



**HAL**  
open science

# Recherche des bases génétiques de caractères de croissance et de réponse au stress hydrique d'une core collection de pommiers, par génétique d'association

Victor Debernardi

► **To cite this version:**

Victor Debernardi. Recherche des bases génétiques de caractères de croissance et de réponse au stress hydrique d'une core collection de pommiers, par génétique d'association. Sciences du Vivant [q-bio]. 2016. hal-02796611

**HAL Id: hal-02796611**

**<https://hal.inrae.fr/hal-02796611v1>**

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.



## **AGROCAMPUS OUEST**

### **CFR Rennes**

Année universitaire : 2015-2016  
Spécialité : Production et Technologie du Végétal  
(ProTeV)  
Parcours : Sélection, Production et Valorisation des  
Végétaux (SPVV)  
Option : Innovation Variétale

### **Mémoire de Fin d'Études**

Master de l'Institut Supérieur des Sciences  
agronomiques, agroalimentaires, horticoles et du  
paysage

# **Recherche des bases génétiques de caractères de croissance et de réponse au stress hydrique d'une core-collection de pommiers, par génétique d'association.**

Par : Victor Debernardi

**Soutenu à Rennes le 15/09/2016**

**Devant le jury composé de :**

Président : Anne Laperche

Maître de stage : Evelyne Costes

Enseignant référent : Maria Manzanares-Dauleux



**Annexe III : Fiche de confidentialité et de diffusion du mémoire**

**Confidentialité :**

Non  Oui si oui :  1 an  5 ans  10 ans

Pendant toute la durée de confidentialité, aucune diffusion du mémoire n'est possible (1).  
A la fin de la période de confidentialité, sa diffusion est soumise aux règles ci-dessous  
(droits d'auteur et autorisation de diffusion par l'enseignant) INRA - Montpellier SupAgro  
Date et signature du maître de stage (2) : 26/08/2016 UMR AGAP

Equipe AFEF

**Droits d'auteur :**

L'auteur (3) autorise la diffusion de son travail

Oui  Non

Si oui, il autorise

- la diffusion papier du mémoire uniquement(4)  
 la diffusion papier du mémoire et la diffusion électronique du résumé  
 la diffusion papier et électronique du mémoire (joindre dans ce cas la fiche de conformité du mémoire numérique et le contrat de diffusion)

Date et signature de l'auteur :

Victor Debernardi 16/09/2016

**Autorisation de diffusion par le responsable de spécialisation ou son représentant :**

L'enseignant juge le mémoire de qualité suffisante pour être diffusé

Oui  Non

Si non, seul le titre du mémoire apparaîtra dans les bases de données.

Si oui, il autorise

- la diffusion papier du mémoire uniquement(4)  
 la diffusion papier du mémoire et la diffusion électronique du résumé  
 la diffusion papier et électronique du mémoire

Date et signature de l'enseignant :

Architecture et Fonctionnement des Espèces Fruitières  
Av. Agropolis TA - A-108 03 - 34398 Montpellier Cedex 05  
Tél. 33 (0)4 67 61 75 03 Fax 33 (0)4 67 61 55 96  
Courriel : prudhon@supagro.inra.fr

(1) L'administration, les enseignants et les différents services de documentation d'AGROCAMPUS OUEST s'engagent à respecter cette confidentialité.

(2) Signature et cachet de l'organisme

(3) Auteur = étudiant qui réalise son mémoire de fin d'études

(4) La référence bibliographique (= Nom de l'auteur, titre du mémoire, année de soutenance, diplôme, spécialité et spécialisation/Option)) sera signalée dans les bases de données documentaires sans le résumé



## Table des matières

### Liste des abréviations

I.Introduction.....	1
1) Biologie du pommier et diversité génétique. ....	1
2) Ressources en marquage moléculaire.....	2
3) Le pommier une espèce cultivée et améliorée.....	4
4) Problématique.....	6
II. MATERIEL ET METHODE.....	7
1) Matériel végétal.....	7
2) Les variables phénotypiques.....	7
3) Structure, apparentement, et déséquilibre de liaison.....	8
4) GWAS.....	10
5) Gènes candidats.....	11
III. RESULTATS.....	12
1) Exploration statistique des variables.....	12
2) Structure et apparentement.....	13
3) GWAS.....	14
4) Gènes candidats.....	16
IV. DISCUSSION.....	17
1) Matériel végétal et dispositif expérimental.....	17
2) Qualité du génotypage.....	17
3) Ressource bioinformatique.....	17
4) Résultats des associations.....	18
5) Discussion sur les gènes candidats.....	18
V. CONCLUSION.....	19

Remerciements

Bibliographie

Annexes



## Abréviation.

BIC: Critère d'Information Bayésien.

BLUP: Best Linear Unbiased Predictor.

ECPGR: European Cooperative Programme for Plant Genetic Resources.

GWAS : Genome Wide Association Studies.

INRA: Institut National de Recherche Agronomique.

IPCC: Intergovernmental Panel on Climate Change.

LD : Déséquilibre de liaison.

MAF : Minor Allele Frequency.

MLM : Modèle linéaire mixte.

MLM\_K : Modèle linéaire mixte prenant en compte l'apparentement entre individus.

MLM\_K\_Q : Modèle linéaire mixte complet (structure et apparentement pris en compte).

QTL : Quantitative Trait Locus.

Res: Résiduelle.

SNP : Single Nucleotide Polymorphisme.

SSR : Single Sequence Repeats.

SWC: Soil Water Content.

WUE: Water Use Efficiency.

WS: Water Stressed.

WW: Well-Watered.



## I. INTRODUCTION :

### 1) Biologie du pommier et diversité génétique.

#### 1.1) Classification.

Le pommier commun (*Malus x domestica* Borkh.) fait partie de la famille des Rosaceae qui comprend de nombreuses espèces importantes économiquement (poirier, cerisier, abricotier, pêcher, fraisier...) et appartient au clade des rosidées comprenant un tiers des plantes à fleurs. Le pommier appartient à la sous-famille des *Maloideae* (= tribu des Pyreae) et au genre *Malus*. La sous-famille des *Maloideae* est caractérisée par un ovaire infère adhérent et par la présence de pépins dans les fruits.

#### 1.2) Diversité et origine du pommier.

Le genre *Malus* a été cultivé très tôt. En effet, des études paléontologiques ont révélé la présence du genre *Malus* à l'ère tertiaire (Challice et Westwood, 1973). Les variétés furent sélectionnées et multipliées très tôt grâce au greffage, connu depuis plus de 2000 ans. La création de nouvelles variétés ne fut initiée qu'à la fin du 18<sup>ème</sup> siècle avec T.A. Knight grâce à l'hybridation contrôlée.

Le genre *Malus* comprend 55 espèces (Harris *et al.*, 2002) et de nombreuses sous-espèces dont la plupart s'intercroisent. Les espèces du genre *Malus* se retrouvent dans différentes régions tempérées du monde. On les rencontre des Balkans à la Chine et au Japon, en passant par le Caucase, le Turkestan, les montagnes de l'Altaï et la Sibérie. Il existe aussi des espèces originaires d'Amérique du Nord comme *M. ioensis* Brit., *M. coronaria* Mill., ou encore *M. angustifolia* Michx. Toutefois le centre de plus grande diversité se situe en Asie, au niveau des forêts du Caucase, du Kazakhstan et du Turkestan. (Gallais et Bannerot, 1995).

Le pommier cultivé fut longtemps appelé *M. pumila* Mill., espèce endémique des Balkans jusqu'au nord des montagnes de l'Altaï. Pourtant, il est vraisemblable que différentes espèces ont contribué à son évolution dont *M. sylvestris* Mill, que l'on retrouve dans toute l'Europe, et *M. baccata* Borkh. qui a contribué à sa résistance au froid. Certains auteurs considèrent que le pommier cultivé dérive plus directement de *M. sieversii* Ledeb. endémique des montagnes d'Asie centrale. Depuis plus d'un demi-siècle, de nombreux travaux sont effectués pour introduire par hybridation chez le pommier des résistances aux maladies provenant de divers espèces telles que, *M. floribunda* Sieb., *M. micromalus* Mak., *M. prunifolia* Borkh., *M. astrosanguinea* Seneid. Tous ces croisements ont conduit à dénommer le pommier cultivé *Malus x domestica* Borkh. Il a été montré que *M. sieversii* endémique du Kazakhstan, est à l'origine de toutes les pommes cultivées (Harris *et al.*, 2002). Ceci a été confirmé grâce à une étude phylogénétique (Velasco *et al.*, 2010) démontrant que *M. sieversii* est le principal ancêtre du pommier cultivé.

#### 1.3) organisation des ressources génétiques sous forme de collection.

Avant le début du 20<sup>ème</sup> siècle, la majorité des cultivars de pommier étaient obtenus par semi direct des pépins, la chance avait alors un rôle prépondérant. A partir du milieu du vingtième siècle des cultivars provenant d'hybridation contrôlée sont apparus sur le marché (Way, *et al.*, 1990, cité par Lassois *et al.*, 2015). La plupart de ces cultivars furent obtenus à partir d'un nombre réduit de progéniteurs ou de mutants. Ils montrent donc un fort niveau de parenté (Noiton et Alspach, 1996). Ceci a mené à une forte baisse de la diversité malgré la forte diversité génétique disponible.

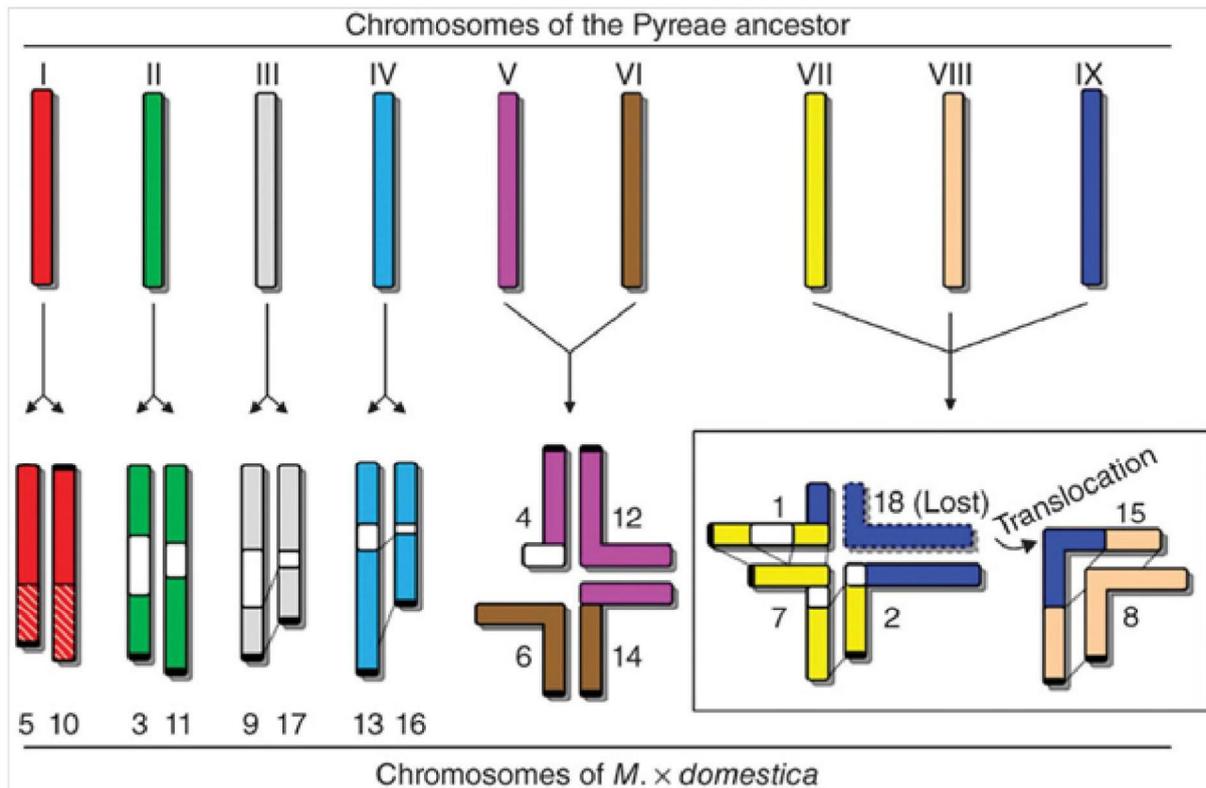


Figure 1: Modèle expliquant le passage de 9 chromosomes chez l'ancêtre du pommier à 18 chromosomes. Une polyploïdisation de l'ensemble du génome suivi par un réarrangement des chromosomes sont ici postulés. Les couleurs en commun indiquent les homologies entre chromosomes. Le blanc indique l'absence d'une contrepartie dupliquée. Les segments hachurés indiquent une homologie partielle.

Il existe plusieurs collections de pommiers à travers le monde. En Europe les plus grandes sont en France, en Allemagne, en Angleterre et en Slovaquie. L'ECPGR (European Cooperative Programme for Plant Genetic Resources) possède une base de données regroupant 21 013 accessions provenant de 34 instituts de 12 pays. D'autres collections importantes existent en Nouvelle Zélande et aux Etats-Unis, notamment à Geneva où sont regroupées plusieurs milliers d'accessions. De nombreuses accessions de pommiers proviennent des forêts du Kazakhstan où il y a une grande diversité, ces forêts étant considérées comme le berceau d'origine du pommier (Harris *et al.*, 2002 ; Gross *et al.*, 2014).

Pour réaliser des études génomiques et afin de réduire le nombre d'individus des core collections ont été mises au point. Une core collection consiste en un nombre limité d'accessions choisies pour représenter, avec un minimum de redondance, la diversité génétique d'une espèce et de ses plantes sauvages apparentées (Frankel *et al.*, 1984). Les trois objectifs majeurs d'une core collection (en terme d'utilisation) sont de disposer d'une représentation aussi large que possible de la diversité génétique, pour pouvoir réaliser des études approfondies sur un ensemble réduit de génotypes, et pouvoir extrapoler les résultats ainsi obtenus dans les collections de base.

La conservation des ressources génétiques de pommier est, en France, menée par des associations d'amateurs et des instituts publics, tel que l'INRA et des conservatoires régionales. A partir de ces collections françaises de pommier une core collection a été réalisée par l'INRA. Elle a été réalisée à partir d'une collection de 2163 accessions, elle regroupe 278 variétés françaises de pommier de table, et représente 90% de la diversité allélique total des ressources génétiques de pommiers français (Lassois *et al.*, 2015).

## 2) Ressources en marquage moléculaire.

### 2.1) Structure du génome et origine génétique.

Le génome du pommier Golden a été séquencé (Velasco *et al.*, 2010). Il s'agit du premier séquençage et analyse d'un génome d'une plante de la famille des Rosaceae. La taille estimée du génome est actuellement de 742,3Mb. Le génome du pommier a un nombre important de séquences répétées, mais il possède le plus faible taux de transposons par rapport aux génomes séquencés d'autres plantes. La densité de gènes est en revanche très élevée par rapport aux autres plantes séquencées, et du fait de sa reproduction majoritairement allogame, le génome du pommier est fortement hétérozygote.

Le nombre chromosomique de base des Maloïdées ( $x=17$ ) est élevé par rapport à celui des autres Rosacées ( $X=7$  chez les Rosoïdées,  $x=8$  chez les Prunoïdées, et  $x=9$  chez les Spiraeoïdées). Cette augmentation du nombre chromosomique chez le pommier est due à un évènement de polyploïdisation récent (50 millions d'années) dans le génome du Pyreae ancêtre du pommier. Plusieurs modèles ont été proposés pour expliquer le nombre élevé de chromosomes des Maloïdées.

L'étude sur le séquençage du génome du pommier supporte l'hypothèse de l'autopolyploïdisation (Velasco *et al.*, 2010). En effet l'étude de génomes de la famille des *Gillenia* ( $x=9$ ) a mis en évidence l'origine monophylétique des Pyreae et *Gillenia*. De plus un modèle expliquant la dérivation du génome jusqu'à  $x=17$  à partir d'un génome à  $x=9$  a été mise au point (figure 1) (Velasco *et al.*, 2010).



## 2.2) Puce à ADN

Dans les années 80 les SSRs (Single Sequence Repeats) ont été découverts et employés comme marqueurs génétiques hautement polymorphiques (Miesfeld *et al.*, 1981). Ces marqueurs ne correspondent qu'à des séquences non codantes. Dans les années 90 de nouveaux marqueurs ont été découverts, il s'agit des SNPs (single nucleotide polymorphisme). Ces derniers sont liés à des séquences non codantes et codantes, ce qui offre la possibilité d'identifier des marqueurs reliés à un trait d'intérêt pour un coût relativement faible par points. Les SNP sont des séquences dont le polymorphisme diffère par une unique paire de base. Les SNPs sont stables, abondants, et distribués uniformément sur le génome (Mesak *et al.*, 2014). Les SNPs étaient d'abord utilisés pour des organismes modèles pour lesquelles les séquences du génome complet étaient disponibles (Hinds *et al.*, 2005). Durant la dernière décennie, des méthodes de génotypage à haut débit ont vu le jour, permettant une meilleure précision de détection des QTLs et des gènes.

