prévisionnel)  
CBN : Conservatoire Botanique National  
CIRAD : Centre de Coopération Internationale en Recherche Agronomique pour le Développement  
cM : CentiMorgan  
CNRS : Centre National de la Recherche Scientifique  
CP : Cross Pollination  
CTIFL : Centre Technique Interprofessionnel des Fruits et Légumes  
DAVEM : Diversité et Adaptation de la Vigne et des Espèces Méditerranéennes  
EMS : Éthyl Méthane Sulfonate  
EST : Expressed Sequence Tag  
GBS : Genotyping by Sequencing  
GDR : Genome Database for Rosaceae  
GIS : Groupement d'Intérêt Scientifique  
ID : Intégration de Données  
IFV : Institut Français de la Vigne et du Vin  
INRA : Institut National de la Recherche Agronomique  
INRIA : Institut National de la Recherche d'Informatique et d'Automatique  
IRD : Institut de Recherche pour le Développement  
ISO : Organisation Internationale de Normalisation  
LG : Linkage Group  
LOD : Logarithm of odds (logarithme des probabilités)  
ML : Maximum Likelihood  
NCBI : National Center for Biotechnology Information  
NGS : New Generation Sequencing

PCR : Polymerase Chain Reaction (Amplification en chaîne par polymérase)

QTL : Quantitative Trait Loci

REML : Restricted Maximum Likelihood

rMQM : restricted Multiple QTLs Mapping

RNAseq : Acide Ribonucléique sequencing

SNP : Single Nucleotide Polymorphism

SSR : Simple Sequence Repeat

UMR : Unité Mixte de Recherche

## INTRODUCTION

L'amélioration génétique des espèces cultivées et notamment des espèces fruitières vise à apporter des solutions adaptées et durables aux arboriculteurs en répondant à leurs impératifs et aux enjeux du marché des fruits en France et à l'étranger. Dans ce contexte global où le producteur français est en concurrence avec des agriculteurs du monde entier, il devient primordial pour lui de pouvoir satisfaire ses consommateurs en proposant des fruits de qualité à un prix compétitif. En parallèle, il doit s'adapter à de nouveaux enjeux sociétaux tels la réduction des produits phytosanitaires. Or, l'arboriculture, et particulièrement la culture de pommiers, est l'un des domaines agricoles consommant le plus de traitements. Ainsi, les diverses substances nécessaires pour assurer une bonne qualité du fruit ainsi qu'un rendement suffisant et régulier d'une année sur l'autre sont coûteuses et leur application est chronophage. Des schémas de sélection ont d'ores et déjà pris en compte le caractère de résistance à la tavelure afin de réduire les traitements. La caractéristique d'alternance de production, qui pourtant représente une charge de travail importante si l'on veut la maîtriser, n'est pas encore prise en compte car elle est très complexe. La plantation de parcelles de variétés dont la production fruitière est régulière permettrait pourtant de réduire les interventions culturales sur les parcelles, tant mécaniques que chimiques, mais aussi d'assurer un rendement en fruits régulier et donc un revenu régulier au producteur. C'est à partir de la compréhension des aspects génétiques qui induisent ce caractère de l'alternance qu'il sera possible de les intégrer à des programmes de sélection variétale et à terme de proposer aux agriculteurs des variétés à la production régulière.

Dans ce contexte, l'équipe mixte AFEF, Architecture et Fonctionnement des Espèces Fruitières, appartenant à l'Unité Mixte de Recherche AGAP, Amélioration Génétique et Adaptation des Plantes, étudie les causes génétiques liées à ce caractère phénotypique de la régularité de production chez deux espèces le présentant : le pommier et l'olivier.

Ce mémoire s'inscrit donc dans l'approfondissement des connaissances sur ce thème par la construction de cartes génétiques permettant de détecter des zones du génome qui y sont liées (QTLs). Le déterminisme génétique de ces deux espèces pour ce caractère d'intérêt agronomique, l'alternance, est développé dans cette étude.

Après une mise en contexte des problématiques de ce mémoire, le matériel utilisé et la méthode mise en place sont décrits pour aboutir à la présentation des résultats obtenus qui sont discutés et induisent de nouvelles perspectives.



## Partie I : Contexte

## **1. Le domaine de la génétique végétale**

### **1.1. La génétique à travers les âges**

Entre 13 000 et 10 000 ans se sont écoulés depuis que la domestication des végétaux a vu le jour, au cours l'ère Néolithique (Purugganan et Fuller, 2009). C'est dans le croissant fertile en Mésopotamie que l'Homme s'est sédentarisé pour la première fois et a opéré ce que l'on appelle aujourd'hui « la Révolution Néolithique ». Sans en avoir toujours conscience, une sélection artificielle des végétaux a été effectuée en se basant sur des caractères phénotypiques d'intérêts agricoles, médicaux, etc... Or, le phénotype d'un individu est le reflet de son génotype en interaction avec son environnement, représenté par la formule ci-dessous (Via et Lande, 1985).

$$P = G + E + G/E$$

Phénotype = Génotype + Environnement + Interaction Génotype/Environnement

Indirectement, ce sont donc les gènes responsables des caractères alors recherchés qui ont bénéficié de cette première sélection empirique basée uniquement sur le phénotype de la plante.

Pourtant, ce n'est qu'au 19<sup>ème</sup> siècle, soit il y a moins de 200 ans, que les fondements de la génétique des populations ont commencé à s'éclaircir grâce au travail fourni par Charles Darwin qui a développé la théorie de l'évolution et de la sélection naturelle qu'il présenta en 1858 puis dans son ouvrage « L'Origine des Espèces » publié en 1859 (Darwin, 1859 ; Lane, 1994 ; Mandal, 2013). A la même période, le moine autrichien Gregor Mendel proposa les premiers principes de transmission des caractères simples, aussi dits mendéliens, en 1865 (Mendel, 1866 ; Heberlé-Bors, 1996 ; Miko, 2008). Les théories de Mendel, s'appuient uniquement sur des observations phénotypiques d'un seul caractère transmissible et mesurable sur plusieurs générations de pois commun (*Pisum sativum*).

En 1869, la nucléine, qui sera plus tard appelée acide désoxyribonucléique ou ADN, est isolée et décrite par Frederick Miescher (Miescher, 1869). C'est en 1909 que le terme « gène » apparaît pour la première fois dans la littérature scientifique pour désigner l'unité de Mendel qui se transmet de générations en générations (Johannsen, 1909). Cependant, il ne sera démontré que l'ADN est un matériel héréditaire qu'en 1944 (Avery et al., 1944). La structure à double hélice de l'ADN sera quant à elle déterminée 9 ans après (Watson et Crick, 1953).

Ce n'est qu'au cours du 20<sup>ème</sup> siècle que les théories de Mendel, qui avaient été relativement oubliées pendant quelques décennies, sont redécouvertes et utilisées dans le cadre de nouveaux travaux scientifiques internationaux. Les découvertes scientifiques autour d'aspects génétiques se sont multipliées au cours du siècle dernier pour atteindre aujourd'hui un haut niveau de connaissances sur l'ADN et la génétique plus généralement. Au travers de ces connaissances et du développement de hautes technologies et biotechnologies, les organismes vivants peuvent être étudiés de plus en plus finement.

### 1.2. La recherche génétique assistée par marqueurs moléculaires

Avec les avancées en matière de génétique moléculaire, de nouvelles techniques de caractérisation génétique se sont développées. La cartographie génétique est ainsi vieille de plus d'un siècle (Morton, 2004). Elle consiste à modéliser les chromosomes physiques de l'être vivant étudié en essayant de les représenter le plus précisément possible.

Les marqueurs moléculaires se fixent sur une partie précise de la séquence nucléotidique (toujours la même) de l'espèce étudiée. Il est alors possible, après séquençage d'une variété par exemple, d'élaborer des cartes génétiques en assemblant ces marqueurs. Ces cartes génétiques serviront de repères pour l'espèce concernée. A partir de cet outil, il est possible de positionner sur ces cartes des zones nommées QTLs (Quantitative Trait Loci) correspondant à un trait phénotypique chez une population (Zeng, 1994). Sous ces zones se trouvent de très nombreux gènes, plusieurs centaines voire milliers. Parmi eux, des gènes dits candidats pourront être détectés. Ces derniers sont potentiellement impliqués dans l'expression du caractère étudié, de manière plus ou moins directe.

Des cartes génétiques ont été confectionnées pour de nombreuses espèces végétales aussi bien annuelles que pérennes. Ainsi, l'orge dispose de plus de quinze cartes réalisées à partir de différents types de marqueurs (USDA, 2014). Pour ce qui est des espèces pérennes, de nombreux arbres cultivés ont également une ou plusieurs cartes de références. En effet, la vigne (Wang et al., 2012), le pêcher (Dirlewanger et al., 1998) et la plupart des arbres économiquement importants en possèdent.

### 1.3. Les enjeux de la recherche en génétique végétale

L'adaptation de la production végétale à l'accroissement de la population mondiale ainsi qu'au dérèglement climatique est l'enjeu actuel majeur qui anime la recherche concernant les végétaux (Trottet, 2008). De plus, le développement rapide de certains pays et ainsi la hausse du niveau de vie de leurs habitants va entraîner une augmentation mondiale de leur consommation de viande, notamment. Or, l'élevage en vue de la production carnée est très consommateur de produits végétaux type fourrages et compléments protéinés destinés à l'alimentation animale. De même, l'utilisation de biomasse agricole pour produire des agrocarburants mobilise des surfaces qui auraient pu être disponibles pour la production de denrées alimentaires. Ainsi, l'une des visées de la recherche en génétique végétale est d'optimiser la production des espèces étudiées en augmentant le rendement à l'hectare de ces plantes tout en les adaptant à des systèmes de cultures à bas intrants. Le but est de permettre à chacun de développer son activité dans le respect de l'autre et de la planète. L'agriculture intensive n'est plus la priorité absolue du secteur agricole français comme cela a pu être le cas après la dernière guerre mondiale. Ce mode de production a laissé place à l'agriculture raisonnée alors que les enjeux environnementaux et relatifs à la santé se sont eux aussi peu à peu imposés (Paillotin, 2000). Ce changement s'est fait parallèlement à la



prise de conscience du consommateur quant à l'importance d'un mode de consommation plus raisonné.

Comprendre les mécanismes en jeu dans les expressions de différents caractères d'intérêts agronomiques, tant sur le plan de la physiologie végétale que sur celui de leur déterminisme génétique va contribuer à avancer les recherches concernant divers phénomènes qui aujourd'hui posent des contraintes aux agriculteurs. Ces nouvelles compétences pourront alors servir de bases pour explorer de nouvelles pistes de systèmes culturaux et de sélection variétale pour pallier ces contraintes.

Le développement rapide des biotechnologies a nécessité une adaptation, d'ailleurs toujours en cours, des réglementations encadrant la propriété industrielle et les notions d'éthique relatives aux sélections et modifications variétales (Girard et Noiville, 2014).

La modification du vivant vise à répondre à des besoins liés à des phénomènes naturels pouvant être vécus comme des contraintes par les principaux acteurs utilisant ces végétaux. Ainsi, divers laboratoires de recherches travaillent sur les aspects génétiques liés à ces caractères agronomiques afin de les comprendre et à terme, de les maîtriser.

## **2. Le centre de recherche**

### **2.1. L'UMR AGAP**

L'équipe AFEF, Architecture et Fonctionnement des Espèces Fruitières, au sein de laquelle j'ai effectué mon stage de fin d'études fait partie d'une unité plus importante : l'UMR AGAP (Amélioration Génétique et Adaptation des Plantes Méditerranéennes et Tropicales). Cette unité mixte de recherche, créée le 1<sup>er</sup> janvier 2011, est sous tutelle de deux organismes de recherches, le CIRAD (Centre de Coopération Internationale en Recherche Agronomique pour le Développement) et l'INRA (Institut National de la Recherche Agronomique), ainsi que d'une unité d'enseignement supérieur, Montpellier SupAgro. De plus, des organismes de recherche tels l'INRIA (Institut National de Recherche en Informatique et en Automatique), l'IFV (Institut Français de la Vigne et du Vin) et le CBN (Conservatoire Botanique National) font également partie de cette structure.

Les treize équipes de l'UMR AGAP regroupent des chercheurs appartenant à ces différents organismes de recherches et d'enseignement supérieur (UMR AGAP, 2012 (a)). Ces équipes sont réparties selon leur objet d'étude ou encore leur domaine de compétence. Ainsi, parmi ces groupes, certains travaillent exclusivement sur le riz (équipe DAR, Développement Adaptatif du Riz), sur la vigne (équipe DAAV, Diversité, adaptation et amélioration de la vigne) ou bien sur la modélisation des végétaux (équipe VP, Virtual Plants) ou encore la bioinformatique (équipe ID, Intégrations de Données).

Le but de cette UMR est de rassembler dans une même structure des unités travaillant toutes en lien avec l'amélioration génétique du végétal. Les membres de cette unité

étudient des espèces d'intérêts agronomiques et biologiques variés tels le sorgho, le riz, la vigne, l'hévéa, le cacaoyer, le palmier à huile, sans oublier le pommier et l'olivier. Bien d'autres espèces sont aussi étudiées dans ce cadre par l'UMR AGAP. Le regroupement de toutes ces équipes vise à augmenter la visibilité des recherches françaises dans ce domaine. Cette unité est certifiée ISO 9001. En effet, l'UMR AGAP vise à apporter à ses travailleurs une qualité des conditions de travail exemplaire et s'engage pour minimiser son impact sur l'environnement dans le cadre d'une stratégie de développement durable (UMR AGAP, 2014).

La participation active de l'INRA au sein de l'UMR offre une visibilité importante à la structure. En 2013, l'INRA comptait au total près de 8500 agents titulaires répartis dans les 17 centres régionaux de l'institut ainsi qu'au siège social. Le budget de cet organisme de recherche national s'élève à plus de 880 millions d'euros alloués aux équipes de recherches chaque année (INRA, 2013). Depuis le milieu des années 2000, l'unité mixte s'intéresse de plus en plus aux aspects génétiques conduisant au phénomène de l'alternance de production des cultures pérennes et qui va constituer le cœur de cette étude au sein de l'équipe AFEF.

## 2.2. L'Equipe AFEF

L'AFEF, Architecture et Fonctionnement des Espèces Fruitières, est une équipe composée de 12 personnes occupant un poste de permanent (Figure 1). A ces permanents s'ajoutent plusieurs doctorants, post-doctorants et stagiaires dont le nombre varie suivant les années ainsi que les périodes de l'année. Basés au centre du CIRAD, à Montferrier-sur-Lez au nord de Montpellier, les chercheurs de cette équipe travaillent, conjointement et en partenariat avec d'autres équipes et organismes, sur le pommier, l'olivier et le manguier, principalement. L'équipe est composée de chercheurs ayant des compétences différentes et complémentaires ce qui permet d'entreprendre des travaux sur des sujets pluridisciplinaires. Ainsi, les statistiques, les aspects génétiques et la physiologie végétale, entre autres, sont des domaines maîtrisés au sein d'AFEF.

Cette équipe est l'auteure de nombreuses publications scientifiques dans divers journaux tels Plos One, Scientia Horticulturae et BMC Plant Biology pour n'en citer que quelques-uns<sup>1</sup>. De plus, d'importantes expérimentations de phénotypage haut-débit ont été menées par AFEF. Ainsi, au cours l'été 2014, « PhénoArch », une manipulation visant à étudier le stress hydrique sur une population de pommiers, a été conduite dans une des serres de l'INRA de Montpellier (Lucas, 2014).

---

<sup>1</sup> Annexe I : Publications clés de l'équipe

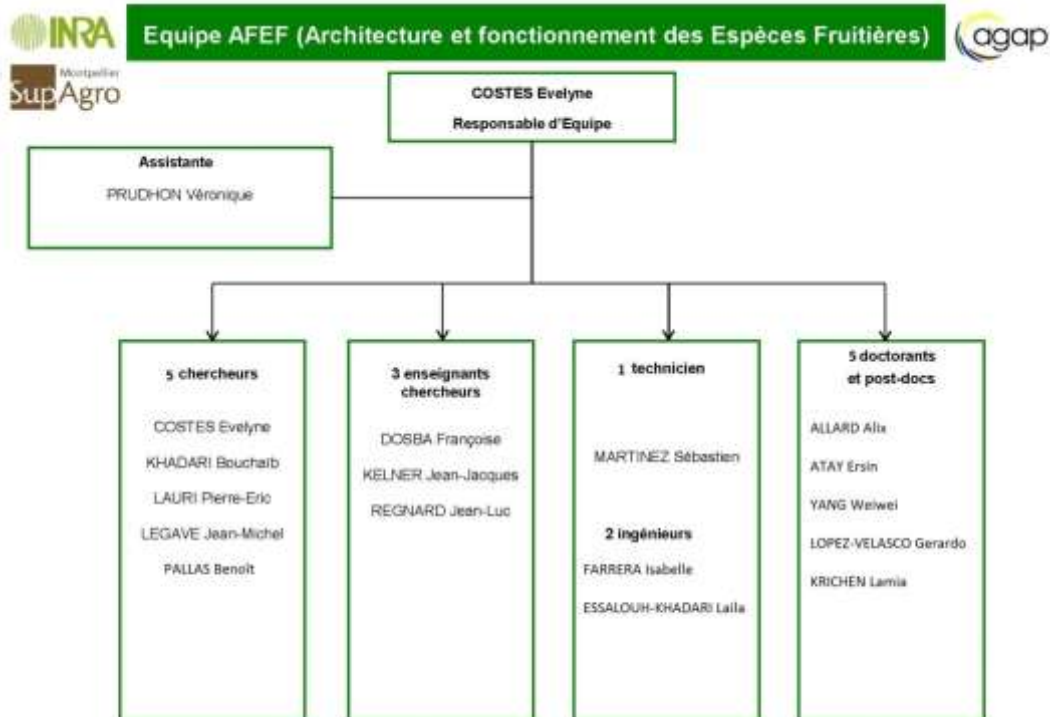


Figure 1 : Organigramme de l'équipe AFEF

### 2.3. Le fonctionnement de l'équipe

L'équipe est localisée dans le bâtiment 3 sur le site du CIRAD à Montpellier, dans un unique couloir. Cela permet une communication facile au sein de l'équipe. Les ordinateurs sont reliés au réseau internet du CIRAD qui donne accès à de nombreux sites de publications scientifiques ainsi qu'à l'intranet de l'organisme.

Cependant, bien que l'AFEF soit réunie dans un même lieu, seules deux autres équipes de l'UMR AGAP se trouvent sur le même site. Les échanges avec d'autres collègues de la même unité mixte mais basés dans les locaux de l'INRA à Montpellier s'en trouvent donc diminués. Cette séparation des équipes en plusieurs lieux est incontestablement un frein au travail en synergie des différents membres de l'UMR.

En revanche, le campus de Lavalette où se trouve le CIRAD est un endroit privilégié pour le secteur de la recherche sur le végétal. En effet, la fondation Agropolis est située toute proche ainsi que deux autres grands centres de recherche, l'IRD (Institut de Recherche pour le Développement) et le CNRS (Centre National de la Recherche Scientifique). Le regroupement de ces différents organismes offre un rayonnement plus important au campus en matière de recherche scientifique et plus précisément dans le domaine de l'agronomie.

En travaillant dans les locaux du CIRAD, l'équipe peut également disposer des laboratoires, des serres et des équipements dans le cadre de ses propres recherches.

L'équipe reçoit ses financements en soumettant ses projets de recherches à des appels à projets provenant de divers organismes comme l'ANR (Agence Nationale de la Recherche) ou de fonds européens, fondations scientifiques et associations caritatives entre autres (Ministère de l'Education nationale, de l'Enseignement supérieur et de la Recherche, 2013).

#### 2.4. Le GIS Fruits

Les GIS, Groupement d'Intérêt Scientifique, sont des groupements contractuels à but non-lucratif, créés à l'initiative du CNRS et visant à promouvoir la coopération scientifique entre organismes de recherches, de développement et de formation concernant un sujet scientifique fédérateur des différentes entités qui le composent (CNRS, 2012).

Le GIS Fruits regroupe de nombreux acteurs de la filière fruitière en France. En l'état, ce groupe rassemble vingt-deux membres dont l'INRA, le Conservatoire Botanique, Montpellier Supagro, le CTIFL (Centre Technique Interprofessionnel des Fruits et Légumes), France Agrimer et Agrocampus Ouest (INRA, 2011). Ce groupe, outre sa fonction de liaison entre les partenaires du secteur fruitier en France, leur propose aussi d'intervenir dans l'organisation de conférences faites par et pour les membres et de constituer des groupes de travail sur différents sujets d'actualité de la filière. Le GIS Fruits se réunit trois fois par ans pour discuter des grandes orientations et des projets à mener. Depuis 2014, le groupement finance des stages Master 2 dans les différents organismes qui le composent. C'est donc dans ce cadre que le stage de fin d'études au sein de l'équipe AFEF de l'UMR AGAP a pu m'être proposé.

La problématique de la régularité de production étant commune à de nombreuses espèces fruitières, trois partenaires du groupement se sont proposés pour apporter leur contribution dans le but de mener une étude sur ce thème.

### 3. L'alternance de production

#### 3.1. Caractérisation du phénomène

De nombreuses espèces fruitières ont la caractéristique au champ d'avoir des populations dont la production passe de « ON » à « OFF » d'une année sur l'autre (Figure 2). Cependant, certaines variétés parmi elles présentent une production stable dans le temps (Monselise et Goldschmidt, 1982).

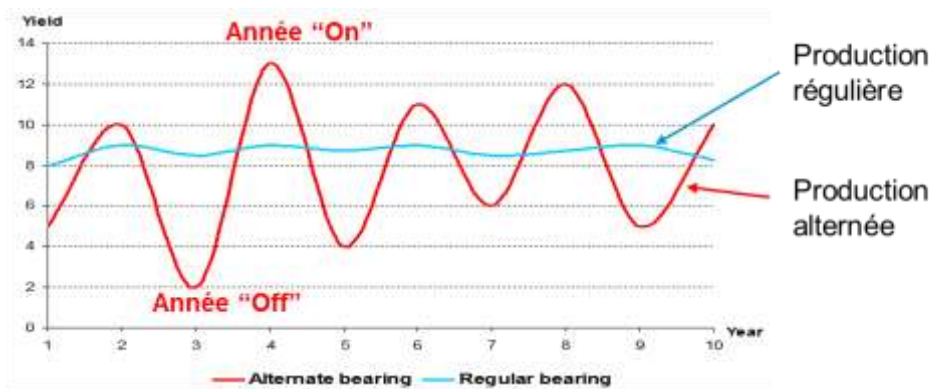


Figure 2 : Représentation de l'alternance de production

Le rendement en fruits d'un arbre peut donc être très élevé une année et très faible l'année suivante. Ces variations ont un fort impact sur le travail ainsi que les revenus de l'arboriculteur. Les années « ON », l'agriculteur a une charge de travail très importante tant au moment de l'éclaircissage que de la récolte notamment, mais il obtiendra une grande quantité de fruits. En revanche, durant les années où ses arbres seront en « OFF », l'agriculteur ne disposera que d'une faible quantité de fruits sur ces mêmes arbres, si prolifiques l'an passé. Ces différences de revenus et de temps de travail entre les exercices comptables peuvent fragiliser une exploitation agricole sur le plan financier et ainsi mettre les exploitants en difficulté certaines années quand ils disposent d'une quantité trop faible de fruits pour répondre à la demande de leurs clients. Par ailleurs, il est compliqué de prévoir à l'avance la masse de travail qui sera demandée afin d'embaucher le nombre de saisonniers adéquat. Inversement, si le marché est inondé par le produit du fait d'une production supérieure à la demande globale, son prix en sera minoré.

L'expression phénotypique de l'alternance résulte en partie des conditions climatiques et environnementales ainsi que de la présence de plusieurs gènes multialléliques sur plusieurs chromosomes (Segura et al., 2006). Il s'agit donc d'un caractère complexe. Une pléiade de gènes et d'allèles peuvent jouer un rôle dans ce processus d'alternance. En effet, la production de fruits d'un arbre va faire intervenir diverses enzymes, hormones et molécules pouvant avoir des fonctions physiologiques en apparence éloignées du mécanisme étudié. A cela s'ajoute un effet du climat et de l'environnement en général plus ou moins prononcé suivant la variété.

### 3.2. Les techniques culturales à disposition pour maîtriser l'alternance

Pour maîtriser ce phénomène d'alternance de production, il existe plusieurs techniques à court terme, consistant à jouer sur des systèmes culturaux adaptés. Cela peut se traduire par une taille spécifique de l'arbre ou l'utilisation d'hormones par exemple, mais ces méthodes doivent être renouvelées chaque année pour être efficaces.

L'éclaircissage, en plus de permettre d'obtenir des fruits de plus gros calibres et de meilleure qualité, tant visuelle que gustative, va favoriser une production de fruit régulière sur l'arbre. En matière d'éclaircissage favorisant un retour des fruits régulier d'une année sur l'autre, la technique de l'extinction est pratiquée (IFPC, 2012). Il s'agit de retirer les coursonnes sur les branches en préférant celles qui sont les plus éloignées du tronc de l'arbre, situées en bout de branche et plutôt sur la partie supérieure de la branche. L'extinction favorise ainsi l'entrée de la lumière dans l'arbre, la croissance des extrémités et le développement de bourgeons placés de façon optimale dans l'arbre. Cette étape peut être partiellement mécanisée et est bien adaptée aux conduites des vergers en agriculture biologique (Larrieu, 2007).

Cette pratique s'accompagne généralement d'un éclaircissage chimique. Des produits composés de différents principes actifs sont à disposition des arboriculteurs. Des hormones de synthèse du type cytokinine et favorisant la chute des bourgeons les plus fragiles peuvent être pulvérisées sur les arbres. D'autres molécules peuvent être utilisées, elles vont être dégradées en éthylène après absorption par l'arbre et cela va provoquer la chute précoce de certains fruits (OMAG SAP., 2011).

L'association de ces deux types de méthodes (mécanique et chimique) permet une bonne maîtrise de la charge mais reste très chronophage et coûteuse en produits chimiques.

A long et moyen terme, la maîtrise de ce phénomène passe par la sélection de variétés présentant le caractère de la régularité de production. C'est en amont de cette sélection qu'intervient l'équipe AFEF. En décortiquant les zones du génome impliquées, des gènes d'intérêts agronomiques détectés pourront ensuite éventuellement être utilisés dans un programme de sélection génomique. Sur les arbres fruitiers et les espèces végétales pérennes, ce processus est plus compliqué que sur les plantes annuelles comme les Poacées ou les Fabacées car beaucoup plus long. Cela peut prendre des années voire des dizaines d'années pour parvenir à une nouvelle variété présentant les expressions phénotypiques recherchées sans pour autant avoir introduit de caractère indésirable. En effet, et spécialement pour le caractère de la régularité de production, il est nécessaire d'attendre non seulement que l'arbre ait atteint l'âge de donner des fruits mais encore plusieurs années pour dépasser l'effet ontogénique. Cet effet correspond à la croissance naturelle de l'arbre, durant plusieurs années, sa production va augmenter régulièrement (Segura et al., 2008). Ainsi, pour caractériser au mieux la fluctuation de production d'un individu, l'optimal est de se fier aux données récoltées sur le plus grand nombre d'années productives possible. Pour cette raison, la sélection d'arbres réguliers est encore très peu enclenchée.

Bien que le pommier et l'olivier soient deux espèces appartenant à deux familles bien distinctes, respectivement les familles des Rosacées et des Oléacées, ces deux arbres fruitiers ont en commun ce caractère agronomique d'alternance de production.

#### **4. Deux espèces : Pommier et Olivier**

##### **4.1. Le Pommier**

###### **4.1.1. Histoire**

Le pommier, scientifiquement connu sous le nom de *Malus domestica* Borkh, est une espèce végétale pérenne et cultivée dans les régions tempérées du monde. Son fruit, la pomme, a toujours eu une place particulière dans l'imaginaire collectif. En effet, la pomme dans la Bible représente le fruit défendu.

Le foyer des pommiers sauvages, situé dans une zone reculée du Kazakhstan, au cœur des montagnes du Tian Shan, fut découvert au début du siècle dernier par le professeur kazakh Aymak Djangaliev. Ce dernier a ensuite voué sa vie à la conservation du berceau de la biodiversité des pommiers (Peix, 2010). Dans ces montagnes préservées pousse l'espèce de pommier sauvage *Malus sieversii*, dans une forêt s'étendant sur plusieurs kilomètres carrés. Certains de ces pommiers sauvages produisent de grosses pommes semblables à celles de nos pommiers cultivés et présentent la caractéristique recherchée de résister à la tavelure. Ces pommiers sauvages sont à l'origine des pommiers domestiques que nous connaissons aujourd'hui. En effet, au travers des premiers voyages le long de la route de la soie, leurs graines ont été transportées et se sont modifiées peu à peu. L'analyse de leurs caractéristiques génétiques a prouvé qu'ils avaient majoritairement contribué au pool génétique du pommier commun (Forsline et Aldwinckle, 2003).

###### **4.1.2. Le marché de la pomme, en France et à l'étranger**

La pomme est l'un des fruits les plus produits et les plus consommés au monde. Chaque année, l'Europe, troisième producteur mondial, produit une quantité de plus de 9.8 millions de pommes. Rien qu'en France, 1.8 millions sont produites et la pomme représente plus de 20% des parts de marché de la consommation de fruits (FNPF, 2012). En moyenne 90% des ménages français en consomment régulièrement. En termes de surfaces allouées à la culture de pommier, 40 000 ha sont cultivés sur le territoire français.

Si la pomme est l'un des fruits les plus consommés et cultivés des régions tempérées c'est en grande partie grâce à sa facilité de propagation et à la qualité de ses fruits (Pereira-Lorenzo et al., 2009). Les pommes sont considérées comme étant des fruits particulièrement bons pour la santé, il est même recommandé de manger au moins une pomme par jour. Un proverbe anglo-saxon résume bien cette idée : « An apple a day keeps the doctor away<sup>1</sup> ». En effet, ce fruit est riche en polyphénols antioxydants et en fibres en plus d'être peu

---

<sup>1</sup> Une pomme par jour tient le médecin à distance

calorique. La pomme est reconnue pour faire chuter le mauvais cholestérol et réduire le risque de diabète de type II (Nathan, 2010).