De nombreuses ressources génétiques en marquage moléculaire sont disponibles chez le pommier domestique. Il existe plusieurs puces à ADN Infinium®, la IRSC 8K et 20K (Chagné *et al.*, 2012 ; Bianco *et al.*, 2014) qui ont été utilisées pour le génotypage de nombreuses populations biparentales (Clark *et al.*, 2014 ; Falginella *et al.*, 2015), et core collection (Kumar *et al.*, 2013). Mais les études de génétique d'association qui utilisent la puce 8K ont mis en évidence la faible densité de marqueurs pour une étude de génétique d'association. De plus le LD (déséquilibre de liaison) décroît rapidement chez le pommier avec un  $r^2 < 0.2$  pour un espacement de 100Kpb pour deux core collections de pommier à cidre et à dessert (Leforestier *et al.*, 2015), et un  $r^2 < 0.13$  pour 100Kpb chez une population diversifiée de pommiers (Kumar *et al.*, 2013). Ce faible LD nécessite l'utilisation d'une plus grande densité de marqueur indépendant lors d'analyses de génétique d'association.

Une puce à ADN de 487 000 SNPs (487K SNP Affymetrix Axiom®) a été récemment développée (Bianco *et al.*, 2016), elle recouvre aussi bien les régions codantes et non-codantes, et comprenant des SNPs rares pour permettre des études de diversité génétique incluant des pommiers non caractérisés. C'est cette puce qui a été utilisée dans cette étude.

Plusieurs versions du génome sont disponibles celle utilisée dans cette étude est la version V3.

## 2.3) Nécessité de prendre en compte la structure de la population en GWAS.

L'association entre variation génotypique et phénotypique a premièrement été abordé par analyse QTL (Doebley, 1991). L'utilisation de populations biparentales en ségrégation a permis un contrôle de l'histoire des croisements et des ségrégations. En revanche, il y a méconnaissance de l'histoire évolutive à l'échelle des populations. L'histoire évolutive des populations peut être limitée à la connaissance de la structure génétique des populations. Les populations soumises à un processus d'échange de gènes, d'isolement, d'adaptation, ou de sélection se différencient et se structurent génétiquement. La structuration est caractérisée par des différences de fréquences alléliques entre populations.

Bien que les études d'association de type GWAS ont le potentiel d'identifier des polymorphismes génétique sous-jacent des caractères agronomique d'intérêt avec une plus grande précision que dans les études QTL basées sur populations bi-parentales, la détection de faux positifs est un problème majeur et peut être partiellement attribué à de fausses associations causées par la structure de la population et à un apparemment inégal parmi des individus d'une même cohorte. L'effet de la structure est un problème pour l'utilisation des

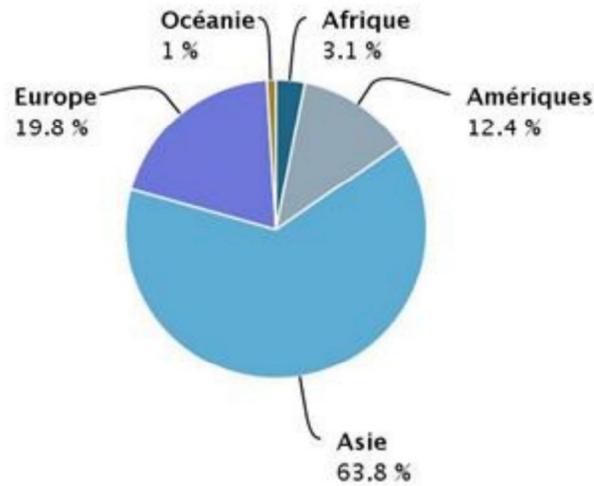


Figure 2 : Part de production de pomme en Asie, Europe, Amériques, Afrique, et Océanie (FAO, moyenne de 2010 à 2014)

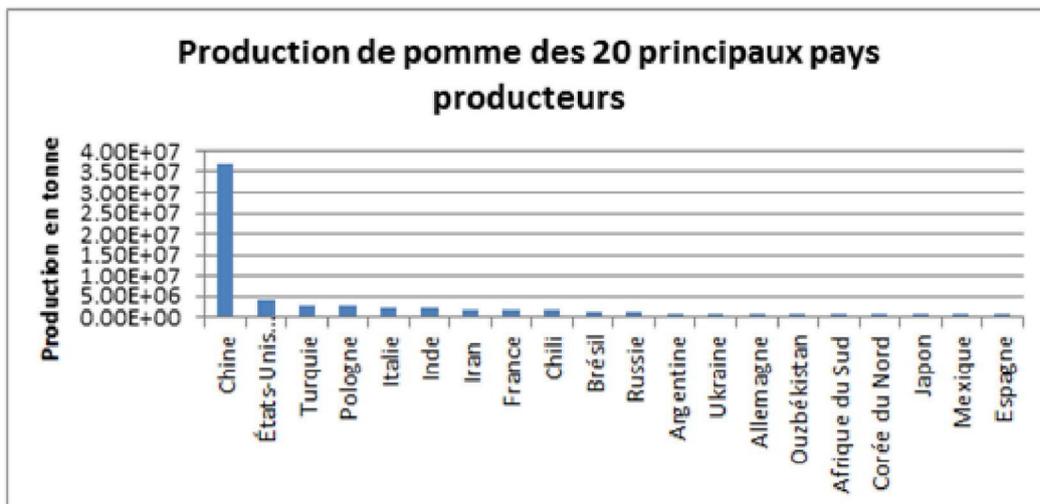


Figure 3: production de pomme des 20 principaux pays producteurs (moyenne de 2010 à 2014) (FAO).

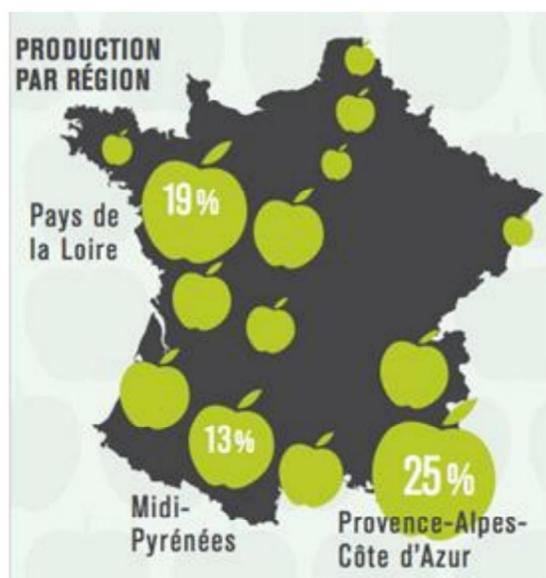


Figure 4: Carte de production de pommes par région (source: Planetoscope, 2016)

approches d'association au niveau de population chez les plantes cultivées (Pritchard et Donnelly 2001). En effet lors d'une analyse d'association entre la variabilité phénotypique et la variabilité génétique au sein d'un panel regroupant des individus de différents groupes génétiquement structurés, la structure peut induire un biais en agissant comme un facteur confondant (Marchini *et al.*, 2004). Un allèle lié à la structure peut alors montrer une corrélation significative avec un caractère différenciant les groupes sans qu'il y ait une réalité biologique d'un lien entre le locus considéré et le caractère. L'association statistique n'est alors pas liée à une réalité génétique. Pour corriger cet effet, les études d'association doivent prendre en compte les ascendances des individus dans leurs analyses, on parle alors « d'association structurée » (Pritchard et Donnelly, 2001). Pour ce faire une méthode simple consiste à incorporer l'ascendance des individus comme une covariable. Dans un panel d'association, le nombre de populations structurées n'est pas connu *a priori*, et la définition des populations sur des bases géographiques ou morphologiques ne correspond pas forcément à une structuration génétique. Des progrès ont été réalisés pour définir la structure des populations à partir de données génotypiques multi-loci (Pritchard *et al.*, 2000), ainsi des outils sont disponibles pour estimer la structure d'une population afin de la prendre en compte dans les analyses GWAS.

### 3) Le pommier une espèce cultivée et améliorée.

#### 3.1) La Production

Le pommier est l'arbre fruitier le plus cultivé des régions tempérées du globe (Velasco *et al.*, 2010). Les principales régions productrices sont, l'Asie et principalement la Chine domine la production avec 63.8% de la production mondiale, suivie par l'Europe avec 19.8%, et les Amériques avec 12.4%. (FAO, 2014) (figure 2 et 3). En Europe la production est dominée par la Pologne, l'Italie et la France, suivis par l'Ukraine, l'Allemagne, et l'Espagne.

En France, la production de pommes a diminué depuis les années 90 passant de 2.9 millions de tonnes en 1992 à 1.7 millions de tonnes en 2013. Les principales régions productrices sont la région Provence-Alpes Côte d'Azur, suivie par les Pays de la Loire, puis par le Midi-pyrénées (Figure 4). Le pommier a un rôle économique majeur en France où il est bien devant les autres fruits avec 22.6% de part de marché. Il existe en France plus de 400 variétés dont seulement une douzaine domine (Golden delicious, Royal Gala, Reine des reinettes, Belle de boskoop...). 38% de la production est exportée, et la filière emploie environ 50 000 personnes faisant de cette filière un secteur clef de l'économie rural.

#### 3.2) Contexte climatique.

Le contexte climatique actuel montre une augmentation des températures ainsi que de la fréquence et de l'intensité des sécheresses. On observe donc une diminution de la disponibilité en eau (IPCC, 2014). De plus, l'agriculture connaît une augmentation de la demande en eau due à une augmentation des surfaces irriguées et de la population mondiale. Or, l'agriculture étant le secteur consommant le plus d'eau au niveau mondial, il est crucial de trouver des moyens pour produire une alimentation demandant moins d'eau.

Sous climat méditerranéen les espèces fruitières sont très cultivées. Or, sous ce climat l'irrigation joue un rôle essentiel. En effet elle permet de sécuriser les rendements et donc les revenus des producteurs. Sous les climats tempérés océaniques qui sont sujets à des précipitations plus régulières, les vergers peuvent assurer une production sans besoin d'irrigation. Mais dans les zones plus arides qui montrent un déficit de pluviométrie printanier et estival, la production et la survie des vergers sont dépendantes de la disponibilité en eau. Et



dans le cas du pommier, les calibres, la qualité et le rendement sont dépendants du pilotage rationnel de l'irrigation (Girona et Naor, 2012).

### 3.3) l'efficience d'utilisation de l'eau.

Dans le contexte du réchauffement global, l'amélioration de l'efficience d'utilisation de l'eau (WUE= water use efficiency) est un trait quantitatif primordial.

L'efficience d'utilisation de l'eau correspond à la pente de la courbe reliant photosynthèse et transpiration (Hopkins, 2003). Elle peut être estimée en calculant la quantité d'eau consommée par unité de CO<sub>2</sub> produit. D'autres méthodes peuvent être utilisées pour estimer cette variable comme par exemple la quantité d'eau consommée par unité de biomasse produite (en poids), c'est cette dernière méthode qui fût utilisée dans cette étude.

Les plantes peuvent développer différentes stratégies pour améliorer cette WUE :

-Eviter les chutes de potentiel hydrique dans la plante, en effet lorsque le sol se dessèche son potentiel hydrique diminue, et lorsque la plante transpire la chute de potentiel hydrique s'aggrave à mesure que le potentiel hydrique du sol est bas, que les conductances hydrauliques caractérisant la capacité de transfert de l'eau à travers la plante sont faible, et que la transpiration est élevée (Marguerit *et al.*, 2014). Si le potentiel hydrique atteint des valeurs négatives extrêmes dans la plante alors l'eau dans les vaisseaux de xylème atteint une telle tension qu'elle peut s'évaporer brutalement et une poche de gaz va se former et obstruer le vaisseau. On parle alors de cavitation (Brodersen *et al.*, 2013).

-Limiter la transpiration, en effet sous contrainte hydrique on observe un ralentissement de la croissance (Cramer *et al.*, 2013) qui s'explique par le fait que la photosynthèse nécessite l'ouverture des stomates pour fournir le CO<sub>2</sub> nécessaire à cette dernière, or cette ouverture des stomates provoque inévitablement une évaporation de l'eau des tissus. Ceci explique pourquoi sous contrainte hydrique on observe une fermeture des stomates (Coupel-Ledru *et al.*, 2014).

-Limiter la diminution de photosynthèse malgré la fermeture des stomates.

-Maintenir l'eau dans les cellules grâce à l'ajustement osmotique. Ceci est réalisé par accumulation de solutés de faible poids moléculaires dans les cellules, qui provoque une baisse du potentiel hydrique de ces cellules et permet un maintien de l'eau dans ces dernières (Patakas et Noitsakis, 1999).

-Le développement d'un système racinaire plus profond pour permettre une exploration optimal du sol et donc puiser l'eau présente en profondeur (Marguerit *et al.*, 2014).

Dans cette étude les plantes utilisées sont celles d'une core collection de l'INRA qui ont grandi en serre puis ont été transférées dans une plateforme de phénotypage (PhenoArch) pour étudier des scions âgés d'un an pour leur accumulation de biomasse, leur transpiration et leur WUE sous condition optimale de croissance et sous condition de stress hydrique (Lopez *et al.*, 2015). Pour chacun des traits étudiés, neuf modèles mixtes ont été évalués en prenant en compte l'effet génétique et l'hétérogénéité spatiale au sein de la plateforme. Les valeurs d'héritabilité indiquent que la biomasse, la WUE, et dans une moindre mesure la transpiration sont sous contrôle génétique. Pour l'ensemble de la collection, la biomasse prend en compte 70% de la variabilité de la WUE (Lopez *et al.*, 2015). Il a donc été montré que la WUE est sous contrôle génétique, c'est pourquoi les croisements en amélioration variétale peuvent permettre d'améliorer cette WUE.

### 3.4) L'amélioration actuelle du pommier.

Plusieurs améliorations importantes ont eu lieu chez le pommier dont certaines au niveau de la résistance aux stress biotiques comme la résistance à la tavelure. La résistance à cette



maladie causée par *Venturia inaequalis* a été permise par l'introgession d'un gène « majeur » de résistance issu de *Malus floribunda*. Il s'agit du gène Vf qui fut introduit en 1914 par Crandall, mais dont l'efficacité ne put être vérifiée quand 1943 après une épidémie (Gessler *et al.*, 2006). Elle est maintenant présente dans plus de 70 cultivars. Les schémas de sélection possèdent classiquement un test à la tavelure et à l'oïdium. De nombreux cultivars résistants à la tavelure et à l'oïdium sont créés mais peu possèdent les qualités organoleptiques nécessaires pour devenir des variétés commercialisées. La recherche de nouveaux géniteurs est donc primordiale car les résistances durables à la tavelure ne se retrouvent pas dans les variétés récentes mais sont nombreuses dans les variétés anciennes (Didelot *et al.*, 2016).

Pour la qualité du fruit, il existe aussi de nombreux programmes de sélection, basés sur l'attrait du fruit, les qualités organoleptiques, et sur l'aptitude à la conservation (Sadok *et al.*, 2015).

En revanche il n'existe que peu d'études sur la résistance aux stress abiotiques comme la résistance au stress hydrique par exemple. On peut toutefois citer une étude sur le déterminisme génétique de la réponse à la contrainte hydrique d'une population d'hybride de pommier (Virlet *et al.*, 2014), ainsi qu'une étude d'analyse de marqueurs de séquences exprimés en réponse aux basses températures et au stress hydrique chez le cultivar Royal Gala (Wisniewski *et al.*, 2008).

#### 4) Problématique.

Le thème de cette recherche est de réaliser une étude d'association sur la collection de pommier de variété française de l'INRA. Cette étude portera sur des caractères de développement du système aérien et notamment les composantes de la biomasse, et sur des caractères de réponse au stress hydrique du sol, à l'échelle des métamères et de la plantes entières (partie aérienne). Le système sous-terrain n'est pas analysé, en effet tous les arbres étudiés sont des scions de un an greffés sur un porte-greffe commun M9. Existe-t-il des loci impliqués dans l'efficacité d'utilisation de l'eau et dans les caractères de développement aérien chez *M. domestica* dans la core collection de pommier de table de l'INRA? Si des loci sont effectivement corrélés avec un trait impliqué dans cette meilleure efficacité d'utilisation de l'eau, quels gènes candidats peuvent être proposés ?

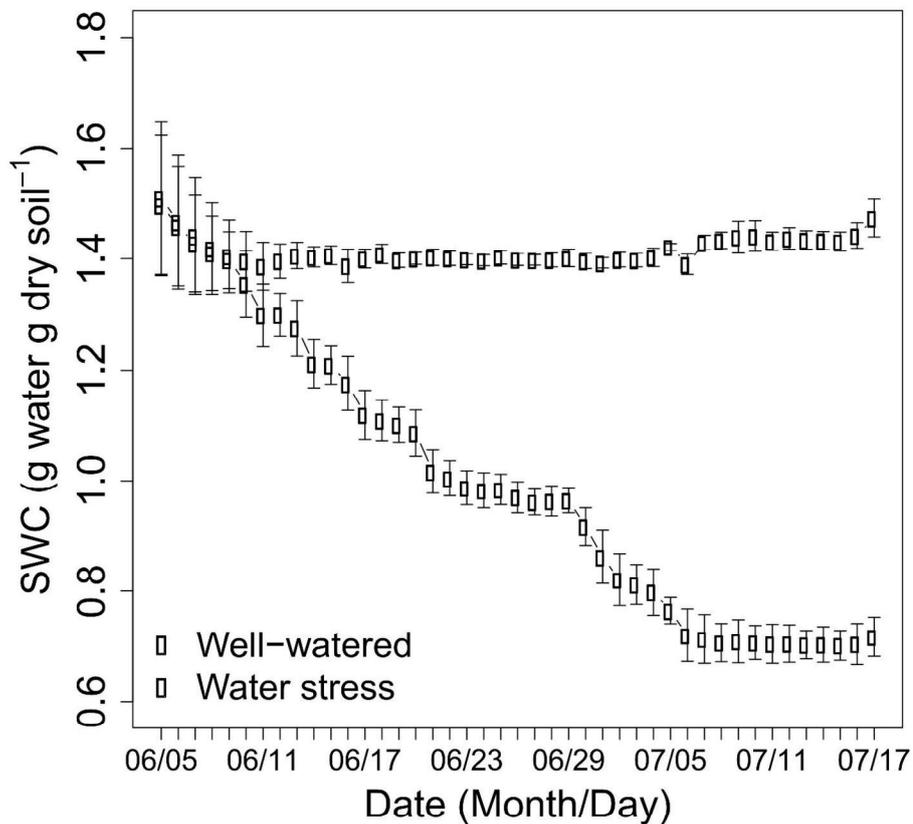


Figure 5 : Contenu en eau du sol (SWC) pour les plantes stressées (WS) (courbe du bas) et pour les plantes non stressées (WW) (courbe du haut). Pour les plantes stressées 2 plateaux de SWC sont observés, le premier autour de SWC=1 correspondant à un stress intermédiaire (celui utilisé dans l'étude) et le second autour de SWC=0.7 correspondant à un stress fort (non utilisé dans l'étude). (Lopez *et al.*, en préparation).

Tableau 1 : Liste des variables morphologiques et physiologiques étudiées et description de leurs obtentions.

Echelle	Variables étudiées	Formule de calcul	Méthode d'obtention
Plante entière	Initial_biomass		photo 3D et allométrie
	Final_biomass		photo 3D et allométrie
	Initial_leaf_area		scan de feuilles
	Final_leaf_area		scan de feuilles
	Number_of_leaves		mesuré
	Height		mesuré
	Basal_shoot_diameter		mesuré
Métamère	Maximum_internode_length		mesuré
	Internode_diameter		mesuré
	Internode_dry_weight		mesuré
	Internode_volume	$\pi \times \left(\frac{IN_{diam}^2}{2}\right) \times IN\_L\_max$	
	Internode_density	$\frac{IN\_DW}{IN\_volume}$	
	Leaf_dry_weight		mesuré
	Leaf_area		mesuré
	Specific_leaf_weight	$\frac{Leaf\_DW}{Leaf\_area}$	
Leaf_shoot_ratio	$\frac{Leaf\_DW}{Leaf\_DW+IN\_DW}$		
Trait physiologique	Plant_transpiration		pesée du pot
	Leaf_transpiration		pesée du pot
	Accumulated_transpiration		pesée du pot
	Water_use_efficiency	$\frac{f\_Bio-i\_Bio}{A\_T}$	

## II. MATERIEL ET METHODE.

### 1) Matériel végétal.

Les plantes utilisées dans cette étude sont des pommiers issus de la core collection de pommier de table français de l'INRA. Il s'agit des mêmes génotypes utilisés dans l'étude sur les variations génétiques lié à des traits morphologiques et à la transpiration de pommier sous condition hydrique non stressante (Lopez *et al.*, 2015). Sur les 278 génotypes de la core collection initialement définie à Angers (Lassois *et al.*, 2015) 179 sont utilisés dans cette étude car tous les individus n'ont pas survécu ou n'ont pas été multipliés et tous n'ont pas été génotypés. Les plantes ont été plantées à la station expérimental « DiaScope » à Montpellier en hiver 2014. Des scions de un an ont été prélevés et greffés sur un même porte-greffe. Il s'agit de PAJAM<sup>®</sup> 2 Cepiland (P2C), un porte greffe nanifiant issu d'une sélection clonale du Paradis Jaune de Metz (série M9) (Masseron *et al.*, 1985). Ce porte greffe a été choisi car il est le plus utilisé dans les vergers de production. Son utilisation permet donc une conduite de culture équivalente à celle des exploitations agricoles. D'avril à juin 2014 les plantes ont été cultivées en serre à « DiaScope » de manière à favoriser la croissance. Les plantes étaient cultivées de manière optimale avec une irrigation de surface trois fois par semaine, une fertilisation avant plantation, et un contrôle sanitaire jusqu'au 3 juin. A partir de cette date les plants ont été transférées dans la plateforme de phénotypage à haut débit PhénoArch où elles ont été cultivées jusqu'au 19 juillet. 8 réplicats ont été utilisés par génotype, quatre cultivés en condition d'irrigation optimale, et les quatre autres cultivés sous stress hydrique. Ce stress est définis par le rapport entre le poids d'eau et le poids de sol sec (SWC= soil water content). Le SWC a été maintenu autour de  $1.4\text{g.g}^{-1}$  pour les plantes non stressées, et a été réduit jusqu'à  $1.0\text{g.g}^{-1}$  pour les plantes stressées, ceci correspond à un stress intermédiaire, puis le stress a été réduit jusqu'à  $0.7\text{g.g}^{-1}$  (Lopez *et al.*, sous presse) (Figure 5). Pour les individus stressés seul le stress intermédiaire a été pris en compte dans cette étude.

### 2) Les variables phénotypiques.

20 variables ont été étudiées pour rendre compte de la production de biomasse du système aérien en lien avec des variables physiologiques liées à l'utilisation de l'eau. La production de biomasse a été décomposée en composantes qui peuvent être distinguées en trois groupes : celles liées à la biomasse de la plante entière, celles liées au métamère, et les variables physiologiques (Transpiration et WUE) (Tableau 1). Pour les plantes cultivées sous stress hydrique les variables liées au métamère correspondent à la moyenne des données recueillies durant la période de stress intermédiaire (SWC= $1\text{g.g}^{-1}$ ). Pour l'ensemble des variables neuf modèles mixtes ont été développés pour prendre en compte l'effet génétique et l'hétérogénéité spatiale au sein de la plateforme. Puis les BLUPs (Best Linear Unbiased Predictor) ont été estimés pour chaque variable après avoir sélectionné le modèle qui minimise le critère d'information bayésien (BIC) (Lopez *et al.*, 2015).

Deux modalités différentes ont été considérées dans cette étude, la première correspond aux plantes non stressées (WW= Well Watered), la deuxième aux plantes stressées (WS = Water stress). Pour chaque variable une corrélation a été réalisée entre les plantes cultivées sous stress hydrique et les plantes non stressées et l'écart à la droite de corrélation a été utilisé pour former un nouveau jeu de données. Ceci a été fait pour corriger la corrélation entre les modalités WW et WS, dans la mesure où une plante plus développée, est plus impactée par le stress hydrique qu'une plus petite.



### 3) Structure, apparentement, et déséquilibre de liaison.

#### 3.1) Les données génotypiques.

Les données génotypiques ont été obtenues à l'aide de la puce à ADN à 487 000 SNPs (487K SNP Affymetrix Axiom®). 275K marqueurs ont été utilisés, il s'agit de marqueurs définis comme étant très robustes et pouvant être utilisés en génétique d'association (Bianco *et al.*, 2016). Le génotypage de la population a été réalisé par l'UMR IRHS de l'INRA d'Angers.

L'analyse de GWAS nécessite au préalable plusieurs filtres sur le jeu de données génotypiques ([www.bioconductor.org](http://www.bioconductor.org)) cela afin de supprimer les données non informatives qui augmentent le temps de traitement et diminuent la puissance de détection des associations.

Un premier filtre a été réalisé au niveau de la fréquence de l'allèle mineur (MAF), afin de ne garder que les marqueurs avec une MAF supérieur à un certain seuil. Ce seuil est défini de manière à avoir un nombre d'individus par classe allélique suffisant afin de diminuer le risque de détection de faux positifs. Le filtre utilisé ici a été défini de manière à avoir un minimum de 5 individus par classe allélique.

Un second filtre a été réalisé pour ne pas avoir de marqueur pour lesquelles des individus seraient totalement homozygotes ou hétérozygotes. En effet l'ensemble des différentes classes alléliques sont nécessaires pour pouvoir comparer l'effet d'un allèle par rapport à un autre. Le taux d'hétérozygotie maximal d'un marqueur a alors été fixé à 95%.

Le nombre de données manquantes par génotype a été vérifié et devant le taux très faible trouvé il n'a pas été nécessaire de réaliser un troisième filtre.

Suite à ces différents filtres, 261 754 SNPs ont été retenus pour cette étude.

#### 3.2) Déséquilibre de liaison.

Le déséquilibre de liaison (LD) a été calculé entre chaque SNP d'un même groupe de liaison, pour pouvoir estimer l'espacement nécessaire entre les SNPs avant le calcul de la matrice d'apparentement. Le calcul a été réalisé à l'aide du programme TASSEL en utilisant un nombre de permutations de 1000 et une fenêtre de 50 SNPs. Le coefficient de Pearson  $R^2$  est calculé entre deux loci de la façon suivante :

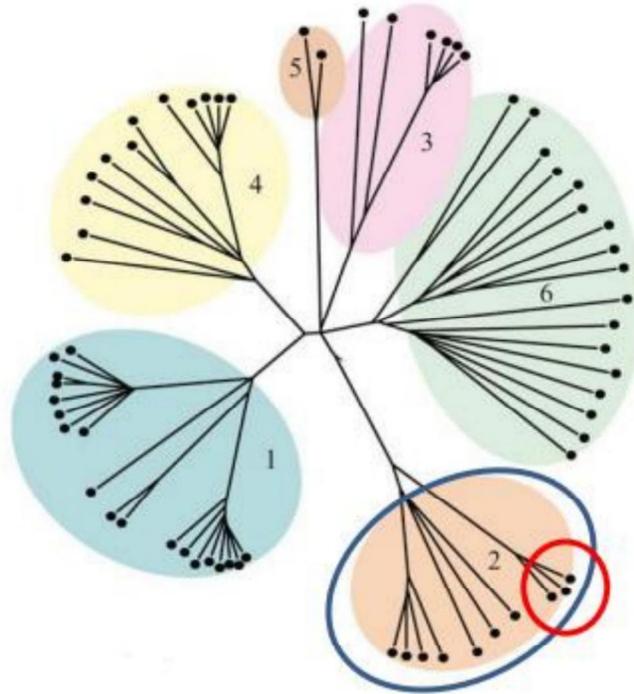
$$R^2 = \frac{D^2}{p(A)p(B)p(a)p(b)}$$

Où :

- D est la différence entre la fréquence observée et celle attendue par l'équilibre de Hardy-Weinberg :  $D_{AB} = p_{AB} - p_A p_B$
- Le dénominateur est le produit des deux paires de fréquences alléliques.

Une fois le  $R^2$  obtenu pour l'ensemble des marqueurs, des classes de distance ont été définies arbitrairement, et pour chaque classe de distance le  $R^2$  moyen a été calculé.

la décroissance du LD est observée en traçant les histogrammes des  $R^2_{\text{moyen}}$  par classe, les classes de distance étant : 0-100-500-1000-2000-3000-4000-5000-10000-50000-100000-200000-300000-400000-500000-600000-700000-800000-900000-1000000-2500000pb



**Figure 6 : Arbre illustrant la structure et l'apparentement au sein d'une population.**

Les différents groupes de structure sont représenté par les ovals colorés, le cercle rouge montre des individus apparentés. ( Cours d' Ingrid Vilmus, 2015)

### 3.3) Structure.

Il y a en GWAS besoin de connaître l'histoire évolutive de la population étudiée, car la structure de la population peut induire un biais lors d'analyse d'association en agissant comme un facteur confondant (Marchini *et al.*, 2004). Il est donc nécessaire de vérifier l'effet de la structuration de la population. La structure est définie par le nombre K de groupes de génomes ancestraux et la probabilité qu'ont ces individus d'appartenir à ces différents groupes.

L'estimation de la structure est basée sur le maximum de vraisemblance et assigne les individus aux groupes en fonction de leur fréquence allélique multi-loci. L'analyse teste différents nombres K de groupes et le choix final est fait à l'aide d'une méthode de validation croisée.

Cette méthode consiste à partitionner l'ensemble des génotypes observés en K groupes. Les groupes sont alors analysés un par un en gardant un des groupes de génotypes et en considérant les autres génotypes comme des données manquantes. Puis à partir du groupe conservé les génotypes masqués sont estimés et l'erreur de prédiction (cross validation error) est estimée par comparaison entre données prédites et données réelles. Le nombre K de génomes ancestraux choisi est celui pour lequel l'erreur de la validation croisée est la plus faible sur le graphique d'erreur de validation croisée en fonction du nombre k de groupes ancestraux.

Pour estimer la structure de la population j'ai utilisé le programme ADMIXTURE (Alexander *et al.*, 2009 ; Zhou *et al.*, 2009) sous la plateforme SNIPLAY. Ce programme a été choisi car il est beaucoup plus rapide que d'autres comme EIGENSTRAT ou STRUCTURE. Ce gain de vitesse permet l'utilisation d'un jeu de marqueurs bien plus important, et permet de corriger la stratification de la population dans les études d'association (Alexander *et al.*, 2009). De plus cet outil permet l'estimation de la structure d'une population non apparentée comme est supposée être la population étudiée ici (Lasso *et al.*, 2015).

### 3.4) Apparentement.

En GWAS, les individus phénotypés sont considérés comme non apparentés, mais des liens de parenté entre individus peuvent être importants. La structure se réfère à une ancestralité commune lointaine entre les individus, l'apparentement se réfère quant à lui à une ancestralité commune récente (Cours d'Ingrid Vilms, 2015) (Figure 6). En effet différents groupes d'individus apparentés peuvent avoir un ancêtre commun éloigné. Pour prendre en compte l'apparentement, une matrice d'identité par état (IBS) a été calculée et utilisée en covariable pour l'analyse d'association. L'apparentement est mesuré par la probabilité qu'à un locus donné deux individus partagent le même allèle issu d'un ancêtre commun (Eding et Meuwissen, 2001). Cette probabilité varie en fonction du degré d'apparentement entre les deux individus. L'apparentement peut être calculé à partir de l'information apportée par les marqueurs en se basant sur le principe d'identité par état. Si à un locus donné deux individus portent le même allèle alors il est probablement hérité d'un ancêtre commun. La méthode utilisée pour ce calcul est la méthode « centrée » (Endelman et Jannink, 2012).

Pour réaliser la matrice d'apparentement les SNPs sont préalablement filtrés pour ne garder que des SNPs éloignés de 10Kpb cela afin d'avoir suffisamment de marqueurs tout en évitant de prendre des SNPs en trop fort déséquilibre de liaison (Zhu *et al.*, 2008 ; Astel et Balding, 2009).



#### 4) GWAS

##### 4.1) Présentation du modèle d'estimation des effets génétiques.

Un test de Shapiro a été réalisé sous un risque  $\alpha=0.05$  pour vérifier la normalité de la distribution des variables. Au vu de la distribution normale et de la nature continue des variables utilisées, un modèle linéaire a été choisi pour réaliser l'étude d'association. Pour pouvoir utiliser l'apparentement des génotypes en effet aléatoire le modèle linéaire mixte est choisi. Inclure les effets aléatoires permet d'améliorer le pouvoir statistique de l'association.

Deux modèles ont été testés. L'un prenant en compte l'apparentement mais pas la structure et l'autre prenant en compte la structure et l'apparentement.

Le modèle complet peut s'écrire :  $y = \mu + Qv + X\beta + Zu + e$

$\mu$  correspond à la moyenne générale,  $Q$  est la matrice de structure,  $v$  est le vecteur des effets du groupe de structure,  $\beta$  est l'effet de l'allèle au marqueur et  $X$  correspond au génotype des individus au marqueur testé.  $\mu$ ,  $Qv$ , et  $x\beta$  sont les effets fixes du modèle,  $Z$  correspond à la matrice d'apparentement,  $u$  correspond à l'effet aléatoire du génotype et  $e$  est le vecteur des résidus.  $u$  et  $e$  sont les effets aléatoires du modèle.

Pour  $i$  le nombre d'individus,  $k$  le nombre de groupes de structure, et  $m$  le nombre de marqueurs, les dimensions des différentes matrices utilisées sont :  $Q(i,k)$ ,  $Z(i,i)$ ,  $X(m,i)$

La dimension des vecteurs utilisés sont :  $y(i)$ ,  $v(i)$ ,  $\beta(m)$ ,  $u(i)$ .

Le second modèle qui ne prend pas en compte la structure s'écrit :  $y = \mu + X\beta + Zu + e$

Le meilleur modèle a été choisi pour chaque variable par comparaison du maximum de vraisemblance et en utilisant le critère d'information bayésien (BIC) (Schwarz, 1978). Le BIC est défini par la formule  $-2 \ln(L) + k \ln(n)$  avec  $L$  le maximum de vraisemblance du modèle estimé,  $k$  le nombre de paramètres estimés et  $n$  la taille de la population. Le modèle avec le plus petit BIC est sélectionné.

Afin de réduire le temps d'analyse et d'améliorer l'adéquation au modèle une compression qui réduit la dimension de la matrice d'apparentement a été utilisée. Le principe de cette méthode est de regrouper les génotypes en clusters en se basant sur la matrice d'apparentement.

Les composantes de la variance génétique et résiduelle ont été ré-estimées à chaque marqueur par une méthode REML (Restricted Maximum Likelihood). Ceci permet d'optimiser le pouvoir de détection d'association (Baptiste Guillon, communication personnelle).