Il existe plus de 9500 variétés de pommes domestiques, cependant, une dizaine seulement se disputent le marché mondial. Partout autour du monde, on peut retrouver sur les étals des primeurs les variétés Golden Delicious, Fuji, Gala, Granny Smith et Braeburn, pour ne citer que celles-ci (Rouèche, 1997).

#### 4.1.3. Caractéristiques botaniques

Le pommier cultivé est une espèce pérenne du genre *Malus*. Il produit l'un des fruits les plus cultivés dans les régions tempérées et des plus consommés dans le monde entier. Cependant, le pommier reste une espèce rustique qui peut supporter des températures allant jusqu'à -17°C suivant la variété (Landouer, 2015) Un pommier peut vivre plus de 100 ans sur son propre pied. En revanche, en verger, sur un porte-greffe, le pommier est remplacé à 15 ans généralement, lorsque sa production diminue.

Cette espèce angiosperme se reproduit par pollinisation croisée car elle est autostérile, en d'autres termes, le pollen qui provient d'un individu ne peut pas féconder ce même individu. Etant entomogame, ce sont les insectes butineurs qui sont chargés de transporter le pollen d'un arbre à un autre. D'un point de vue anatomique, les fleurs du pommier, de couleur blanche ou rosée, sont composées de 5 pétales et 5 sépales. Toutes les fleurs contiennent à la fois des organes mâles (étamines portant le pollen) et femelles (gynécées renfermant l'ovaire), elles sont donc hermaphrodites. La floraison du pommier commun s'étend de mai à juin (Tela-Botanica, 2012). La récolte s'étend du début d'été à l'hiver suivant les variétés.

#### 4.1.4. Recherches menées sur le pommier

De nombreuses équipes, en France et dans le monde, s'intéressent au pommier. Etant l'un des fruits les plus consommés, sa compréhension par les chercheurs afin d'induire l'évolution de pratiques culturelles ou des variétés est primordiale.

Des cartes génétiques ont d'ores et déjà été réalisées pour plusieurs variétés et suivant plusieurs méthodes. De plus, le génome du pommier de la variété Golden Delicious a été entièrement séquencé (Velasco et al., 2010) ce qui permet de donner une référence physique aux cartes et aux zones repérées sur ces cartes.

Dans l'équipe AFEF, des études sont menées sur différents aspects du développement du pommier. Ainsi, l'utilisation des ressources en eau par la plante, sa croissance et son architecture au sein d'une même population sont des paramètres étudiés. Ces points d'intérêts sont également regardés avec un angle plus génétique de manière à saisir les différences entre populations et entre individus d'une même population.

Des programmes de sélection sont également mis en place en France et à l'étranger pour le pommier afin de proposer sur le marché de nouvelles pommes présentant des caractères d'intérêts, gustatifs ou agronomiques. En France, c'est à la station INRA d'Angers que se font



les recherches en sélection sur le pommier (Laurens, 2013). C'est d'ailleurs là-bas que la pomme « Ariane », résistante à la tavelure, et permettant donc de réduire les traitements phytosanitaires contre la tavelure en verger, a vu le jour.

#### 4.2. L'olivier

##### 4.2.1. Histoire de l'olivier

L'olivier commun, scientifiquement nommé *Olea europaea*, est une espèce fruitière de la famille des Oléacées, de l'ordre des Lamiales plus généralement. Originaire d'Europe méridionale ainsi que du Proche-Orient (Landouer, 2015), des oliviers sauvages, aussi appelés oléastres, s'y développaient déjà 60 000 ans avant J-C. Cette espèce est aujourd'hui présente sur tout le pourtour méditerranéen et dans le monde entier, là où le climat lui est favorable. En France, des traces de cette culture ont été retrouvées dans les Pyrénées Orientales datant de 5500 ans avant Jésus Christ (Afidol, 2014).

L'Antiquité fut l'âge d'or de l'olivier. L'huile d'olive était alors un produit phare du commerce international. Utilisée pour l'énergie, l'alimentation, au cours des rites chrétiens et dans la fabrication de cosmétiques notamment, la demande pour cette marchandise était très forte. Au Moyen-âge, l'huile d'olive était encore largement utilisée, mais principalement par l'Eglise et les familles les plus aisées. Cette huile était alors rare et chère, considérée comme très précieuse. Les lampes à huile sur les autels n'étaient alimentées que par cette huile, suivant les instructions des Saintes Ecritures (Museo dell'Olivo, 2012).

Au XXème siècle, la découverte de l'électricité ainsi que celle d'huiles végétales moins chères issues de graines comme l'huile de tournesol, ou de colza, éclipsent pour un temps l'huile d'olive qui est plus chère et est alors de moins en moins consommée. D'autres espèces fruitières comme la vigne ou les pêchers viennent remplacer les oliviers dans les vergers de la zone méditerranéenne.

Ce n'est qu'au cours des années 70 que les bienfaits de l'huile d'olive dans la gastronomie méditerranéenne sont mis en évidence. La consommation d'huile d'olive au quotidien dans la cuisine préviendrait notamment les risques cardiovasculaires en limitant la formation de mauvais cholestérol et son accumulation dans les artères notamment (Horrocks et Yeo, 1999).

##### 4.2.2. Le marché de l'olive en France et à l'étranger

Plus de 3 098 000 tonnes d'huile d'olive sont produites chaque année dans le monde dont plus de 70% en Europe, soit 2 180 500 tonnes (France Agrimer, 2012). Avec une production d'huile d'olive dépassant le million de tonnes par an soit environs 36%, l'Espagne en est le premier producteur mondial devant l'Italie qui le talonne à 20% de la production mondiale.

C'est sur le pourtour méditerranéen que se trouvent 90% des pays producteurs. Cependant, la culture d'olivier, dont la demande se développe de plus en plus, s'est finalement implantée aux Etats-Unis ainsi qu'en Amérique du Sud et au Moyen Orient.

En termes de consommation d'huile, l'Europe, à elle seule, représente le premier marché mondial avec plus de la moitié de l'huile d'olive produite qui y est consommée.

A la production d'huile d'olive s'ajoute la production mondiale de 2.5 millions de tonnes d'olives de table en 2012/2013 (Conseil Oléicole Internationale, 2014). Cette utilisation des olives demande un travail important car la récolte n'est pas automatisée et doit se faire à la main.

#### 4.2.3. Physiologie de l'olivier

L'olivier, très rustique, s'adapte à des climats extrêmes mais supporte mal l'humidité. Cet arbre présente une longévité exceptionnelle, les plus vieux spécimens pouvant être âgés de plus de deux millénaires. La floraison est induite en mai. 2% des fleurs environ vont donner des fruits. L'olivier est une espèce anémogame, ce ne sont pas les insectes qui seront en charge du transport du pollen mais bien le vent (Pesson et Louveaux, 1984). Pour cette raison, les fleurs d'olivier sont petites, légères, en grand nombre afin de pouvoir se déplacer loin au gré du vent et peu colorées car elles n'ont pas pour vocation d'attirer les insectes butineurs. Les fruits muriront au cours de l'été et seront récoltés en automne pour les olives vertes et en novembre voire décembre pour les olives noires qui auront acquis cette couleur sombre en parvenant à maturité.

La croissance de l'olivier est très lente même si celui-ci peut atteindre jusqu'à 15m de hauteur.

#### 4.2.4. Recherches menées sur l'olivier

Contrairement au pommier, l'olivier ne dispose pas encore du séquençage complet de son génome. Celui-ci est en cours de réalisation, mené par un consortium international mais devrait se poursuivre jusqu'en 2016 (CSIC, 2013).

L'expansion du marché de l'olivier dans le monde entier ajoute un nouveau défi. En effet, l'huile d'olive est considérée comme un produit d'exception et sa popularité grandissante pose des questions de traçabilité. De plus en plus, des problèmes qualitatifs et des tromperies sont recensés. Il peut s'agir d'huile d'olive coupée avec une huile moins chère, des huiles rances désodorisées sont également mises sur le marché et un consommateur néophyte ne verra pas l'escroquerie. Des recherches sont donc menées en ce sens, l'Union Européenne a lancé un appel à projets pour prévoir des mesures de traçabilité et d'authentification de l'huile d'olive d'ici à 2020 (Commission Européenne, 2013).

La recherche sur l'olivier portant sur la sélection assistée par marqueurs se développe depuis plusieurs années et notamment afin d'obtenir des variétés un peu plus précoces (León et al.,

2007). D'autres équipes travaillent sur des programmes de sélection pour adapter les réponses des oliviers aux nouvelles techniques culturales et à la mécanisation mises en place (Lavee et al., 2014). Des cartes génétiques existent également pour l'olivier (La Rosa et al., 2003 ; Wu et al., 2004 ; El Aabidine et al., 2009). De plus, des recherches sont menées dans le but de caractériser les ressources génétiques de l'espèce à l'aide de marqueurs moléculaires (Gomes et al., 2012). Cependant, l'olivier étant une espèce un peu plus marginalement étudiée que le pommier, un nombre moins important de données est répertorié

Dans l'équipe AFEF, ce sont l'architecture de l'arbre ainsi que ses données de production et de rendement qui sont étudiées par des aspects génétiques.

Les études antérieures liées à ces deux espèces présentant le caractère commun de l'alternance de production ont amené des réflexions sur ce sujet et ont permis de définir les problématiques traitées dans la partie II.

## Partie II : Problématiques

Le phénomène de l'alternance de production a un impact sur la production annuelle de nombreux arboriculteurs. Des techniques culturales sont à disposition pour limiter l'expression de ce caractère mais des études plus approfondies s'appuyant sur des éléments de génétique pourraient permettre de maîtriser durablement et en amont l'alternance de production.

Depuis le milieu des années 2000, l'équipe AFEF conduit des recherches sur la régularité de production du pommier et de l'olivier. Cette thématique d'étude s'est traduite, dès 2008, par le lancement d'une thèse menée par Baptiste Guitton concernant le contrôle génétique de l'alternance chez le pommier sur la population « Starkrimson x Granny Smith » (Guitton, 2011). La démarche s'est poursuivie en 2010 par le commencement de la thèse d'Inès Ben Sadok portant sur l'olivier et abordant notamment le rendement et la variation de production d'une année sur l'autre de cet arbre symbolique de la Méditerranée (Ben Sadok, 2013). Le croisement étudié, « Olivière x Arbequina », est également celui qui a été étudié dans le cadre du stage que j'ai réalisé. La thèse d'Alix Allard, débutée en 2012, est l'étude génétique de plusieurs populations de pommier présentant un lien de parenté et dont fait partie « X3263 x Belrène » qui est aussi le matériel utilisé pour ce mémoire. Plusieurs caractères sont étudiés pour chacun de ces croisements de manière à pouvoir comparer les résultats et obtenir des données précises quant aux zones génomiques impliquées pour le même caractère chez les différentes populations.

C'est dans ce contexte d'étude suivie pour la compréhension en profondeur de ce caractère phénotypique d'alternance chez ces deux espèces, le pommier et l'olivier, que ce stage de six mois a pu être mis en place. Dans un premier temps, le sujet a été proposé au GIS Fruits qui, pour la première fois en 2014, souhaitait financer des stages d'intérêts pour ses partenaires. Dans cette optique, trois partenaires du GIS au minimum devaient apporter leur collaboration au projet. Ici, l'INRA, Montpellier SupAgro ainsi que le Conservatoire Botanique National méditerranéen de Porquerolles se sont rassemblés pour mettre ce sujet à l'étude. Ce stage a donc visé à s'intéresser à un enjeu économique important pour les arboriculteurs ainsi que pour toute la filière fruitière en France. Le GIS Fruits, en plus du soutien financier apporté au stage de master 2, a pu permettre la mise en commun de compétences relatives au sujet.

Ce stage de fin d'étude proposé par l'INRA dans l'équipe Architecture et Fonctionnement des Espèces Fruitières s'inscrit dans une étude de longue durée. Ce travail est positionné dans le cadre d'une présélection de gènes candidats sur l'olivier et le pommier et pourra éventuellement aboutir sur la sélection de variétés d'intérêts, à la production régulière. Effectuer des analyses sur ces deux espèces distinctes permettra par la suite de les comparer pour ces caractères de rendement et d'alternance en partant de bases communes.

La mission portée par ce stage consiste donc à détecter les QTLs (Quantitative Trait Loci) intervenant dans le processus de variation de la production suivant les années et ceci dans le

but de repérer des gènes d'intérêts agronomiques impliqués. Ces gènes pourront ensuite, éventuellement et dans une étude ultérieure, être sélectionnés ou contre-sélectionnés afin de créer des variétés ayant une production régulière en « ON ».

Ce travail de présélection est nécessaire avant tout commencement d'une sélection génétique. La compréhension des mécanismes liés au caractère pourra alors accélérer le processus en ne sélectionnant que les individus présentant les zones repérées lors de la présélection comme jouant un rôle dans l'alternance de production.

L'élaboration de nouvelles cartes génétiques ainsi que la détection de QTLs liés à l'alternance va, non seulement contribuer à l'avancement des connaissances scientifiques dans le domaine de la génétique de ces deux espèces cultivées, mais également permettre de valider et préciser des résultats déjà obtenus il y a quelques années, avec de plus anciens outils.

Plusieurs questions s'imposent ainsi, dans la lignée des travaux précédents, des projets de recherches à venir ainsi que pour répondre aux besoins des arboriculteurs et des acteurs de la filière fruitière.

Retrouve-t-on les QTLs liés à l'alternance détectés sur la population « Starkrimson x Granny » sur d'autres populations ?

Retrouve-t-on les QTLs du rendement déjà détectés sur « Olivière x Arbequina » sur la nouvelle carte développée, saturée en marqueurs ?

Va-t-on également trouver des QTLs liés à l'alternance chez l'olivier ?

Les nouvelles cartes génétiques saturées vont-elles permettre de détecter plus de QTLs pour les deux espèces ?



## Partie III : Matériel et Méthodes



## 1. Schéma méthodologique

La figure 3 ci-dessous représente la méthode suivie dans l'objectif de parvenir aux résultats escomptés. Les parties grisées correspondent à des étapes qui interviennent dans la méthodologie, en amont de ce stage, et effectuées par les techniciens de l'équipe AFEF depuis 2005. Les parties blanches entourées d'un liseré vert ont été accomplies au cours de la mission de ce stage.

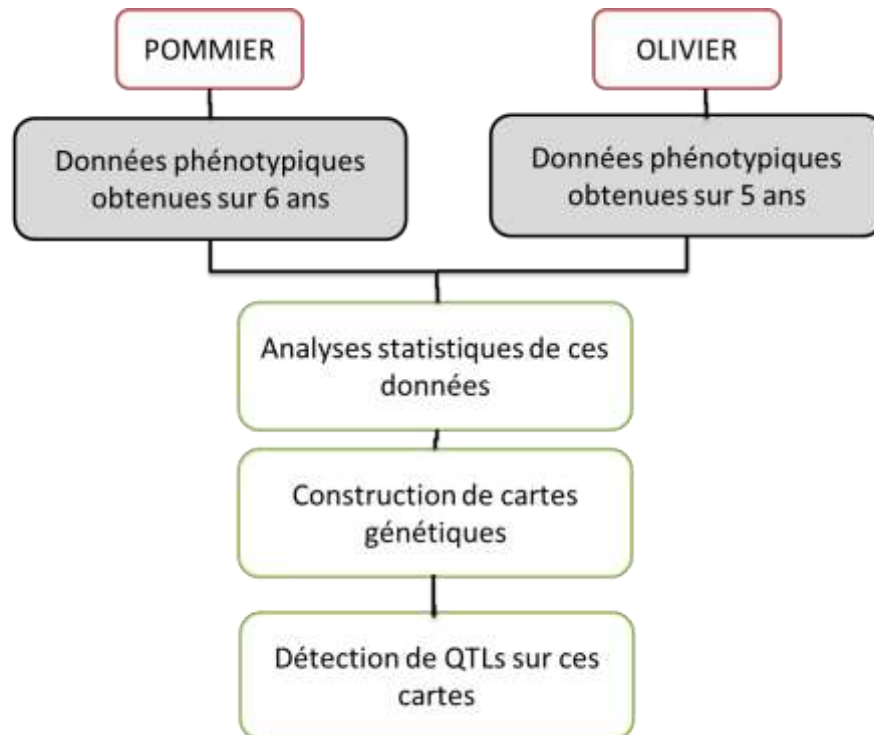


Figure 3: Schéma méthodologique du projet

## 2. Les populations étudiées

### 2.1. Caractéristiques de la population de pommier

La population en ségrégation biparentale étudiée « X3263 x Belrène » est un croisement entre deux variétés de pommiers. Le parent femelle est X3263 et le parent mâle pollinisateur est le cultivar « Belrène ». Tous les individus génotypés appartiennent donc à une seule et même fratrie (full-sib family). Il s'agit d'une population de type CP (Cross-Pollinated) qui désigne la population issue du croisement entre deux génotypes distincts. La population de pommier « X3263 x Belrène » comporte 322 individus dont 50 ont deux arbres par génotype. On a donc 272 génotypes uniques sur toute la population. La décision de faire deux arbres par génotype pour uniquement 50 génotypes parmi les 272 est justifiée par l'importance de

disposer de plus de génotypes distincts pour une meilleure résolution de l'étude. La population est plantée à l'unité expérimentale Diascope de Melgueil, non loin de Mauguio, dans la région Languedoc Roussillon (INRA, 2009).

Tous les individus de cette population croisée, les répétitions incluses, ont été greffés sur le porte-greffe « Pajam 1 » en 2005 et plantés dans des conditions semblables. Pajam 1, aussi nommé Paradis jaune de Metz 1 (Lancep), est un porte-greffe peu vigoureux mais productif et nanifiant. L'arbre greffé sur Pajam 1 peut atteindre environ deux mètres de hauteur. Son système racinaire est, en revanche, plutôt faible et nécessite un support pour l'arbre. D'ordinaire, sa durée de vie se situe autour de 25 ans mais son entrée en production est rapide (Dumas, 2014).

Le parent pollinisateur, Belrène, est une variété enregistrée en 1970 et issue de mutations du cultivar Reine des Reinettes (Spiegel-Roy, 1990). Elle a été obtenue par mutation chimique en exposant son parent à des doses d'hormones de croissance EMS (Éthyl Méthane Sulfonate) ayant induit des mutations (IAEA, 2010). Par rapport à la variété Reine des Reinettes, son mutant Belrène est caractérisé par la précocité de sa première mise à fruit, ses gros fruits très colorés et son rendement plus faible. Elle présente également une forte tendance à l'alternance de production.

Le parent femelle, X3263, est lui-même un hybride conçu à la station expérimentale de l'INRA d'Angers (Celton, 2011). C'est un croisement entre la variété « Red Winter » et « X3177 ». Cette dernière est également une variété hybride, croisement opéré à l'INRA d'Angers entre les cultivars « Idared » et « Prima ». A l'inverse de la variété Belrène, X3263 a une production de fruits relativement régulière tous les ans (Parveaud et al., 2010).

Aucune intervention de taille n'a eu lieu sur cette parcelle. Les arbres poussent sans être taillés, afin d'imiter leurs conditions naturelles de développement et ainsi de pouvoir étudier leur architecture en conditions naturelles, sans l'intervention humaine. Seuls des traitements phytosanitaires sont effectués sur ces populations. La première mise à fruit des pommiers plantés en 2005 a eu lieu après trois ans. En 2008, des données phénotypiques concernant le rendement des arbres sont obtenues sur la parcelle.

Les données phénotypiques ont été recueillies, depuis 2008 et tous les ans jusqu'à aujourd'hui, à la station expérimentale Diascope du domaine de Melgueil à Mauguio. Pour chaque arbre, le nombre de fruits sur l'arbre et le nombre de fruits au sol ont été comptés. Puis, les fruits ont été pesés, ce qui a permis de déduire le poids de fruits par arbre.

Le nombre total de fruits par arbre et par année de 2008 à 2014 représente donc la donnée majeure pour estimer l'alternance au sein d'une population en ségrégation bi-parentale.



Figure 4: Population "X3263 x Belrène"

## 2.2. Caractéristiques de la population d'olivier

Le croisement entre les deux variétés d'olivier, Olivière (parent femelle) et Arbequina (parent pollinisateur mâle) est planté au même endroit que la population de pommier « X3263 x Belrène », à la station expérimentale de Melgueil à Mauguio.

Olivière est une variété dont les organes reproducteurs mâles sont stériles, son pollen n'est pas fonctionnel (Besnard et al., 2000). Originaire des Pyrénées Orientales, c'est l'une des variétés les plus cultivées en France. En effet, à croissance rapide et très productive, elle donne une huile de qualité et très aromatique. Ses olives peuvent également être consommées comme olives de table quand elles sont noires mais la production est majoritairement transformée en huile. Olivière est peu sujette à l'alternance de production lorsqu'elle est bien entretenue et suffisamment taillée (Moutier et al., 2004).

La variété Arbequina, utilisée ici comme pollinisatrice, est une variété d'origine espagnole, catalane plus précisément, et dont les fruits donnent une huile douce dans laquelle on retrouve des arômes de pommes mûres et de coings. Cette variété est peu vigoureuse mais se caractérise par une production de fruits plutôt régulière d'une année sur l'autre.

Le croisement de ces deux variétés a produit 240 arbres dont 160 ont été plantés à raison de six répétitions par génotype sur la parcelle d'expérimentation en mai 2005 selon six blocs randomisés.

En 2010 cependant, un arbre sur deux a été déplanté de manière à laisser aux individus restants un espace suffisant pour se développer. Le dispositif expérimental est donc composé de trois répétitions pour chaque génotype, réparties sur les six blocs.

Les oliviers de la population « Olivière x Arbequina » présents sur cette parcelle ne sont pas taillés, tout comme les pommiers du croisement « X3263 x Belrène » et ce pour les mêmes raisons. L'architecture de l'arbre et son développement sans l'intervention de taille y sont étudiés. Cependant, plantés il y a une décennie, ces oliviers ont à présent des branches et un

feuillage denses qui rendent très difficile la récolte des fruits à des fins scientifiques pour déterminer le rendement des arbres. Il a donc été décidé de tailler une première fois cette population de manière à éclaircir son feuillage et permettre la récolte de ses fruits à partir de l'automne 2015.

Le rendement de cette parcelle d'oliviers n'a pu être déterminé que sur 6 ans. En effet, l'année 2014 a été marquée par une attaque de mouches sur les fruits encore immatures. Toute la récolte de l'unité expérimentale a été perdue, tout comme pour une très grande partie des olives des oléiculteurs français (Le Point, 2014).



Figure 5 : Variété "Olivière"

X



Figure 6 : Variété "Arbequina"

### **3. Les données génétiques pour ces deux espèces**

Plusieurs types d'informations génétiques sont à disposition pour étudier ces populations de pommiers et d'oliviers.

#### **3.1. Les marqueurs SNP**

##### **3.1.1. Définition**

Un marqueur SNP (Single Nucleotide Polymorphism) correspond à un point de polymorphisme nucléotidique sur l'ADN qui est une structure à double hélice, composée de bases nucléotidiques reliées entre elles. Ces bases azotées qui fonctionnent par paires se succèdent sur toute la longueur de l'ADN. L'Adénine et la Guanine ainsi que la Thymine et la Cytosine constituent ces paires. Le polymorphisme nucléotidique est le possible échange de ces bases en un point précis sur l'ADN. Ce sont ces différences entre individus qui sont mises en évidence par les marqueurs SNP (Gnis, 2014).

### 3.1.2. Les marqueurs sur la population issue du croisement X3263 x Belrène

Les données SNP ont été obtenues dans le cadre du programme de recherche FruitBreedomics (Projet européen, 2010). Ce projet, porté par un consortium international composé de 26 membres, vise à développer des méthodes d'amélioration des espèces fruitières en utilisant les dernières avancées technologiques.

Le séquençage a été effectué grâce à une puce 9K (9000 marqueurs positionnés) sur 175 descendants de la population hybride issue du croisement entre les variétés X3263 et Belrène. Ce séquençage a été réalisé grâce à la technologie Infinium II de la société Illumina (Illumina, 2006). La puce a été construite pour identifier des marqueurs à intervalles réguliers sur le génome, et ainsi le couvrir au maximum sans laisser de « gaps » (grands espaces sur le chromosome non marqué, où l'information manque). La fréquence de l'allèle le moins représenté est comprise entre 0.1 et 0.5. Le but de l'utilisation de cette technique est de couvrir des zones où se trouvent des gènes candidats (Durel, 2012). Par cette méthode de détection par puce, il y a peu de données manquantes pour chaque individu.

Après avoir filtré la matrice pour n'obtenir que les SNP les plus fiables, il reste 2649 marqueurs SNP qui serviront à élaborer des cartes génétiques représentant les 17 chromosomes du pommier.

### 3.1.3. Les marqueurs pour Olivière x Arbequina

Les marqueurs SNP ont été détectés pour cette espèce dans le cadre du programme ARCAD, Agropolis Resource Centre for Crop Conservation, Adaptation and Diversity (Arcad, 2010). Le projet porté par ce programme consiste à étudier l'histoire évolutive des espèces méditerranéennes et tropicales, leur domestication ainsi que les possibilités d'adaptation de ces plantes aux nouveaux enjeux climatiques. Douze espèces sont étudiées en parallèle dans ce cadre dont les pérennes comme la vigne, l'olivier ainsi que l'hévéa.

Les marqueurs ici utilisés ont été obtenus après séquençage du transcriptome de l'olivier. Le transcriptome correspond à des séquences de l'ADN transcrites à partir des ARNs messagers. Il ne comporte que les séquences codantes, plus précisément des exons des gènes. Les séquences non codantes ne sont alors pas marquées.

Les données de séquence du transcriptome de l'olivier ont été obtenues par séquençage selon les nouveaux procédés de type NGS (Next Generation Sequencing) en utilisant le séquenceur NGS de type HiSeq (Illumina, 2011). Les ARNs messagers ont été extraits à partir de boutons floraux non ouverts en vue de maximiser le nombre de gènes pour le séquençage.

La profondeur utilisée pour détecter ces marqueurs est de 8X. Ainsi, les marqueurs retenus ont été détectés huit fois sur le transcriptome. Cette précision de 8X est un critère de validation du marqueur détecté sur le transcriptome. La profondeur peut varier, plus elle est élevée et plus le séquençage est précis. Une profondeur de 8X est considérée comme un niveau suffisant pour la validation des données SNP obtenues. Cependant, ce niveau de

validation conduit à des données manquantes chez des descendants pour chaque marqueur. Par ailleurs, ces marqueurs étant positionnés uniquement sur les parties codantes de l'ADN, les zones qui codent pour des gènes, ils ne sont pas répartis de manière homogène sur le chromosome.

Les données SNP obtenues à la suite de ce séquençage ont par la suite été filtrées grâce à un pipeline élaboré par l'équipe ID (Intégration de Données). Le premier paramètre à prendre en compte dans le filtre est la longueur de la séquence du transcriptome assemblée. La question était de savoir pour quelle longueur de séquence, il y aurait le plus de marqueurs positionnés. En assemblant des transcrits (ou reads) de 100 paires de bases nucléotidiques, il a été possible de reconstituer des séquences de transcriptome. La limite à la reconstruction a été placée à 700 paires de base. Cette longueur correspond au plus grand nombre de séquences assemblées (Figure 7).

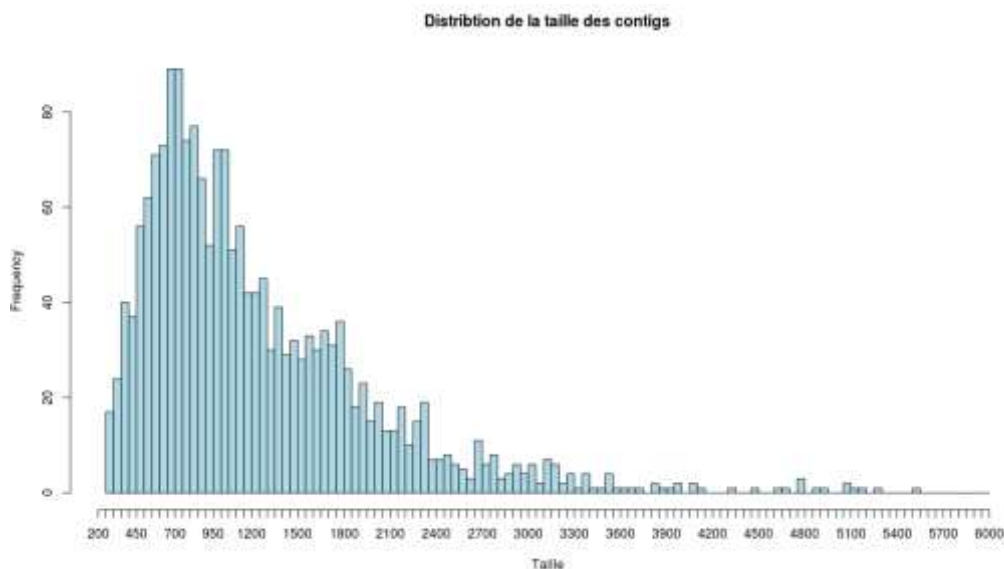


Figure 7 : Histogramme représentant la fréquence associée à chaque taille de séquence de transcriptome

En faisant passer la matrice par le filtre élaboré, tous les marqueurs pour lesquels les données manquaient pour plus de 50% des individus génotypés ont été enlevés.

Un filtre pour retirer les marqueurs présentant une distorsion de ségrégation trop importante a également été mis en place pour nettoyer la matrice des marqueurs suspects.

Après filtrage, il reste dans la matrice 2893 marqueurs SNP ayant un nombre acceptable de données manquantes et qui ne sont pas trop distordus.

### 3.2. Les marqueurs SSR

#### 3.2.1. Définition

Le marqueur SSR, ou marqueur microsatellite, est un court fragment d'ADN dont la séquence est répétée en tandem, 1, 2, 3 ou 4 bases nucléotidiques (Campbell M.A., 2001). Des amorces peuvent amplifier des séquences d'ADN avec un motif microsatellite PCR (Polymerase Chain Reaction). Les fragments d'ADN migrent alors sur un gel d'électrophorèse. Les combinaisons alléliques de ces marqueurs ainsi que leur type de ségrégation, s'ils sont bialléliques, trialléliques ou quadrialéliques, sont détectés sur le gel ou à l'aide d'un séquenceur suivant la taille des fragments d'ADN amplifiés. Ainsi, si pour un individu diploïde, on observe deux fragments sur le gel, il s'agit alors d'un marqueur SSR avec deux allèles.