##### 4.2) Seuil de significativité des associations.

Les résultats d'association sont présentés sous la forme d'un Manhattan plot avec en abscisse la position le long du génome, et en ordonné le  $-\log(p)$  où  $p$  représente la p-value de chaque SNP testé sous l'hypothèse  $H_0$  : le SNP n'a pas d'effet sur la variable phénotypique.

Une fois les effets alléliques estimés il faut définir un seuil de significativité d'association entre les marqueurs et les variables.

Le seuil de Bonferroni (Bonferroni ; 1950) classiquement utilisé en GWAS pour déterminer la significativité d'une association est défini comme  $-\log_{10}(0.05/m)$  avec  $m$  le nombre de marqueurs. Ce seuil dépend donc du nombre de marqueurs utilisés. Pour un nombre important



de marqueurs le seuil devient alors trop strigent pour détecter des associations. C'est pourquoi de nombreuses études en GWAS utilisent des seuils définis arbitrairement afin de se dégager du bruit de fond tout en captant des associations (Moriya *et al.*, 2016). Ici le nombre de marqueur utilisé est grand (273K SNPs). Au vu des premiers Manhattan Plot, le seuil a été défini arbitrairement à  $-\log(p)=5$ .

Un second seuil plus bas a aussi été considéré à  $-\log(p)=3.5$ . Ce seuil va permettre de comparer les associations des variables fortement corrélées. En effet il est possible qu'on observe une association significative dépassant le seuil haut pour une variable, mais que pour une autre fortement corrélée on ne retrouve pas cette association. Le seuil bas va alors permettre de voir si on observe bien une association pour le même locus même si cette dernière n'est pas significative.

Pour déclarer une association comme significative il faut aussi prendre en compte l'isolement du SNP. En effet un SNP hautement significatif qui se retrouve isolé sur le Manhattan plot est sûrement un faux positif car le déséquilibre de liaison devrait faire que les SNPs proches soient également au-dessus du seuil de significativité.

Pour le chromosome 18 (correspondant au SNP dont la position est méconnue) les associations trouvées devront être vérifiées selon la même procédure une fois la position des SNPs établie (en cours).

#### 5) Gènes candidats.

Une fois des associations détectées une approche gène candidat a été réalisée afin de trouver quels gènes se trouvent dans l'intervalle entre les marqueurs significatifs.

##### 5.1) Recherche des gènes

L'analyse d'association a permis de détecter la position des SNPs associés significativement avec un des traits d'intérêts. Une fois cette position connue, un intervalle de 1000pb a été définie (500pb avant et après le SNP) quand un unique SNP dépasse le seuil de significativité. Quand plusieurs SNPs dépassent le seuil dans une même région chromosomique, les positions extrêmes ont été utilisées comme bornes de l'intervalle. Ces intervalles ont été recherchés dans la base de données génétiques des Rosaceae ([rosaceae.org](http://rosaceae.org)) pour trouver les contigs associés et faire le lien avec la version V1 du génome du pommier (Velasco *et al.*, 2010) (mieux annotée que la version V3 ([rosaceae.org](http://rosaceae.org))). Une fois les contigs trouvés il est possible de trouver les gènes (les MDPs) dans les intervalles définis.

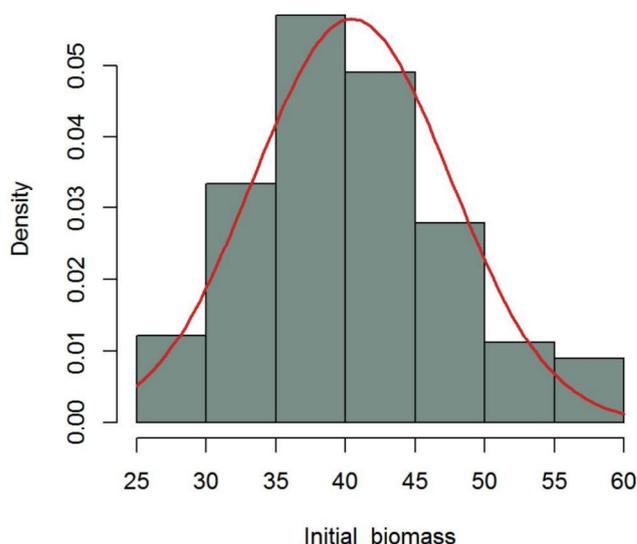
##### 5.2) Recherche de la fonction des gènes

Toujours dans la base de données des Rosaceae, l'homologie entre les gènes trouvés sur la version V1 du génome du pommier (les MDP) et les gènes d'*arabidopsis thaliana* est disponible. Cela permet d'obtenir pour chaque MDP présent dans l'intervalle la correspondance avec les gènes d'*arabidopsis thaliana*. Puis ces gènes ont été recherchés dans la base de données génétique de cette plante modèle (TAIR : [arabidopsis.org](http://arabidopsis.org)) afin d'obtenir leur fonction connue.

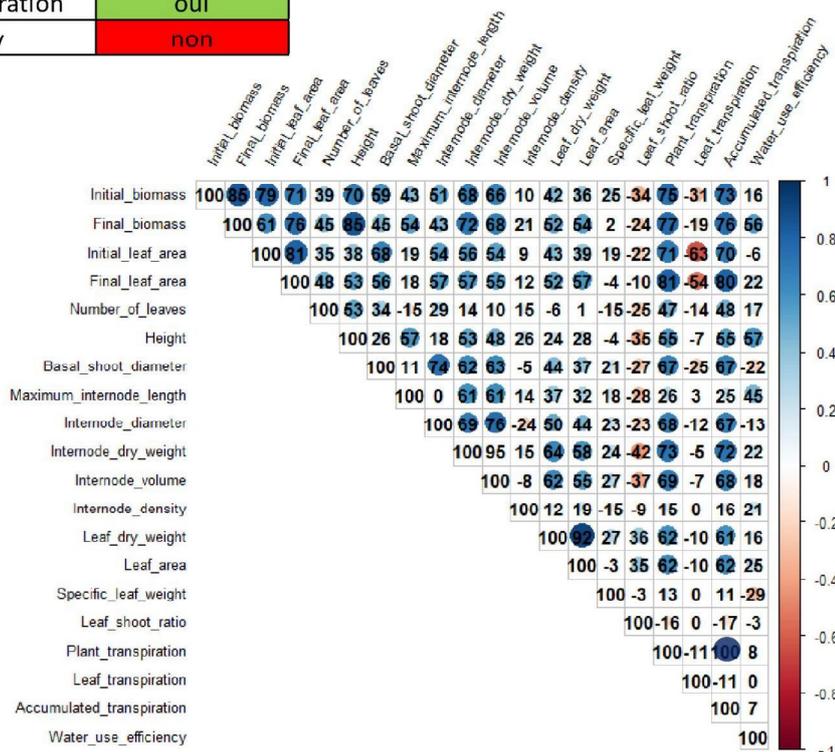
**Tableau 2 : Vérification de la distribution normale des variables par un test statistique de Shapiro au risque  $\alpha=0.05$**

Variables	Normalité (Test de Shapiro pvalue<0.05)
Initial_biomass	non
Final_biomass	oui
Initial_leaf_area	non
Final_leaf_area	non
Number_of_leaves	oui
Height	oui
Basal_shoot_diameter	oui
Maximum_internode_length	non
number_of_leaves	oui
Internode_dry_weight	non
Internode_volume	oui
Internode_density	oui
Leaf_dry_weight	oui
Specific_leaf_area	oui
Specifiq-leaf_weight	oui
Leaf_shoot_ratio	non
Plant_transpiration	oui
Leaf_transpiration	oui
Accumulated_transpiration	oui
Water_use_efficiency	non

**Histogramme de Initial\_biomass**



**Figure 7: Histogramme de distribution de la biomasse initiale, la courbe rouge représente la courbe de normalité**



**Figure 8: Corrélations entre les différentes variables étudiées exprimées en pourcentage.**

**Tableau 3 : Déséquilibre de liaison moyen pour une distance supérieur à 100Kpb, et 200Kpb quand le seuil de  $R^2=0.2$  n'est pas atteint.**

Chromosome	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
$R^2$ moyen pour une distance supérieure à 100kpb	0.228	0.146	0.219	0.197	0.175	0.230	0.190	0.206	0.217	0.174	0.200	0.172	0.267	0.200	0.242	0.251	0.152

### III. RESULTATS

#### 1) Exploration statistique des variables.

##### 1.1) Vérification de la normalité.

La vérification de la normalité de distribution des variables a été réalisée à l'aide d'un test de Shapiro (Tableau 2).

La normalité a été observée pour la majorité des variables. Pour certaines variables, le test de Shapiro indique une distribution non normale. Cependant, l'observation de leur distribution par rapport à la courbe de normalité montre qu'elles n'en sont pas très éloignées. Elles ont été considérées normales dans la suite de l'étude. Par exemple la distribution de l' « Initial\_biomass » (Figure 7). Les autres histogrammes de distribution des variables pour lesquelles le test de shapiro indique une distribution non normale sont présentés en annexe 1.

##### 1.2) Analyse de corrélation entre les variables.

Les coefficients de corrélation ont des valeurs élevées entre certaines variables, notamment celles qui diffèrent uniquement par la date d'acquisition des résultats (Initial\_biomass et Final\_biomass par exemple). On observe également une forte corrélation pour les variables de même type, par exemple les variables décrivant les feuilles sont bien corrélées entre elles. Il en est de même pour les variables décrivant les entre-nœuds. En effet une corrélation de 0.95 est observée entre le volume des entre-nœuds et le poids sec de ces derniers. Une corrélation de 1 est observée entre la transpiration foliaire totale et la transpiration accumulée (Figure 8).

Les fortes corrélations entre les variables ont permis de diminuer leur nombre de variables. Ainsi seule une des deux dernière variables citées sera analysée car leur information est redondante. Ceci a également permis de vérifier les associations entre génotypes et phénotypes. En effet, des variables très corrélées devraient montrer des associations communes. Cette vérification a permis d'écarter d'éventuels faux positifs lors de l'analyse de génétique d'association.

##### 1.3) Déséquilibre de liaison.

On observe un  $R^2_{\text{moyen}}$  autour de 0.2 pour une distance supérieure à 100Kpb pour la majorité des différents chromosomes (Tableau 3). Il a été choisi d'espacer les SNPs de seulement 10Kpb pour le calcul de la matrice d'apparement car en espaçant les SNPs de 100Kpb trop de SNPs étaient supprimés.

Pour certains chromosomes (3, 6, 7, 10, 12, 13, et 16), il existe un fort LD à grande distance. Ceci est dû au fait que lors du génotypage certains marqueurs ont été mal placés sur le génome et ont été mis en extrémité des chromosomes. Ceci fausse alors l'observation du LD à grande distance. C'est pourquoi les SNPs associés significativement à une des variables et positionnés en extrémité des chromosomes n'ont pas été considéré dans la recherche de gènes candidats.

### Cross-validation plot

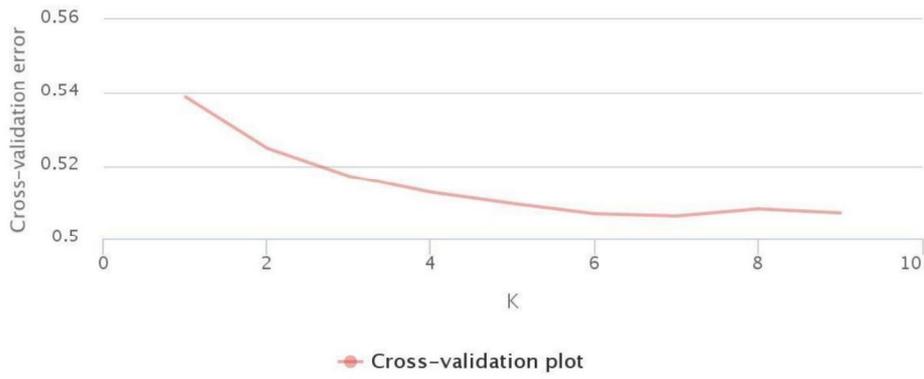


Figure 9 : graphique de l'erreur de la validation croisé en fonction du nombre K de groupes ancestraux

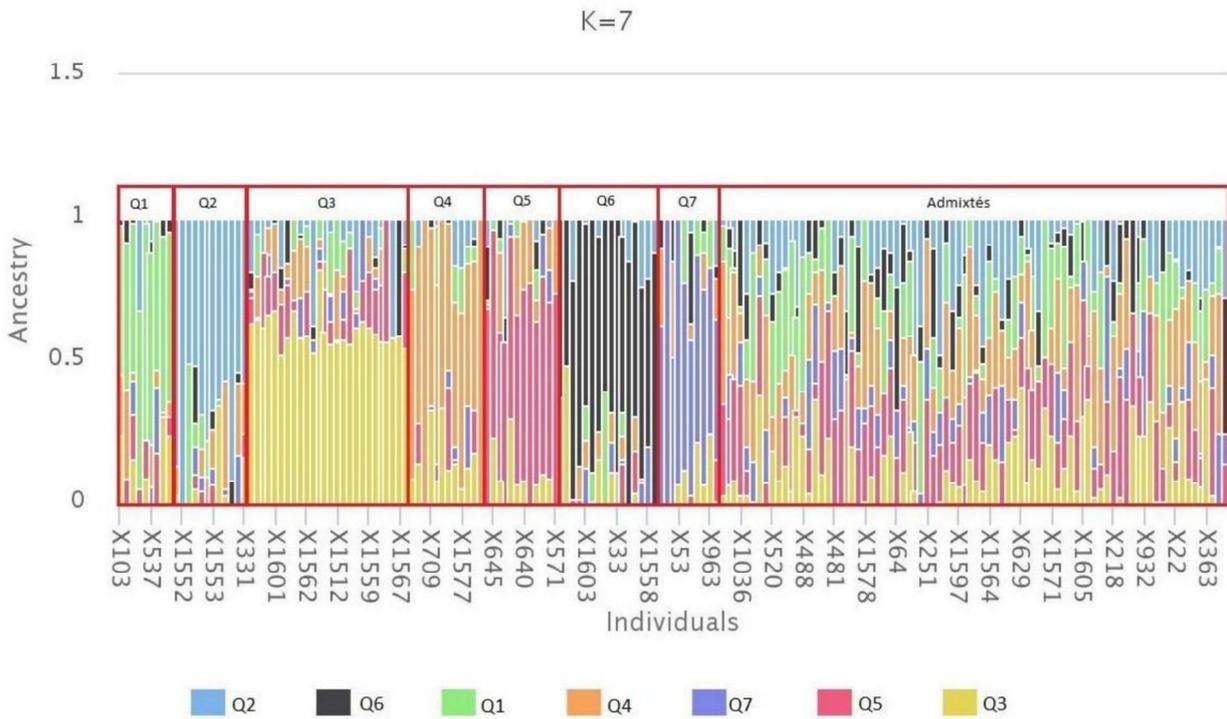


Figure 10 : Proportion d'ascendance des 179 individus étudiés pour K=7 groupes ancestraux.

Highcharts.com

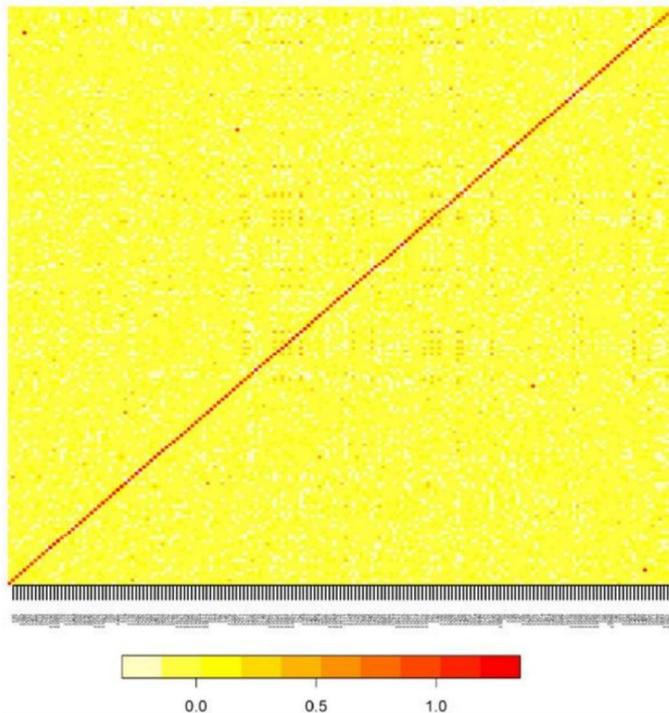


Figure 11: Matrice d'apparement, le rouge indique les individus fortement apparentés, le blanc le jaune les individus faiblement apparentés

## 2) Structure et apparentement

### 2.1) Structure

En figure 9 est présenté le graphique de validation croisée (erreur de validation croisée en fonction du nombre  $K$  de génomes ancestraux). Le nombre  $K$  qui minimise l'erreur de validation croisée est égal à 7, nombre à partir duquel un plateau apparaît. On choisit donc 7 comme nombre de génomes ancestraux pour notre population.