#### 3.2.2. Leur utilisation

Pour la population de pommier « X3263 x Belrène », les marqueurs microsatellites ont été utilisés pour l'analyse de 186 individus, soit 11 de plus que pour les marqueurs SNP. Concernant le croisement d'olivier « Olivière x Arbequina », 147 arbres sont génotypés à l'aide des marqueurs SSR.

Pour le pommier comme pour l'olivier, l'intérêt d'ajouter des microsatellites aux cartes génétiques faites à partir de marqueurs SNP réside dans la possibilité de comparer ensuite ces nouvelles cartes à d'autres cartes qui intègrent également les mêmes marqueurs SSR. Ils vont permettre de faire le lien entre ces différentes cartes pour la même espèce et notamment d'informer sur le sens de lecture des QTLs sur les groupes de liaisons formés pour la nouvelle cartographie.

## 4. La cartographie génétique

### 4.1. Le logiciel Joinmap 4.1

#### 4.1.1. La matrice d'entrée

Les cartes génétiques sont élaborées grâce au logiciel Joinmap 4.1 (Van Ooijen J.W., 2006). Ce logiciel a été développé par des chercheurs néerlandais à l'université de Wageningen. Une matrice contenant la liste des individus et leur information génétique pour chaque marqueur sert de fichier de base pour le logiciel. Les marqueurs sont utilisés avec leur information génétique pour chaque individu génotypé. Six types de ségrégation des marqueurs sont possibles selon les différents allèles qui composent le locus comme décrit dans la Table 1.

Ainsi, lorsque le locus est bi-allélique et que ce locus est hétérozygote sur le parent femelle, alors le marqueur aura une ségrégation de type <lmxll>. Le parent femelle présente deux allèles différents au total sur ses deux chromosomes. A l'inverse, si cette hétérozygotie intervient chez le parent mâle, la ségrégation sera notée <nnxnp>. Dans le cas où les deux parents sont hétérozygotes mais pour les mêmes allèles, alors le marqueur sera noté <hkxhk>. Lorsque le marqueur est placé sur un gène tri-allélique, les deux parents sont alors hétérozygotes et on parlera d'un marqueur <efxeg>. Enfin, la ségrégation <abxcd> est la plus recherchée car elle donne le plus d'informations génétiques, les deux parents sont hétérozygotes pour ce marqueur et les descendants pourront posséder les allèles « ac », « ad », « bc » et « bd ».

Table 1: Les différents types de ségrégations possibles et leurs génotypes associés

Type de ségrégation	Génotypes possibles	Description
<abxcd>	ac, ad, bc, bd, --	Locus hétérozygote chez les deux parents, quatre allèles
<efxeg>	ee, ef, eg, fg, --	Locus hétérozygote chez les deux parents, trois allèles
<hkxhk>	hh, hk, kk, --	Locus hétérozygote chez les deux parents, deux allèles
<lmxll>	ll, lm, --	Locus hétérozygote chez le premier parent
<nnxnp>	nn, np, --	Locus hétérozygote chez le second parent
<b>Remarques :</b>		
1. Chaque lettre représente un allèle distinct; "-" représente un allèle inconnu		
2. Le logiciel est indifférent à l'ordre des allèles dans le code (hk équivaut à kh)		

#### 4.1.2. Composition de groupes de liaisons

Une fois les combinaisons alléliques de tous les marqueurs pour tous les individus prises en compte dans le logiciel, une fonction de repérage des erreurs est enclenchée pour détecter les possibles erreurs de ségrégation et les fautes de frappe. Elles sont ainsi corrigées avant le début des calculs.

La matrice est alors prête à être séparée par groupes de marqueurs selon leur proximité et jusqu'à un niveau de maximum de vraisemblance (Logarithm Of Odds) qui peut atteindre le LOD 25. Plus ce paramètre est élevé et plus les marqueurs groupés sont liés. Les groupes sont calculés pour chaque palier à partir de LOD 2. Les groupes sont retenus à partir du LOD 4 pour les plus stables et jusqu'au LOD 7.

Cette étape de groupement va permettre de mettre à l'écart les marqueurs les moins pertinents, les plus éloignés des autres, ainsi que d'élaborer des groupes de liaisons supposés correspondre aux chromosomes de l'espèce étudiée. Ainsi, 23 groupes ont été définis pour l'olivier qui compte en effet 23 paires de chromosomes (Figure 8).

Pour le pommier, la matrice de marqueurs a pu être séparée en 17 groupes de liaisons correspondant relativement bien aux groupes de liaisons physiques déterminés grâce au



séquençage complet du génome de la variété de pommier Golden Delicious (Velasco et al., 2010). Ce séquençage comportant environ 25% d'erreur, il était attendu que certains marqueurs n'y correspondent pas tout à fait et se placent sur un autre groupe. Le groupement des marqueurs s'est donc fait en observant la stabilité des groupes de marqueurs supposés être sur le même chromosome jusqu'au LOD 25.

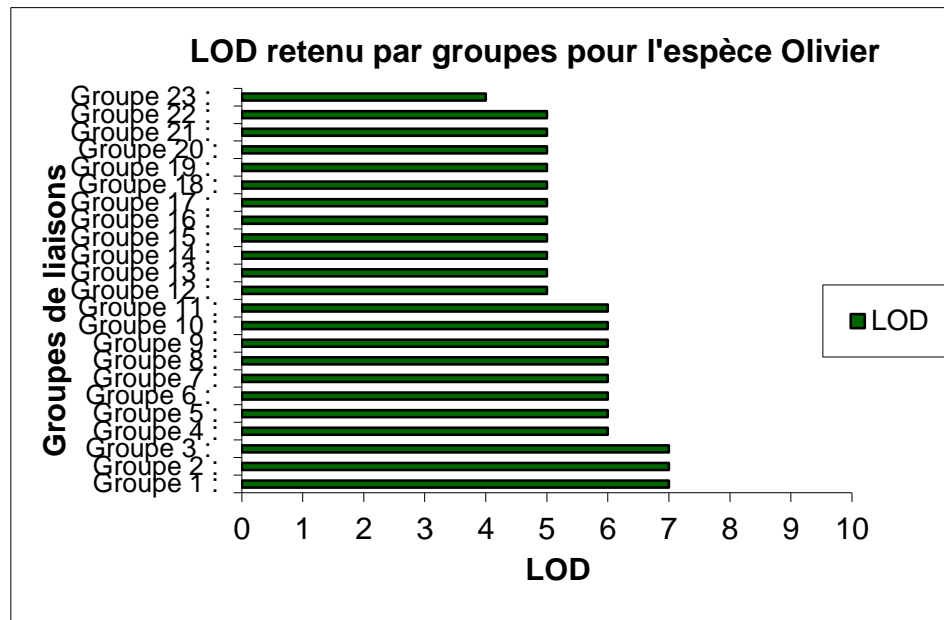


Figure 8 : Niveaux de LOD retenus par groupes pour l'espèce olivier

#### 4.2. Méthode de création de cartes

##### 4.2.1. Carte SNP et carte SSR

Pour élaborer les cartes génétiques des pommiers descendants du croisement « X3263 x Belrène » et pour celles des oliviers issus de « Olivière x Arbequina », la même méthode de cartographie a été appliquée. Une carte contenant les marqueurs SNP et une carte distincte contenant les marqueurs SSR ont été réalisées en premier lieu.

D'après des cartes génétiques effectuées précédemment chez le pommier, les 17 groupes de liaison ont une taille qui varie entre 60 et 90 centimorgans chacune, à l'exception du groupe de liaison 15 qui correspond au chromosome le plus long du génome du pommier et mesure plus de 100 centimorgans (Antanaviciute et al., 2012).

Concernant l'olivier, la taille des chromosomes de cette espèce est bien supérieure à celle des chromosomes du pommier. les groupes de liaison peuvent atteindre 150 cM (Ben Sadok I. et al., 2013).

Les cartes sont alors élaborées groupe par groupe par le logiciel JoinMap. Un algorithme de régression avec la fonction de Kosambi, développée par le mathématicien indien Damodar Dharmananda Kosambi, est appliqué (Kosambi D.D., 1943). Cet algorithme, associé à cette fonction, consiste à calculer, marqueur par marqueur, le taux de recombinaison entre eux.

Cet outil est adapté pour les plantes pérennes ne présentant pas de lignées fixes homozygotes. Trois « rounds », ou tours, sont ainsi calculés pour chaque groupe. Le tour 1 enlève une grande quantité de marqueurs, les moins pertinents. Le second round va alors permettre d'en replacer un certain nombre en fonction de ceux déjà positionnés. Enfin, la troisième carte effectuée lors du troisième tour prend en compte tous les marqueurs sans exception. Cette dernière est à considérer à titre indicatif car le placement de certains marqueurs est forcé par l'algorithme. C'est donc la seconde carte générée par l'algorithme que l'on retient. Un nombre considérable de marqueurs est placé et ce ne sont que les plus pertinents.

Après réalisation de ces étapes, on obtient alors deux jeux de cartes génétiques différents pour les deux espèces, l'un contenant des marqueurs microsatellites et le second contenant des marqueurs SNP.

#### 4.2.2. La carte optimisée

Les matrices contenant respectivement les marqueurs SNP et microsatellites (SSR) ont, tout d'abord, été utilisées séparément dans le logiciel afin de créer deux projets de cartes différents. Créer une carte par type de marqueur va permettre d'avoir une idée de l'ordre des marqueurs ainsi que de la taille de la carte génétique par groupe en situation optimale, lorsque les marqueurs sont similaires.

La carte optimisée correspond à la carte contenant à la fois les marqueurs SNP et les marqueurs SSR imbriqués au milieu des SNP. Les matrices SSR et SNP sont combinées. Pour les individus génotypés uniquement pour les marqueurs microsatellites, on considère des données manquantes notées « -- » pour les SNP.

Après groupement des marqueurs tous confondus en 17 groupes pour le pommier et 23 pour l'olivier comme précédemment, la même méthodologie, avec l'algorithme de régression et la fonction de Kosambi, est appliquée. Chaque groupe de liaison forme une carte ne retenant que les marqueurs qui présentent un réel lien entre eux.

Certaines cartes paraissent artificiellement allongées, allant jusqu'à 180 centiMorgans. Pour résoudre ce problème, les doubles recombinaisons suspectes ont été supprimées et les cartes recalculées de manière à en réduire la taille. Les doubles recombinaisons peuvent correspondre à des recombinaisons multiples entre les chromosomes lors de la méiose mais peuvent aussi être dues à des erreurs. De même, ces doubles recombinaisons sont remplacées par des données manquantes « -- », en blanc dans la Table 2. La couleur bleue représente la partie du chromosome issue de l'un des parents tandis que la couleur orange représente la partie héritée de l'autre parent. Ainsi, chez l'individu X3263 x Belrene\_40, une recombinaison entre les chromosomes 3 des deux parents a eu lieu au niveau des marqueurs GD\_02143\_L3\_PA et R\_10537812\_L3\_PA entre 30.412 et 49.4 cM et est matérialisée par le changement de couleur. En effet, on observe une variation dans la phase.

Ce sont au niveau des marqueurs microsatellites insérés (en rouge dans la Table 2) que se situent souvent les doubles recombinaisons suspectes. Ces erreurs sont repérables en se fiant aux couleurs attribuées aux phases. Ici les bandes bleues au milieu d'un bloc jaune, qui correspondent à un changement de phase peu probable ont donc été remplacées par des données manquantes.

Table 2: Détermination des phases des marqueurs pour deux individus de la population "X3263 x Belrène"

Locs	Ségrégation	Phase	Classification	Position	x3263XBelrene_40	x3263XBelrene_41
R_33128436_L14_PA	<nnxnp>	{-0}	(nn,np)	1.242	nn	np
R_38448471_L3_PA	<nnxnp>	{-0}	(nn,np)	2.513	nn	np
GD_02143_L3_PA	<nnxnp>	{-1}	(nn,np)	30.412	np	nn
R_10537812_L3_PA	<nnxnp>	{-1}	(nn,np)	49.4	nn	nn
R_10504649_L3_PA	<nnxnp>	{-0}	(nn,np)	50.606	np	np
R_10506816_L3_PA	<nnxnp>	{-1}	(nn,np)	51.004	nn	nn
R_9974411_L3_PA	<nnxnp>	{-1}	(nn,np)	54.756	nn	nn
GDsnp01990	<nnxnp>	{-1}	(nn,np)	54.756	nn	nn
R_9950780_L3_PA	<nnxnp>	{-1}	(nn,np)	54.756	nn	nn
R_9469274_L3_PA	<nnxnp>	{-1}	(nn,np)	57.638	nn	nn
R_9104035_L3_PA	<nnxnp>	{-1}	(nn,np)	59.306	nn	nn
R_8499613_L3_PA	<nnxnp>	{-1}	(nn,np)	61.186	nn	nn
Ref_CH03g07_L3	<abxcd>	{-0}	(nn,np)	62	np	--
R_7180717_L3_PA	<nnxnp>	{-1}	(nn,np)	64.537	nn	nn
R_29854787_L3_PA	<nnxnp>	{-1}	(nn,np)	65.466	nn	nn
Ref_GD_SNP00588_L3	<nnxnp>	{-1}	(nn,np)	70	nn	--
R_2659646_L3_PA	<nnxnp>	{-1}	(nn,np)	79.108	nn	nn
R_2604385_L3_PA	<nnxnp>	{-1}	(nn,np)	79.108	nn	nn
R_18474457_L3_PA	<nnxnp>	{-1}	(nn,np)	82.562	nn	np
R_1057933_L3_PA	<nnxnp>	{-1}	(nn,np)	83.138	nn	np
Ref_CH03e03_L3	<abxcd>	{-1}	(nn,np)	83.138	--	--
R_1122860_L3_PA	<nnxnp>	{-0}	(nn,np)	83.138	np	nn
GD_00506_L3_PA	<nnxnp>	{-1}	(nn,np)	84.294	nn	np
R_744263_L3_87_3	<nnxnp>	{-0}	(nn,np)	86.634	np	nn

Cette carte peut être qualifiée d'optimisée car elle ne regroupe que des informations fiables et propres, tous les marqueurs pouvant contenir des erreurs ayant été écartés.

Ainsi, plusieurs méthodes sont mises en œuvre pour nettoyer la matrice de données et obtenir les cartes les plus fiables possibles. En calculant trois cartes, trois « rounds » par groupes de liaisons, l'algorithme de régression opère une première sélection. De plus, un tri des marqueurs est fait manuellement et empiriquement en observant les cartes. Ainsi, des marqueurs anormalement distordus ou qui allongent anormalement la carte, isolés en bout de groupe, peuvent présenter une erreur de génotypage ou avoir été mal groupés. Il est alors préférable de les écarter des calculs pour obtenir une cartographie la plus fiable possible.

#### 4.2.3. La carte élaborée à partir de marqueurs bialléliques

La carte optimisée a été créée grâce à des marqueurs SNP multialléliques reconstruits à partir de marqueurs bialléliques sur le transcriptome.

Une carte composée uniquement de ces marqueurs bialléliques a donc été élaborée de manière à comparer les deux méthodes. La méthode classique est celle utilisant les marqueurs bialléliques, l'autre méthode étant inédite.

La méthodologie est la même que pour la carte optimisée à l'exception des marqueurs utilisés. La matrice d'entrée de joinmap 4.1 contient 7729 marqueurs SNP à deux allèles soit <nnxnp>, <lmxll> et <hkxhk>. Les marqueurs SSR sont également inclus dans cette carte.

Afin de permettre d'autres éventuelles comparaisons entre les cartes élaborées pour l'olivier, deux autres cartographies ont été réalisées et les méthodologies utilisées sont décrites en Annexe II et Annexe III.

### **5. Analyses statistiques appliquées aux données génétiques et phénotypiques des deux populations**

#### 5.1. Le logiciel R

Dans le cadre de ce projet, toutes les données ont été analysées avec le logiciel libre R. Ce programme, créé en 1993 à partir du langage de statistiques S, est développé depuis par une équipe permanente de statisticiens et de programmeurs ainsi que par toute une communauté internationale d'utilisateurs du logiciel (R FOUNDATION, 2015).

Ce logiciel offre donc des possibilités très larges en terme de statistiques. Des paquets très variés de formules statistiques sont disponibles gratuitement sur le site internet du logiciel. Ils permettent l'analyse de données allant des modèles les plus basiques à des modèles très spécifiques tels ceux appliqués aux données génétiques et en passant par des modèles mixtes.

La version du logiciel utilisée pour les statistiques est la version 3.1.1 (BROMAN K.W. et al., 2003). Mise à disposition des utilisateurs le 10 juin 2014, c'est l'une des versions les plus récentes et donc des plus complètes parmi celles sorties à ce jour.

#### 5.2. Statistiques appliquées aux données génétiques

Les matrices contenant les données SNP du pommier et de l'olivier ont été analysées sous R. Il a été possible d'utiliser directement la matrice joinmap grâce à la fonction « readcross » du package qtl de R. L'extension de cette fonction pour Mapqtl et Joinmap a été développée par Thimothée Flutre, chercheur dans l'équipe DAAV (UMR AGAP, INRA de Montpellier).

Cependant, cette fonction récente de « readcross » n'étant pas encore disponible sur R par téléchargement sur internet, il a fallu passer par le cluster de calcul du CIRAD pour accéder à la version expérimentale. Ce cluster est formé de plusieurs ordinateurs indépendants capables de traiter un calcul simultanément et ainsi de dépasser la capacité maximale d'un ordinateur isolé (Migale, 2014).

De plus, le logiciel MobaXterm a permis de générer les graphiques et de faciliter l'utilisation de R sur le cluster de calcul en proposant une interface ergonomique ainsi que des outils pour passer facilement les documents du réseau du CIRAD à l'ordinateur (Mobatek, 2015).

Le taux d'erreur de génotypage, le nombre et la position des données manquantes sur les groupes de liaisons des chromosomes ainsi que des cartes génétiques ont été déterminés par le biais de cet outil statistique. Après avoir calculé le taux d'erreur de génotypage des données, les cartes génétiques ont pu être estimées une nouvelle fois en prenant en compte ce paramètre, de manière à valider la cohérence de la carte. De plus, les taux de recombinaisons des marqueurs par groupe de liaison sont également représentés graphiquement.

### 5.3. Analyses préliminaires sur les données phénotypiques

Ces données ont également été analysées grâce au logiciel libre R, toujours sous la version 3.1.1 et en utilisant R Studio (R Studio, 2014). R Studio est un logiciel qui permet une utilisation optimale de R. En effet, il propose une interface ergonomique permettant de charger les données, écrire et modifier les scripts, lancer les analyses de données et récupérer les résultats, matrices et graphiques, le tout sur une seule et même page.

Tout d'abord, des analyses statistiques préliminaires consistant à connaître les jeux de données ont été effectuées, notamment en calculant des histogrammes de distribution des données ainsi que des box-plots.

La première étape a donc été de vérifier la normalité des données avec un test de Shapiro-Wilk (Shapiro S.S et Wilk, 1965). La normalité étant extrêmement rare dans ce type d'expérimentation, les données du rendement ont été transformées grâce à la fonction racine carrée, en ajoutant une colonne supplémentaire à la matrice, mais n'ont toujours pas atteint la normalité.

Les données ont ensuite pu être séparées en sous-matrices par année afin d'étudier leurs caractéristiques séparément. Par cette méthode, des corrélations positives ou négatives entre les différentes années ont pu être calculées grâce à un test de corrélation en utilisant la fonction de Spearman qui est un test non-paramétrique, plus adapté lorsque les données ne suivent pas la loi normale, comme ici.

Le Levene's test a quant à lui permis de déterminer l'homoscédasticité des variances, autrement dit, si les variances liées à la production de fruits étaient homogènes entre les

années (Levene H., 1960). Ce test est également adapté dans le cas de données non-normales.

#### 5.4. Elaboration de modèles mixtes

##### 5.4.1. Caractéristiques des modèles

L'élaboration de modèles mixtes avec R a été réalisée en utilisant la fonction *lmer* contenue dans le paquet *lme4* (Bates D. et al., 2011). Différents modèles mixtes ont pu être appliqués dans le but de déterminer les paramètres les plus importants et le modèle qui correspond le plus aux données. Ces modèles, regroupant plusieurs paramètres ainsi que leurs interactions ont été calculés en utilisant l'algorithme du REML (Restricted Maximum Likelihood).

Un modèle mixte est un modèle statistique dans lequel sont pris en compte des effets fixes ainsi que des effets aléatoires pour une variable. Les effets fixes correspondent aux paramètres que l'on veut étudier et qui sont les mêmes pour toute la population. A ceux-ci vont s'ajouter des effets aléatoires qui sont liés à la dépendance incertaine de nos données avec d'autres paramètres (Pinheiro J.C. and Bates D.M., 2000).

Le modèle le plus adapté a été choisi en fonction de la valeur de son AIC, Akaike information criterion, et de son BIC, Bayesian information criterion (Burnham K.P. and Anderson D.R., 2004). Plus ces deux indicateurs sont faibles et plus le modèle correspond aux données analysées.

##### 5.4.2. Le modèle retenu

Trois modèles ont ici été testés pour les deux espèces. Le premier considère l'année en effet fixe ainsi que le génotype en effet aléatoire. Dans le second cas, aucun effet fixe n'a été considéré. Enfin, pour le troisième modèle, les années sont prises en compte comme des effets fixes pour la variable du rendement de l'arbre. L'effet génotype ainsi que l'effet de l'interaction entre le génotype et l'année sont incorporés au modèle en tant qu'effets aléatoires.

Après avoir validé ce dernier modèle en observant son AIC et son BIC, la variance  $\gamma$  a ensuite été ajoutée à son tour.

$$Y_{ij} = \mu + A_j + U_{ij}$$

Avec  $Y_{ij}$  : Observation du rendement

$\mu$  : Effet Année en effet fixe

$A_j$  : Effet génotype en effet aléatoire

$U_{ij}$  : Interaction entre l'année et le génotype

### 5.4.3. Calcul de l'héritabilité d'un caractère

Une héritabilité au sens large est calculée pour chaque trait étudié.

$$H^2 = \frac{VG}{VG + \left(\frac{VR}{R}\right)}$$

VG : Variance génétique

VR : Variance phénotypique

R : Le nombre moyen de répétitions par génotype

### 5.5. Calculs des indices de régularités

#### 5.5.1. Indice de régularité de production (BBI)

C'est en calculant le Biennial Bearing Index (BBI) qu'il est possible de caractériser le comportement de production d'un individu au fil des années (Hoblyn et al., 1936). Sa valeur est comprise entre 0 et 1, 0 étant l'indice pour lequel la production est toute à fait régulière et 1 celui pour lequel la production alterne d'une année sur l'autre allant d'un extrême à l'autre.

La formule pour calculer le BBI des descendants issus du croisement « Olivière x Arbequina » prend en compte les données phénotypiques de rendement pour 6 ans, à partir de 2008 et jusqu'à 2013. Quant au pommier, les données ont été comptabilisées sur sept ans, de 2008 à 2014 grâce à la formule suivante :

$$BBI = \frac{\sum_{i=2008}^{2014} (|y_{i+1} - y_i|) / (y_i + y_{i+1})}{n + 1}$$

Avec  $y_i$  : Le rendement en année  $i$

$y_{i+1}$  : Le rendement en année  $i+1$

$n$  : Le nombre d'années récoltées

Pour les génotypes avec deux arbres chacun, c'est la moyenne de la production totale par année qui a été prise en compte.

### 5.5.2. Estimation de la régularité de production à partir des résidus normés

Deux indices ont été développés dans le but de corriger les biais d'estimation imputés au BBI (Durand J-B. et al., 2013). En effet, le BBI peut entraîner une confusion entre la véritable alternance de production d'un génotype et la tendance de production de ce même génotype dans le temps.

Le BBI\_res\_norm permet d'estimer l'alternance de production au sein d'une population tout en normalisant les résidus. Il diffère du BBI commun en cela. Une fois les résidus normalisés, la tendance de production est minimisée pour ne laisser que l'estimation de la régularité de production.

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

Avec  $Y_{g,r,t}$  : Le rendement  $Y$  de la réplication  $r$  du génotype  $g$  à l'année  $t$ .

$\hat{\varepsilon}_{g,r,t}$  : Le résidu empirique de la tendance du modèle en année  $t$ .

$T_{g,r}$  : Nombre de mesures par réplication  $r$  du génotype  $g$

Le BBI\_pas\_norm, quant à lui, associé au BBI\_res\_norm, va donner une estimation de la quantité produite au cours du temps. Plus les écarts de production sont importants d'une année sur l'autre et plus le BBI\_pas\_norm sera élevé.

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

L'indice d'autocorrélation des résidus représente l'écart des résidus avec la courbe de tendance du rendement de l'individu. Auto-corrélés négativement, cela signifie que pour deux années consécutives, les résidus seront situés d'un côté et de l'autre de la droite linéaire de tendance. Positivement, les résidus ne seront pas systématiquement d'un côté et de l'autre de la droite de tendance.

Ces trois indices sont conjointement examinés pour appréhender au mieux le degré d'alternance et de production d'un individu.



## 5.6. Les BLUPs

Les BLUPs, Best Linear Unbiased Prediction, ont été développés en 1950 (Henderson, 1950). Ils sont calculés avec des modèles mixtes en appliquant la fonction *ranef*, diminutif de Random Effect, contenue dans le paquet lme4 de R. La transformation en Blups des différents traits étudiés permet d'estimer les effets aléatoires (Robinson G.K., 1991).

Auparavant, les différents traits étudiés, calculés à l'arbre et non au génotype pour plus de précision, ont dû être transformés grâce à un modèle mixte comprenant l'intercept en effet fixe et le génotype en effet aléatoire avec ce modèle :

$$Y_{ij} = \mu + A_j$$

Avec  $Y_{ij}$  : Observation du rendement

$\mu$  : Intercept en effet fixe

$A_j$  : Effet génotype en effet aléatoire

A partir de ces données ramenées à une seule valeur par génotype, les BLUPs sont calculés pour permettre de détecter des QTLs pour les différents caractères.

## 6. Méthode de détection de QTLs (Quantitative Traits Loci)

### 6.1. Le logiciel MapQTL 6

La détection de QTLs sur les cartes génétiques déjà élaborées est une étape effectuée grâce à MapQTL 6. Ce logiciel a été développé par la même équipe néerlandaise qui a réalisé Joinmap 4.1, utilisé pour la cartographie génétique dans ce projet (Van Ooijen J.W., 2009).

Les données génétiques, les cartes correspondantes ainsi que les données phénotypiques pour les mêmes individus sont rassemblées dans un même projet de MapQTL. La matrice contenant les informations liées au génotype des individus doit être au format « .loc » et le codage sous forme <abxcd>, <efxeg>, <hkhk>, <lmxll> et <nnxnp>. Ces valeurs sont d'ailleurs exportées depuis joinmap 4.1. Les données par marqueur et par individu sont prises en compte dans ce projet QTL par ce fichier. Le fichier des cartes génétiques est quant à lui au format « .map » et informe sur les marqueurs présents dans les différents groupes de liaisons ainsi que sur leur position sur la carte génétique en centiMorgan. Il correspond à la cartographie génétique élaborée sous Joinmap 4.1. Enfin, les données phénotypiques sont en format « .qua ». On y trouve les individus avec l'information brute qui leur correspond ou encore les BLUPs par génotypes calculés grâce aux modèles mixtes et aux scripts R.

Une fois ces 3 fichiers mis dans un même projet, il est possible de lancer des analyses de détection de QTLs.

## 6.2. Analyses sous MapQTL 6

Une analyse d'Interval Mapping est effectuée en premier. Cette analyse permet d'identifier les loci pour lesquels le LOD du caractère étudié est le plus élevé. Un LOD significatif (la valeur pour laquelle le LOD devient significatif va varier selon le caractère et la population) permet de détecter la présence ou non d'un QTL à cet endroit. Des graphiques représentant le niveau de LOD par position en cM sur le groupe sont générés. Afin de savoir à partir de quelle valeur le LOD est significatif, il est nécessaire de lancer un test de permutation. Ce test va indiquer, groupe par groupe et pour l'ensemble du génome quel est le LOD à prendre en compte pour un risque d'erreur de 5%. Une recherche des co-facteurs des QTLs est ensuite effectuée, empiriquement tout d'abord, en notant les marqueurs qui présentent les LODs les plus élevés puis grâce à la fonction « ACS » ou « Automatic Cofactor Selection ». L'intérêt de l'identification de ces cofacteurs, en effectuant un test de rMQM (restricted Multiple QTLs Mapping), est de les retirer des analyses de LOD sur les groupes et ainsi de faire apparaître d'autres QTLs. La sélection de co-facteurs peut être répétée de manière à détecter tous les QTLs. Les QTLs sont définis à LOD-1 et LOD -2 autour du pic de LOD du QTL. Les marqueurs à ces positions sont notés pour définir les bornes du QTL.

## 6.3. Analyses complémentaires sur les QTLs

Pour chaque QTL, les effets alléliques mâles ( $A_m$ ) et femelles ( $A_f$ ) ainsi que la dominance ( $D$ ) ont été calculés avec les formules suivantes :

$$A_f = [(\mu_{ac} + \mu_{ad}) - (\mu_{bc} + \mu_{bd})] / 4$$

$$A_m = [(\mu_{ac} + \mu_{bc}) - (\mu_{ad} + \mu_{bd})] / 4$$

$$D = [(\mu_{ac} + \mu_{bd}) - (\mu_{ad} + \mu_{bc})] / 4$$

$\mu_{ac}$ ,  $\mu_{ad}$ ,  $\mu_{bc}$  et  $\mu_{bd}$  sont les moyennes phénotypiques estimées, associées avec les quatre classes génotypiques possibles : ac, bc, ad et bd. Ces moyennes font partie des informations délivrées lors des analyses effectuées sur MapQTL 6.