La proportion d'ascendance des 179 individus étudiés pour  $K=7$  groupes de génomes ancestraux est présentée en figure 10. La proportion d'ascendance est la part de génome des individus provenant des génomes ancestraux. Chaque génotype est représenté par une barre verticale divisé en  $K=7$  segments représentant la quantité d'ascendance des sept groupes dans son génome. On observe qu'une grande partie de la population est admixtée avec une probabilité d'appartenir à un des groupes ancestraux toujours inférieurs à 50%. Ces individus admixtés représentent 49% de la population étudiée. La grande majorité des individus correspondent à un mélange de différents génomes ancestraux. En effet seulement 9% des individus ont une probabilité d'appartenance à un des groupes ancestraux supérieurs à 0.75. Aucun groupe bien défini n'est donc observé, tous les individus ayant dans leur génome un mélange de différents génomes ancestraux, la population ne semble donc pas fortement structurée en groupes distincts. Les individus semblent donc de par ce mélange apparentés les uns aux autres et partagent différents génomes ancestraux. On peut supposer de part leur origine commune, qu'ils sont issus d'un nombre limité de fondateurs, et de ce fait il est probable que ces individus soient apparentés les uns aux autres.

### 2.2) Apparentement.

La matrice d'apparentement a été calculée par IBS (identité par état) (Figure 11). Les individus fortement apparentés sont représentés en rouge et les individus faiblement apparentés sont représentés en jaune. On observe que la plupart des individus sont faiblement apparentés les uns aux autres, hormis deux paires qui le sont fortement. Les individus X8409 et X8411 sont fortement apparentés, ainsi que les individus X9437 et X8223. Ce fort apparentement concerne la Reinette de Maurs avec la Reinette sanguine du Rhin, et la Gewurtzleiken avec la Platet. Une douzaine d'individus ont un apparentement de l'ordre de 0.5, il s'agit des individus X8229 avec X9122, X9130, X1314, X2953 ; X9122 avec X1314, X9090 ; X0640 avec X8392 ; X1225 avec X0421 ; X8244 avec X0898 ; X0036 avec X9090 ; X1314 avec X2953 ; X8203 avec X8209. Les apparentements sont donc très variables au sein de la population, et cette matrice n'est donc pas à négliger et doit d'être prise en compte lors de l'analyse d'association.

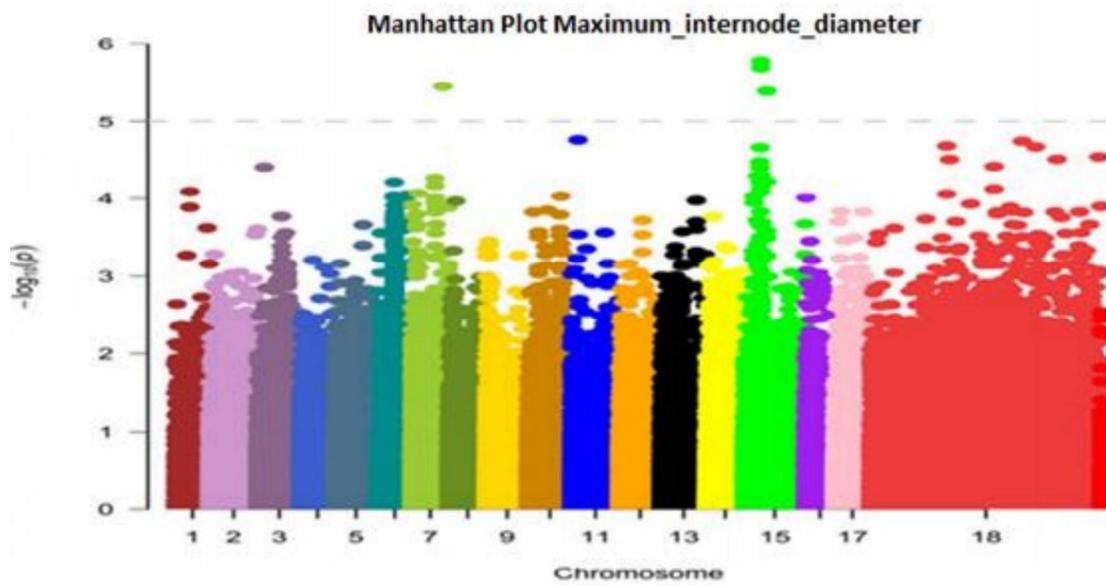


Figure 12 : Manhattan plot de la variable Maximum\_internode\_diameter en condition non stressante, illustré pour l'ensemble des chromosomes

Manhattan\_MLM\_K10\_Ch7\_Maximum\_Internode\_diameter

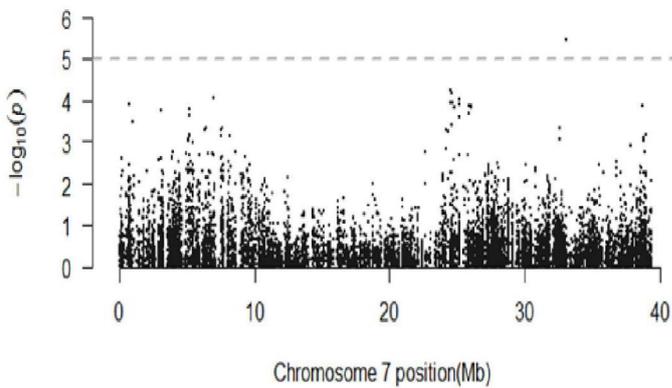


Figure 13 : Manhattan plot de la variable Maximum\_internode\_diameter du chromosome 7 en condition non stressante.

Manhattan\_MLM\_K10\_Ch15\_Maximum\_Internode\_diameter

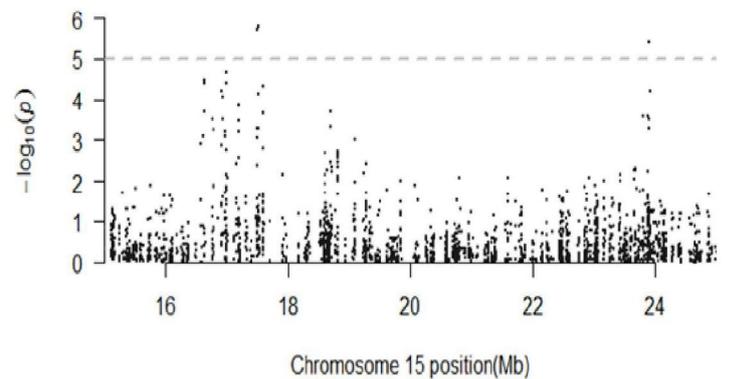


Figure 14 : Manhattan plot de la variable Maximum\_internode\_diameter du chromosome 15 (zoomé sur les association) en condition non stressante.

Manhattan\_MLM\_K10\_Ch15\_Basal\_shoot\_diameter

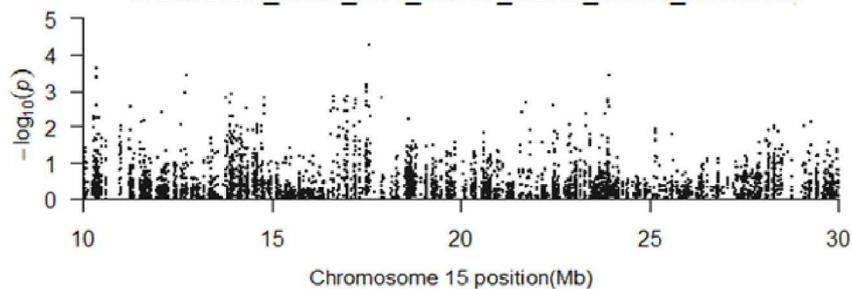


Figure 15 : Manhattan plot de la variable Basal\_shoot\_diameter du chromosome 15 en condition non stressante.

### 3) GWAS

#### 3.1) Analyse des variables obtenues en condition non stressante.

##### 3.1.1) Choix des modèles en condition non stressante.

Pour la quasi-totalité des variables le modèle prenant en compte l'apparement a été sélectionné. Le modèle complet prenant en compte la structuration en 7 groupes et l'apparement est sélectionné pour seulement quatre variables (Final\_biomass, Initial\_leaf\_area, Final\_leaf\_area, et Accumulated\_transpiration). Un tableau récapitulatif des modèles choisis est disponible en annexe 2.

##### 3.1.2) Analyse des associations.

Pour un seuil placé à  $-\log(p)=5$  des associations significatives entre génotypes et phénotypes ont été observées pour 13 variables sur 20. Il s'agit de : Plant\_transpiration, accumulated\_transpiration, initial\_biomass, final\_biomass, initial\_leaf\_area, number\_of\_leaves, leaf\_area, leaf\_dry\_weight, maximum\_internode\_length, internode\_diameter, internode\_volume, specific\_leaf\_weight, et leaf\_shoot\_ratio.

La variable Maximum\_internode\_diameter est traitée en exemple, les autres variables ont été analysées de façon similaire.

Le Manhattan plot de la variable Maximum\_internode\_diameter pour l'ensemble des chromosomes, indique trois associations significatives sur les chromosomes 7 et 15 (Figure 12).

Le Manhattan plot de la variable Maximum\_internode\_diameter sur le chromosome 7, indique que l'association qui dépasse le seuil de significativité concerne un unique SNP isolé. Il a donc été considéré comme faux positif (Figure 13).

Le Manhattan plot de cette même variable sur le chromosome 15, indique des associations significatives au niveau de deux loci (aux alentours de 17Mb et de 24Mb). La présence de plusieurs SNP dépassant le seuil ainsi que la présence de chandelle conforte l'association observée (Figure 14)

Le diamètre maximal des entre-nœuds est corrélé à 0.74 avec la variable Basal\_shoot\_diameter. Or bien qu'aucune association ne dépasse le seuil de significativité pour cette dernière variable, des chandelles correspondant à une augmentation locale du  $-\log(p)$  peuvent être observées sur son Manhattan plot au niveau du chromosome 15 pour des positions similaires, où des SNPs atteignent un  $-\log(p)$  de 4. Ceci conforte notre association (Figure 15).

Un tableau récapitulatif des différentes associations en condition non stressante est présenté en annexe 3a et 3b. On observe 15 associations significatives pour les variables liées à la plante entière, 39 associations significatives pour les variables liées aux métamères, et pour les variables correspondant aux traits physiologiques seul 3 SNPs sont associés significativement. Pour l'ensemble des associations la significativité n'est pas très forte (de l'ordre de  $-\log(p)=5.2$ ). Seul un SNP est associée très significativement à une variable, il s'agit de Affx-114058120 qui est associé à l'aire foliaire initiale sur le chromosome 9 avec un  $-\log(p)=6.59$ .

**Tableau 4 : Nombre d'associations significatives trouvées par variable en condition de stress hydrique.**

	Variable	SNP associé significativement
Plante entière	Initial_biomass	1
	Final_biomass	3
	Initial_leaf_area	2
	Final_leaf_area	0
	Number_of_leaves	2
	Height	5
	Basal_shoot_diameter	0
Métamère	Maximum_internode_length	0
	Internode_diameter	2
	Internode_dry_weight	3
	Internode_volume	0
	Internode_density	1
	Leaf_area	4
	Leaf_dry_weight	1
	Specific_leaf_weight	1
	Leaf_shoot_ratio	2
Trait physiologique	Plant_transpiration	0
	Leaf_transpiration	4
	Accumulated_transpiration	0
	Water_use_efficiency	0

**Tableau 5 : Coefficient de corrélation de la régression linéaire des variables entre la condition de non stress et la condition de stress hydrique.**

Variables	R <sup>2</sup> entre WW et WS
Initial_biomass	0.36
Final_biomass	0.40
Initial_leaf_area	0.39
Final_leaf_area	0.39
Number_of_leaves	0.25
Height	0.36
Basal_shoot_diameter	0.13
Maximum_internode_length	0.52
Internode_diameter	0.16
Internode_dry_weight	0.26
Internode_volume	0.25
Internode_density	0.10
Leaf_dry_weight	0.42
Specific_Leaf_area	0.44
Specific_leaf_weight	0.29
Leaf_shoot_ratio	0.31
Plant_transpiration	0.17
Leaf_transpiration	0.50
Accumulated_transpiration	0.17
Water_use_efficiency	0.36

**Tableau 6 : Valeur de BIC pour les variables des résiduelles, le jaune correspond au modèle sélectionné (modèle qui minimise le BIC)**

	MLM_K	MLM_K_Q
Initial_biomass_res	1042.956	1063.305
Final_biomass_res	1192.061	1209.018
Initial_leaf_area_res	1975.972	1961.871
Final_leaf_area_res	2161.827	2144.222
Number_of_leaves_res	564.4149	604.8103
Height_res	977.6353	998.9788
Basal_shoot_diameter_res	42.58841	95.98439
Maximum_internode_length_res	879.4997	903.7101
Internode_diameter_res	16.82865	74.85625
Internode_dry_weight_res	-604.3316	-527.8578
Internode_volume_res	-324.2216	-261.8378
Internode_density_res	-928.1216	-840.3378
Leaf_dry_weight_res	-524.8616	-452.6078
Leaf_area_res	1011.503	1031.939
Specific_leaf_weight_res	-2227.072	-2089.958
Leaf_shoot_ratio_res	-869.4616	-778.1578
Plant_transpiration_res	1030.595	1053.598
Leaf_transpiration_res	-1139.072	-1041.45
Accumulated_transpiration_res	2146.529	2128.46
Water_use_efficiency_res	-1871.572	-1744.958

### 3.2) Analyse des variables obtenues en condition de stress hydrique.

#### 3.2.1) Choix des modèles en condition de stress hydrique.

Le modèle qui prend en compte la matrice d'apparement a été retenu pour la majorité des variables. Le modèle prenant en compte la matrice de structure en sept groupes et l'apparement est choisie pour 4 variables sur 20 (Initial\_leaf\_area, Final\_leaf\_area, Maximum\_internode\_length, et Accumulated\_transpiration). On remarque que le modèle prenant en compte matrice et apparement a été sélectionné pour trois même variable entre la modalité WW et WS (Initial\_leaf\_area, Final\_leaf\_area, et Accumulated\_transpiration) Cependant après observation des Manhattan plots certaines associations ont un  $-\log(p)$  plus grand avec le modèle prenant en compte la matrice de structure par rapport au modèle prenant en compte l'apparement, et des associations significatives sont observés bien que ce ne soit pas le modèle qui minimise le BIC. Pour ces variables le modèle aboutissant à des associations plus significatives est sélectionné. Ceci est le cas pour les variables : specific\_leaf\_area, et specific\_leaf\_weight. Pour la variable leaf\_area\_initial des associations différentes ont été trouvées en fonction du modèle, les résultats des deux modèles sont alors étudiés.

#### 3.2.2) Analyse des associations.

7 variables n'ont pas montré d'association avec des marqueurs pour le seuil de significativité défini. Ces variables sont : Final\_leaf\_area, Basal\_shoot\_diameter, Maximum\_internode\_length, Internode\_volume, Plant\_transpiration, Accumulated\_transpiration, et Water\_use\_efficiency. Pour l'ensemble des autres variables des associations ont été trouvées (Tableau 4).

Un tableau récapitulatif des différentes associations en condition de stress hydrique est présenté en annexe 4a et 4b. Treize associations significatives ont été trouvées pour les variables à l'échelle de la plante entière. Douze pour les variables à l'échelle des métamères. Une seule association significative est observée pour les variables de traits physiologiques, cette association concerne la transpiration foliaire (Tableau 4).

Aucune co-localisation n'est observée entre les associations significatives des variables en condition non stressé et en condition de stress hydrique. De plus la valeur de  $-\log(p)$  est pour l'ensemble des associations de comprise entre 5 et 5.7. Aucun SNP n'est associé à une variable avec une valeur hautement significative ( $-\log(p) > 6$ ).

Les corrélations entre les variables en condition WW et WS sont faibles (Tableau 5). En effet la corrélation la plus forte n'a un coefficient de corrélation  $R^2$  que de 0.5. On peut supposer que le changement d'hydratation à modifié la dynamique des variables ce qui pourrait expliquer que les associations ne ressortent pas sur les même loci dans ces deux conditions.