## 7. Recherche des gènes candidats

Une fois les QTLs détectés, l'étape suivante a été de rechercher les gènes positionnés sous ces zones génomiques. Pour cela, les bornes des QTLs, qui ont été définies lors de l'analyse par les marqueurs situés au LOD-2 par rapport au pic de LOD, sont entrées dans la base de données Genome Database for Rosaceae pour le pommier (GDR, 2012). Le marqueur est donc entré dans une barre de recherche qui va trouver sa position exacte sur la séquence du génome. Une fois connues les positions des marqueurs servant de bornes pour un QTL, la zone précise de ce QTL sur le chromosome physique est renseignée. Sur cette même base de

données, un outil supplémentaire est à disposition. Ce fichier contient, pour chaque chromosome du pommier, les gènes présents, leur position sur le chromosome et l'organisme sur lequel ils ont été mis en évidence au départ, en l'occurrence, *Arabidopsis thaliana*. De plus, un identifiant est indiqué pour chaque gène avec une courte description de sa fonction si celle-ci est connue.

Ainsi, grâce aux positions précises des QTLs qui ont été trouvées grâce à GDR, toute la séquence de gènes entre ces deux positions et qui peut correspondre au caractère étudié est connue. Les identifiants des annotations présentes sous les QTLs sont ensuite listées sur un outil présent sur internet et classant les annotations des gènes d'*Arabidopsis thaliana* suivant leur fonction (TAIR, 2005).

Pour l'olivier, les bornes des QTLs sont définies de la même manière mais la recherche de gènes candidats diffère. Chaque marqueur SNP a été annoté par l'équipe ID au moment du séquençage, leur fonction est donc connue. Il est nécessaire de regarder au cas par cas la fonction présente sous les marqueurs qui composent le QTL.

## Partie III : Résultats

# 1. La cartographie génétique

## 1.1. La carte génétique du pommier

### 1.1.1. Les cartes génétiques

Le pommier, *Malus domestica* Borkh., possède 17 chromosomes diploïdes. Les marqueurs utilisés pour la cartographie de la population hybride « X3263 x Belrène » sont classés en 17 groupes de liaisons<sup>1</sup>. Ces cartes comprennent des données SNP obtenues grâce à une puce 9K et des données plus anciennes, pour la plupart microsatellites mais aussi SNP, utilisés dans le cadre de l'élaboration de cartes génétiques pour d'autres populations de pommier tel le croisement « Starkrimson x Granny Smith ». Ces marqueurs plus anciens qui servent de lien avec d'autres cartes sont représentés en rouge sur la figure ci-dessous qui montre trois groupes de liaisons types.

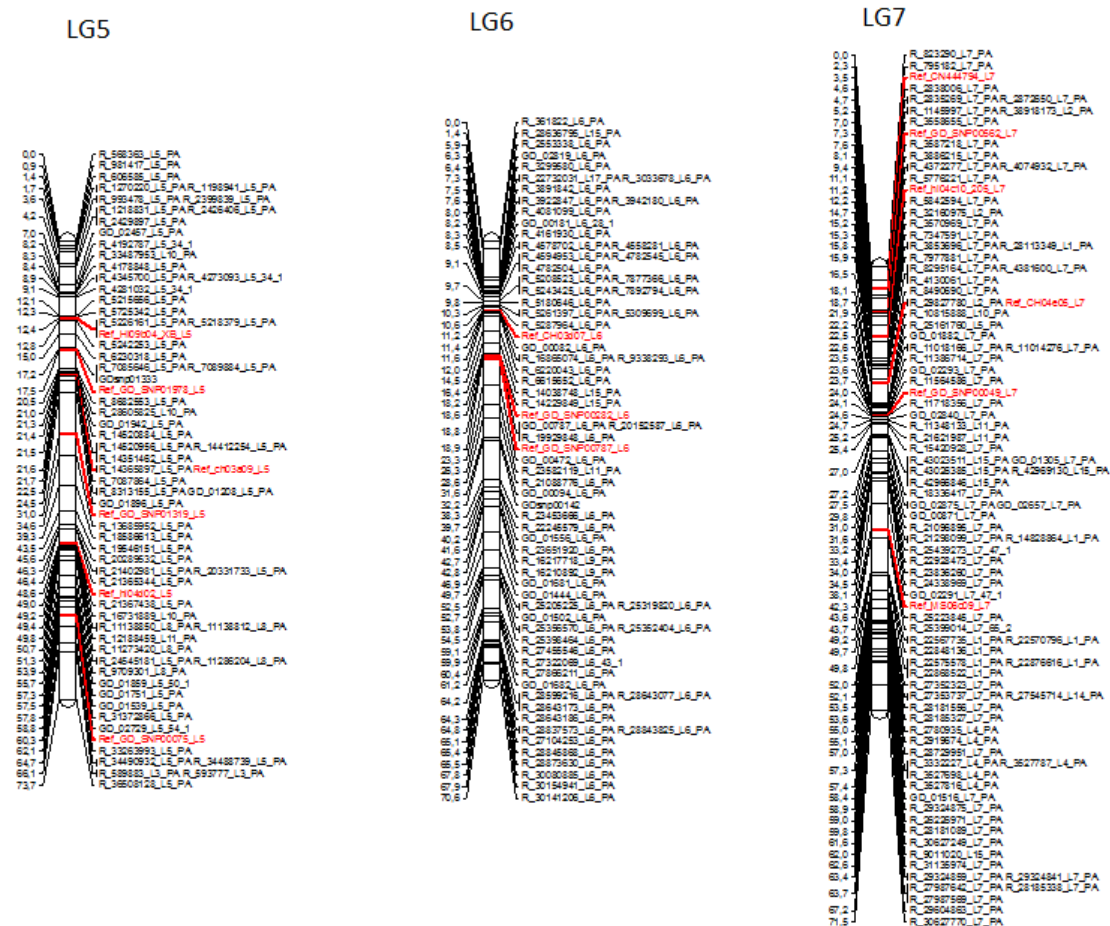


Figure 9 : Groupes de liaisons 5, 6 et 7 de la population "X3263 x Belrène"

<sup>1</sup> Voir Annexe IV : Cartographie complète de la population de pommier « X3263 X Belrène » avec positionnement des QTLs

### 1.1.2. Statistiques sur les cartes

Des statistiques ont pu être calculées sur la cartographie élaborée à partir du croisement « X3263 x Belrène ».

Table 3: Statistiques sur la cartographie de la population de pommier

X3263 x Belrène	
Général :	17 groupes
Nb de loci totaux :	1695
Longueur moyenne :	67.509
Longueur min :	47.015
Longueur max :	106.371
Distance min moyenne :	0.002
Distance max moyenne :	6.362
Distance moyenne :	0.727
Nombre de loci moyen:	99.706
Distance min :	0.001
Distance max :	12.862
Nombre min de loci :	50
Nombre max de loci :	171
Longueur totale	1147.649 cM

Avec une moyenne de plus de 67 marqueurs par groupe espacés de 0.727 centiMorgans, les cartes génétiques de cette population sont saturées et laissent peu de zones non couvertes. La longueur totale de la carte atteint 1147.649 cM. Ce résultat est en conformité avec les autres cartes élaborées à ce jour à partir des populations hybrides de pommier. En moyenne, chaque groupe de liaison mesure 67.509 cM.

## 1.2. La carte génétique de l'olivier

### 1.2.1. La carte optimisée

L'olivier est une espèce qui compte 23 chromosomes diploïdes. Ainsi, les marqueurs développés à partir du transcriptome ont pu, grâce à la méthodologie employée et décrite dans la partie Matériel et Méthodes, former 23 groupes de liaison correspondant au nombre de chromosomes de l'olivier<sup>1</sup>. Dans un même esprit que pour la carte pommier, les marqueurs SSR, servant de ponts entre différentes cartes construites dans différentes équipes, sont en rouge. Les marqueurs SSR vont notamment servir ici à faire correspondre ces groupes avec ceux de la carte préalablement construite (Ben Sadok et al., 2013).

---

<sup>1</sup> Voir Annexe III : Cartographie complète de la population d'olivier « Olivière x Arbequina » avec positionnement des QTLs

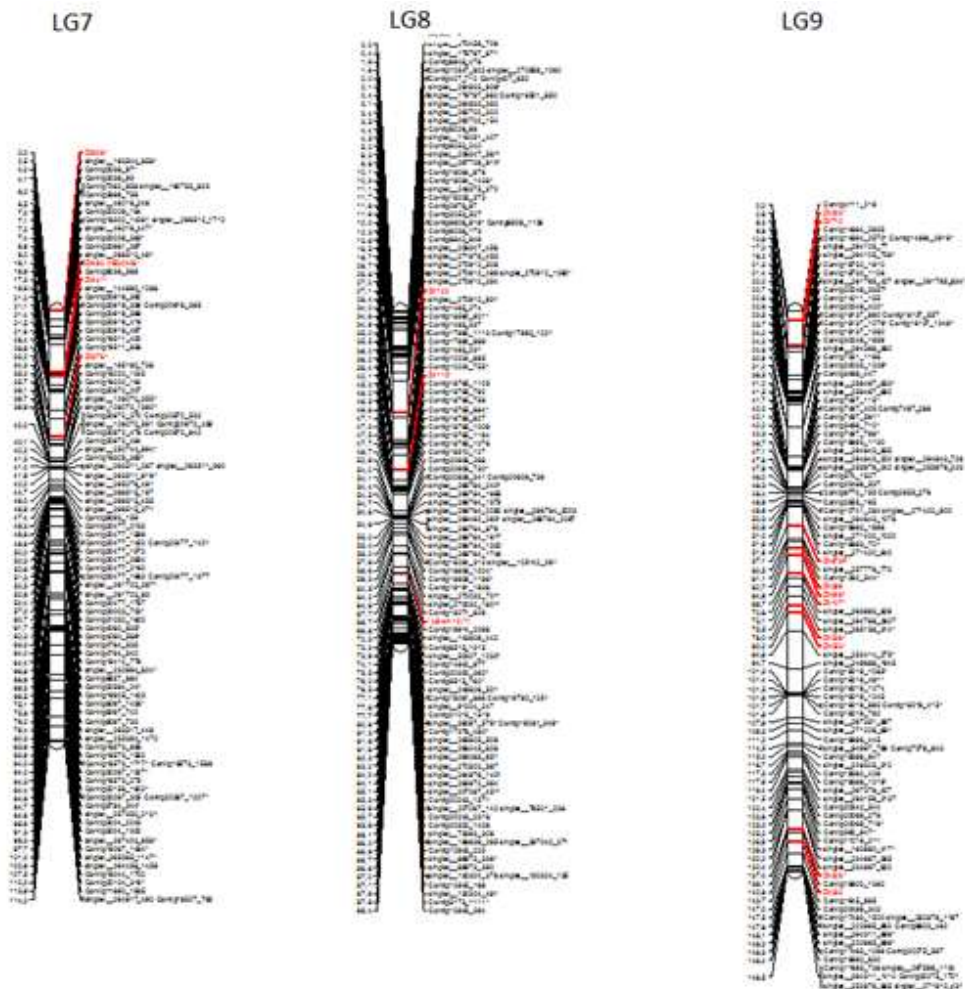


Figure 10 : Groupes de liaisons 7, 8 et 9 de la population "Olivière x Arbequina"

### 1.2.2. Statistiques sur les cartes

Des statistiques concernant cette cartographie ont pu être effectuées. Cela permet d'évaluer la pertinence globale des résultats en regardant les distances entre marqueurs, la taille totale de la carte comparée à la taille connue du génome et le nombre de marqueurs utilisés notamment.

La distance maximale moyenne entre marqueurs dépasse 5cM. Elle peut s'expliquer par la nature des marqueurs qui proviennent des séquences codantes. Il n'est donc pas surprenant de découvrir des zones sur lesquelles aucun marqueur ne se s'est positionné. Cela indique que cette zone est probablement non codante. La distance moyenne entre les marqueurs positionnés est égale à 1.001 cM. Ce chiffre est un peu plus élevé que pour le pommier du fait du nombre plus important de zones moins couvertes. La longueur totale de la carte s'élève à 2346.846 cM avec une longueur moyenne par groupe de liaison de 102.037 cM.

Pour cette cartographie, la méthodologie a été élaborée en reconstruisant des marqueurs multialléliques. Ainsi, plusieurs marqueurs bialléliques se positionnant exactement au même endroit, ont été assemblés.

### 1.2.3. La différence avec la cartographie basée sur des marqueurs bialléliques

Les marqueurs bialléliques rassemblés pour construire les marqueurs multialléliques de la carte optimisée ont été utilisés pour faire une carte biallélique de manière à comparer les deux méthodes (Table 4).

Les mêmes statistiques que pour les cartes précédentes ont été effectuées. Avec cette méthode, la cartographie totale de l'olivier passe de 2347 cM à 2517 et est donc allongée de 170 cM. De plus, 1266 marqueurs supplémentaires sont utilisés dont seulement 68 marqueurs microsatellites soit 1.83%. Sur la carte sur laquelle sont positionnés les marqueurs reconstruits, 88 soit 3.6% des marqueurs totaux sont des SSR.

Table 4: Statistiques comparatives concernant les cartes "multialléliques" et "bialléliques" élaborées pour l'olivier

<b>Olivière x Arbequine</b>			
		Multiallélique	Biallélique
<b>Général :</b>		23 groupes	23 groupes
<b>Nb de marqueurs totaux</b>		2456	3722
<b>Longueur moyenne :</b>		102.037	109.456
<b>Longueur min</b>		53.334	61.708
<b>Longueur max</b>		163.511	164.97
<b>Distance min moyenne :</b>		0.004	0.001
<b>Distance max moyenne :</b>		8.282	9.227
<b>Distance moyenne :</b>		1.001	0.717
<b>Nombre de loci moyen:</b>		107	161.826
<b>Distance min :</b>		0.001	0.001
<b>Distance max :</b>		13.454	18.492
<b>Nombre min de loci :</b>		49	69
<b>Nombre max de loci :</b>		183	282
<b>Longueur totale</b>		2346.846 cM	2517.48

## 2. Résultats des analyses statistiques

### 2.1. Analyse des données génétiques

Les données SNP pour le pommier ne présentent qu'un faible taux d'erreur de génotypage : 0.0005%. Elles sont donc fiables pour être utilisées en cartographie et n'ont également que peu de données manquantes par marqueur.

Le graphique ci-dessous présente le taux de recombinaison des marqueurs par groupe de liaison pour la population « X3263 x Belrène ». La droite est linéaire et centrée. La couleur



rouge indique un fort taux de recombinaison des marqueurs, la couleur bleu foncé un faible taux. Les recombinaisons les plus fortes se font au sein d'un même groupe de liaison, puisqu'il s'agit de la recombinaison entre paire de chromosomes durant la méiose.

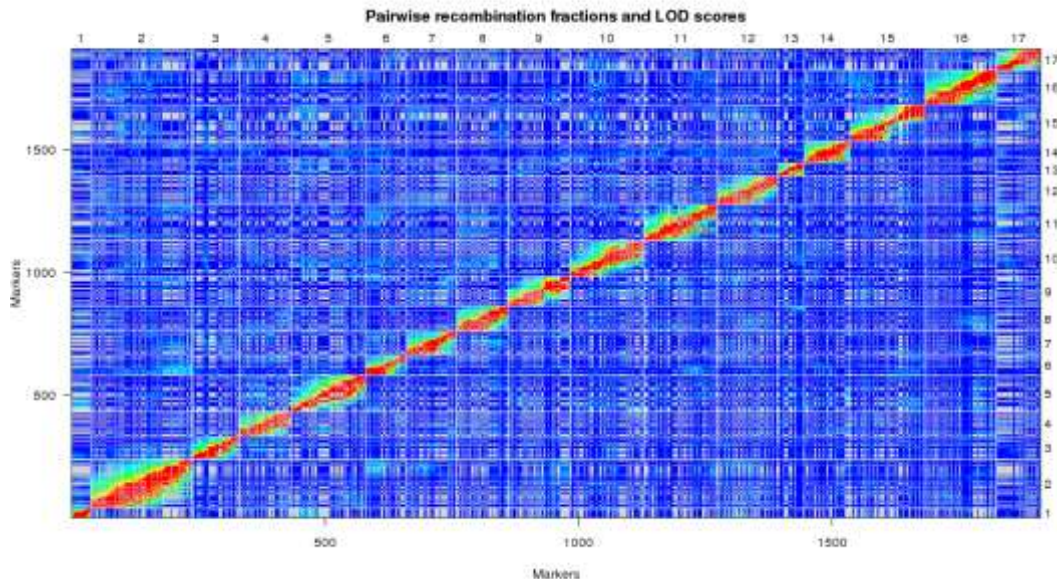


Figure 11 : Taux de recombinaisons par groupes et entre groupes de liaison pour les données du pommier

Pour l'olivier, les données, obtenues de manière différente grâce au transcriptome, ont un taux d'erreur de 0.0025%, donc un peu plus élevé que chez le pommier mais acceptable car toujours relativement faible. En revanche, pour ces données, des marqueurs qui ont des données manquantes pour 50% d'individus sont acceptés.

Ci-dessous, la figure représente les taux de recombinaisons entre les 23 groupes de liaisons d'« Olivière x Arbequina ».

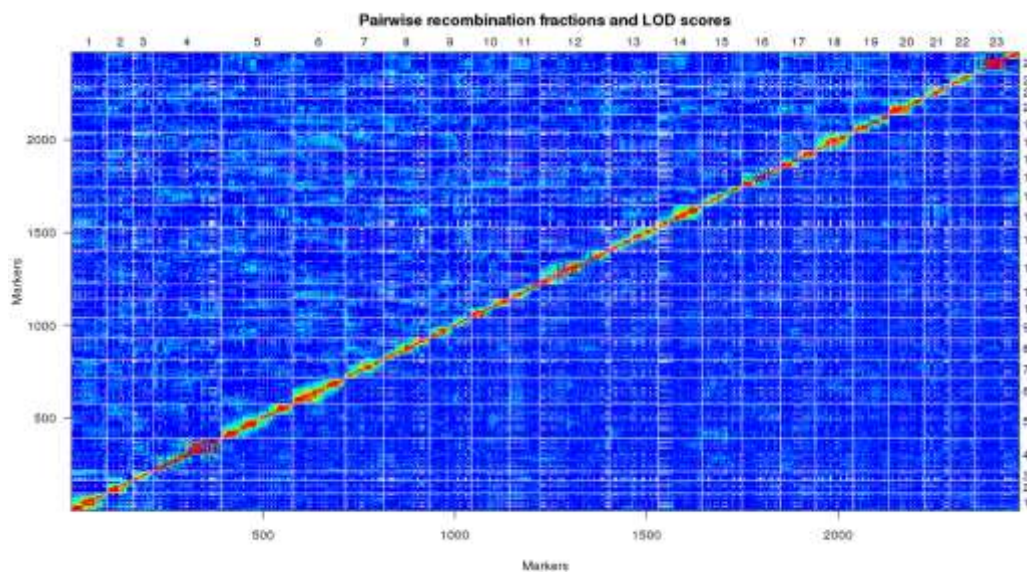


Figure 12 : Taux de recombinaisons par groupes et entre groupes de liaison pour les données de l'olivier

## 2.2. Analyse des données phénotypiques

Les matrices de données phénotypiques pour l'olivier et le pommier ont été analysées parallèlement. En effet, elles contiennent les mêmes informations pour les deux espèces. Pour chaque génotype, les données de production par année et par répétition sont répertoriées.

### 2.2.1. Les différentes années de phénotypage

Les "Box-plots" ci-dessous présentent la répartition de la production de pommes (Figure 13) et d'olives (Figure 14) sur, respectivement, sept et six ans.

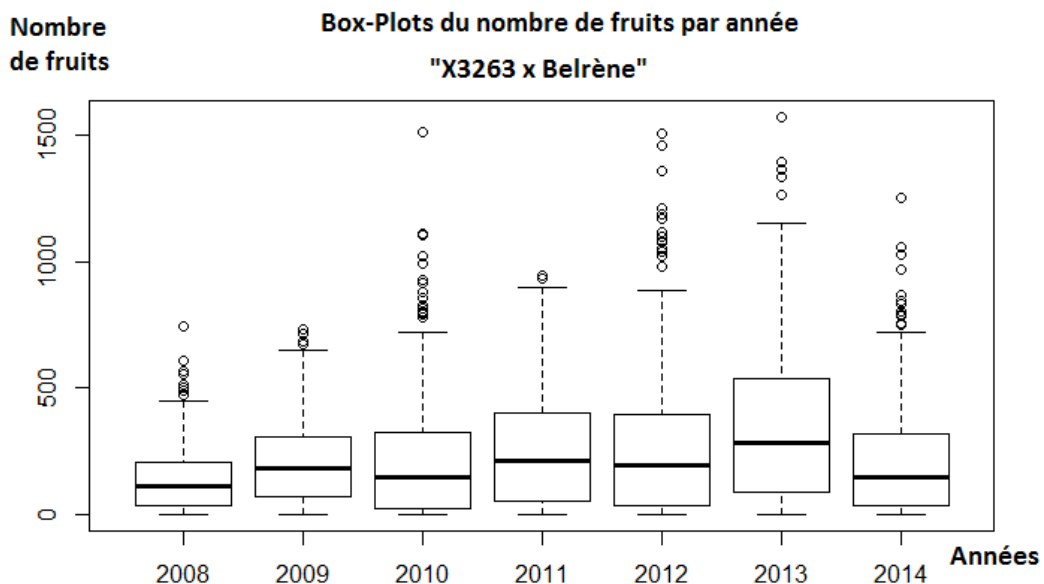


Figure 13 : "Box-plots" du nombre de fruits par année pour "X3263 x Belrène"

En 2008, les oliviers du croisement étudié n'ont pas tous atteint leur maturité et n'ont donc pas commencé à produire d'olives. C'est pour cette raison que la médiane du « Box-plot » se trouve très proche de zéro.

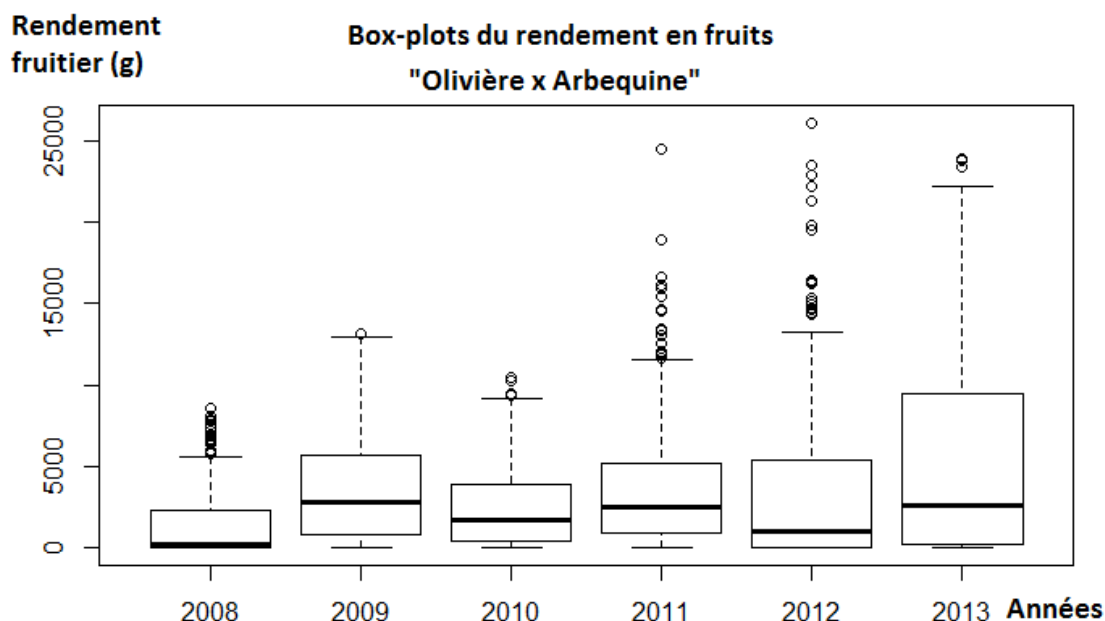


Figure 14 : "Box-plots" du nombre de fruits par année pour "Olivière x Arbequina"

### 2.2.2. Héritabilité des caractères

Pour chaque caractère et pour chaque espèce, l'héritabilité au sens large a été calculée. Cette héritabilité est considérée comme significative au-dessus du seuil de 0,2 (Gallais A., 2009).

La Table 5 récapitule les héritabilités calculées  $H^2$  :

Table 5 : Héritabilités des différents caractères pour les deux espèces

	Pommier	Olivier
<b>Rendement</b>	0.574	0.44
<b>BBI</b>	0.73	0.38
<b>BBI_res_norm</b>	0.71	0.55
<b>BBI_pas_norm</b>	0.67	0.52
<b>Autocor_resid</b>	0.63	0.27

### 2.2.3. Interaction entre le génotype, le rendement et l'année

Pour les deux espèces étudiées, une relation est observée entre ces trois variables. La Figure 15 représente cette interaction chez la population « X3263 x Belrène ». Malgré le grand nombre de génotypes dont les données se croisent, une tendance à l'alternance de production est clairement observée. Par exemple, le génotype 9 qui a le plus haut rendement en 2010 voit sa production de fruits baisser drastiquement en 2011 où l'arbre n'a presque plus de fruits, et remonter en 2012 jusqu'à 1500 fruits. Ce génotype montre une amplitude de rendement d'une année sur l'autre remarquable. A l'inverse certains

génotypes présentent une production faible en 2010 et forte en 2011. Cette tendance est illustrée par les piques sur le graphique. En revanche, il est plus difficile de remarquer les génotypes réguliers sur cette figure, tant les informations sur le graphique sont denses.

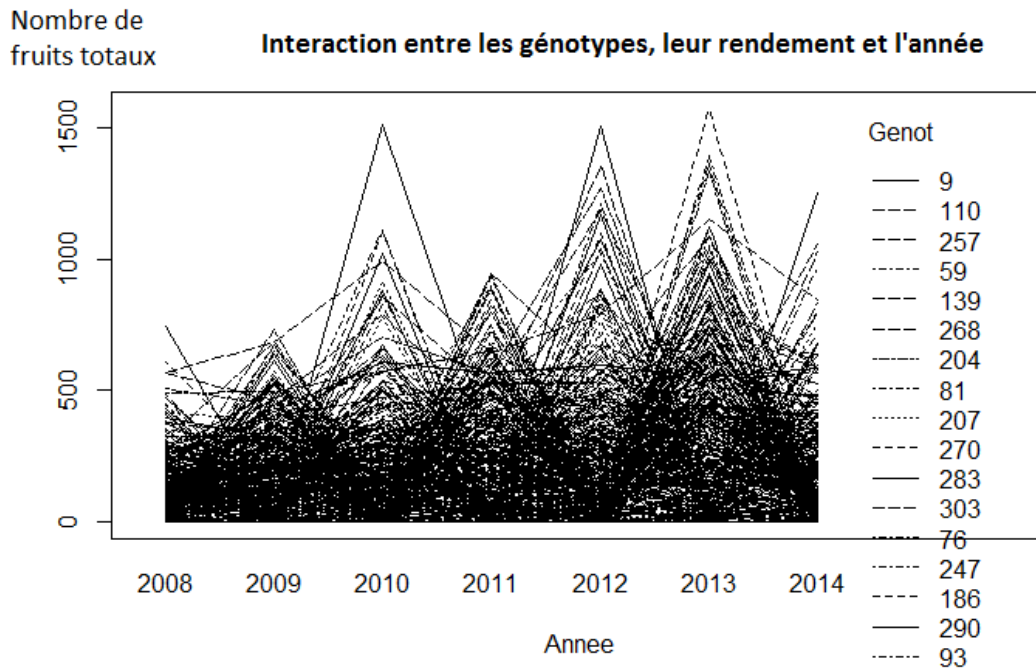


Figure 15 : Interaction entre les génotypes de pommier, leur rendement et l'année

On peut noter chez l'olivier un comportement des génotypes similaire à celui chez le pommier (Figure 16).

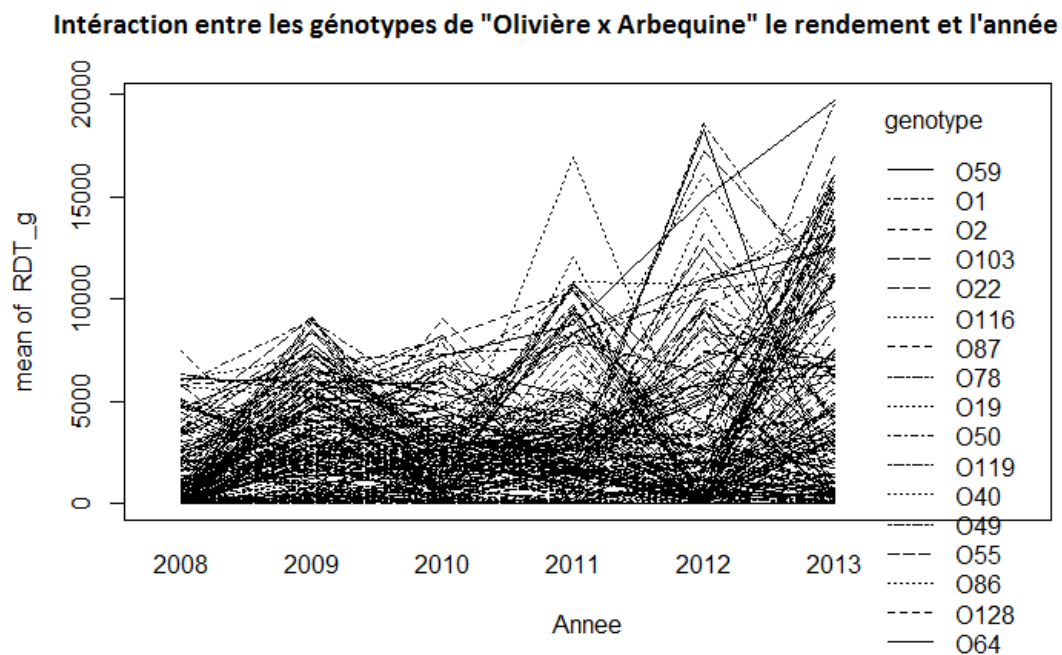


Figure 16 : Interaction entre les génotypes d'olivier, leur rendement et l'année

#### 2.2.4. Trois types de comportements des géotypes

La régularité de production est estimée à partir de plusieurs indices : Le BBI (Biennial Bearing Index), le BBI\_res\_norm (Biennial Bearing Index calculé sur les résidus normés du modèle), le BBI\_pas\_norm (ratio entre la production de deux années successives et la moyenne de la production sur le total des années) et l'autocorrélation des résidus. Ces quatre indicateurs permettent une bonne estimation de l'alternance.

Le paramètre "Autocor" exprimé en fonction du paramètre "BBI\_res\_norm" permet de rendre compte de la distribution des géotypes (chaque point sur le graphique correspond à un géotype) et de pouvoir les classer selon leur comportement. Plus le BBI\_res\_norm est élevé et l'autocorrélation des résidus est faible et plus le géotype est alternant. A l'inverse, un BBI\_res\_norm proche de zéro avec une forte autocorrélation des résidus est caractéristique des individus à la production régulière (Figures 17 et 18).

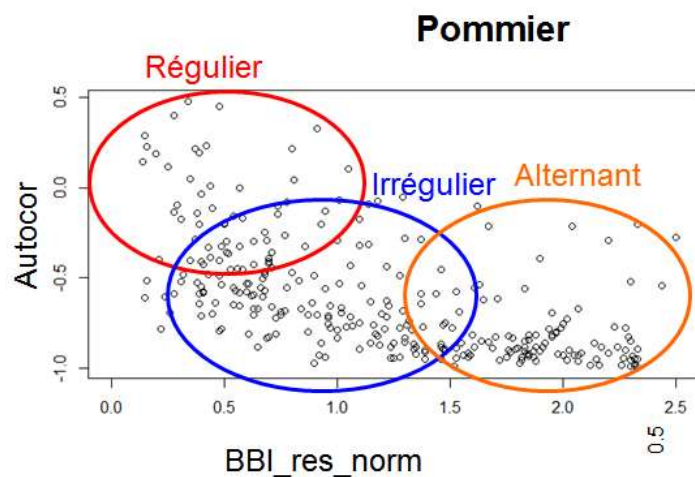


Figure 17: Comportements des géotypes de pommier

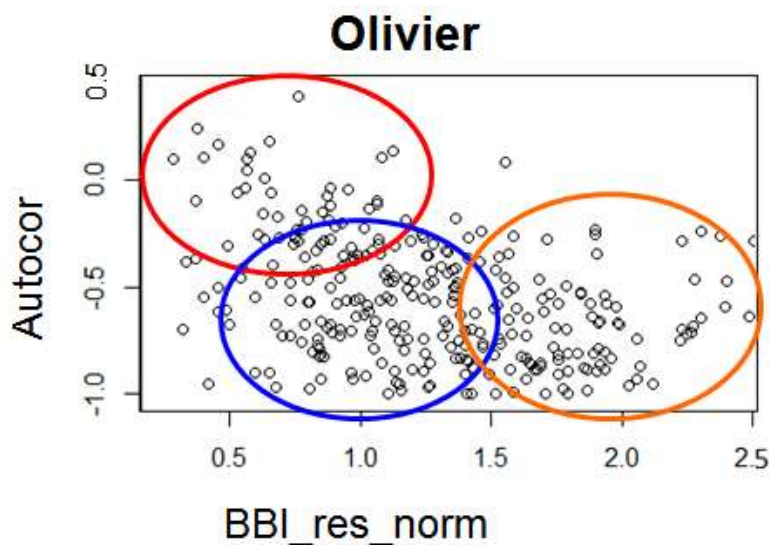


Figure 18 : Comportements des géotypes d'olivier

Le comportement régulier d'un individu se caractérise par une production similaire à celle présentée sur le graphique ci-dessous (Figure 19). La production de fruits augmente progressivement d'année en année due à l'effet ontogénique et se stabilise en 2013, ici autour de 500 fruits produits.

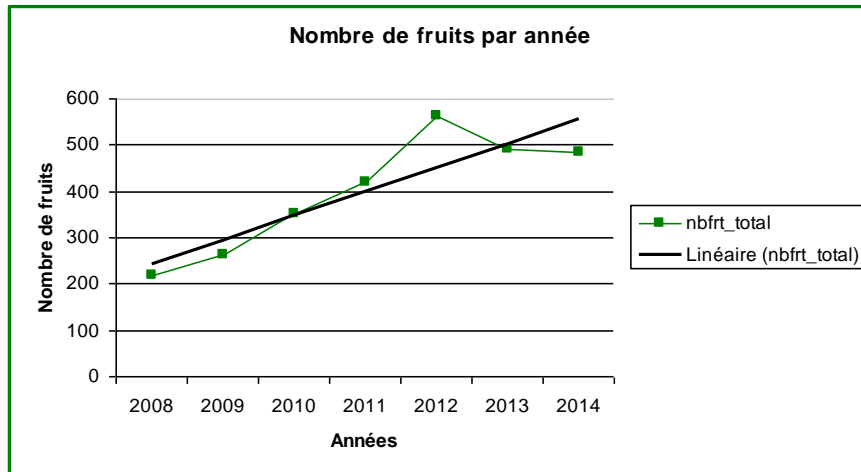


Figure 19 : Exemple de comportement de production d'un individu "régulier"

La Figure 20 représente quant à elle le comportement typique d'un génotype alternant. Le nombre de fruits produit passe de plusieurs centaines à moins de cent voire zéro l'année suivante. Ici, un effet ontogénique est observé puisque les années « On », la production est de plus en plus forte.

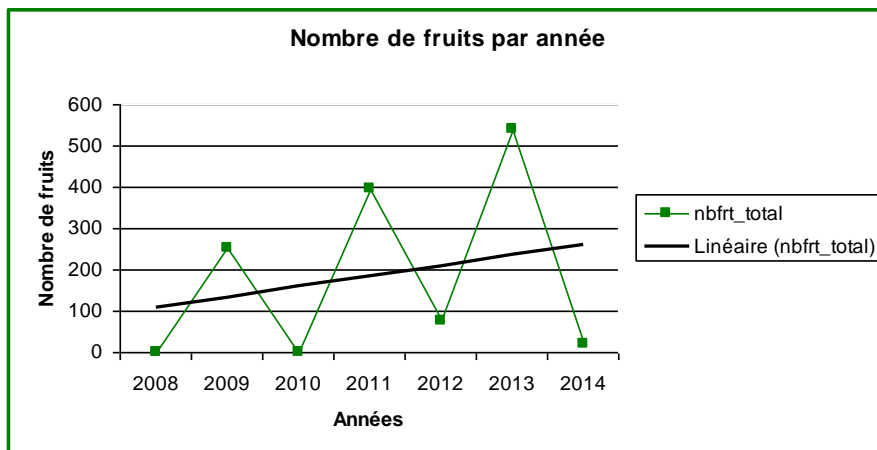


Figure 20 : Exemple de comportement de production d'un individu "alternant"

#### 2.2.5. Différence entre BBI res\_norm et BBI commun

Si le BBI\_res\_norm a été calculé en plus du BBI commun c'est que celui-ci entraîne une confusion entre la tendance de production sur plusieurs années et le caractère de l'alternance lui-même.

La Figure 21 informe sur les valeurs de BBI\_res\_norm, BBI\_pas\_norm, la moyenne de la production, l'autocorrélation des résidus et le BBI pour les génotypes représentés sur les graphiques qui leur sont associés.

Le second génotype représenté sur la Figure 21 est le plus alternant. Ses années « On », la production va de 400 fruits/an à presque 1000. Les années « Off » en revanche, la production de ce génotype est quasiment nulle. Bien qu'alternant également, le premier génotype de la Figure 21 présente une amplitude de rendement plus faible d'une année sur l'autre. Pourtant, son BBI est plus élevé. Cet exemple illustre bien le fait que le BBI est un indice qui peut prêter à confusion. Ici, disposer du BBI\_res\_norm et du BBI\_pas\_norm prend tout son intérêt. Ainsi, ces deux indices sont plus élevés pour le second génotype, le plus alternant des deux.

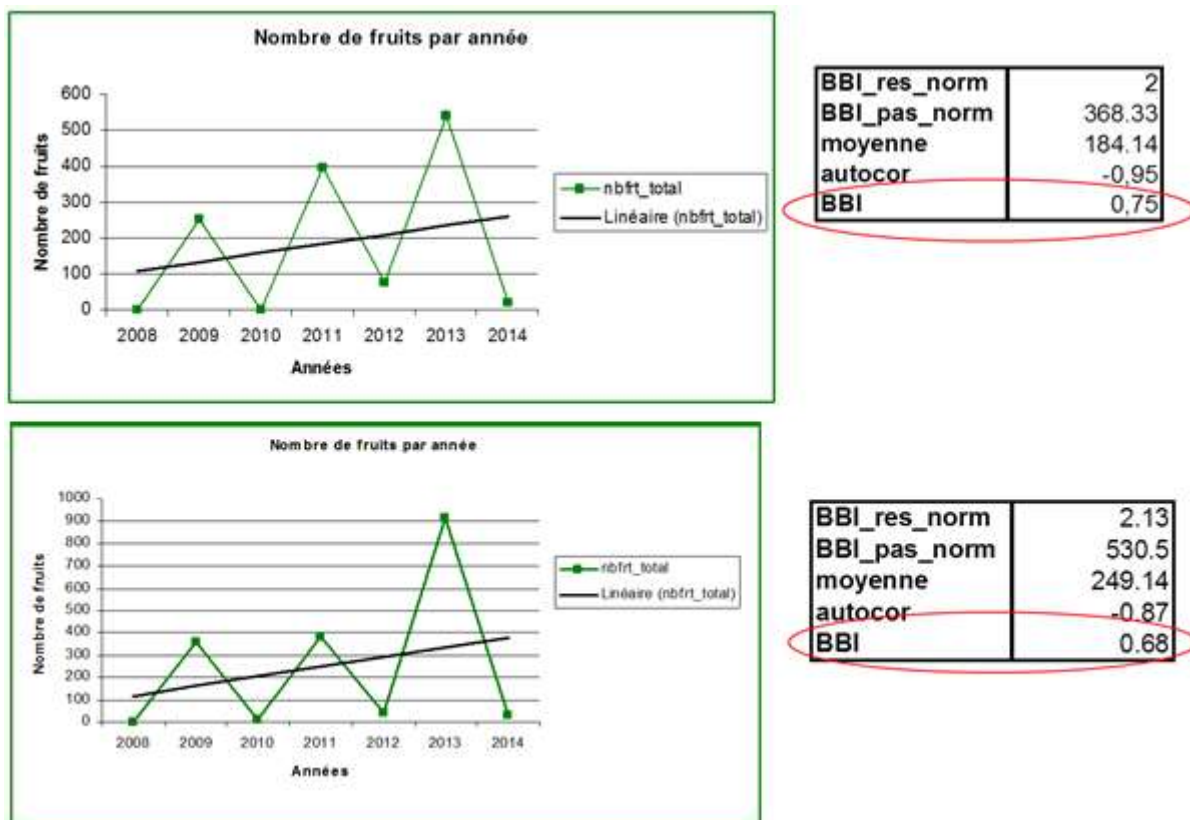


Figure 21 : Comparaison des indices de régularités entre deux génotypes alternants

### 3. QTLs détectés

Pour l'olivier comme pour le pommier, des QTLs correspondant au rendement ainsi qu'à la régularité de production ont pu être détectés.

### 3.1. Les QTLs détectés sur X3263 x Belrène

#### 3.1.1. Informations sur les QTLs

La Table 7 est un récapitulatif des QTLs positionnés sur les cartes génétiques de la population de pommier étudiées. Cinq variables ont été prises en compte dans ce projet : le Rendement, le BBI, le BBI avec les résidus normalisés, le BBI non pondéré et enfin l'autocorrélation des résidus du modèle (Table 6). Pour chacune de ces variables, plusieurs QTLs ont été détectés.

Les co-facteurs, marqueurs pour lesquels le LOD est le plus élevé dans la zone, ont été retenus à partir d'un LOD supérieur ou égal à 4.35. Le  $R^2$  correspond au pourcentage de la variable expliqué par ce QTL. Ainsi, plus ce paramètre est élevé et plus cette zone joue un rôle important dans l'expression du caractère pour laquelle elle est détectée. Le groupe de liaison (LG) sur lequel se trouve cette zone est aussi renseigné.

Pour finir, l'effet allélique dominant ainsi que l'effet mâle, l'effet femelle et la dominance sont déterminés.

Table 6 : Descriptions des variables utilisées pour "X3263 x Belrène"

X3263 x Belrène		
Caractère	Nom de la variable	Formule
Rendement	RDT	Nombre de fruits
Biennial Bearing Index	BBI_genot	
BBI avec résidus normalisés	BBI_res_norm	
BBI non pondéré	BBI_pas_norm	
Autocorrélation des résidus	autocor_resid	

Table 7 : Liste des QTLs détectés pour chaque variable chez "X3263 x Belrène"

Variable	LG	LOD	R <sup>2</sup> (%)	Cofactor	Allelic effect	Af	Am	D
RDT	5	4.66	11.1	Ref_h04d02_L5	Af	-1.025451	-0.812904	-0.05015
	10	14.04	29.9	R_11114150_L12_PA	Af, Am, D	1.3930795	1.3711305	-1.119313
	13	4.17	10	R_3856174_L13_PA	Af	1.08975335	-0.39289165	-0.24570865
BBI_res_norm	3	5.86	9.4	Ref GD SNP00194_L3	Am	-0.13878475	0.03043625	0.00246225
	8	4.99	7.8	GDsnp01670	Af	-0.14396208	0.03332943	0.00473158
	10	8.13	19.1	R_14019319_L14_PA	Af, D	-0.13815133	-0.06228603	0.15411918
	10	6.85	16.3	R_17046031_L10_PA	Af	-0.15832818	-0.05637133	0.08448833
	11	4.44	6.2	Ref_CH04a12_XB_L11	Af, Am	-0.0606377	0.0795678	-0.0187806
BBI_pas_norm	3	8.51	16.9	R_10663755_L3_PA	Af	-54.443975	14.983425	-22.203825
	15	4.02	7.9	GD_00134_L15_PA	Af, Am	-36.2384225	22.1829275	-12.5091775
autocor_resid_norm	3	4.92	9.3	R_9462180_L3_PA	Af, D	0.06085385	-0.0164403	0.006185
	12	5.93	11.4	GDsnp00334	Af	0.06779873	-0.01650618	-0.01639068
	13	4.75	11.6	Ref_AU223486_X_L13	Am	-0.0173885	-0.07059455	-0.00715345
BBI_genot	3	3.88	9.8	R_9462180_L3_PA	Af	-0.08533675	0.04364625	0.02991575
	8	4.35	10.8	GDsnp01670	Af	0.09215825	0.03862775	0.00997275
	10	6.14	14.8	R_16718103_L10_PA	Af, Am, D	-0.08586125	-0.03589575	0.06989075
	10	10.46	23.9	R_11114150_L12_PA	D	-0.07977175	-0.04860275	0.12872425





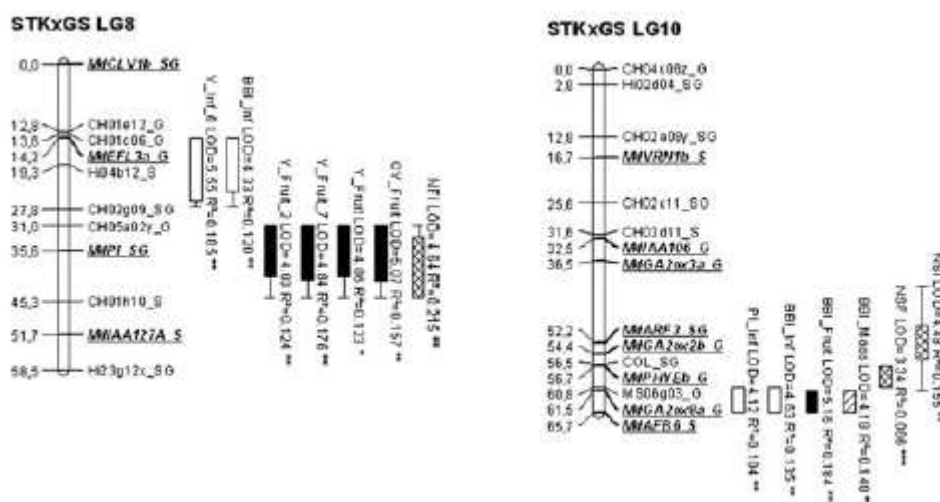


Figure 23 : Détection des QTLs sur les groupes 8 et 10 de la population "Starkrimson x Granny" (Guitton et al., 2012)

Sur la Figure 23, la cartographie génétique a été construite avec les marqueurs SSR en commun avec la nouvelle carte saturée. Des QTLs avaient été détectés pour un nombre plus important de caractères : outre le rendement et le BBI, des QTLs des variables comme le rendement en inflorescences, entre autres, ont été positionnées.

### 3.2. Les QTLs détectés sur Olivière x Arbequina

#### 3.2.1. Informations sur les QTLs

La Table 9 informe sur les caractères qui ont servi pour l'analyse QTL et leur notation. La seule différence entre les variables utilisées pour le pommier et celles de l'olivier réside dans le calcul du rendement. Ici, ce n'est pas le nombre de fruits qui est utilisé mais le poids total des fruits (Table 8).

Table 8 : Descriptions des variables utilisées pour "Olivière x Arbequina"

Olivière x Arbequina		
Caractère	Nom de la variable	Formule
Rendement	RDT	Masse totale des fruits
Biennial Bearing Index	BBI_genot	
BBI avec résidus normalisés	BBI_res_norm	
BBI non pondéré	BBI_pas_norm	
Autocorrelation des residus	autocor_resid	

Les groupes de liaisons 2, 4 et 6 (en bleu, orange et vert dans la Table 9) reviennent régulièrement lors de la détection de QTLs. Ici, des QTLs d'intérêts ont été retenus à partir d'un LOD de 3,8.

Table 9 : Liste des QTLs détectés pour chaque variable chez "Olivière x Arbequina"

Variable	LG	LOD	R <sup>2</sup> (%)	Cofactor	Allelic effect	Af	Am	D
RDT	2	5.4	14.2	singlet_268334_114	Af	2.78156675	0.39142825	-0.94813325
	2	4.32	11.7	Contig4358_530	D	1.74173275	1.59124275	2.86764225
	4	5.2	18.2	Contig19837_1221*	Af, Am	-3.6400175	-3.5840775	-2.1804925
	6	4.37	14.2	Contig19892_596	Af, D	3.3764258	0.5533357	-3.0016092
	9	6.41	12.6	singlet_253414_275*	Af, Am, D	2.9881325	-3.0563675	2.6790975
	14	5.03	10.2	UDO17*	Af, Am, D	2.141335	2.81815	-3.674375
	17	5.54	14.5	Contig16308_886	Af	3.1002265	1.1986185	-0.8446335
	18	4.26	11.6	singlet_265849_829	D	0.4666875	-0.1477025	-2.3239925
BBI_res_norr	2	3.93	13.6	Contig6863_178	Af, Am, D	-0.07393855	0.03289665	-0.0214016
	4	4.47	23	Contig17690_983	Am	-0.02254875	0.13716775	-0.00116025
	4	4.3	18.7	Contig9813_1570	Af, Am	0.03936338	0.08087913	-0.01466323
	6	3.8	13.9	Contig20870_352	D	-0.02262103	-0.02092893	-0.08441568
	19	3.85	13.4	Contig20952_324	Am, D	-0.0043824	0.02366465	0.04889925
BBI_pas_norr	5	5.88	18.7	singlet_73079_75	Am, D	-75.780025	206.594025	-333.427475
	6	3.82	14.9	Contig19892_596	Af, D	263.9554	27.00105	-224.6121
	16	3.88	13	Contig12665_812	Af, D	-147.66785	-85.07635	216.26665
	18	5.32	17.2	singlet_248695_170	Af, Am	-216.70985	73.73965	-155.45165
autocor_resid	3	4.1	14.2	Contig5865_251	Am	0.0015809	-0.0280894	0.00228835
	8	5.3	24.2	singlet_248375_372	Am, Af	-0.00558845	-0.008492	7.67E-05
	13	5.21	17.5	singlet_165108_413	Af, D	-0.00646827	0.00095205	-0.00581974
BBI_genot	4	4.04	10	singlet_77367_374	Af, Am	0.04893225	0.06771725	-0.03070275
	16	4.67	24.1	Contig18349_537	Af, D	0.060691	0.0388265	0.062549
	17	4.5	10	Contig447_153*	Af, Am	-0.0756135	0.032781	-0.00031
	19	4.66	13.3	Contig18541_237	Am	-0.01579475	0.08386025	-0.02881325
	22	4.54	10.1	Contig10871_1401	D	0.0231515	0.0021275	260.07228

### 3.2.2. Positionnement des QTLs sur les cartes génétiques

Sur la Figure 24 sont représentés les groupes de liaison 2,4 et 6 et les QTLs qui leur sont associés.

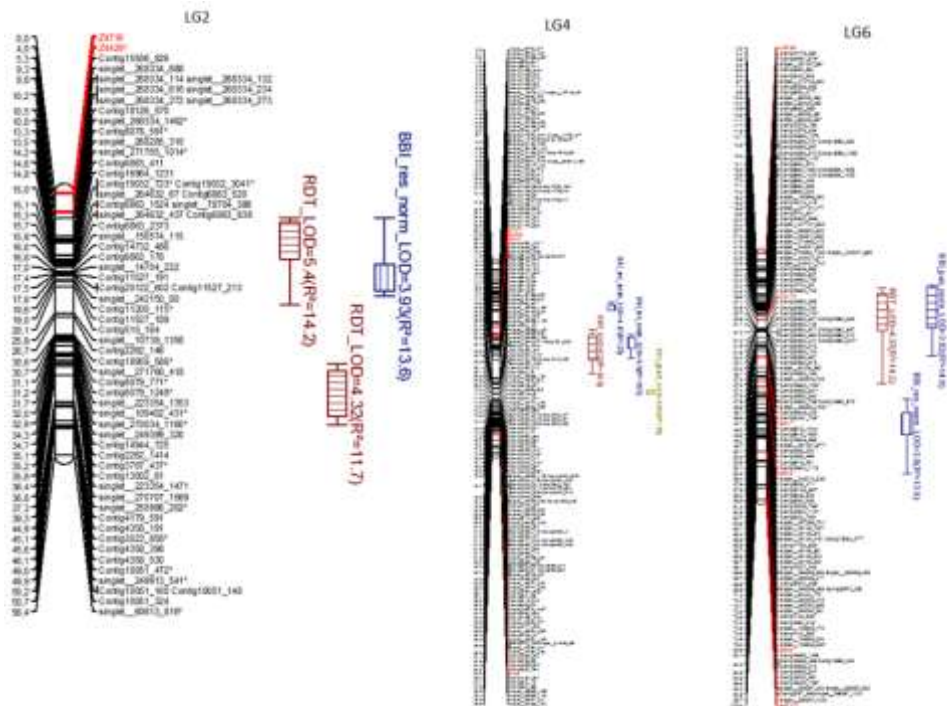


Figure 24: Positionnement des QTLs sur les groupes 2, 4 et 6 des cartes génétiques élaborées pour "Olivière x Arbequina"

### 3.2.3. Comparaison avec les résultats antérieurs (Ben Sadok et al., 2013)

C'est sur le groupe de liaison 9 de la population « Olivière x Arbequina » qu'avait été détecté le QTL lié au rendement (Ben Sadok et al., 2013 ; Figure 25). Cette carte génétique de 2012 est constituée de marqueurs AFLP ainsi que d'une partie des marqueurs microsatellites utilisés pour la nouvelle carte saturée dans le cadre de ce stage. Pour la même population, le même groupe de liaison mesurait environ 45 cM, alors qu'il est de 148.5 cM pour la nouvelle carte saturée.

La Figure 25 montre que la position du QTL détecté lors de cette analyse est exactement la même que dans le cas de l'étude précédente (Ben Sadok et al., 2013). En revanche, du fait de l'ajout de marqueur SNP, la position en centiMorgan a changé. En effet le QTL est placé à 75 cM sur la carte saturée en SNP et à 35 cM sur la carte AFLP.

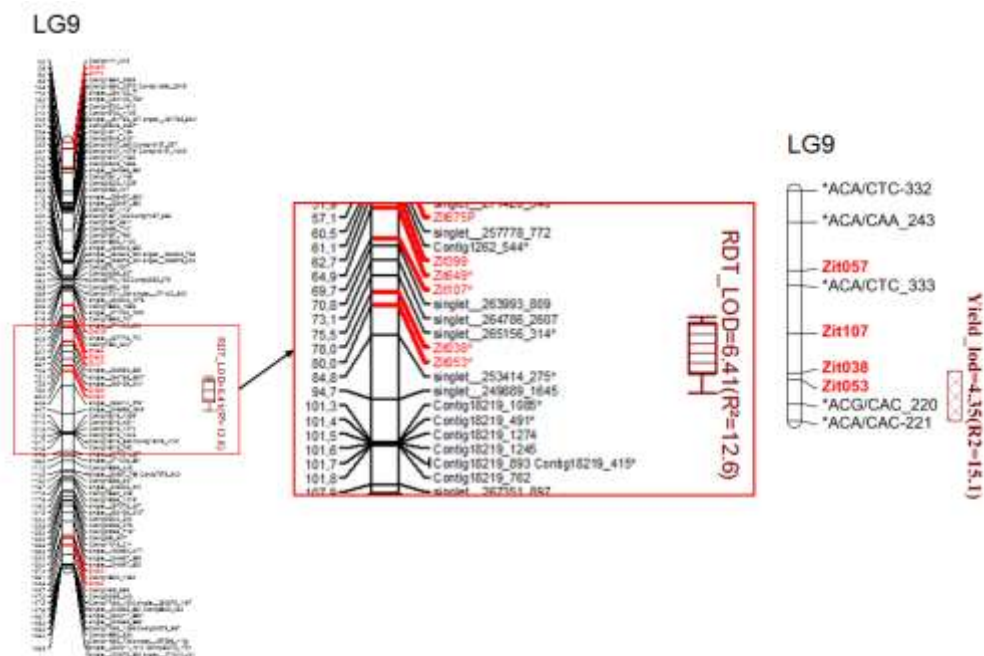


Figure 25 : Comparaison entre QTL de rendement détecté sur le groupe 9 de la nouvelle carte saturée et celui détecté sur la carte AFLP (Ben Sadok et al., 2013)

#### 4. Les gènes candidats impliqués

La distance en cM qui délimite un QTL sur un groupe de liaison peut correspondre à un très grand nombre de gènes codant pour diverses protéines et responsables de mécanismes parfois très différents.

Pour le pommier, il est possible de classer les gènes et les annotations suivant leur fonction connue sur l'espèce *Arabidopsis* qui a été étudiée de manière très approfondie par la communauté scientifique. La Table 10 recense le nombre de gènes et d'annotations par catégorie fonctionnelle et suivant leur catégorie majeure (Composant cellulaire, fonction

moléculaire ou mécanisme biologique) pour le QTL de rendement du groupe de liaison 10 chez « X3263 x Belrène ».

Table 10: Nombre de gènes présents sous le QTL de rendement du groupe de liaison 5 du pommier par grandes catégories fonctionnelles

Keyword Category	Functional Category	Annotation Count	Gene Count
GO Cellular Component	other intracellular components	256	<a href="#">138</a>
GO Cellular Component	other cytoplasmic components	249	<a href="#">161</a>
GO Cellular Component	nucleus	198	<a href="#">161</a>
GO Cellular Component	other membranes	166	<a href="#">103</a>
GO Cellular Component	chloroplast	141	<a href="#">81</a>
GO Cellular Component	plasma membrane	104	<a href="#">98</a>
GO Cellular Component	plastid	80	<a href="#">45</a>
GO Cellular Component	mitochondria	65	<a href="#">55</a>
GO Cellular Component	cytosol	61	<a href="#">55</a>
GO Cellular Component	other cellular components	54	<a href="#">54</a>
GO Cellular Component	extracellular	52	<a href="#">43</a>
GO Cellular Component	cell wall	25	<a href="#">21</a>
GO Cellular Component	ribosome	25	<a href="#">15</a>
GO Cellular Component	Golgi apparatus	24	<a href="#">19</a>
GO Cellular Component	ER	14	<a href="#">12</a>
GO Cellular Component	unknown cellular components	12	<a href="#">12</a>
GO Molecular Function	other binding	165	<a href="#">131</a>
GO Molecular Function	transferase activity	137	<a href="#">91</a>
GO Molecular Function	protein binding	105	<a href="#">90</a>
GO Molecular Function	hydrolase activity	99	<a href="#">70</a>
GO Molecular Function	other enzyme activity	84	<a href="#">58</a>
GO Molecular Function	kinase activity	78	<a href="#">39</a>
GO Molecular Function	unknown molecular functions	74	<a href="#">74</a>
GO Molecular Function	nucleotide binding	72	<a href="#">66</a>
GO Molecular Function	DNA or RNA binding	69	<a href="#">63</a>
GO Molecular Function	transporter activity	58	<a href="#">31</a>
GO Molecular Function	receptor binding or activity	37	<a href="#">19</a>
GO Molecular Function	transcription factor activity	31	<a href="#">31</a>
GO Molecular Function	nucleic acid binding	16	<a href="#">16</a>
GO Molecular Function	other molecular functions	14	<a href="#">14</a>
GO Molecular Function	structural molecule activity	11	<a href="#">11</a>
GO Biological Process	other cellular processes	994	<a href="#">296</a>
GO Biological Process	other metabolic processes	738	<a href="#">286</a>
GO Biological Process	protein metabolism	195	<a href="#">127</a>
GO Biological Process	cell organization and biogenesis	190	<a href="#">78</a>
GO Biological Process	response to stress	189	<a href="#">105</a>
GO Biological Process	developmental processes	176	<a href="#">89</a>
GO Biological Process	response to abiotic or biotic stimulus	173	<a href="#">92</a>
GO Biological Process	transport	151	<a href="#">78</a>
GO Biological Process	other biological processes	135	<a href="#">88</a>
GO Biological Process	transcription,DNA-dependent	105	<a href="#">62</a>
GO Biological Process	signal transduction	96	<a href="#">61</a>
GO Biological Process	unknown biological processes	86	<a href="#">86</a>
GO Biological Process	DNA or RNA metabolism	56	<a href="#">26</a>
GO Biological Process	electron transport or energy pathways	24	<a href="#">18</a>

En revanche, pour l'olivier, les fonctions correspondant aux zones génomiques des marqueurs sont regardées SNP par SNP en ciblant ceux présents sous les QTLs d'intérêts. Quelques fonctions et molécules ont pu être répertoriées mais de manière moins exhaustive que pour la population de pommier. La fonction notée correspond à celle trouvée sous le cofacteur du QTL (Table 11).