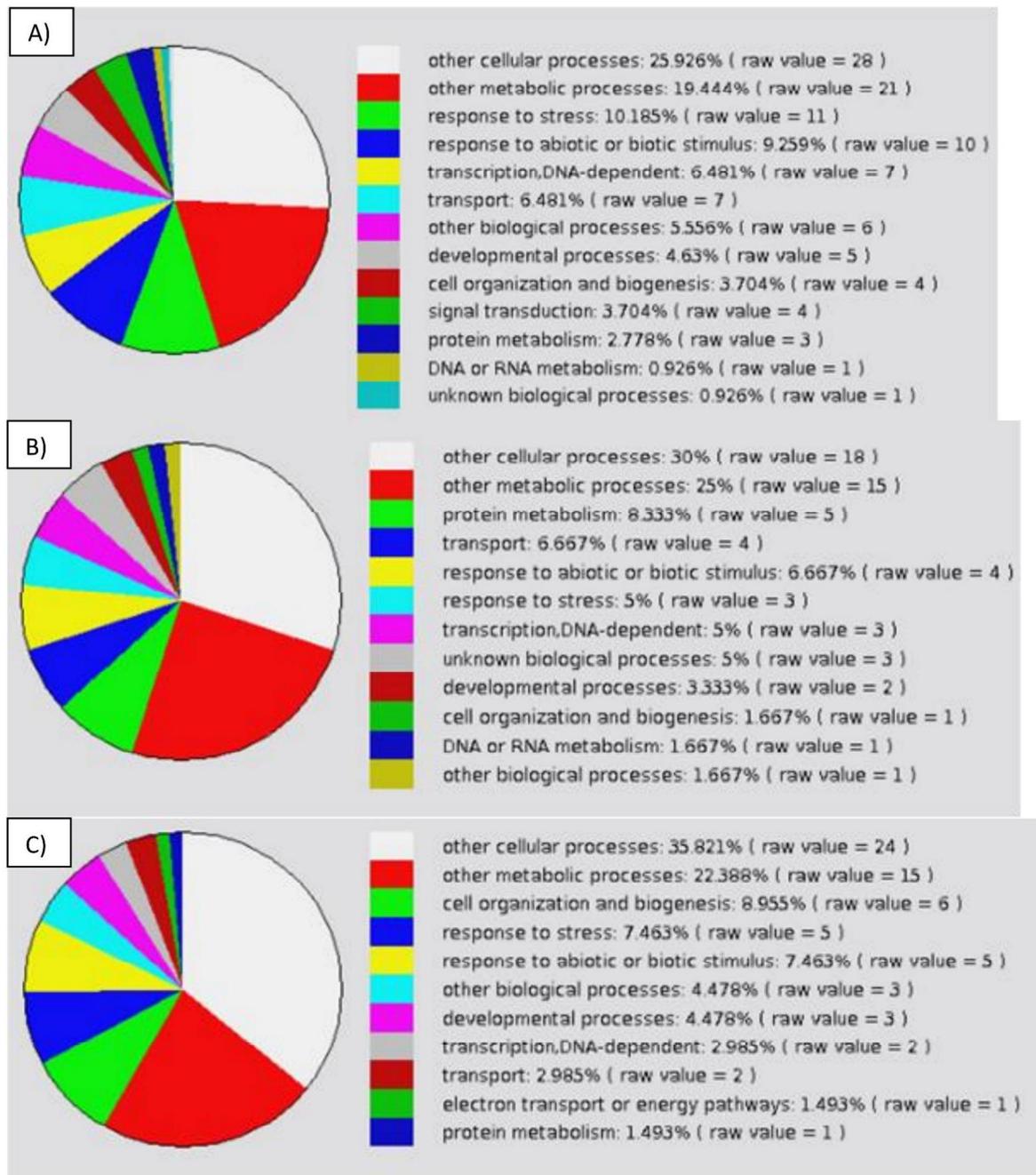
### 3.3) Analyse des variables obtenues sur les valeurs résiduelles.

#### 3.3.1) Choix des modèles.

Le modèle complet prenant en compte l'apparement et la structure (MLM\_K\_Q) est sélectionné pour seulement trois variables sur vingt (Initial\_leaf\_area, Final\_leaf\_area, et Accumulated\_transpiration). Pour ces trois variables le modèle MLM\_K\_Q est sélectionné pour les trois modalités. Pour les autres variables c'est le modèle ne prenant en compte que l'apparement (MLM\_K) qui est sélectionné (Tableau 6).

**Tableau 7 : Nombre d'associations significatives trouvées pour les résiduelles**

	Variables	SNPs associés significativement
Plante entière	Initial_biomass_res	1
	Final_biomass_res	0
	Initial_leaf_area_res	4
	Final_leaf_area_res	0
	Number_of_leaves_res	2
	Height_res	0
	Basal_shoot_diameter_res	0
Métamère	Maximum_internode_length_res	2
	Internode_diameter_res	0
	Internode_dry_weight_res	1
	Internode_volume_res	1
	Internode_density_res	0
	Leaf_area_res	13
	Leaf_dry_weight_res	4
	Specific_leaf_weight_res	1
	Leaf_shoot_ratio_res	9
Trait physiologique	Plant_transpiration_res	1
	Leaf_transpiration_res	3
	Accumulated_transpiration_res	1
	Water_use_efficiency_res	1



**Figure 16 : Ontologie des processus biologique des gènes candidats trouvés pour : A) la modalité WW, B) la modalité WS, C) les résiduelles**

### 3.3.2) Analyse des associations.

Six variables sur les vingt n'ont pas montré d'association avec des marqueurs pour le seuil de significativité défini. Ces variables sont : Final\_biomass, Final\_leaf\_area, Height, Basal\_shoot\_diameter, Internode\_diameter, Internode\_density. Pour l'ensemble des autres variables des associations ont été trouvées (Tableau 7).

Sept associations significatives ont été trouvées pour les variables à l'échelle de la plante entière, 31 pour les variables à l'échelle des métamères, et 6 associations significatives sont observées pour les variables de traits physiologiques (Tableau 7).

Une co-localisation d'association est observée pour l'ensemble des modalités sur le chromosome 11 pour trois variables Internode\_dry\_weight, Leaf\_area, et Leaf\_shoot\_dry\_weight, Plusieurs co-localisations sont observées entre le jeu de données des résiduelles et WW sur le chromosome 3, 9, 10 en prenant un intervalle de 1Mb maximum. De plus une co-localisation exacte (même position à la base près) est observée avec le jeu de données des variables obtenues en condition de stress hydrique. Cette co-localisation concerne la variable Leaf\_shoot\_ratio et se situe sur le chromosome 11 pour deux SNPs aux positions respectives de 11.9 et 10.0 Mb (Annexe 5a et 5b)

La majorité des associations significatives ont une valeur de  $-\log(p)$  comprise entre 5.0 et 5.5. Certains SNPs sont associés aux variables de manière plus significative, avec une valeur de  $-\log(p)$  proche de 6, il s'agit des SNPs Affx-113983729, Affx-113755966 respectivement associés aux variables Initial\_leaf\_area et Leaf\_area.

#### 4) Gènes candidats.

De nombreux gènes candidats ont été trouvés pour chaque modalité. 152 MDPs ont été trouvés, sur ces 152 MDPs, le lien avec les gènes d'*Arabidopsis* a permis d'identifier 121 accessions de gènes d'*Arabidopsis*, et sur ces 121 gènes 45 gènes différents aux fonctions connus ont été identifiés. Pour la modalité WW 22 gènes différents ont été trouvés, 14 pour WS et 12 pour les résiduelles. En annexe 6 sont présentés les gènes candidats replacés sur les chromosomes, le rouge représente la modalité résiduelle, en vert la modalité non stressée, et en bleu la modalité de stress hydrique (Annexe 6 a,b,c,d).

Pour toutes les modalités (WW, WS, et résiduelle) les ontologies des processus biologiques majoritairement représentés sont des processus métaboliques. On observe des gènes en lien avec le développement (5 pour WW, 2 pour WS, et 3 pour les résiduelles). Des gènes sont aussi en rapport avec la biogénèse (4 pour WW, 1 pour WS, et 6 pour les résiduelles). Et au niveau de la réponse aux stress, 11 ont été trouvés pour la modalité WW, 3 pour WS, et 5 pour les résiduelles (Figure 16). Plusieurs gènes candidats trouvés ont pour fonction une activité kinase. Des gènes de résistance au stress hydrique ont été trouvés (COR414-TM1, ATBPM4, MBF1C, PLDEPSILON, CYCD6;1). Ainsi que plusieurs gènes liés au cytosquelette (famille des tubulines).



## IV. DISCUSSION

### 1) Matériel végétal et dispositif expérimental.

L'expérimentation dont ont été tirées les données a été réalisée en conditions contrôlées, ce qui permet un meilleur contrôle de la croissance. Le besoin en eau des plantes a été mesuré de manière spécifique pour chaque individu par pesée des pots et comparaison entre pesées successives. Ceci a permis un meilleur contrôle des conditions de croissance qu'en plein champ.

Tous les individus étudiés sont greffés sur un même porte greffe ce qui permet une homogénéisation du développement racinaire, mais induit un effet du porte greffe sur le greffon, car tous les génotypes ne réagissent pas de façon similaire à la greffe (Daunay *et al.*, 1986).

L'utilisation d'une core collection a permis de s'assurer d'une maximisation de la variabilité allélique tout en évitant les redondances, ainsi l'apparement observé dans la population est faible. Cependant, la population étudiée est constituée de variétés françaises. L'origine commune des individus bien qu'il existe plusieurs climats contrastés en France, est une limite à la diversité. Or en étude de génétique d'association des individus contrastés sont nécessaires pour observer des associations significatives. Une étude sur une collection plus large, par exemple européenne permettrait d'augmenter la puissance de détection des associations en augmentant la taille de la population ainsi que la diversité génétique.

Pour les plantes cultivées sous contrainte hydrique, certaines variables ont été calculées sur une période donnée correspondant à un stress hydrique modéré. Cette période temporelle réduite a sans doute induit une différence amoindrie entre les phénotypes des individus par rapport aux variables obtenues sur les individus non stressés. L'augmentation de cette période de stress aurait probablement permis d'obtenir des phénotypes plus contrastés.

### 2) Qualité du génotypage.

Le génotypage des individus a été réalisé à l'aide d'une puce à ADN à haute densité de marqueurs spécialement conçu pour des études de génétique d'association, cela a permis d'éviter une phase d'imputation des données manquantes qui aurait été nécessaire de réaliser avec une technologie de génotypage par séquençage (GBS). Ceci a permis un gain de temps non négligeable et une confiance plus grande dans le génotypage. De plus cette puce à une densité de marqueur largement supérieure aux précédentes puces utilisées dans les études précédentes (Chagné *et al.*, 2012 ; Bianco *et al.*, 2014) permettant une meilleure précision des associations.

Les SNPs utilisés sont robuste (Bianco *et al.*, 2016), et sont liés à des séquences codantes et non codantes, ce qui permet la génétique d'association, en comparaison à l'utilisation de SSRs qui correspondent uniquement à des séquences non codantes.

### 3) Ressource bioinformatique.

L'assemblage du génome du pommier actuellement disponible présente des limites. En effet le chromosome 18 correspond aux SNPs qui n'ont pas pu être attribués à un chromosome (par manque de données de cartographie ou parce que ces données étaient contradictoires). Or des associations ont été trouvées pour ces SNPs. Ce mauvais positionnement induit une limite quant à la significativité des résultats d'associations obtenus avec ces SNPs. D'autres SNPs n'ont pas pu être attribués à un contig ou un scaffold de la version 3 du génome, ils proviennent des puces de génotypages précédentes et ont été placés sur le chromosome 20.

**Tableau 8: Gènes candidats trouvés pour chacune des modalités et des variables. Le vert représente les gènes impliqués dans la tolérance au stress hydrique et le jaune représente les gènes impliqués dans la synthèse du cytosquelette.**

Variables	Modalité	SNP	Chromosome	Position(b)	MDP	Ontologi avec arabidopsis	Protéine
Plant_transpiration	ww	Affx-114153760	5	1019213.00	MDP0000286920	<a href="#">AT4G04320.2</a>	malonyl-CoA decarboxylase family protein
Accumulated_transpiration	ww	Affx-103176750	5	3205756.00	MDP0000334968	<a href="#">AT5G60860.1</a>	AtRABA1f
Accumulated_transpiration	ww	Affx-113824981	10	33519729.00	MDP0000221699	<a href="#">AT1G29395.1</a>	COR414-TM1
					MDP0000334947	<a href="#">AT1G29395.1</a>	
					MDP0000334945	<a href="#">AT1G29395.1</a>	
Initial_biomass	ww	Affx-114101793	5	24002024.00	MDP0000505414	<a href="#">AT5G36930.1</a>	disease resistance protein (TIR-NBS-LRR class), putative
Final_biomass	ww	Affx-114197363	5	3628204.00	MDP0000173149	<a href="#">AT5G46070.1</a>	GTP binding / GTPase
	ww	Affx-114042657	17	19657231.00	MDP0000170174	<a href="#">AT3G60830.1</a>	ATARP7 (ACTIN-RELATED PROTEIN 7); structural constituent of cytoskeleton
Initial_leaf_area	ww	Affx-114058120	9	2689938.00	MDP0000231545	<a href="#">AT3G01490.1</a>	protein kinase, putative
	ww	Affx-113832920	9	3143448.00	MDP0000738776	<a href="#">AT2G30050.1</a>	transducin family protein
	ww	Affx-113811653	10	35900864.00	MDP0000439604	<a href="#">AT4G23160.1</a>	protein kinase family protein
Specific leaf area	ww	Affx-113945078	14	419899.00	MDP0000846455	<a href="#">AT3G03710.1</a>	RIF10
	ww	Affx-113803803	14	1029395.00	MDP0000322056	<a href="#">AT3G03740.1</a>	ATBPM4
1026448.00				MDP0000322057	<a href="#">AT1G28760.1</a>		
1014217.00				MDP0000235446	<a href="#">AT3G03740.1</a>		
Maximum_internode_length	ww	Affx-113846029	3	2365673.00	MDP0000224337	<a href="#">AT1G43760.1</a>	unknown protein
	ww	Affx-114097432	17	8924699.00	MDP0000212059	<a href="#">AT4G03270.1</a>	CYCD6;1
Internode_diameter	ww	Affx-114018803	15	17495097.00	MDP0000803538	<a href="#">AT3G03940.1</a>	protein kinase family protein
	ww	Affx-113896328	15	17506855.00	MDP0000565846	<a href="#">AT3G24500.1</a>	MBF1C
	ww			MDP0000129942	<a href="#">AT3G24500.1</a>		
	ww	Affx-113827098	15	23911344.00	MDP0000296744	<a href="#">AT5G45130.1</a>	RHA1
Specific_leaf_weight	ww	Affx-114150611	3	14139050.00	MDP0000453715	<a href="#">AT2G28070.1</a>	ABC
	ww	Affx-113865143	10	10977657.00	MDP0000168595	<a href="#">AT4G34430.2</a>	CHB3
Leaf_shoot_ratio	ww	Affx-113994247	11	42515292.00	MDP0000161241	<a href="#">AT5G44160.1</a>	NUC (nutcracker)
	ww			MDP0000209426	<a href="#">AT5G44160.1</a>		
	ww	Affx-113884013	16	4913610.00	MDP0000222043	<a href="#">AT2G03690.1</a>	Coq4
Final_biomass	WS	Affx-114010611	7	34360933	MDP0000188253	<a href="#">AT5G16660.1</a>	unknown protein
Initial_leaf_area	WS	Affx-114114430	13	12870928	MDP0000135101	<a href="#">AT2G02760.1</a>	ATUBC2
	WS	Affx-114211307	10	35607414	MDP0000792811	<a href="#">AT1G28260.2</a>	SMG7
Height	WS	Affx-113764380	10	27589373	MDP0000260772	<a href="#">AT1G55790.1</a>	ATP binding / nucleotide binding / phenylalanine-tRNA ligase
	WS				MDP0000187640	<a href="#">AT4G23160.1</a>	
	WS	Affx-113916813	11	39371318	MDP0000321069	<a href="#">AT4G23160.1</a>	protein kinase family protein
	WS	Affx-113940902	11	42228379	MDP0000186799	<a href="#">AT4G23160.1</a>	binding / catalytic/ oxidoreductase
	WS				MDP0000307311	<a href="#">AT4G03140.1</a>	
Shoot_dry_weight	WS	Affx-113812487	5	32064344	MDP0000248475	<a href="#">AT5G64360.4</a>	DNAJ
Leaf_dry_weight	WS	Affx-114017122	2	11933928	MDP0000211452	<a href="#">AT2G38290.1</a>	ATAMT2
Specific_leaf_area	WS	Affx-114130653	11	42409619	MDP0000901262	<a href="#">AT1G03020.1</a>	glutaredoxin
	WS				MDP0000244891	<a href="#">AT1G03020.1</a>	
	WS				MDP0000155448	<a href="#">AT3G62930.1</a>	
	WS				MDP0000901262	<a href="#">AT1G03020.1</a>	
	WS				MDP0000244891	<a href="#">AT1G03020.1</a>	
	WS				MDP0000155448	<a href="#">AT3G62930.1</a>	
Specific_leaf_weight	WS	Affx-114181357	11	42490729	MDP0000127798	<a href="#">AT3G62900.1</a>	zinc ion binding
Leaf_shoot_ratio	WS	Affx-113922200	13	5572816	MDP0000297771	<a href="#">AT3G59970.1</a>	MTHFR1
Leaf_shoot_ratio	WS	Affx-113987487	11	11952218	MDP0000282919	<a href="#">AT1G07670.1</a>	calcium-transporting
Leaf_transpiration	WS	Affx-114122565	2	7790084	MDP0000592543	<a href="#">AT5G35550.1</a>	TT2
Number_of_leaves	res	Affx-114004122	11	8550965	MDP0000487114	<a href="#">AT2G29410.1</a>	MTPB1
	res				MDP0000597116	<a href="#">AT2G29410.1</a>	
	res				Affx-113765943	14	
Maximum_internode_length	res	Affx-114179844	8	126330	MDP0000770495	<a href="#">AT2G38010.1</a>	ceramidase family protein
	res	Affx-114133571		126816	MDP0000770495	<a href="#">AT2G38010.1</a>	
Leaf_dry_weight	res	Affx-113994730	6	9874898	MDP0000239006	<a href="#">AT5G47435.1</a>	formyltetrahydrofolate deformylase, putative
Specific_leaf_area	res	Affx-113898375	10	9658839	MDP0000173207	<a href="#">AT4G04850.2</a>	member of Putative potassium transporter family
	res				MDP0000399197	<a href="#">AT4G23160.1</a>	
Specific_leaf_weight	res	Affx-114186775	13	37088986	MDP0000419719	<a href="#">AT4G23160.1</a>	protein kinase family protein
	res				MDP0000715916	<a href="#">AT4G23160.1</a>	
	res				MDP0000647900	<a href="#">AT4G23160.1</a>	
	res				MDP0000691979	<a href="#">AT4G23160.1</a>	
	res				MDP0000298635	<a href="#">AT3G02380.1</a>	
Leaf_shoot_ratio	res	Affx-113832660	9	5555010	MDP0000140484	<a href="#">AT3G02380.1</a>	COL2
	res				MDP0000241871	<a href="#">AT3G02380.1</a>	
	res				Affx-113987487	11	
Plant_transpiration	res	Affx-114063227	3	43317151	MDP0000181167	<a href="#">AT1G55180.1</a>	PLDEPSILON (PHOSPHOLIPASE D ALPHA 4); phospholipase D
Leaf_transpiration	res	Affx-114203619	10	24578770	MDP0000706980	<a href="#">AT2G30570.1</a>	tubulin family protein
	res	Affx-114176443	11	38843121	MDP0000372691	<a href="#">AT1G34060.1</a>	PSBW
							alliinase family protein

Mais aucune association n'a été trouvée avec ces SNPs.