Table 11 : Annotations de gènes sous les marqueurs qui forment les QTLs détectés chez l'olivier

Variable	Groupe	LOD	Annotation		
<b>RDT</b>		2	5.4	pectinestérase	
		2	4.32	ADP ribosylation factor	
		4	5.2	protéine sgt1	
		9	6.41	Homeobox-leucine	
		18	4.26	somatic	embryogenesis
<b>BBI_res_norm</b>		2	3.93	staphylococcal	nuclease
		4	4.47	leucine	
		4	4.3	catalase	isozyme
		6	3.8	polyadenylate-binding	protein
<b>BBI_pas_norm</b>		5	5.88	hevamine-a	os=hevea
		16	3.88	perakine	reductase
		18	5.32	bi1-like	protein
<b>autocor_resid_norm</b>		3	4.1	40s	ribosomal
		13	5.21	copper	transport

Cette série de résultats peut être sujette à une discussion concernant les limites des analyses pratiquées. En outre, divers approfondissements pourraient y être apportés de manière à les affiner et pour aller plus loin dans l'étude tant des caractères que des deux espèces.



## Partie IV : Discussion



Deux principaux résultats ont donc été obtenus lors de ce stage. Le premier correspond aux nouvelles cartes génétiques élaborées pour les populations de pommiers « X3263 x Belrène » et d'oliviers « Olivière x Arbequina ». Les analyses statistiques des données phénotypiques effectuées sur ces deux croisements et en particulier par l'approche de génétique quantitative qui a permis la détection de QTLs sur les cartes constituent le second résultat issu de ce travail.

Après avoir exposé les limites des approches méthodologiques et validé leur utilisation dans ce projet, les résultats seront discutés et amèneront des propositions d'améliorations ainsi que des perspectives plus larges pouvant être envisagées.

## **1. Validation et limites des approches utilisées et des résultats obtenus**

### **1.1. Limites et validation de la méthode**

#### **1.1.1. Le phénotypage**

Les deux parcelles ont été plantées la même année et sont entretenues dans des conditions similaires. Cependant, la collecte des données phénotypiques n'a pas été tout à fait réalisée dans les mêmes conditions pour les deux espèces. Le rendement de la population de pommiers « X3263 x Belrène » a été estimé sur sept ans contre six ans pour la population d'oliviers « Olivière x Arbequina ». En effet, en 2014, toute la production de la parcelle d'oliviers a été perdue suite à une attaque de mouches. De plus, la variable « Rendement » a été calculée et estimée de deux manières différentes. Pour le pommier, le rendement correspond au nombre total de fruits produits par l'arbre tandis que pour l'olivier c'est une estimation du poids total qui est prise en compte. La variable retenue pour l'olivier nécessite une double estimation car il s'agit du poids total et du nombre de fruits total de l'arbre. Ainsi, les données utilisées pour les analyses statistiques ne sont pas totalement homogènes entre les deux espèces, ce qui n'exclut pas un biais dans l'étude.

#### **1.1.2. La cartographie génétique**

L'algorithme de régression associé à la fonction de Kosambi a été utilisé pour les deux espèces. Il compare deux à deux tous les marqueurs groupes par groupes pour déterminer lesquels sont les plus proches physiquement par leurs caractéristiques génétiques et leur phase. Lors de l'étape d'élaboration de la carte génétique, certains marqueurs éloignés peuvent naturellement présenter une double recombinaison avec d'autres marqueurs et ainsi présenter la même phase alors qu'ils sont éloignés de plusieurs dizaines de centiMorgans l'un de l'autre et se situent au cœur d'une portion phasée différemment. Ces recombinaisons peuvent également être artificielles car des erreurs de phases peuvent exister, ce qui influe sur la taille des cartes génétiques élaborées.

Enfin, à l'exception des erreurs bien visibles telles les doubles recombinaisons improbables, certaines erreurs non corrigées restent susceptibles de fausser les cartes.

### 1.1.3. Les analyses statistiques

Les données phénotypiques expérimentales utilisées dans le cadre des analyses ont posé un problème de normalité. En effet, les relevés de la population de pommiers comme de l'olivier suivaient plutôt la loi de Poisson que la loi Normale comme le montre la Figure 26. La fréquence est très concentrée entre zéro et cent, ce qui n'est pas représentatif d'une distribution normale.

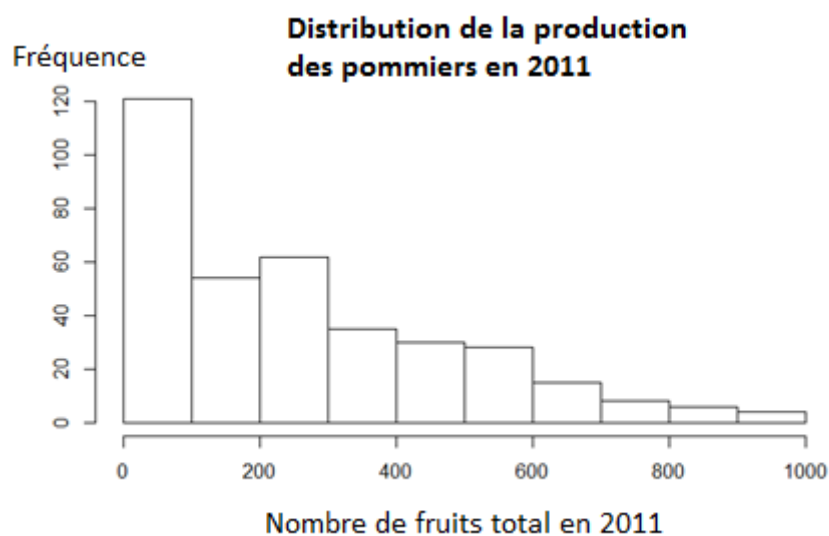


Figure 26 : Histogramme de répartition des géotypes pour le nombre de fruits total récolté en 2011 pour la population "X3263 x Belrène"

Après plusieurs tentatives infructueuses de normalisation des données, il a fallu travailler à l'aide de fonctions non-paramétriques. C'est une approche de statistiques adaptée à des distributions de données non-normales.

Ces fonctions non paramétriques qui ont vocation à effectuer des tests similaires à ceux effectués sur des données normales par des tests paramétriques sont cependant bien moins puissantes. De plus, les données collectées pendant l'année 2008 ont pu avoir un effet de biais dans les analyses. Les populations ayant été plantées en 2005, les arbres commençaient à peine leur production en 2008 et certains n'étaient pas encore matures. Un certain nombre de géotypes présentent des données manquantes pour le rendement cette année-là parce qu'ils ne sont pas encore entrés en phase de production.

Les données manquantes présentes dans les matrices pour certaines années ou certains géotypes constituent une limite aux analyses statistiques. En effet, une absence de

données ne peut pas être remplacée par zéro. Il est nécessaire de la prendre en compte en adaptant les tests. Les individus présentant des données manquantes de rendement pour plus de deux années ont été écartés des tests. Un nombre trop élevé de données manquantes ne permet pas de calculer correctement les différents BBI. Au-delà de ce seuil, la formule perd son sens car la variation de rendement est calculée en comparant les années deux à deux.

#### 1.1.4. La détection de QTLs

Parmi les QTLs qui ont été détectés sur les populations « X3263 x Belrène » comme « Olivière x Arbequina », ceux correspondant aux différents indices de régularité sont redondants car ils sont liés par nature. En effet, un même script sur le logiciel R calcule à la suite, génotype par génotype, le BBI\_res\_norm, le BBI\_pas\_norm et l'autocorrélation des résidus avec des informations communes. Ces trois indices, à lire séparément, sont donc issus d'une même analyse et sont utiles pour la compréhension de l'alternance mais n'apporte pas d'informations supplémentaires en tant que QTLs.

L'algorithme permettant de détecter les QTLs peut trouver des résultats différents suivant la sélection de cofacteurs effectuée. En effet, après avoir sélectionné les cofacteurs, c'est-à-dire les marqueurs présentant les LOD les plus élevés lors de la première détection, une nouvelle détection de QTLs va permettre de faire apparaître des zones qui n'étaient pas significatives auparavant. Ces opérations peuvent être répétées plusieurs fois de manière à repérer le maximum de QTLs. Si certaines positions sont identifiées pour plusieurs variables analysées, il s'agit alors d'un gage de fiabilité et sans doute de zones jouant un rôle important dans l'expression du caractère.

#### 1.1.5. La recherche de gènes candidats

La recherche de gènes candidats nécessite un long travail de génomique et de bioinformatique. En effet, les QTLs mesurent plusieurs centiMorgan. Or, des centaines d'annotations de gènes sont situées sous chacune de ces unités de cartographie génétique. Des analyses plus approfondies basées sur une approche de génomique et de bioinformatique permettraient de repérer quels sont les gènes potentiellement liés aux caractères examinés.

Cette approche a été initiée dans le cadre de ce stage mais la réalisation de l'étude nécessite au moins un nouveau travail de Master. D'autre part, les informations de la version actuelle de la séquence de référence du pommier (Velasco et al., 2010) sont en cours de révision et une nouvelle version devrait être disponible en cours d'année 2015 ce qui devrait améliorer la recherche des gènes candidats (GDR, 2012).

## 1.2. Discussion et validation des résultats obtenus

### 1.2.1. Les cartes génétiques

Les nouvelles cartes génétiques élaborées pour les deux espèces sont avant tout un outil génétique pour la détection de QTLs et les analyses génétiques. Les techniques et les marqueurs utilisés ont permis de construire des cartes précises mais qui pourront être améliorées par la suite. Elles ne doivent pas être considérées comme totalement représentatives et exactes quant à la position en cM des marqueurs.

L'une des limites de ces cartes réside dans le fait que les marqueurs microsatellites ne sont pas répartis de manière homogène entre les groupes de liaisons et sur ces groupes. Ainsi, sur la cartographie de la population « Olivière x Arbequina », aucun marqueur SSR n'est positionné sur le groupe de liaison 13. De ce fait, ce groupe ne pourra pas être relié avec une autre carte génétique pour comparer ses QTLs. Pour la carte du pommier, en revanche, les anciens marqueurs SSR et SNP servant de ponts avec d'autres cartes ont bien été insérés entre les SNP obtenus avec la puce 9K sur tous les groupes de liaisons sans exception.

Pour les deux populations, la carte génétique saturée en marqueurs SNP a cependant réellement permis d'apporter une information complémentaire par rapport aux cartes moins pourvues en marqueurs réalisées à ce jour, et ce sur plusieurs aspects. Le premier aspect amélioré est la répartition des marqueurs sur les cartes qui représentent les chromosomes. Ainsi, des zones non couvertes par les marqueurs lors des précédentes cartographies ont pu l'être grâce à celle-ci. De plus, le grand nombre de marqueurs utilisés a permis de préciser les groupes de liaisons. En effet, l'algorithme utilisé étudie les liens entre les marqueurs. Plus il y a de données et plus les liaisons trouvées vont être fortes et nombreuses entre les marqueurs d'un même groupe. Dans la carte génétique précédemment construite chez l'olivier (Ben Sadok et al., 2013), les marqueurs AFLP et SSR avaient été séparés en 26 groupes de liaisons, alors que dans la présente étude, les marqueurs SSR et SNP ont été rassemblés en 23 groupes de liaisons correspondant aux 23 chromosomes de l'olivier.

Deux cartes génétiques ont été élaborées chez l'olivier. La première a été réalisée en utilisant des marqueurs SNP bialléliques et d'autres reconstruits sur la base de leur positionnement physique (sur la même séquence de transcriptome) et de leur association en phase (taux de recombinaison). La seconde carte a été construite en utilisant les marqueurs SNP sous forme biallélique. Entre ces deux cartes, une différence de longueur de 170cM est mise en évidence. La carte optimisée construite avec les marqueurs multialléliques est légèrement plus courte, elle mesure 2346.8cM contre 2517cM pour la carte biallélique. Ces valeurs étant déjà relativement élevées pour une cartographie génétique, la longueur la plus faible, c'est-à-dire celle de la carte optimisée avec les marqueurs multialléliques, est retenue. En outre, lors de la cartographie génétique, il est bien plus pratique d'utiliser la matrice contenant les marqueurs reconstruits. En effet, le très grand nombre de marqueurs bialléliques (autour de 7000) rend très long le calcul des cartes génétiques pour chaque

groupe de liaison. L'algorithme doit alors analyser parfois plus de 300 marqueurs par groupe et cela peut prendre un temps considérable avant d'obtenir les trois cartes correspondant aux trois « rounds » opérés par Joinmap 4.1.

Quant à la population de pommiers « X3263 x Belrène », il s'agit de la première carte génétique réalisée pour ce croisement. La carte complète est un peu plus courte en comparaison avec d'autres cartes génétiques élaborées pour d'autres populations mais la longueur des cartes obtenues reste cependant raisonnable. La taille des cartes pour chaque groupe varie entre 47.015 cM pour la plus courte et 106.371cM pour la plus longue. La carte de 47cM est sans doute incomplète, en effet, les cartes génétiques du pommier mesurent généralement entre 60 et 90cM à l'exception de celle représentant le chromosome 15 (Liebhard et al., 2002 ; Kenis et Keulemans, 2004). C'est la carte de 106.371cM qui correspond ici au groupe de liaison 15. Ce chromosome est plus long que les autres du fait d'une translocation des chromosomes ancestraux 8 et 9 (Velasco et al., 2010). Une partie du chromosome ancestral 8 s'est insérée au milieu du 9 lors de la duplication des chromosomes du pommier, il y a 30 à 45 millions d'années, donnant ainsi un chromosome 15 agrandi.

### 1.2.2. Les résultats statistiques

Des analyses statistiques ont été effectuées sur les données génétiques et phénotypiques disponibles pour les deux populations expérimentales.

L'analyse des données génétiques a montré que la matrice des données SNP de la population « X3263 x Belrène » était légèrement plus fiable que celle du croisement « Olivière x Arbequina ». En effet, les marqueurs du pommier présentent seulement 0.0005% d'erreurs contre 0.0025% d'erreurs pour les marqueurs de l'olivier définis sur le transcriptome de l'ARN. Les deux matrices utilisées pour la construction de la cartographie génétique sont donc relativement propres et présentent peu d'erreurs. Pour les deux populations, les taux de recombinaisons lors de la méiose entre différents groupes de liaisons et au sein de ces groupes semblent corrects. Les zones en rouge représentent les zones de fortes recombinaisons tandis que celles en bleu recombinent peu.

Les analyses de base sur les données phénotypiques ont donné des résultats assez semblables pour les deux espèces. Les "Box-Plots" témoignent d'un effet ontogénique pour les deux populations. Globalement, la production de fruits augmente régulièrement d'années en années, suivant la croissance de l'arbre.

Le calcul de l'héritabilité des caractères révèle que toutes les variables étudiées sont héréditaires. Cependant, les données phénotypiques de l'olivier le sont dans une moindre mesure par rapport à celles du pommier. Cela peut s'expliquer par la différence entre la variable rendement considérée pour le pommier (le nombre total de fruits) et celle prise pour l'olivier (le poids du total des fruits). La variable étudiée dans ce deuxième cas est bien plus complexe et donc moins héréditaire. Comme elles restent au-dessus du seuil de 0.2, les

variables étudiées sont tout de même considérées comme suffisamment hérissables pour être analysées en complément de données génétiques.

Les graphiques représentant les interactions entre les génotypes, les années et le rendement, permettent de visualiser aisément le comportement global des parcelles. Néanmoins, l'analyse de ces graphiques reste assez superficielle étant donné le grand nombre de génotypes qui apparaissent et qui sont difficilement distinguables.

Trois comportements de génotypes ont malgré tout été identifiés après de plus amples analyses statistiques. Les groupes formés sur les nuages de points ne sont cependant pas absolus. Ils sont formés empiriquement selon le critère du placement des génotypes sur le graphique en fonction du BBI calculé avec les résidus normés et de l'autocorrélation de ces résidus. Trois types distincts sont mis en évidence : alternant, régulier et irrégulier.

### 1.2.3. Les QTLs identifiés

Dans la Table 9 récapitulant les QTLs repérés sur la nouvelle carte de la population « Olivière x Arbequina », les QTLs sont relativement dispersés en comparaison de ceux détectés sur « X3263 x Belrène ».

Le QTL de rendement trouvé sur le chromosome 9 de l'olivier est positionné au niveau des mêmes marqueurs SSR sur lesquels un QTL avait été détecté en 2012 malgré la différence entre les cartes (Ben Sadok et al., 2013). Cette similitude valide la pertinence des deux cartes génétiques réalisées pour cette population. La position en centiMorgan n'est toutefois pas la même, cela est dû à la plus grande précision de la carte saturée en marqueurs.

Pour le pommier, les QTLs sont plus concentrés, ils se situent principalement sur les groupes de liaisons 3, 8 et 10. Cependant, ces QTLs trouvés sur la nouvelle carte ne correspondent que partiellement à ceux détectés sur la population « Starkrimson x Granny Smith » (Guitton et al., 2012). Bien que certaines zones d'intérêts semblent communes et se positionnent sur les mêmes groupes de liaisons, elles se situent à plusieurs dizaines de centiMorgans d'écart par rapport à l'ancienne carte. Cette différence est sans doute due, non seulement à la plus grande saturation de la nouvelle carte SNP mais également au changement de population. Ces deux croisements ne présentent aucun lien de parenté.

Tous les QTLs détectés n'ont pas la même importance dans l'expression des caractères observés. La valeur du  $R^2$  indique le degré explicatif du QTL. Ainsi, certaines zones détectées à un même LOD n'expliquent pas toutes autant la variable étudiée. Il est donc préférable de se fier au pourcentage de variation expliqué par le QTL pour en connaître l'importance.

De même, des différences sont observées entre le pommier et l'olivier. Aucun QTL n'est relevé au-delà du LOD 5.88 pour l'olivier, ce qui est relativement faible en comparaison du pommier pour lequel un QTL est détecté au LOD 14.04 comme représenté sur la Figure 27. Ce QTL, situé sur le groupe de liaison 10, explique à lui seul 29.9% de la variation du caractère « Rendement ». Cette valeur explicative est relativement élevée et souligne l'importance de ce QTL.

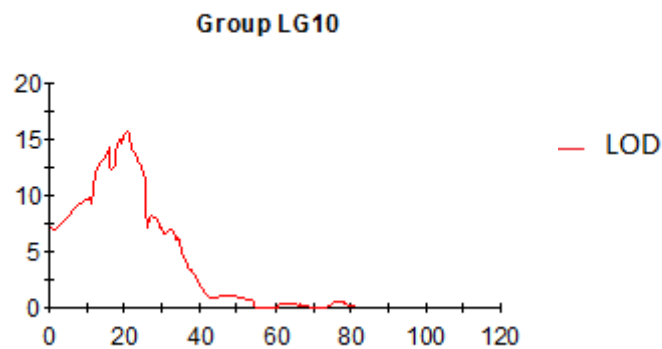


Figure 27 : Valeur du LOD en fonction de la position en cM sur le groupe de liaison 10 du pommier de la descendance "X3263 x Belrène"

#### 1.2.4. Les gènes candidats

Dans cette première analyse, il n'a pas été possible de déterminer précisément quels étaient les gènes ayant un véritable rôle dans les caractères d'alternances et de rendements. Seules des catégories de gènes chez le pommier ont pu être relevées. Les résultats sont répertoriés selon de grandes catégories et le nombre de gènes parmi ces groupes sont comptés. Même si cette information permet de savoir quelle catégorie de gènes est la mieux représentée sous le QTL étudié, cela reste assez superficiel et ne permet pas de désigner des gènes ou fonctions en particulier qui justifieraient la présence d'un QTL à cet emplacement.

Pour l'olivier, des annotations de gènes sur les marqueurs ont été relevées. En revanche, l'état d'avancement actuel des recherches ne permet pas vraiment d'aller plus loin dans la recherche des gènes impliqués pour cette espèce.

Les résultats obtenus suite le travail effectué appellent plusieurs propositions susceptibles de les améliorer.

## 2. Propositions d'amélioration

### 2.1. Ré-estimation de la variable « Rendement » chez l'olivier

Afin d'homogénéiser les données phénotypiques, la variable « Rendement » considérée pour l'olivier comme étant le poids total du nombre de fruits total par arbre, peut être ré-estimée. Les fiches élaborées lors des récoltes contiennent notamment le poids total de la récolte et le poids de 100 fruits rapporté au poids au fruit par génotype. La Table 12 illustre les différentes données disponibles par génotype, par répétition et par année :

Table 12: Données phénotypiques relevées par arbre et par année pour la population "Olivière x Arbequina »

Numarbre	R	genotype	Année	poids recolte (g)	nb fruits sol	poids de 100 fruits (g)	poids moyen 1fruit (g)	poids au sol(g)	RDT (g)
2667	1	O1	2011	880	400	406	4.06	1624	2504

Ainsi, un nouveau rendement composé uniquement du nombre total de fruits récoltés et au sol peut être calculé avec les données à disposition.

Nombre de fruits total = RDT (g) / poids moyen de 1 fruit (g)

A titre d'exemple : 2504 g / 4.06 g = 616.75 fruits

L'arbre 2667, répétition 1 du génotype O1, a donc produit 617 fruits en 2011.

Une nouvelle colonne peut alors être ajoutée au tableau et calculée automatiquement comme montré sur la Table 13.

Table 13 : Données phénotypiques disponibles pour l'olivier avec le Nombre total de fruits ajouté

Numarbre	R	Génotype	Année	Poids recolte (g)	Nb fruits au sol	Poids de 100 fruits (g)	Poids moyen de 1 fruit (g)	Poids au sol(g)	RDT (g)	Nb de fruits total
2667	1	O1	2011	880	400	406	4.06	1624	2504	617

En remplaçant l'ancienne variable de rendement (RDT\_g) par le nombre de fruits total dans les analyses statistiques avec R, il est possible que les résultats changent. En effet, les analyses concernant le rendement comme celles relatives à l'alternance utilisent cette variable pour les calculs. Il n'est donc pas impossible qu'avec ce changement, l'héritabilité du caractère du rendement d' « Olivière x Arbequina » se trouve améliorée et que les QTLs soient modifiés, peut-être moins dispersés sur et entre les groupes de liaisons.

## 2.2. Identification des gènes plus précisément

La découverte des gènes sous les QTLs qui a été effectuée a permis d'avoir un premier aperçu des informations génétiques présentes sous les zones génomiques détectées.

Pour la population de pommiers « X3263 x Belrène », une liste exhaustive de gènes potentiellement impliqués a pu être élaborée. De manière à aller plus loin, des gènes impliqués dans la réponse de la plante à divers stimuli, dans la synthèse d'enzymes ou le transport de sucres ou d'hormones jouant un rôle dans l'alternance de production pourront être identifiés. En effet, des hormones telles les auxines, les cytokinines, les gibérellines mais aussi l'acide abscissique et l'éthylène peuvent intervenir à un moment ou à un autre dans l'expression du rendement et de la régularité de production chez l'arbre fruitier et ainsi induire l'année « ON » ou « OFF » de la plante (Hanke et al., 2007 ; Guitton et al., 2013). Les gènes qui codent pour ces enzymes qui régulent le métabolisme des hormones, leur perception au niveau du méristème apical des bourgeons ou encore qui permettent d'en assurer le transport, pourront être repérés. La connaissance des mécanismes en jeu sera améliorée.



La même méthodologie utilisée pour définir les gènes candidats pour la floraison et le rendement chez « Starkrimson x Granny Smith » peut être mise en œuvre (Guitton et al., 2011). Pour cela, une liste exhaustive des gènes impliqués dans l'induction et la différenciation florale du pommier et des gènes liés aux hormones et potentiellement impliqués dans l'alternance de production a été réalisée. Des séquences de protéines agissant dans le développement des fleurs chez *Arabidopsis thaliana* sont ensuite extraites de la base de données NCBI (<http://www.ncbi.nlm.nih.gov/>) (NCBI, 2014). Un certain nombre de gènes impliqués dans la réponse de la plante, la synthèse et le transport des auxines, cytokinines ainsi que de l'acide abscissique, l'éthylène et les gibbérellines a également été sélectionné.

La liste de gènes sous les QTLs les plus déterminants dans l'explication du caractère doivent être comparés à la liste de gènes potentiellement impliqués dans le processus d'induction floral qu'ont fait Guitton et al. (2012) et qui étaient positionnés en partie sous les QTLs de la population « Starkrimson X Granny Smith ». De plus, des données d'expression de gènes ont également été produites à partir d'une puce micro-array représentant l'ensemble des gènes du pommier (Thèse Guitton, 2011). La comparaison de ces données devrait permettre de mieux appréhender les mécanismes et les gènes candidats liés à l'alternance pour cette espèce.

Dans le cas de la population « Olivière x Arbequina », il faudra attendre le séquençage complet du génome de l'olivier pour obtenir une liste exhaustive et précise des gènes potentiellement impliqués et leur position sur le génome.

### **3. Les perspectives de l'étude**

#### **3.1. Perspectives imminentes**

##### **3.1.1. Changements sur les parcelles**

En 2014, après la perte de la récolte des olives du fait de l'attaque de mouches de l'olivier (*Bactrocera oleae*), il a été décidé de réaliser une opération de taille sur les oliviers de la parcelle à Mauguio. Plantés en 2005, ceux-ci n'avaient depuis lors jamais été taillés. En n'intervenant pas sur la parcelle, le but était de pouvoir observer leur croissance et leur architecture dans un environnement naturel. Cependant, la collecte des fruits par les équipes de terrain de l'INRA à des fins de recherche sur l'architecture ainsi que sur la production est devenue de plus en plus difficile au fil des années. Les branches et le feuillage des individus sont devenus de plus en plus denses rendant ainsi impraticable l'accès à certaines branches et certains fruits. Cette taille va faciliter la récolte des fruits mais en revanche va entraîner une rupture dans la continuité du phénotypage sur la parcelle.

### 3.1.2. Un article scientifique en construction

La méthode mise en place pour construire la cartographie génétique avec les marqueurs SNP sur la population d'oliviers « Olivière x Arbequina » va faire l'objet d'une publication scientifique en cours de montage par l'équipe AFEF. En effet, la technique de reconstruction des marqueurs bialléliques en marqueurs multialléliques est inédite. Développée par l'équipe ID de l'UMR AGAP et testée au cours de ce stage durant l'étape de la cartographie génétique, il a été mis en évidence, en comparant les deux méthodes, que cette technique apporte une vraie plus-value pour la cartographie génétique. Un article détaillant cette nouvelle méthode basée sur la reconstruction des marqueurs détectés sur le transcriptome de l'olivier va donc être rédigé par Bouchaïb Khadari, chercheur de l'équipe AFEF, en partenariat avec Gauthier Sarah et Hajar Chouiki de l'équipe ID. La publication de cette technique pourra permettre à d'autres équipes de s'en servir pour optimiser leur cartographie en réduisant le nombre de marqueurs tout en augmentant l'information contenue dans les marqueurs dits « reconstruits ». Ces derniers regroupent en un seul marqueur multiallélique plusieurs marqueurs bialléliques. Grâce à cette méthode, les cartes génétiques de la population « Olivière x Arbequina » sont un peu plus réduites que lorsque seuls des marqueurs bialléliques sont utilisés et semblent plus fiables.

### 3.1.3. Utilisation des cartes du pommier dans une thèse

Les cartes élaborées pour la population « X3263 x Belrène », en plus de servir à la détection de QTLs de rendement et d'alternance de production fruitière pour ce projet et de permettre la comparaison avec les résultats déjà obtenus sur la population « Starkrimson x Granny Smith », vont pouvoir être utilisées dans le cadre de la thèse d'Alix Allard, en 3<sup>ème</sup> année de doctorat. Cette thèse porte sur l'étude de plusieurs populations de pommiers présentant des liens familiaux. Des QTLs correspondant à d'autres caractères vont également pouvoir être détectés sur cette même carte.

De plus, les caractères de rendement et d'alternance vont potentiellement être étudiés sur les autres populations de pommiers pour étayer la comparaison de ces variables phénotypiques entre des populations connectées. Cela peut être intéressant pour observer les différences et les points communs entre les zones génomiques détectées pour ces caractères.

## 3.2. Perspective à long terme : La sélection de variétés régulières

La compréhension des mécanismes génétiques qui conduisent à l'alternance de production chez le pommier et l'olivier est primordiale pour envisager de planter un jour des variétés régulières de ces deux espèces.

L'alternance est un phénomène déjà maîtrisé par les diverses opérations culturales qui peuvent être menées sur les parcelles des arbres présentant ce caractère. Cependant,

proposer des variétés régulières en amont de la filière permettrait un gain de temps dans l'utilisation du matériel agricole et des produits phytosanitaires nécessaires à la gestion du phénomène. Des variétés naturellement régulières existent déjà mais ne correspondent pas aux variétés de pommes les plus populaires.

La sélection de nouvelles variétés de pommiers en France se déroule à l'INRA d'Angers. Contrairement aux espèces annuelles dont les générations se succèdent d'une année sur l'autre, pour les espèces pérennes, il faut attendre plusieurs années avant qu'une nouvelle génération entre en production. Ainsi, tout le processus de sélection jusqu'au lancement de la nouvelle variété peut prendre jusqu'à 20 ans. Une sélection assistée par marqueurs serait donc envisageable pour le caractère de l'alternance, associé à d'autres caractères gustatifs et qualitatifs pour obtenir un fruit qui puisse concurrencer les pommes vedettes du marché tout en facilitant le travail du producteur.

Les oliviers, quant à eux, ne sont pas sélectionnés en France mais le sont dans d'autres pays méditerranéens comme l'Espagne, la Grèce et l'Italie (Fabbri et al., 2009). Ainsi, des équipes de chercheurs sélectionnent des oliviers présentant une bonne quantité et une bonne qualité d'huile mais les aspects d'alternance n'ont pas encore été pris en compte dans les schémas de sélection (Hatzopoulos et al., 2002). Pour cette espèce également, des variétés comme Olivière, le parent femelle de la population étudiée pour ce projet, présentent une production de fruits régulière. Pour les variétés alternantes, la conduite culturale est ici aussi primordiale.

La compréhension du mécanisme induisant l'alternance par positionnement de QTLs et de gènes candidats impliqués dans l'expression du caractère du rendement permettra d'incorporer ces variables dans les schémas de sélections assistés par marqueurs moléculaires.

Pour le moment, il y a peu d'applications concrètes et immédiates de l'étude, cependant, il s'agit d'un travail effectué en amont et nécessaire tant pour l'avancement des connaissances scientifiques dans ce domaine que pour une éventuelle sélection de variétés présentant les caractéristiques de régularités souhaitées.

## CONCLUSION

Les progrès en génétique au cours du siècle dernier ont ouvert la voie à des études approfondies sur la face cachée du vivant, le patrimoine génétique. Le phénomène de l'alternance, commun à de nombreuses espèces pérennes dont le pommier et l'olivier, est déterminé génétiquement de manière complexe par de nombreux mécanismes qui entrent en jeu. Les différentes zones génomiques impliquées dans ce caractère ont été mises en évidence sur les nouvelles cartes génétiques saturées élaborées. Cet outil a permis une comparaison avec d'autres résultats obtenus précédemment tout en améliorant la précision de la détection des QTLs. Des catégories de gènes ont ainsi pu être listés sous les zones d'intérêt repérées. Le déterminisme génétique des caractères de rendement et d'alternance est caractérisé par ces zones. Une recherche approfondie des gènes responsables de l'alternance permettra d'affiner ces résultats.

Une fois les gènes impliqués validés et leurs fonctions connues pour les deux espèces étudiées, les différences et les similitudes entre elles pourront être comparées. L'alternance de la production est un caractère très stable et commun à des espèces fruitières même lorsqu'elles sont de familles différentes. Ici, le pommier, issu de la famille des Rosacées, et l'olivier de la famille des Oléacées présentent tous les deux ce trait phénotypique. L'une des hypothèses possible est alors que l'alternance de production est régie génétiquement par des gènes très stables sur le génome qui seraient restés les mêmes depuis leur dernier ancêtre commun. La validation de cette théorie faciliterait la sélection par marqueurs assistée de l'olivier et du pommier dans un premier temps, suivis éventuellement par d'autres espèces présentant le même déterminisme pour la régularité de production. La poursuite des recherches dans ce domaine confirmera ou infirmera cette hypothèse.



## Bibliographie

AFIDOL, 2014. Histoire [en ligne]. Disponible sur :  
<http://www.afidol.org/component/content/article/4-grand-public/olivier/12-histoire>  
(consulté le 16/02/2015).

ANTANAVICIUTE L., FERNANDEZ-FERNANDEZ F., JANSEN J., BANCHI E., EVANS K.M., VIOLA R., VELASCO R., DUNWELL J.M., TROGGIO M. and SARGENT D.J., 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, 10 p.

ARCAD, 2010. Agropolis Resource Center for Crop Conservation, Adaptation and Diversity (ARCAD) [en ligne]. Disponible sur :  
<http://umr-agap.cirad.fr/domaines-thematiques/dynamique-de-la-diversite-et-domestication/projet-arcad> (consulté le 17/03/2015).

AVERY O.T., MACLEOD C.M. and MCCARTY M., 1944. Studies of the chemical nature of the substance inducing transformation of pneumococcal types, Induction of transformation by a deoxyribonucleic acid fraction isolated from pneumococcus type III. *Journal of Experimental Medicine*, 79, pp. 137-158.

BATES D., MAECHLER M. and BOLKER B., 2011. lme4: Linear mixed-effects models using Eigen and syntax. *R package version 0.999375-39*.

BEN SADOK I., 2013. « Developmental changes of the Olive tree from juvenility to maturity and Genetic basis of vegetative and reproductive traits », Thèse de doctorat en biologie sous la direction de Evelyne Costes et Ahmed rebai, Montpellier : Ecole doctorale SIBAGHE, 261 p.

BEN SADOK I., CELTON J-M, ESSALOUH L, EL AABIDINE A.Z., GARCIA G, MARTINEZ S., GRATI-KAMOUN N., REBAI A., COSTES E. and KHADARI B., 2013. QTL Mapping of Flowering and Fruiting Traits in Olive, *PLoS One*, 8 (5), 16p.

BESNARD G., KHADARI B., VILLEMUR P. and BERVILLE A., 2000. Cytoplasmic male sterility in the olive (*Olea europaea* L.). *Theoretical and Applied Genetic*, 100, pp. 1018-1024.

BROMAN K.W., WU H., SEN S and CHURCHILL G.A., 2003. R/qtl: QTL mapping in experimental crosses. *Bioinformatics*, 19, pp. 889-890.

BURNHAM K.P. and ANDERSON D.R., 2004. Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods and Research*, 33 (2), pp. 261-304.

CAMPBELL M.A., 2001. Microsatellite DNA Methodology [en ligne]. Disponible sur :  
<http://www.bio.davidson.edu/genomics/method/microsatellite.html> (consulté le 23/02/15).

CELTON J-M., MARTINEZ S., JAMMES M-J., BECHTI A., DUREL C-E., SALVI S., LEGAVE J-M. and 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 Physiologist*, 192, pp. 378-392.

CNRS, 2012. Groupement d'intérêt scientifique (GIS) [en ligne]. Disponible sur : <https://www.dgdr.cnrs.fr/daj/partenariat/partenariat/structcontractuelles/gis3.htm> (consulté le 19/01/2015).

COMMISSION EUROPÉENNE, 2013. Plan d'action sur le secteur de l'huile d'olive de l'UE. Document de travail des services de la DG AGRI, 8 p.

CONSEIL OLEICOLE INTERNATIONAL, 2014. Chiffres du marché mondial des olives de table [en ligne]. Disponible sur : [http://www.internationaloliveoil.org/estaticos/view/132-world-table-olive-figures?lang=fr\\_FR](http://www.internationaloliveoil.org/estaticos/view/132-world-table-olive-figures?lang=fr_FR) (consulté le 25/04/2015).

CSIC, 2013. Premier séquençage complet du génome de l'olive [en ligne]. Disponible sur : <http://www.bulletins-electroniques.com/actualites/74626.htm> (consulté le 16/02/2015).

DARWIN C., 1859. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*, London : John Murray, Albemarle Street, 502 p.

DIRLEWANGER E., PRONIER V., PARVERY C., ROTHAN C., GUYE A. and MONET R., 1998. Genetic linkage map of peach [*Prunus persica* (L.) Batsch] using morphological and molecular markers. *Theoretical Applied Genetics*, 97, pp. 888-895.

DUMAS P., 2014. Pommier pajam 1 [en ligne]. Disponible sur : <http://www.fruitiers.net/fiche.php?NumFiche=1111> (consulté le 19/02/2015).

DURAND J-B, GUITTON B., PEYHARDI J., HOLTZ Y., GUEDON Y., TROTTIER C. and 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(16), pp. 5099-5113.

DUREL C-E., LASSERRE-ZUBER P., LE VAN A., LE CAM B., CAFFIER V. et SAPOUKHINA N., 2012. Durabilité de la résistance du pommier à la tavelure (*Venturia inaequalis*). *Le sélectionneur français*, 63, pp. 73-76.



EL AABIDINE A. Z., CHARAFI J., GROUT C., DOLIGEZ A., SANTONI S., MOUKHLI A., JAY-ALLEMAND C., EL MODAFAR C. and KHADARI B., 2009. Construction of a Genetic Linkage Map for the Olive Based on AFLP and SSR Markers. *Crop Science*, 50 (6), 2291-2302.

FABBRI A., LAMBARDI M. and OZDEN-TOKATLI Y., 2009. Olive Breeding, In : JAIN S.M. and PRIYADARSHAN P.M., 2009. Breeding Plantation Tree Crops: Tropical Species, New York : Springer, pp. 423-465.

FNPF, 2012. Chiffres clés [en ligne]. Disponible sur : <http://www.fnpfuits.com/sites/fnpfweb/chiffresK/productions/#Pomme> (consulté le 13/02/2015).

FORSLINE P.L. et ALDWINCLE H.A., 2003. Evaluation of *Malus sieversii* seedling populations for disease resistance and horticultural traits. *Acta Horticulturae*, 663, pp. 529-534.

FRANCE AGRIMER, 2012. Le marché de l'huile d'olive, 2011/2012 [en ligne]. Disponible sur : <http://www.franceagrimer.fr/content/download/17461/136571/file/Le%20march%C3%A9%20de%20huile%20d%27olive.pdf> (consulté le 24/04/2015).

GALLAIS A., 2009. *Hétérosis et variétés hybrides en amélioration des plantes*, Versailles : Edition QUAE, 356p.

GDR, 2012. *Malus x domestica* Whole Genome v1.0 Assembly & Annotation [en ligne]. Disponible sur : [http://www.rosaceae.org/species/malus/malus\\_x\\_domestica/genome\\_v1.0](http://www.rosaceae.org/species/malus/malus_x_domestica/genome_v1.0) (consulté le 16/03/2015).

GIRARD F. et NOIVILLE C., 2014. *Biotechnologies végétales et propriété industrielle*, Paris : La documentation Française, 187p.

GNIS, 2014. Les marqueurs SNP [en ligne]. Disponible sur : <http://www.gnis-pedagogie.org/biotechnologie-amelioration-marqueur-snp.html> (consulté le 08/03/2015).

GNIS, 2015. Les marqueurs AFLP [en ligne]. Disponible sur : <http://www.gnis-pedagogie.org/biotechnologie-amelioration-marqueur-aflp.html> (consulté le 06/03/2015).

GOMES S., MARTINS-LOPES P. And GUEDES-PINTO H., 2012. Olive Tree Genetic Resources Characterization Through Molecular Markers. In: Genetic Diversity in Plants. Rijeka : InTech Europe, 498 p.

GUITTON B., 2011. Genetic control of Biennial Bearing in Apple, Thèse de doctorat en biologie fonctionnelle des plantes sous la direction de Evelyne Costes et David Chagné, Montpellier : Ecole Doctorale SIBAGHE, 236 p.

GUITTON B., KELNER J.J., VELASCO R., GARDINER S.E., CHAGNÉ D., and COSTES E., 2012. Genetic control of biennial bearing in apple. *Journal of Experimental Botany*, 63(1), pp. 131-149.

GUITTON B., CHAGNE D. et COSTES E., 2013. Génétique de l'alternance de production chez le pommier : vers la sélection de cultivars ayant une production régulière et la compréhension des mécanismes physiologiques responsables de l'induction florale [en ligne]. Disponible sur : [http://jejardine.org/images/stories/Les%20coulisses%20de%20la%20floraison-B\\_Guitton.pdf](http://jejardine.org/images/stories/Les%20coulisses%20de%20la%20floraison-B_Guitton.pdf) (consulté le 12/04/2015)

HANKE M.V., FLACHOWSKY H., PEIL A. and HATTASCH C., 2007. No Flower no Fruit – Genetic Potentials to Trigger Flowering in Fruit Trees. *Genes, Genomes and Genomics*, 1(1), 20p.

HATZOPOULOS P., BANILAS G., GIANNOULIA K., GAZIS F., NIKOLOUDAKIS N., MILIONI D. and HARALAMPIDIS K., 2002. Breeding, molecular markers and molecular biology of the olive tree. *European Journal of Lipid Science and Technology*, 104, pp. 574-586.

HEBERLE-BORS E., 2001. Génie Génétique, une histoire, un défi, Paris : INRA Editions, 290 p.

HENDERSON C.R., 1950. Estimation of genetic parameters. *The Annals of Mathematical Statistics*, 9, 309 p.

HOBLYN T, GRUBB N, PAINTER A. and WATES B., 1936. Studies in biennial bearing. *Journal of Pomology Horticultural Science*, 14, pp. 39-76.

HORROCKS L.A. and YEO Y.K., 1999. Health benefits of docosahexaenoic acid (DHA). *Pharmacological Research*, 40, pp. 211-225.

IAEA, 2010. Belrène [en ligne]. Disponible sur : <http://mvgs.iaea.org/Search.aspx?ID=270> (consulté le 19/02/2015).

IFPC, 2012. Etude des différents modes de conduite de l'arbre [en ligne]. Disponible sur : <http://www.ifpc.eu/programmes-de-recherche/verger/etude-de-differents-modes-de-conduite-de-larbre.html> (consulté le 20/01/2015).

ILLUMINA, 2006. Infinium® II Assay Workflow. San Diego : Illumina Inc., 2 p.

ILLUMINA, 2011. *HiSeq™ Sequencing Systems*, San Diego : Illumina Inc., 4 p.

INRA, 2009. UE0398 MELGUEIL DIASCOPE - Domaine Expérimental de Melgueil [en ligne]. Disponible sur : <http://annuaire.inra.fr/afficherStructure.action?code=0398&type=SO> (consulté le 19/02/2015).

INRA, 2011. Le Groupement d'Intérêt Scientifique Fruits [en ligne]. Disponible sur : <http://www.gis-fruits.org/> (consulté le 17/01/2015)

INRA, 2013. Les chiffres clés [en ligne]. Disponible sur : <http://institut.inra.fr/Reperes/Chiffres> (consulté le 19/01/2015).

JOHANNSEN W., 1909. *A Century of DNA*. Cambridge : MIT Press, 118 p.

KENIS K. and KEULEMANS J., 2004. Genetic linkage maps of two apple cultivars (*Malus domestica* Borkh.) based on AFLP and microsatellite markers. *Molecular Breeding*, 15 (2), pp. 205-219.

KOSAMBI D.D., 1943. The estimation of map distances from recombination values. *Annals of Eugenics*, 12 (1), pp. 172-175.

LA ROSA R., ANGIOLILLO A., GUERRERO C., PELLEGRINI M., RALLO L., BESNARD G., BERVILL A., MARTIN A. and BALDONI L., 2003. A first linkage map of olive (*Olea europaea* L.) cultivars using RAPD, AFLP, RFLP and SSR markers. *Theoretical and Applied Genetics*, 106, 1273-1282.

LANE J.A., 1994. History of Genetics Timeline [en ligne]. Disponible sur : <http://www.accessexcellence.org/AE/AEPC/WWC/1994/geneticstln.php> (consulté le 03/02/2015)

LANDOUER P-Y., 2015. Pommier [en ligne]. Disponible sur : <http://www.lesarbres.fr/pommier.html> (consulté le 16/02/2015).

LARRIEU J-F., 2007. Maîtriser l'alternance des variétés de pommes en bio. *Réussir Fruits et Légumes*, 264, pp. 42-43.

LAURENS F., 2013. Histoire, génétique et sélection variétale du pommier (à couteau), *FormaVie* 5/04/2013, 70p.

LAVEE S., AVIDAN B. and BEN-ARI G., 2014. Trends in Breeding New Olive Varieties in Israel for Quality and Economic Management. *Agricultural Science*, 5, pp. 701-709.

LE POINT, 2014. La mouche de l'olive compromet la production oléicole en Provence [en ligne]. Disponible sur : [http://www.lepoint.fr/economie/la-mouche-de-l-olive-compromet-la-production-oleicole-en-provence-02-12-2014-1886094\\_28.php](http://www.lepoint.fr/economie/la-mouche-de-l-olive-compromet-la-production-oleicole-en-provence-02-12-2014-1886094_28.php) (consulté le 06/03/2015).

LEON L., LA ROSA R., BARRANCO D. and RALLO L., 2007. Breeding for Early Bearing in Olive. *HortScience*, 42(3), 499-502.

LEVENE H., 1960. Robust testes for equality of variances. In Contributions to Probability and Statistics. Stanford : I. Olkin, Stanford University Press, pp. 278-292.

LUCAS P., 2014. Phenoarch [en ligne]. Disponible sur : [https://www.phenome-fppn.fr/phenome\\_eng/Facilities/Montpellier-Controlled/PHENOARCH](https://www.phenome-fppn.fr/phenome_eng/Facilities/Montpellier-Controlled/PHENOARCH) (consulté le 12/02/2015).

LIEBHARD R., KOLLER B., GIANFRANCESCHI L. and GESSLER C., 2002. Creating a saturated reference map for the apple (*Malus domestica* Borkh.) genome. *Theoretical and Applied Genetics*, 106 (8), pp. 1497-1508.

MANDAL A., 2013. History of Genetics [en ligne]. Disponible sur : <http://www.news-medical.net/health/History-of-Genetics.aspx> (consulté le 19/01/2015).

MENDEL G., 1866. Versuche uber Pflanzenhybriden. *Verhandlungen des naturforschenden Vereines in Brünn, Bd. IV für das Jahr 1865*, Abhandlungen, pp. 3-47.

MIESCHER F., 1869. Letter I; to Wilhelm His; Tübingen, February 26th, 1869. In: HIS W., et al. (Eds.), Die Histochemischen und Physiologischen Arbeiten von Friedrich Miescher—Aus dem wissenschaftlichen Briefwechsel von F. Miescher. Leipzig : F.C.W. Vogel, 1, pp. 33-38.

MIKO I., 2008. Gregor Mendel and the principles of inheritance. *Nature Education*, 1(1), 134 p.

MIGALE, 2014. Le cluster de calcul [en ligne]. Disponible sur : <http://migale.jouy.inra.fr/?q=cluster> (consulté le 06/03/2015).

MINISTERE DE L'EDUCATION NATIONALE, DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE, 2013. Le financement de la recherche [en ligne]. Disponible sur : <http://www.enseignementsup-recherche.gouv.fr/cid56386/qui-finance-la-recherche.html> (consulté le 19/01/2015).

MOBATEK, 2015. MobaXterm [en ligne]. Disponible sur : <http://mobaxterm.mobatek.net/> (consulté le 17/01/2015).

MONSELISE S.P. and GOLDSCHMIDT E.E., 1982. Alternate Bearing in Fruit Trees. *Horticultural Reviews*, 4, pp. 128-173.

MORTON N.E., 2004. History of genetic mapping. In : *Encyclopedia of Genetics, Genomics, Proteomics and Bioinformatics, Part 2. Genomics*, Hoboken : John Wiley & Sons, Ltd, 115 p.

MUSEO DELL'OLIVO, 2012. L'olivier à travers 6000 années d'Histoire [en ligne]. Disponible sur : <http://www.museodellolivo.com/fr/frstor1.htm> (consulté le 16/02/2015).

NATHAN P., 2010. La pomme, ses atouts santé. Paris : Editions Expressions Santé, 4p.

NCBI, 2014. Databases [en ligne]. Disponible sur : <http://www.ncbi.nlm.nih.gov/> (consulté le 3/04/2015).

OMAG SAP., 2011. L'éclaircissage des pommiers, *Dossier Technique*, Mollégès : OMAG SAP, 4p.

PAILLOTIN G., 2000. L'agriculture raisonnée. Rapport au Ministre de l'Agriculture et de la Pêche, Paris : INRA, 57p.

PARVEAUD C-E., GOMEZ C., LIBOUREL G., ROMET L., WARLOP F., SIMON S., BRUN L., POUZOULET D., DELEBECQ A., TOURNANT L., OSTE S. and LAURENS F., 2010. Assessment of the susceptibility to pests and diseases of 36 apple cultivars in four low-input organic orchards in France. 14th International Conference on Organic Fruit Growing, Ecofruit. February 2010, Hohenheim University, 7p.

PEIX C., 2010. Aux origines de la pomme ou le jardin d'Eden retrouvé [DVD]. Strasbourg : Seppia.

PINHEIRO J.C. and BATES D.M., 2000. *Mixed-Effects Models in S and S-Plus (Statistics and Computing)*. New York : Springer, 528p.

PEREIRA-LORENZO S., RAMOS-CABRER A.M. and FISCHER M., 2009. Breeding Apple (*Malus x Domestica* Borkh). In : JAIN S.M. and PRIYADARSHAN P.M., 2009. *Breeding Plantation Tree Crops: Temperate Species*. New York : Springer, pp. 33-83.

PESSON P. et LOUVEAUX J., 1984. *Pollinisation et productions végétales*. Versailles : Editions QUAE, 640 p.

PROJET EUROPEEN, 2010. Fruit Breedomics [en ligne]. Disponible sur :

[http://ec.europa.eu/research/bioeconomy/agriculture/projects/fruit\\_breedomics\\_en.htm](http://ec.europa.eu/research/bioeconomy/agriculture/projects/fruit_breedomics_en.htm)  
(consulté le 26/02/2015).

PURUGGANAN M.D. et FULLER D.Q., 2009. The nature of selection during plant domestication. *Nature*, 457, pp. 843-848.

R FOUNDATION, 2015. The R Project for Statistical Computing [en ligne]. Disponible sur : <http://www.r-project.org/> (consulté le 17/01/2015).

R STUDIO, 2014. R Studio [en ligne]. Disponible sur : <http://www.rstudio.com/> (consulté le 06/03/2015).

ROBINSON G.K., 1991. That BLUP Is a Good Thing: The Estimation of Random Effects. *Statistical Science*, 6 (1), pp. 15-51.

ROUECHE, 1997. Etat des ressources génétiques du pommier [en ligne]. Disponible sur : <http://pomologie.com/pomme1/datapom/types.htm> (consulté le 10/02/2015).

SEGURA V., CILAS C., LAURENS F. and COSTES E., 2006. Phenotyping progenies for complex architectural traits: a strategy for 1-year-old apple trees (*Malus x domestica* Borkh.). *Tree Genetics & Genomes*, 2 (3), pp. 140-151.

SEGURA V., CILAS C. and 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, pp. 302-314.

SHAPIRO S.S. and WILK M.B., 1965. An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, 52 (3/4), pp. 591-611.

SPIEGEL-ROY P., 1990. Economic and agricultural impact of mutation breeding in fruit trees. *Mutation Breeding Review*, 5, 26 p.

TAIR, 2005. GO annotation search, functional categorization and download [en ligne]. Disponible sur : <http://www.arabidopsis.org/tools/bulk/go/index.jsp> (consulté le 17/03/2015).

TELA-BOTANICA, 2012. *Malus domestica* Borkh [en ligne]. Disponible sur : <http://www.tela-botanica.org/bdtfx-nn-40744> (consulté le 10/03/2015).

TROTTE M., 2008. La génétique végétale, quels enjeux pour demain ? In : Mardis de l'Espace des sciences, Rennes, Espaces de Sciences, 7 octobre 2008. Rennes : INRA.

UMR AGAP, 2012 (a). Amélioration génétique et adaptation des plantes méditerranéennes et tropicales [en ligne]. Disponible sur : <http://umr-agap.cirad.fr/> (consulté le 12/02/2015).

UMR AGAP, 2012 (b). Publications clés de l'équipe [en ligne]. Disponible sur : <http://umr-agap.cirad.fr/equipes-scientifiques/architecture-et-fonctionnement-des-especes-fruiteres/publications-cles> (consulté le 22/03/2015).

UMR AGAP, 2014. Une démarche qualité [en ligne]. Disponible sur : <http://umr-agap.cirad.fr/l-unite/demarche-qualite> (consulté le 16/02/2015).

USDA, 2014. Interactive Maps on GrainGenes [en ligne]. Disponible sur : [http://wheat.pw.usda.gov/ggpages/map\\_shortlist.html](http://wheat.pw.usda.gov/ggpages/map_shortlist.html) (consulté le 11/02/2015).

VAN OOIJEN J.W., 2006. JoinMap®4.1, Software for the calculation of genetic linkage maps in experimental populations. Wageningen : Kyazma B.V., 63 p.

VAN OOIJEN J.W., 2009. MapQTL®6, Software for the mapping of quantitative trait loci in experimental populations of diploid species. Wageningen : Kyazma B.V., 64 p.

VIA S. and LANDE R., 1985. Genotype-Environment Interaction and the Evolution of Phenotypic Plasticity. *Evolution*, 39 (3), pp. 505-522.

WATSON J.D. and CRICK F.H.C., 1953. A Structure for Deoxyribose Nucleic Acid. *Nature*, 171, pp. 737-738.

WU S-B., COLLINS G. and SEDGLEY M., 2004. A molecular linkage map of olive (*Olea europaea* L.) based on RAPD, microsatellite, and SCAR markers. *Genome*, 47 (1), pp. 26-35.

ZENG Z-B., 1994. Precision Mapping of Quantitative Trait Loci. *The Genetics Society of America*, 136, pp. 1457-1468.

## TABLE DES FIGURES

Figure 1 : Organigramme de l'équipe AFEF .....	18
Figure 2 : Représentation de l'alternance de production .....	20
Figure 3: Schéma méthodologique du projet .....	32
Figure 4: Population "X3263 x Belrène" .....	34
Figure 5 : Variété "Olivière" .....	35
Figure 6 : Variété "Arbequina" .....	35
Figure 7 : Histogramme représentant la fréquence associée à chaque taille de séquence de transcriptome .....	37
Figure 8 : Niveaux de LOD retenus par groupes pour l'espèce olivier .....	40
Figure 9 : Groupes de liaisons 5, 6 et 7 de la population "X3263 x Belrène" .....	52
Figure 10 : Groupes de liaisons 7, 8 et 9 de la population "Olivière x Arbequina" .....	54
Figure 11 : Taux de recombinaisons par groupes et entre groupes de liaison pour les données du pommier .....	56
Figure 12 : Taux de recombinaisons par groupes et entre groupes de liaison pour les données de l'olivier .....	56
Figure 13 : "Box-plots" du nombre de fruits par année pour "X3263 x Belrène" .....	57
Figure 14 : "Box-plots" du nombre de fruits par année pour "Olivière x Arbequina" .....	58
Figure 15 : Interaction entre les génotypes de pommier, leur rendement et l'année .....	59
Figure 16 : Interaction entre les génotypes d'olivier, leur rendement et l'année .....	59
Figure 17: Comportements des génotypes de pommier .....	60
Figure 18 : Comportements des génotypes d'olivier .....	60
Figure 19 : Exemple de comportement de production d'un individu "régulier" .....	61
Figure 20 : Exemple de comportement de production d'un individu "alternant" .....	61
Figure 21 : Comparaison des indices de régularités entre deux génotypes alternants.....	62
Figure 22 : Positionnement des QTLs sur les groupes 3, 8 et 10 des cartes génétiques élaborées pour "X3263 x Belrène" .....	64
Figure 23 : Détection des QTLs sur les groupes 8 et 10 de la population "Starkrimson x Granny" (Guitton et al., 2012).....	65
Figure 24: Positionnement des QTLs sur les groupes 2, 4 et 6 des cartes génétiques élaborées pour "Olivière x Arbequina" .....	66
Figure 25 : Comparaison entre QTL de rendement détecté sur le groupe 9 de la nouvelle carte saturée et celui détecté sur la carte AFLP (Ben Sadok et al., 2013) .....	67
Figure 26 : Histogramme de répartition des génotypes pour le nombre de fruits total récolté en 2011 pour la population "X3263 x Belrène" .....	73
Figure 27 : Valeur du LOD en fonction de la position en cM sur le groupe de liaison 10 du pommier de la descendance "X3263 x Belrène" .....	78



## TABLE DES TABLEAUX

Table 1: Les différents types de ségrégations possibles et leurs génotypes associés .....	39
Table 2: Détermination des phases des marqueurs pour deux individus de la population "X3263 x Belrène" .....	42
Table 3: Statistiques sur la cartographie de la population de pommier .....	53
Table 4: Statistiques comparatives concernant les cartes "multialléliques" et "bialléliques" élaborées pour l'olivier.....	55
Table 5 : Héritabilités des différents caractères pour les deux espèces.....	58
Table 6 : Descriptions des variables utilisées pour "X3263 x Belrène" .....	63
Table 7 : Liste des QTLs détectés pour chaque variable chez "X3263 x Belrène" .....	63
Table 8 : Descriptions des variables utilisées pour "Olivière x Arbequina" .....	65
Table 9 : Liste des QTLs détectés pour chaque variable chez "Olivière x Arbequina" .....	66
Table 10: Nombre de gènes présents sous le QTL de rendement du groupe de liaison 5 du pommier par grandes catégories fonctionnelles .....	68
Table 11 : Annotations de gènes sous les marqueurs qui forment les QTLs détectés chez l'olivier .....	69
Table 12: Données phénotypiques relevées par arbre et par année pour la population "Olivière x Arbequina » .....	79
Table 13 : Données phénotypiques disponibles pour l'olivier avec le Nombre total de fruits ajouté.....	79

## TABLE DES ANNEXES

ANNEXE I : PUBLICATIONS CLES DE L'EQUIPE AFEF (UMR AGAP, 2012 (B)) .....	iii
ANNEXE II : ELABORATION DE LA CARTE COMPARATIVE POUR L'OLIVIER .....	v
ANNEXE III : REEDITION DE LA CARTE AFLP .....	vi
ANNEXE IV : CARTOGRAPHIE COMPLETE DE LA POPULATION DE POMMIER « X3263 X BELRENE » AVEC POSITIONNEMENT DES QTLs .....	vii
ANNEXE V : CARTOGRAPHIE COMPLETE DE LA POPULATION D'OLIVIER « OLIVIERE X ARBEQUINA » AVEC POSITIONNEMENT DES QTLs .....	xi



## Annexes

## ANNEXE I : PUBLICATIONS CLES DE L'ÉQUIPE AFEF (UMR AGAP, 2012 (B))

- [1] Khadari B, Charafi J, Moukhli A, Ater M. 2008. Substantial genetic diversity in cultivated Moroccan olive despite a single major cultivar: a paradoxical situation evidenced by the use of SSR loci. *Tree Genetics and Genomes* 4: 213-221
- [2] Stephan J, Sinoquet H, Donès N, Haddad N, Talhouk S, Lauri PÉ. 2008. Light interception and partitioning between shoots in apple cultivars influenced by training. *Tree Physiology* 28: 331-342.
- [3] Massonnet C, Regnard JL, Lauri PÉ, Costes E, Sinoquet H. 2008 Unraveling the relative contributions of foliage distribution and leaf functions to branch capacities in two apple cultivars. *Tree Physiology* 28: 665-678.
- [4] Lauri PÉ, Simon S, Brun L, Morel K, Combe F, Defrance H, Hemptinne JL, Hucbourg B, Regnard JL. 2011. Le pommier et ses bioagresseurs – Les composantes dynamique et structurelle de l'architecture de l'arbre modulent les dynamiques d'infestation et d'infection. *Innovations Agronomiques* 15: 65-77.  
[http://www.inra.fr/ciag/revue/volume\\_15\\_septembre\\_2011](http://www.inra.fr/ciag/revue/volume_15_septembre_2011)
- [5] Segura V, Cilas C, Costes E. 2008 Dissecting apple tree architecture into genetic, ontogenetic and environmental effects. I: Mixed linear modelling of repeated spatial and temporal measures. *New Phytologist* 178: 302-314.
- [6] Ripetti V, Escoute J, Verdeil JL, Costes E. 2008 Shaping the shoot: the relative contribution of cell number and cell shape to variations in internode length between parent and hybrid apple trees. *Journal of Experimental Botany* 59: 1399-1407.
- [7] Khadari B, El Aabidine AZ, Grout C, Ben Sadok I, Doligez A, Moutier N, Santoni S, Costes E. 2010. Genetic Linkage Map of the Olive Tree Based on a Pseudo-Test Cross 'Olivière' x 'Arbequina' and AFLP, ISSR and SSR Markers. *Journal of Horticultural Science and Biotechnology* 135: 1-8.
- [8] Lauri PÉ, Gorza O, Cochard H, Martinez S, Celton JM, Ripetti V, Lartaud M, Bry X, Trottier C, Costes E. 2011. Genetic determinism of anatomical and hydraulic traits within an apple progeny. *Plant and Cell Environment* 34: 1276–1290.