De plus une partie des scaffolds attribués à des chromosomes dans la version V1 du génome n'avaient pas de localisation sur ces mêmes chromosomes dans la version V3, ils ont alors été placés en fin de chromosomes. Cette position n'est donc pas fiable (Hélène Muranty, communication personnelle). Ceci a causé problème dans l'étude du LD en créant du LD à haute distance. Malgré cela le LD calculé dans cette étude est en accord avec la bibliographie, (Leforestier *et al.*, 2015 ; Kumar *et al.*, 2013).

Le repositionnement des SNPs du chromosome 18 est en cours de réalisation par l'équipe de Bucher de l'UMR IRHS de l'INRA d'Angers en collaboration avec FEM en Italie. Cette équipe est aussi en train de réaliser une nouvelle version du génome basé sur un double haploïde, qui devrait grandement améliorer l'assemblage et l'annotation du génome.

#### 4) Résultats des associations

Des associations significatives pour l'utilisation de l'eau (WUE, transpiration, et variables foliaires) ont été trouvées sur l'ensemble des chromosomes à l'exception des chromosomes 1 et 4. Contrairement à la bibliographie (Virlet *et al.*, 2015) où des QTLs liés à l'utilisation de l'eau ont été trouvés sur moins de chromosomes (le 1, 3, 5, 6, 8, 9, 10, 11, 13, et 14), pour une population biparentale en ségrégation. Les SNPs associés aux variables Accumulated\_transpiration, Final\_biomass, et Plant\_transpiration en condition WW sur le chromosome 5 sont compris dans les QTLs trouvés par N. Virlet. On observe aussi une co-localisation des SNPs associés à la variable Leaf\_dry\_weight des résiduelles sur le chromosome 6 au position 8.82Mb et 9.87Mb avec un QTL trouvé dans l'étude de N. Virlet. Pour la variable Leaf\_shoot\_ratio de la modalité WW une co-localisation avec un QTL est trouvée sur le chromosome 9 à 31.72Mb et avec cette même variable sur les résiduelles à 21.67Mb. Une autre co-localisation est observée sur le chromosome 10 entre le SNP associé à la variable Plant\_transpiration des résiduelles et un QTL de N. Virlet.

La co-localisation des associations trouvées dans cette étude avec les QTLs trouvés par N. Virlet (2015) n'est pas totalement sûre car la version du génome utilisé n'est pas la même, les positions ne correspondent probablement pas exactement. Il faudrait repositionner les associations obtenues sur la version V1 du génome sur la base des SNPs pour confirmer ces correspondances.

Pour les variables liées au développement morphologique (liées aux inter-nœuds) des associations ont été trouvées sur les chromosomes 5, 9, 10, 11, 14, et 16. La comparaison des associations trouvées avec la bibliographie est réalisée à l'aide d'une étude de recherche de QTLs associés à des traits architecturaux complexes sur le pommier (Segura *et al.*, 2007). Dans cette étude des QTLs ont été trouvés en lien avec les variables décrivant les entre-nœuds sur les chromosomes 3, 7, 10, et 16. Pour les chromosomes en commun (10 et 16) aucune co-localisation n'est trouvée entre les associations obtenues et les QTLs trouvés dans l'étude de V. Segura

#### 5) Discussion sur les gènes candidats

Plusieurs gènes candidat impliqués dans la tolérance au stress hydrique ont été trouvés (Tableau 8). En effet COR414-TM1 est impliqué dans les réponses cellulaires à la sécheresse (Breton *et al.*, 2003). Le gène MBF1C qui a été trouvé sur le chromosome 15, est un facteur de transcription dont l'expression constitutive induit une forte tolérance à la sécheresse chez *Arabidopsis thaliana* (Suzuki *et al.*, 2005). ATBPM4 est lui aussi référencé comme



s'exprimant en réponse au stress hydrique sur la base de données génomique d'*Arabidopsis* (TAIR)

Un gène responsable de la production de trichomes chez *Arabidopsis thaliana* a été trouvé sur le chromosome 2 associé à la variable Leaf\_transpiration, il s'agit de TT2 (Johnson *et al.*, 2002). Or les trichomes induisent une diminution de la transpiration ce qui est un avantage en condition de sécheresse.

Bien que l'étude ne porte pas sur le système racinaire, un gène codant pour une phospholipase (PLDEPSILON) impliqué dans le stress osmotique et le développement racinaire (TAIR) a été trouvé sur le chromosome 13.

De plus plusieurs gènes candidats trouvés ont une fonction kinase, et particulièrement une fonction kinase serine thréonine. Or ces kinases sont impliquées dans les stress abiotiques et le stress hydrique en particulier (Kulik *et al.*, 2011). Une autre forme de kinase a été trouvée. Il s'agit de CYCD6;1 qui est une kinase cycline-dépendante aussi associée à la tolérance au stress hydrique (Kitsios et Doonan, 2011)

Un gène impliqué dans le transport des ions calcium a été trouvé sur le chromosome 11. Or ces ions sont responsables de l'ouverture des stomates en agissant sur les cellules de garde, ce qui est intéressant vis-à-vis du stress hydrique. (Laanemets *et al.*, 2003 cité par Virlet, 2014).

Des gènes de développement ont aussi été trouvés, en particulier un gène impliqué dans le cytosquelette (tubuline) (MDP0000451570), deux autres gènes impliqués dans la constitution du cytosquelette ont été trouvés. Il s'agit de SMG7 et ATARP7 (protéine relié à l'actine).

## V. CONCLUSION

En conclusion ce stage s'est révélé être un véritable travail de prospection. En effet il s'agit de la première étude de génétique d'association sur des variables liées à la sécheresse. Ceci est particulièrement intéressant dans le contexte actuel du réchauffement global, dans lequel la recherche de nouveaux génotypes plus tolérants à la sécheresse est une priorité. De plus plusieurs gènes candidats liés à la tolérance au stress hydrique ont été trouvés.

Ce travail ouvre de nombreuses perspectives. Tout d'abord, il serait intéressant d'utiliser une population plus grande et plus diversifiée issue par exemple d'une collection européenne afin d'améliorer la puissance de détection des associations. Les individus étudiés dans cette étude sont jeunes ils n'ont pas atteint l'âge de fructification. Il serait donc intéressant de réaliser cette étude sur des plants matures afin d'incorporer des données liées à la production en lien avec le stress hydrique.

Une étude complémentaire sur la physiologie des feuilles sur les génotypes les plus contrastés serait également intéressante pour voir comment celle-ci est impactée par le stress hydrique.

Enfin une étude transcriptomique basée sur le différentiel d'expression des gènes candidats proposés chez deux génotypes contrastés a été réalisée et permettra sans doute d'en savoir plus sur l'importance de ces gènes vis-à-vis du stress hydrique.



## REMERCIEMENTS

Un grand merci à mes maître de stage Evelyne et Benoit pour m'avoir accueilli au sein de l'équipe AFEF et m'avoir permis de vivre cette belle aventure.

Je remercie mes professeures Maria Manzanares-Dauleux et Anne Laperche pour m'avoir formé et pour avoir accepté ce sujet de stage.

Je tiens à remercier particulièrement Baptiste pour son aide et sa pédagogie.

Je remercie l'ensemble de l'équipe pour l'ambiance générale et les conseils qu'ils ont tous pus me prodiguer.

Je remercie les autres stagiaires de l'équipe pour l'amitié qu'ils m'ont accordée

Merci à Juliette ma copine pour son soutien moral.

Je dédie ce travail à mes parents qui j'en suis sûre seraient fiers de moi.



## Références bibliographiques

Alexander et al (2005), Admixture 1.3 Software Manual.

Astle, W., & Balding, D. J. (2009). Population structure and cryptic relatedness in genetic association studies. *Statistical Science*, 451-471.

Bianco, L., Cestaro, A., Linsmith, G., Muranty, H., Denance, C., Théron, A., ... & Larger, S. (2016). Development and validation of the Axiom® Apple480K SNP genotyping array. *The Plant Journal*.

Bianco, L., Cestaro, A., Sargent, D. J., Banchi, E., Derdak, S., Di Guardo, M., ... & Laurens, F. (2014). Development and validation of a 20K single nucleotide polymorphism (SNP) whole genome genotyping array for apple (*Malus× domestica* Borkh). *PLoS One*, 9(10), e110377.

Bonferroni, C. (1950). Sulle medie multiple di potenze. *Bollettino dell'Unione Matematica Italiana*, 5(3-4), 267-270.

Breton, G., Danyluk, J., Charron, J. B. F., & Sarhan, F. (2003). Expression profiling and bioinformatic analyses of a novel stress-regulated multispinning transmembrane protein family from cereals and *Arabidopsis*. *Plant physiology*, 132(1), 64-74.

Brodersen, C., & McElrone, A. (2013). Maintenance of xylem network transport capacity: a review of embolism repair in vascular plants. *Frontiers in plant science*, 4, 108.

Chagné, D., Crowhurst, R. N., Troglio, M., Davey, M. W., Gilmore, B., Lawley, C., ... & Velasco, R. (2012). Genome-wide SNP detection, validation, and development of an 8K SNP array for apple. *PLoS one*, 7(2), e31745.

Challice, J. S., & Westwood, M. N. (1973). Numerical taxonomic studies of the genus *Pyrus* using both chemical and botanical characters. *Botanical Journal of the Linnean Society*, 67(2), 121-148.

Clark, M. D., Schmitz, C. A., Rosyara, U. R., Luby, J. J., & Bradeen, J. M. (2014). A consensus 'Honeycrisp' apple (*Malus× domestica*) genetic linkage map from three full-sib progeny populations. *Tree Genetics & Genomes*, 10(3), 627-639.

Coupel-Ledru, A., Lebon, É., Christophe, A., Doligez, A., Cabrera-Bosquet, L., Péchier, P., ... & Simonneau, T. (2014). Genetic variation in a grapevine progeny (*Vitis vinifera* L. cvs Grenache× Syrah) reveals inconsistencies between maintenance of daytime leaf water potential and response of transpiration rate under drought. *Journal of experimental botany*, eru228.

D.H. Alexander, J. Novembre, and K. Lange.(2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*

Daunay, M., Malet, P., Jullian. E., (1986) Influences réciproques des racines et des parties aériennes sur le rendement de l'aubergine (*Solanum melongena* L.) : résultats de greffages inter-variétaux. *Agronomie, EDP Sciences*, 6 (3), 293-297.

Didelot, F., Caffier, V., Orain, G., Lemarquand, A., & Parisi, L. (2016). Sustainable management of scab control through the integration of apple resistant cultivars in a low-fungicide input system. *Agriculture, Ecosystems & Environment*, 217, 41-48.

Doebley, J., & Stec, A. (1991). Genetic analysis of the morphological differences between maize and teosinte. *Genetics*, 129(1), 285-295.

Eding, H., & Meuwissen, T. H. E. (2001). Marker-based estimates of between and within population kinships for the conservation of genetic diversity. *Journal of Animal Breeding and Genetics*, 118(3), 141-159.

Endelman, J. B., & Jannink, J. L. (2012). Shrinkage estimation of the realized relationship matrix. *G3: Genes| Genomes| Genetics*, 2(11), 1405-1413.

Falginella, L., Cipriani, G., Monte, C., Gregori, R., Testolin, R., Velasco, R., ... & Tartarini, S. (2015). A major QTL controlling apple skin russeting maps on the linkage group 12 of 'Renetta Grigia di Torriana'. *BMC plant biology*, 15(1), 150.

Frankel, O. H. (1984). Genetic perspectives of germplasm conservation. *Genetic manipulation: impact on man and society*. Cambridge University Press, Cambridge, 161-170.

Gallais, A., & Bannerot, H. (1995) Amélioration des espèces végétales cultivées. Objectifs et critères de sélection. 580-585.

Gessler, C., Patocchi, A., Sansavini, S., Tartarini, S., & Gianfranceschi, L. (2006). *Venturia inaequalis* resistance in apple. *Critical Reviews in Plant Sciences*, 25(6), 473-503.

Gross, B. L., Henk, A. D., Richards, C. M., Fazio, G., & Volk, G. M. (2014). Genetic diversity in *Malus* × *domestica* (Rosaceae) through time in response to domestication. *American journal of botany*, 101(10), 1770-1779.

H. Zhou, D. H. Alexander, and K. Lange.(2009) A quasi-Newton method for accelerating the convergence of iterative optimization algorithms. *Statistics and Computing*

Harris, S. A., Robinson, J. P., & Juniper, B. E. (2002). Genetic clues to the origin of the apple. *TRENDS in Genetics*, 18(8), 426-430.

Hinds, D. A., Stuve, L. L., Nilsen, G. B., Halperin, E., Eskin, E., Ballinger, D. G., ... & Cox, D. R. (2005). Whole-genome patterns of common DNA variation in three human populations. *Science*, 307(5712), 1072-1079.

Hopkins, 2003, *physiologie végétale*. pp. 204-205

Johnson, C. S., Kolevski, B., & Smyth, D. R. (2002). TRANSPARENT TESTA GLABRA2, a trichome and seed coat development gene of Arabidopsis, encodes a WRKY transcription factor. *The Plant Cell*, 14(6), 1359-1375.

Kitsios, G., & Doonan, J. H. (2011). Cyclin dependent protein kinases and stress responses in plants. *Plant signaling & behavior*, 6(2), 204-209.

Kulik, A., Wawer, I., Krzywińska, E., Bucholc, M., & Dobrowolska, G. (2011). SnRK2 protein kinases—key regulators of plant response to abiotic stresses. *Omic: a journal of integrative biology*, 15(12), 859-872.

Kumar, S., Garrick, D. J., Bink, M. C., Whitworth, C., Chagné, D., & Volz, R. K. (2013). Novel genomic approaches unravel genetic architecture of complex traits in apple. *BMC genomics*, 14(1), 1.

Lassois, L., Denancé, C., Ravon, E., Guyader, A., Guisnel, R., Hibrand-Saint-Oyant, L., ... & Durel, C. E. (2015). Genetic diversity, population structure, parentage analysis, and construction of core collections in the french apple germplasm based on SSR markers. *Plant Molecular Biology Reporter*, 1-18.

Leforestier, D., Ravon, E., Muranty, H., Cornille, A., Lemaire, C., Giraud, T., ... & Branca, A. (2015). Genomic basis of the differences between cider and dessert apple varieties. *Evolutionary applications*, 8(7), 650-661.

Lopez, G., Pallas, B., Martinez, S., Lauri, P. É., Regnard, J. L., Durel, C. É., & Costes, E. (2015). Genetic variation of morphological traits and transpiration in an apple core collection under well-watered conditions: towards the identification of morphotypes with high water use efficiency. *PloS one*, 10(12), e0145540.

Marchini, J., Cardon, L. R., Phillips, M. S., & Donnelly, P. (2004). The effects of human population structure on large genetic association studies. *Nature genetics*, 36(5), 512-517.

Marguerit, E., Coupel-Ledru, A., Barrieu, F., Bert, P. F., Doligez, A., Lebon, E., ... & Ollat, N. (2014). Architecture génétique de réponses au déficit hydrique chez la Vigne. *Innovations Agronomiques*, 38, 49-65.

Masseron, A., Edin, M., & Mandrin, J. F. (1985). Etude du Paradis Jaune de Metz porte-greffe du pommier, et selection de 2 nouveaux types: Pajam 1 (Lancep) et Pajam 2 (Cepiland). *Cahiers du CTIFL (France)*. no. 13.