- [9] Costes E, Smith C, Renton M, Guédon Y, Prusinkiewicz P, Godin C. 2008. MappleT: Simulation of Apple Tree Development Using Mixed Statistical and Biomechanical Models. *Functional Plant Biology* 35: 936-950.
- [10] Legave JM, Farrera I, Alméras 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 and Biotechnology* 83: 76-84.

## ANNEXE II : ELABORATION DE LA CARTE COMPARATIVE POUR L'OLIVIER

Une troisième carte a été élaborée à partir des données génétiques de l'olivier « Olivière x Arbequina ». Pour celle-ci, le plus grand nombre possible de marqueurs microsatellites a été gardé. En effet, la nature même de ces marqueurs les rend plus sujets à des erreurs de génotypage que ne le sont les marqueurs SNP. De ce fait, environ la moitié d'entre eux n'a pas été placée sur la carte optimisée, plus précise, qui servira pour la détection de QTLs.

Cependant, ces marqueurs sont primordiaux pour rendre le travail utilisable par la communauté internationale qui travaille sur ce thème. En effet, ils ont le grand intérêt de pouvoir être communs à plusieurs cartes génétiques, élaborées dans le cadre d'autres projets. Ces cartes pourront alors être comparées et permettront de recouper des informations obtenues parallèlement, grâce aux autres marqueurs.

Suivant une méthode de cartographie génétique similaire à celle des cartes optimisées, des cartes comprenant des SNP et des SSR ont été réalisées. La différence est que seuls les SNP les plus pertinents ont été entrés dans la matrice de départ. Ainsi, les marqueurs SNP déjà écartés pour la carte optimisée n'ont pas été pris en compte mais tous les microsatellites l'ont été. 23 cartes ont été fabriquées par ce biais pour répondre à cet autre besoin de lien fort entre les cartes.

## ANNEXE III : REEDITION DE LA CARTE AFLP

En 2012, une carte génétique pour la même population d'oliviers en ségrégation biparentale était élaborée dans l'équipe, au cours du doctorat d'Inès Ben Sadok. Cette carte contenait une partie des marqueurs microsatellites utilisés pour la nouvelle carte, mais également 330 marqueurs AFLP, Amplified Fragment-Length Polymorphism (Ben Sadok et al., 2013).

Ces marqueurs sont obtenus en plusieurs étapes. Des enzymes de restriction vont couper l'ADN sur des séquences nucléotidiques précises, toujours les mêmes selon les enzymes utilisées en formant ainsi des fragments de différentes tailles. Des amorces, dont les séquences sont également connues, sont ensuite ajoutées aux extrémités de ces fragments d'ADN précédemment obtenus. Ces fragments amorcés sont alors amplifiés par la technique de PCR. Des séquences d'oligonucléotides complémentaires sont ensuite ajoutées à certains fragments, uniquement ceux présentant les bases spécifiques s'amorçant avec ces oligonucléotides. Cette étape permet de réduire le nombre de fragments réellement amplifiés de manière à rendre les résultats plus visualisables.

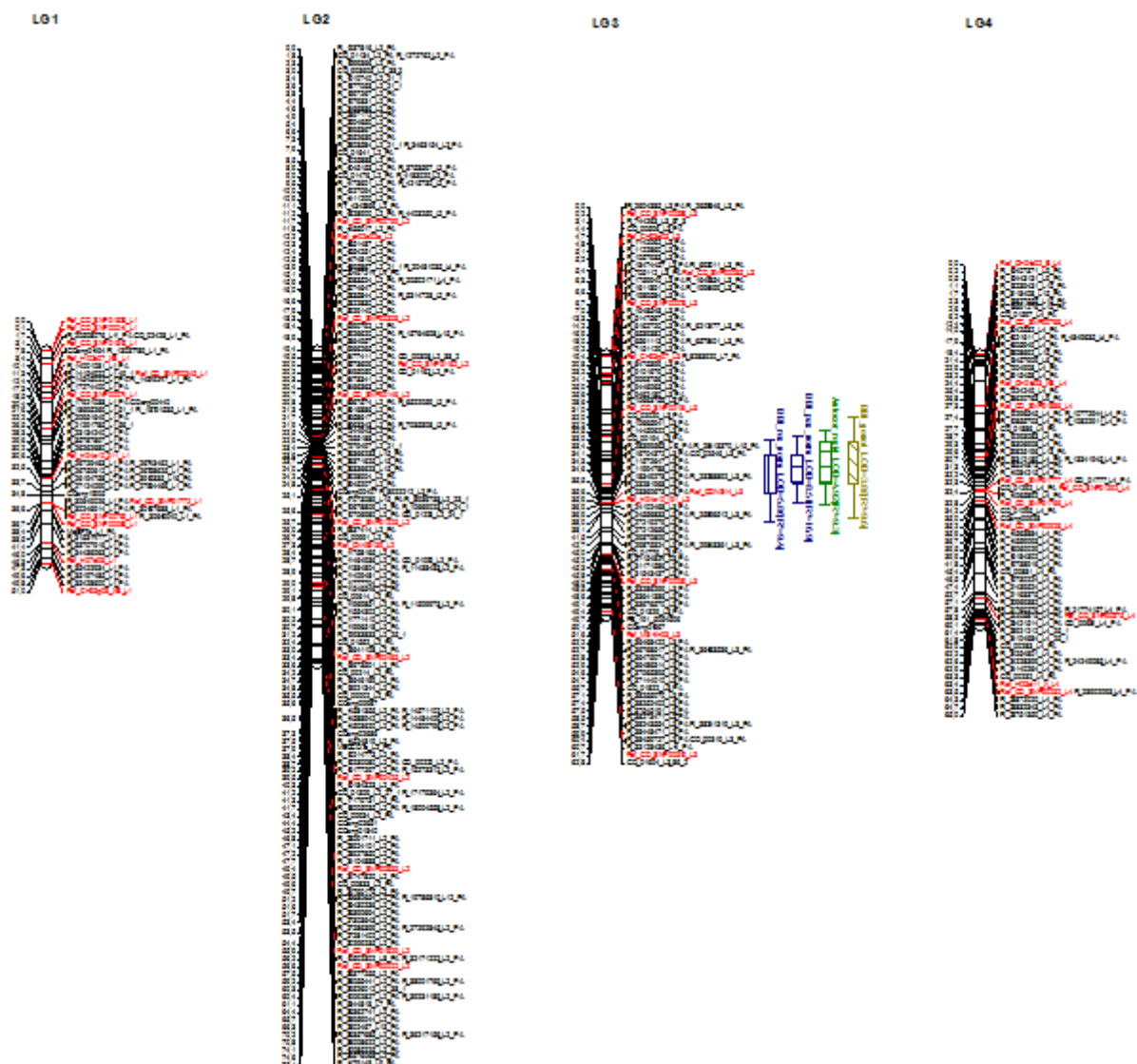
Une fois que l'ADN de 147 individus pris au hasard dans la population « Olivière x Arbequina » a subi cette digestion par les enzymes de restriction en des sites spécifiques puis l'amorçage des fragments, une électrophorèse va permettre de lire les résultats sur un gel.

Pour un marqueur donné, on notera sa présence ou son absence pour chaque individu. Si une bande d'ADN se trouve toujours à la même hauteur sur le gel, il s'agit alors du même marqueur. En observant la ségrégation du marqueur sur la descendance, il est possible d'en induire si le marqueur est hétérozygote chez l'un des parents seulement ou bien chez les deux. De plus, le génotypage des parents Olivière et Arbequina a permis de savoir chez quel parent est présent le polymorphisme.

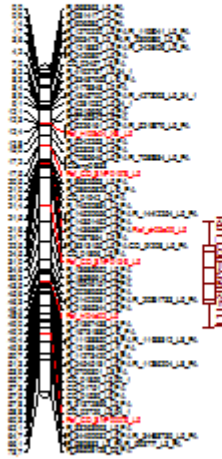
La carte AFLP ainsi mise à jour en y ajoutant la soixantaine de marqueurs microsatellites obtenus ultérieurement, il a été possible de faire le lien avec le travail déjà mené sur le sujet. De plus, 147 individus de la population sont génotypés pour les marqueurs AFLP tandis que seuls 80 le sont pour la carte SNP. Cette seconde carte AFLP peut préciser les analyses faites avec les données SNP.



# ANNEXE IV : CARTOGRAPHIE COMPLETE DE LA POPULATION DE POMMIER « X3263 X BELRENE » AVEC POSITIONNEMENT DES QTLs



LO6



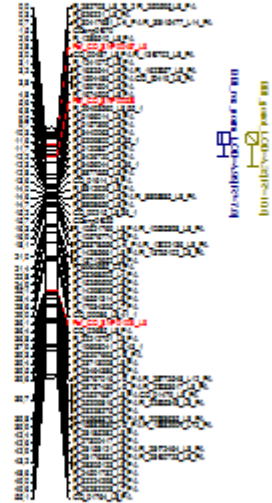
LO8



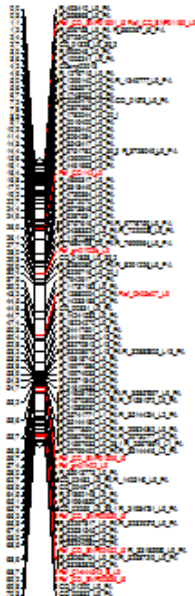
LO7



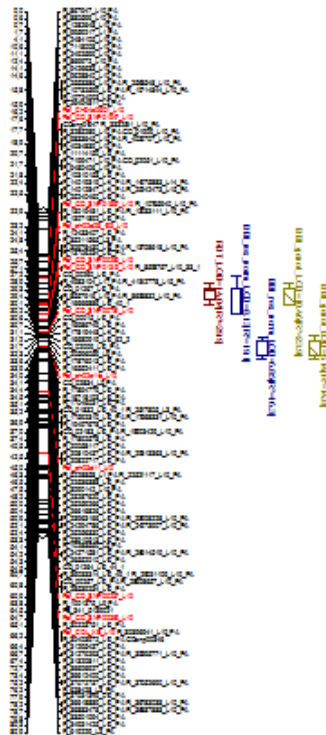
LO8



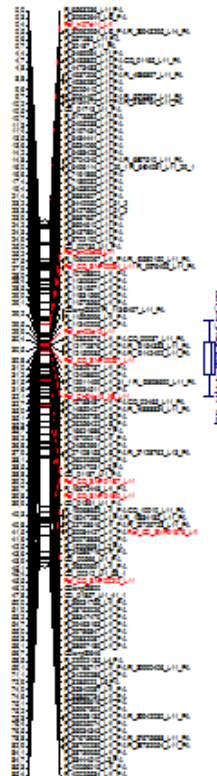
LO9



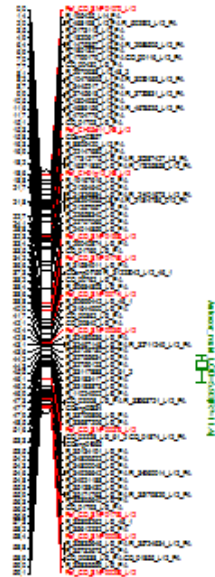
LO10



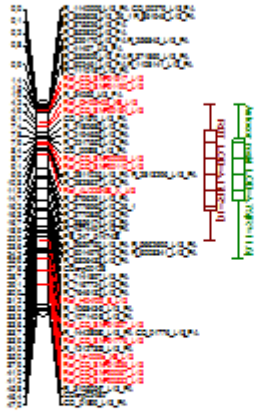
LO11



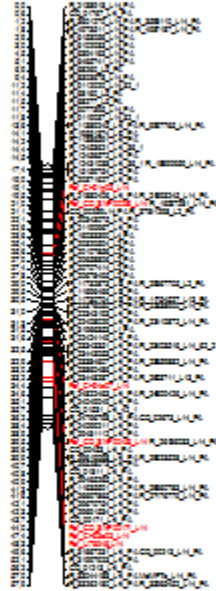
LO12



LG13



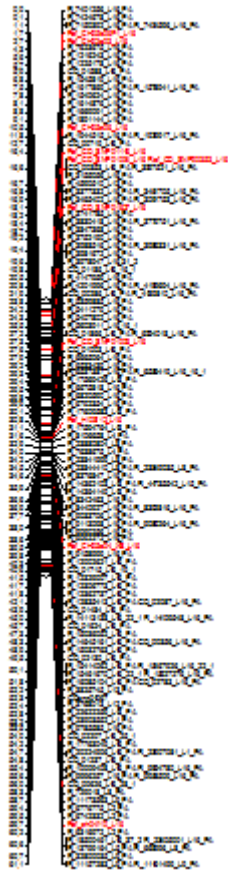
LG14



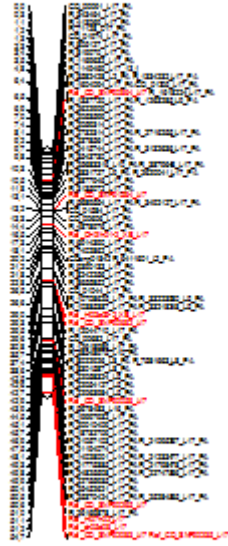
LG16



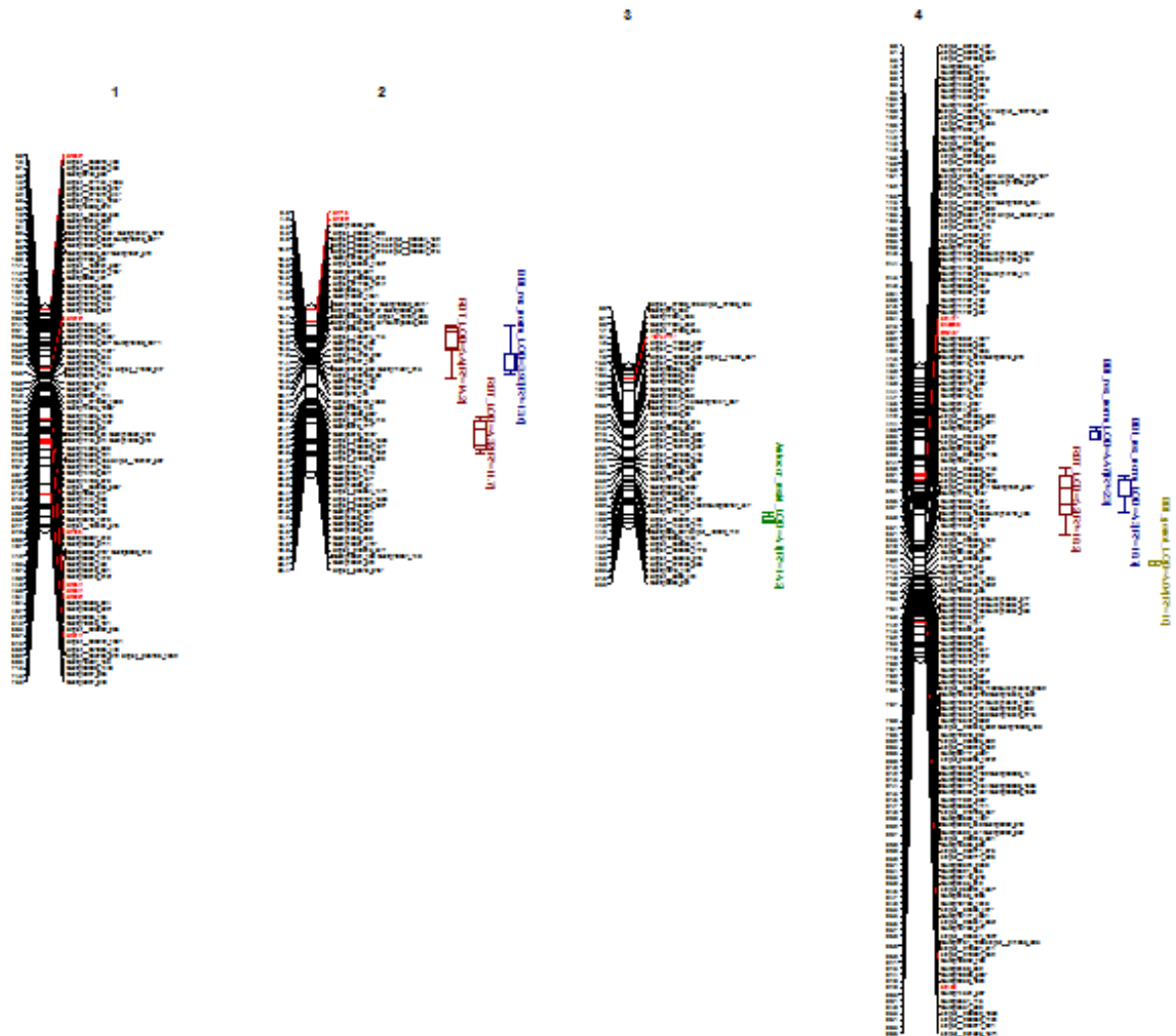
LG16

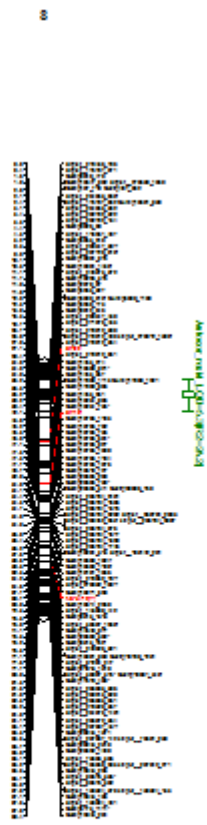
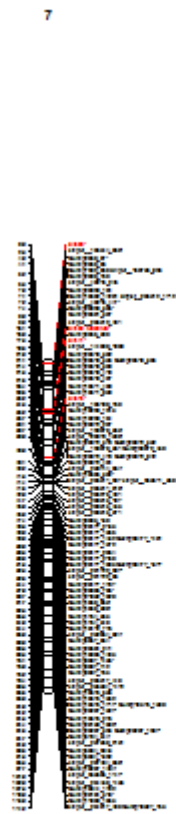
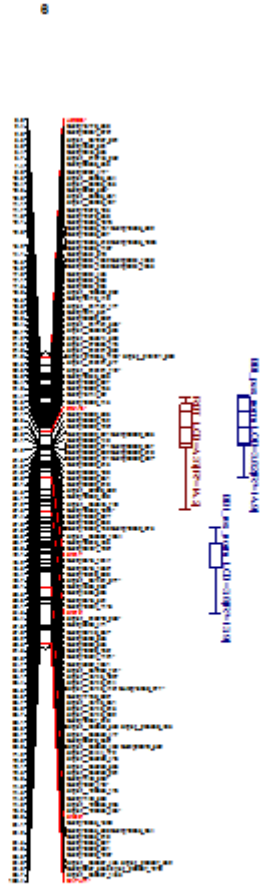


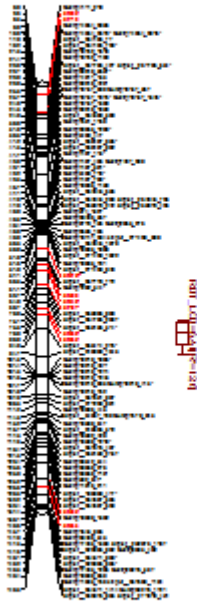
LG17

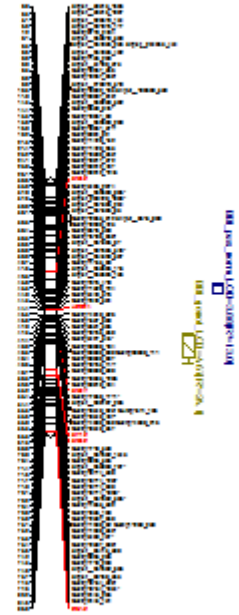
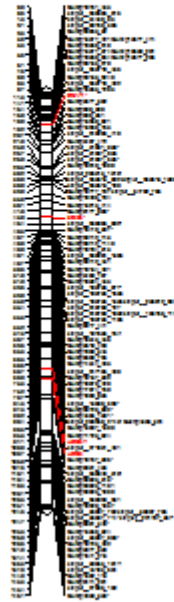
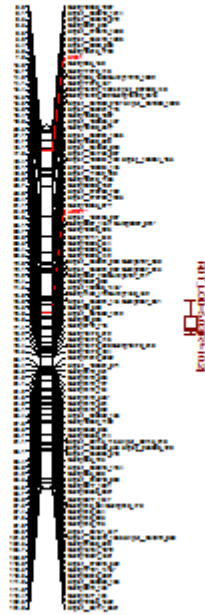


# ANNEXE V : CARTOGRAPHIE COMPLETE DE LA POPULATION D'OLIVIER « OLIVIERE X ARBEQUINA » AVEC POSITIONNEMENT DES QTLs

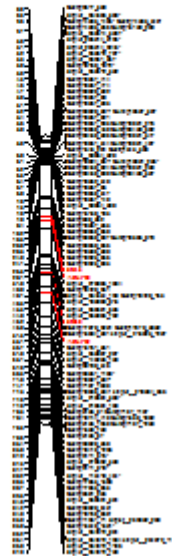
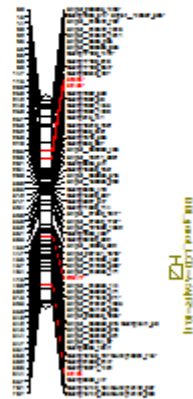
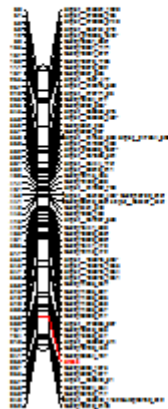
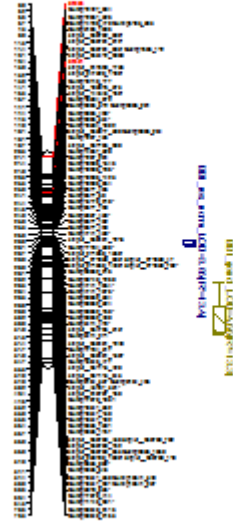
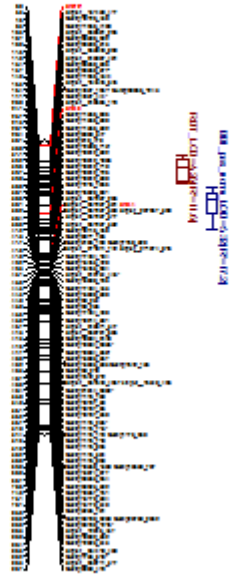
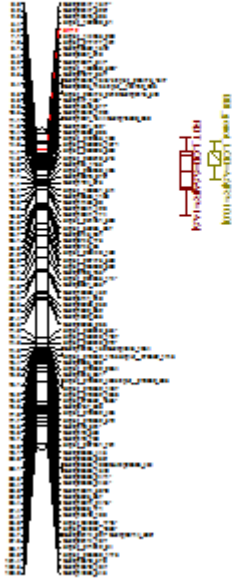












# TABLE DES MATIERES

Résumé.....	3
Sommaire .....	5
Remerciements .....	7
Table des sigles et abréviations.....	9
Introduction.....	11
Partie I : Contexte.....	13
1. Le domaine de la génétique végétale .....	14
1.1. La génétique à travers les âges .....	14
1.2. La recherche génétique assistée par marqueurs moléculaires .....	15
1.3. Les enjeux de la recherche en génétique végétale.....	15
2. Le centre de recherche.....	16
2.1. L'UMR AGAP.....	16
2.2. L'Equipe AFEF.....	17
2.3. Le fonctionnement de l'équipe.....	18
2.4. Le GIS Fruits.....	19
3. L'alternance de production .....	19
3.1. Caractérisation du phénomène .....	19
3.2. Les techniques culturales à disposition pour maîtriser l'alternance .....	20
4. Deux espèces : Pommier et Olivier .....	22
4.1. Le Pommier .....	22
4.1.1. Histoire.....	22
4.1.2. Le marché de la pomme, en France et à l'étranger .....	22
4.1.3. Caractéristiques botaniques.....	23
4.1.4. Recherches menées sur le pommier .....	23
4.2. L'olivier.....	24
4.2.1. Histoire de l'olivier.....	24
4.2.2. Le marché de l'olive en France et à l'étranger .....	24
4.2.3. Physiologie de l'olivier .....	25
4.2.4. Recherches menées sur l'olivier .....	25
Partie II : Problématiques.....	27

Partie III : Matériel et Méthodes.....	31
1. Schéma méthodologique .....	32
2. Les populations étudiées.....	32
2.1. Caractéristiques de la population de pommier .....	32
2.2. Caractéristiques de la population d'olivier .....	34
3. Les données génétiques pour ces deux espèces.....	35
3.1. Les marqueurs SNP .....	35
3.1.1. Définition .....	35
3.1.2. Les marqueurs sur la population issue du croisement X3263 x Belrène .....	36
3.1.3. Les marqueurs pour Olivière x Arbequina .....	36
3.2. Les marqueurs SSR .....	38
3.2.1. Définition .....	38
3.2.2. Leur utilisation .....	38
4. La cartographie génétique .....	38
4.1. Le logiciel Joinmap 4.1 .....	38
4.1.1. La matrice d'entrée .....	38
4.1.2. Composition de groupes de liaisons.....	39
4.2. Méthode de création de cartes .....	40
4.2.1. Carte SNP et carte SSR.....	40
4.2.2. La carte optimisée .....	41
4.2.3. La carte élaborée à partir de marqueurs biallèles .....	43
5. Analyses statistiques appliquées aux données génétiques et phénotypiques des deux populations.....	43
5.1. Le logiciel R.....	43
5.2. Statistiques appliquées aux données génétiques.....	43
5.3. Analyses préliminaires sur les données phénotypiques.....	44
5.4. Elaboration de modèles mixtes .....	45
5.4.1. Caractéristiques des modèles.....	45
5.4.2. Le modèle retenu.....	45
5.4.3. Calcul de l'héritabilité d'un caractère .....	46
5.5. Calculs des indices de régularités .....	46
5.5.1. Indice de régularité de production (BBI) .....	46

5.5.2.	Estimation de la régularité de production à partir des résidus normés .....	47
5.6.	Les BLUPs .....	48
6.	Méthode de détection de QTLs (Quantitative Traits Loci).....	48
6.1.	Le logiciel MapQTL 6 .....	48
6.2.	Analyses sous MapQTL 6.....	49
6.3.	Analyses complémentaires sur les QTLs .....	49
7.	Recherche des gènes candidats .....	49
Partie III : Résultats.....		51
1.	La cartographie génétique .....	52
1.1.	La carte génétique du pommier.....	52
1.1.1.	Les cartes génétiques .....	52
1.1.2.	Statistiques sur les cartes .....	53
1.2.	La carte génétique de l'olivier.....	53
1.2.1.	La carte optimisée .....	53
1.2.2.	Statistiques sur les cartes .....	54
1.2.3.	La différence avec la cartographie basée sur des marqueurs bialléliques ....	55
2.	Résultats des analyses statistiques .....	55
2.1.	Analyse des données génétiques.....	55
2.2.	Analyse des données phénotypiques .....	57
2.2.1.	Les différentes années de phénotypage .....	57
2.2.2.	Héritabilité des caractères.....	58
2.2.3.	Interaction entre le génotype, le rendement et l'année .....	58
2.2.4.	Trois types de comportements des génotypes .....	60
2.2.5.	Différence entre BBI_res_norm et BBI commun.....	61
3.	QTLs détectés .....	62
3.1.	Les QTLs détectés sur X3263 x Belrène .....	63
3.1.1.	Informations sur les QTLs .....	63
3.1.2.	Positionnement des QTLs sur les cartes génétiques .....	64
3.1.3.	Comparaison avec les résultats antérieurs (Guitton et al., 2012).....	64
3.2.	Les QTLs détectés sur Olivière x Arbequina.....	65
3.2.1.	Informations sur les QTLs .....	65
3.2.2.	Positionnement des QTLs sur les cartes génétiques .....	66

3.2.3. Comparaison avec les résultats antérieurs (Ben Sadok et al., 2013) .....	67
4. Les gènes candidats impliqués .....	67
Partie IV : Discussion .....	71
1. Validation et limites des approches utilisées et des résultats obtenus.....	72
1.1. Limites et validation de la méthode .....	72
1.1.1. Le phénotypage .....	72
1.1.2. La cartographie génétique.....	72
1.1.3. Les analyses statistiques.....	73
1.1.4. La détection de QTLs .....	74
1.1.5. La recherche de gènes candidats .....	74
1.2. Discussion et validation des résultats obtenus.....	75
1.2.1. Les cartes génétiques .....	75
1.2.2. Les résultats statistiques .....	76
1.2.3. Les QTLs identifiés .....	77
1.2.4. Les gènes candidats .....	78
2. Propositions d'amélioration.....	78
2.1. Ré-estimation de la variable « Rendement » chez l'olivier .....	78
2.2. Identification des gènes plus précisément .....	79
3. Les perspectives de l'étude .....	80
3.1. Perspectives imminentes .....	80
3.1.1. Changements sur les parcelles .....	80
3.1.2. Un article scientifique en construction .....	81
3.1.3. Utilisation des cartes du pommier dans une thèse.....	81
3.2. Perspective à long terme : La sélection de variétés régulières .....	81
Conclusion .....	83
Bibliographie .....	85
Table des figures.....	95
Table des tableaux.....	96
Table des annexes .....	97
Annexes .....	ii
Annexe I : Publications clés de l'équipe AFEF (UMR AGAP, 2012 (b)) .....	iii
Annexe II : Elaboration de la carte comparative pour l'olivier .....	v

Annexe III : Réédition de la carte AFLP .....	vi
Annexe IV : Cartographie complète de la population de pommier « X3263 x Belrène » avec positionnement des QTLs.....	vii
Annexe V : Cartographie complète de la population d'olivier « Olivière x Arbequina » avec positionnement des QTLs.....	xi
Table des matières .....	xvi