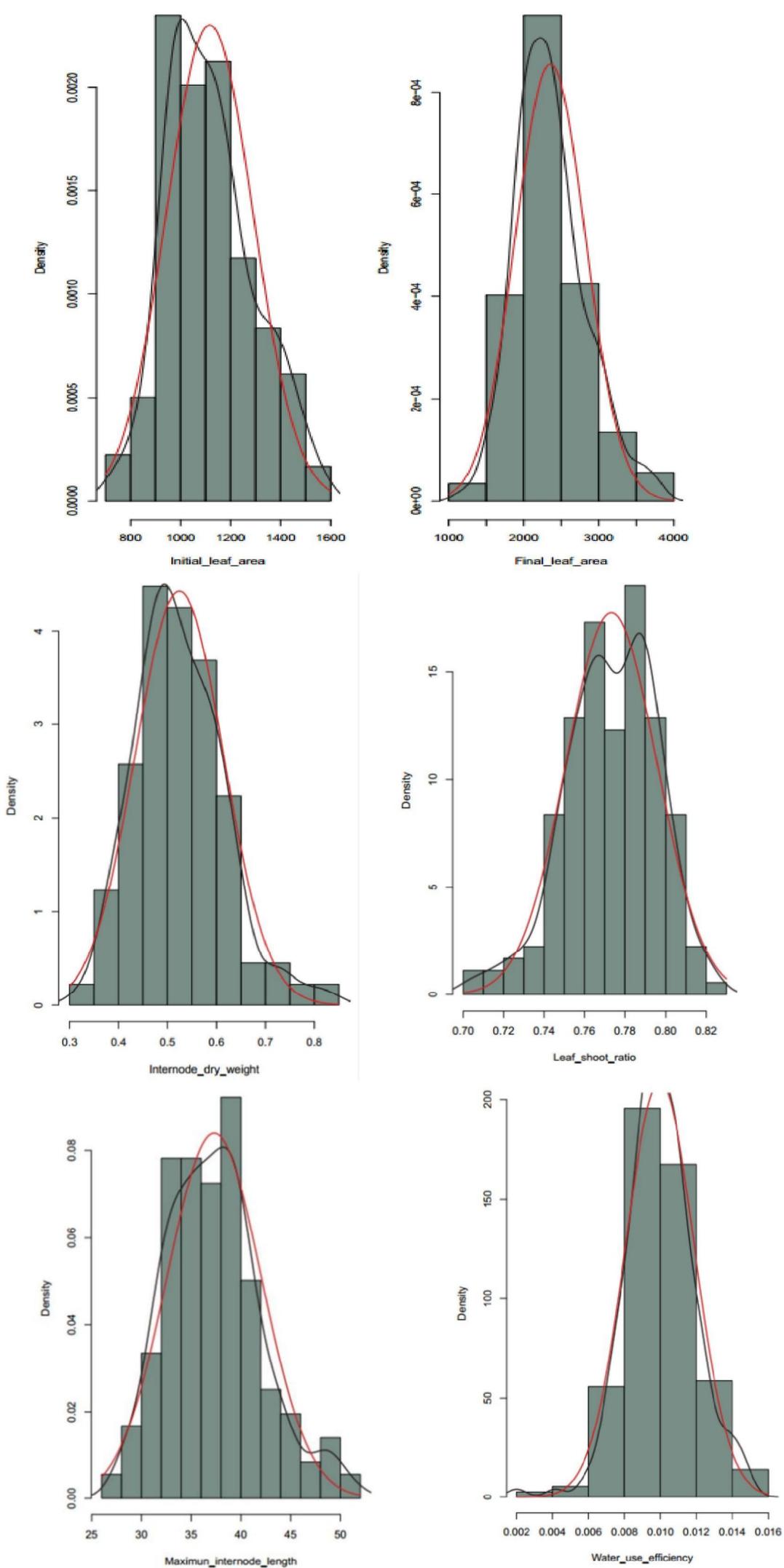
Mesak, F., Tatarenkov, A., Earley, R. L., & Avise, J. C. (2014). Hundreds of SNPs vs. dozens of SSRs: which dataset better characterizes natural clonal lineages in a self-fertilizing fish?. *Frontiers in Ecology and Evolution*, 2, 74.

Miesfeld, R., Krystal, M., & Amheim, N. (1981). A member of a new repeated sequence family which is conserved throughout eucaryotic evolution is found between the human  $\delta$  and  $\beta$  globin genes. *Nucleic acids research*, 9(22), 5931-5948.

- Moriya, S., Kuniyoshi, M., Okada, K., Iwanami, H., Iwata, H., Minamikawa, M., ... & Matsumoto, T. (2016). Identification of QTLs for Flesh Mealiness in Apple (*Malus domestica* Borkh.). *The Horticulture Journal*, (0).
- Naor, A., & Girona, J. (2012). Apple. Crop yield response to water. Rome: FAO (Food and Agriculture Organization of the United Nations), 332-345.
- Noiton D, Alspach P (1996) Founding clones, inbreeding, coancestry and status number of modern apple cultivars. *J Am Soc Hortic Sci* 121, 773–782
- Patakas, A., & Noitsakis, B. (1999). Osmotic adjustment and partitioning of turgor responses to drought in grapevines leaves. *American journal of enology and viticulture*, 50(1), 76-80.
- Pritchard, J. K., & Donnelly, P. (2001). Case-control studies of association in structured or admixed populations. *Theoretical population biology*, 60(3), 227-237.
- Sadok, I. B., Tiecher, A., Galvez-Lopez, D., Lahaye, M., Lasserre-Zuber, P., Bruneau, M., ... & Laurens, F. (2015). Apple fruit texture QTLs: year and cold storage effects on sensory and instrumental traits. *Tree Genetics & Genomes*, 11(6), 1-20.
- Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2), 461-464.
- Segura, V., Denancé, C., Durel, C. E., & Costes, E. (2007). Wide range QTL analysis for complex architectural traits in a 1-year-old apple progeny. *Genome*, 50(2), 159-171.
- Suzuki, N., Rizhsky, L., Liang, H., Shuman, J., Shulaev, V., & Mittler, R. (2005). Enhanced tolerance to environmental stress in transgenic plants expressing the transcriptional coactivator multiprotein bridging factor 1c. *Plant Physiology*, 139(3), 1313-1322.
- Velasco, R., Zharkikh, A., Affourtit, J., Dhingra, A., Cestaro, A., Kalyanaraman, A., ... & Salvi, S. (2010). The genome of the domesticated apple (*Malus [times] domestica* Borkh.). *Nature genetics*, 42(10), 833-839.
- Virlet, N. (2014). Phénotypage haut débit par imagerie multispectrale au verger: étude du déterminisme génétique de la réponse à la contrainte hydrique d'une population d'hybrides de pommier (*Malus x domestica* Borkh.) (Doctoral dissertation, Montpellier, SupAgro).
- Virlet, N., Costes, E., Martinez, S., Kelner, J. J., & Regnard, J. L. (2015). Multispectral airborne imagery in the field reveals genetic determinisms of morphological and transpiration traits of an apple tree hybrid population in response to water deficit. *Journal of experimental botany*, erv355.
- Way R et al. (1990) Apples. In: Moore J, Ballington J (eds) Genetic resources of temperate fruit and nut crops. International society for horticultural science, Wageningen, Netherlands, 1–62

Wisniewski, M., Bassett, C., Norelli, J., Macarasin, D., Artlip, T., Gasic, K., & Korban, S. (2008). Expressed sequence tag analysis of the response of apple (*Malus x domestica* 'Royal Gala') to low temperature and water deficit. *Physiologia plantarum*, 133(2), 298-317.

Zhu, C., Gore, M., Buckler, E. S., & Yu, J. (2008). Status and prospects of association mapping in plants. *The plant genome*, 1(1), 5-20.



**Annexe 1 : Histogramme de distribution des variables non normales d'après le test de Shapiro. La courbe rouge représente la normalité, on observe une distribution très proche de celle-ci.**

	MLM_k	MLM_k_Q
Initial_biomass	1094.878	1107.065
Final_biomass	1460.465	1458.942
Initial_leaf_area	2143.563	2122.022
Final_leaf_area	2459.536	2427.068
Number_of_leaves	857.5518	882.9477
Height	1380.799	1382.823
Basal_shoot_diameter	265.1123	316.4496
Maximum_internode_lengt	977.8022	984.2742
Internode_diameter	206.7983	261.3158
Internode_dry_weight	-357.9478	-291.9592
Internode_volume	-51.14584	6.176787
Internode_density	-1116.238	-1018.989
Leaf_dry_weight	-315.2778	-244.9092
Leaf_area	1249.061	1261.556
Specific_leaf_weight	-2331.838	-2202.089
Leaf_shoot_ratio	-850.9778	-761.5892
Plant_transpiration	1433.803	1440.994
Leaf_transpiration	-917.2278	-826.6092
Accumulated_transpiration	2614.266	2577.613
Water_use_efficiency	-1721.138	-1610.089

**Annexe 2 : Tableau des valeurs de BIC pour la sélection du modèle en condition non stressante. Le jaune représente le modèle sélectionné.**

Variable	Modèle	SNP	Chromosome	Position (Mb)	"-log(p)	génotype	effet allélique	effectif
Initial_biomass	MLM_K	Affx-114101793	5	24.00	5.24	AA	5.89E+00	133
						AG	0.00	41
						GG	1.08E-01	5
		Affx-113978244	17	36.22	5.17	AA	-3.99E-01	107
						AG	0.00	61
						GG	1.01E+01	11
Final_biomass	MLM_K_Q	Affx-114197363	5	3.63	5.55	GG	1.50E+01	106
						GA	0.00	64
						AA	2.69E+02	9
		Affx-114141227	5	5.64	5.81	AA	1.71E+00	101
						AG	0.00	68
						GG	3.68E+01	10
Affx-114042657	17	19.66	5.83	GG	4.42E-01	132		
				GT	0.00	42		
				TT	1.73E+01	3		
Initial_leaf_area	MLM_K_Q	Affx-114058120	9	2.69	6.59	AA	3.27E+00	18
						GA	0.00	71
						GG	4.88E+00	90
		Affx-113832920	9	3.14	5.37	AA	3.90E+00	71
						AC	0.00	81
						CC	2.95E+00	27
		Affx-114070808	9	0.48	5.27	CC	7.76E+00	16
						CT	0.00	70
						TT	1.20E+01	93
		Affx-103180387	9	0.36	5.09	AA	6.33E+01	20
						AG	0.00	75
						GG	1.43E+02	84
		Affx-113811653	10	35.90	5.28	AA	-1.21E+02	113
						AC	0.00	58
						CC	8.97E+01	8
		Affx-113935715	13	39.90	5.56	CC	7.53E+01	16
						TC	0.00	69
						TT	1.54E+02	94
		Affx-114055889	18	80.05	5.04	CC	1.44E+02	90
						TC	0.00	69
						TT	9.52E+01	20
		Affx-113914792	18	190.13	5.53	AA	-1.13E+02	81
						GA	0.00	86
						GG	-1.83E+02	11
Leaf_dry_weight	MLM_K	Affx-113803803	14	1.03	6.05	AA	9.43E-02	11
						AG	0.00	71
						GG	7.98E-02	97
		Affx-113898300	14	1.03	5.99	AA	9.82E-02	10
						AG	0.00	72
						GG	7.85E-02	97
		Affx-114067067	14	1.01	5.83	CC	7.79E-02	99
						CT	0.00	70
						TT	9.80E-02	10
		Affx-114218217	14	1.02	5.78	CC	7.77E-02	98
						CT	0.00	70
						TT	9.79E-02	10
Affx-114095136	14	1.02	5.33	CC	7.52E-02	100		
				CT	0.00	68		
				TT	9.68E-02	10		
Affx-113945078	14	0.42	5.39	AA	7.63E-02	100		
				AC	0.00	73		
				CC	7.23E-02	5		
Affx-113981911	18	155.68	5.23	CC	9.45E-02	10		
				CT	0.00	70		
				TT	7.40E-02	99		
Maximum_internode_length	MLM_K	Affx-113846029	3	2.37	5.44	AA	-1.70E+00	111
						AG	0.00	60
						GG	6.22E+00	8
		Affx-114175694	17	8.99	5.95	GG	-2.96E+00	158
						GT	0.00	18
						TT	9.02E+00	3
		Affx-114097432	17	8.92	5.42	AA	-3.15E+00	110
						AC	0.00	57
						CC	1.73E+00	12
		Affx-113953930	17	9.31	5.26	CC	2.82E+00	2
						CT	0.00	40
						TT	-3.99E+00	137
		Affx-114229759	17	9.18	5.12	GG	-4.00E+00	148
						GT	0.00	31
						TT	NA	0
		Affx-114031960	17	9.18	5.12	TT	-4.00E+00	148
						CT	0.00	31
						CC	NA	0
		Affx-114128447	17	9.18	5.12	GG	-4.00E+00	148
						GT	0.00	31
						TT	NA	0
		Affx-114116563	17	9.32	5.12	GG	-4.00E+00	148
						AG	0.00	31
						AA	NA	0
Affx-114025951	18	219.92	5.20	CC	-4.11E+00	147		
				CT	0.00	30		
				TT	NA	0		

Annexe 3a : Récapitulatif des associations trouvées pour la modalité WW (condition non stressante)

Variable	Modèle	SNP	Chromosome	Position (Mb)	<sup>-</sup> log(p)	génotype	effet allélique	effectif		
Internode_diameter	MLM_K	Affx-114018803	15	17.50	5.69	CC	3.08E-01	69		
						CT	0.00	87		
						TT	-1.36E-01	21		
		Affx-113896328	15	17.51	5.77	AA	-2.68E-01	18		
						AG	0.00	94		
						GG	2.65E-01	67		
		Affx-113827098	15	23.91	5.39	CC	-3.97E-01	12		
						CT	0.00	84		
						TT	2.15E-01	82		
Internode_volume	MLM_K	Affx-113833230	10	25.35	5.11	AA	5.98E-02	89		
						AC	0.00	75		
						CC	2.85E-01	15		
Specific_leaf_weight	MLM_K	Affx-114150611	3	14.14	5.37	CC	1.71E-04	133		
						CT	0.00	42		
						TT	-5.92E-04	4		
		Affx-113883620	3	14.25	5.26	CC	3.28E-04	154		
						CT	0.00	25		
						TT	NA	0		
		Affx-113795978	3	13.76	5.09	CC	1.95E-04	137		
						CT	0.00	39		
						TT	-5.96E-04	3		
		Affx-114170075	3	14.71	5.09	GG	1.95E-04	137		
						GA	0.00	39		
						AA	-5.96E-04	3		
		Affx-113952831	3	43.32	5.71	CC	-2.81E-04	4		
						CT	0.00	23		
						TT	3.04E-04	152		
		Affx-113984169	3	43.44	5.09	GG	1.95E-04	137		
						GA	0.00	39		
						AA	-5.96E-04	3		
		Affx-113865143	10	10.98	5.29	CC	2.73E-04	136		
						CT	0.00	40		
						TT	6.10E-04	3		
		Affx-114111730	10	11.17	5.21	GG	NA	0		
						GT	0.00	21		
						TT	3.54E-04	158		
		Affx-113819617	10	10.82	5.21	AA	3.22E-04	152		
						AG	0.00	26		
						GG	8.93E-04	1		
		Affx-114230160	18	106.38	5.40	GG	2.94E-04	153		
						AG	0.00	22		
						AA	-2.86E-04	4		
		Affx-114051128	18	167.09	5.34	CC	2.65E-04	133		
						CT	0.00	41		
						TT	4.95E-04	5		
		Affx-113966126	18	167.10	5.34	GG	2.65E-04	133		
						AG	0.00	41		
						AA	4.95E-04	5		
		Affx-113939383	18	195.43	5.35	GG	NA	0		
						GT	0.00	21		
						TT	3.57E-04	158		
		Leaf_shoot_ratio	MLM_K	Affx-113886770	9	31.72	5.57	CC	2.21E-02	159
								CT	0.00	19
								TT	-4.44E-02	1
				Affx-113994247	11	42.52	5.90	AA	2.36E-02	151
								AC	0.00	27
								CC	3.52E-02	1
				Affx-113884013	16	4.91	5.19	AA	1.43E-02	105
								AG	0.00	72
								GG	4.68E-02	2
				Affx-113938219	18	184.49	5.22	GG	2.16E-02	149
								GT	0.00	29
								TT	3.34E-02	1
		Affx-114133441	18	187.54	5.37	CC	-1.66E-02	88		
						CT	0.00	79		
						TT	-8.34E-03	12		
Plant_transpiration	MLM_K	Affx-114153760	5	1.02	5.38	AA	-1.74E+01	41		
						AC	0.00	98		
						CC	-1.02E+00	40		
Accumulated_transpiration	MLM_K_Q	Affx-103176750	5	3.21	5.06	TT	3.97E+01	112		
						TG	0.00	59		
						GG	1.33E+03	8		
		Affx-113824981	10	33.52	5.00	CC	-6.51E+02	127		
						CT	0.00	45		
						TT	-1.06E+02	7		
		Affx-114225279	17	10.05	5.11	AA	-5.54E+02	68		
						AC	0.00	89		
						CC	-4.50E+02	22		
Final_leaf_area	MLM_K_Q	Affx-113843645	3	43.55	5.70	GG	-1.09E+01	128		
						GA	0.00	46		
						AA	-5.85E+00	5		
Number_of_leaves	MLM_K	Affx-113941995	17	25.31	5.08	CC	3.32E+00	158		
						CT	0.00	18		
						TT	-4.97E+00	1		
Leaf_area	MLM_K	Affx-113945078	14	0.42	5.39	AA	8.29E+00	100		
						AC	0.00	73		
						CC	8.49E+00	5		

Annexe 3b : Récapitulatif des associations trouvées pour la modalité WW (condition non stressante)

Variable	modèle	Marqueur	Chromosome	Position (Mb)	p	-log(p)	Forme allélique	Effet allélique	Effectif
Initial_biomass	MLM_K	Affx-113814752	7	26.82	1.13E-05	4.95	CC	-8.48E-01	42
							AC	0.00	99
							AA	6.18E+00	35
Final_biomass	MLM_K	Affx-113853649	7	33.80	2.78E-06	5.56	GG	9.35E+00	27
							GT	0.00	84
							TT	-3.73E+00	66
		Affx-114010611	7	34.36	5.70E-06	5.24	AA	-8.18E+00	86
							AG	0.00	80
							GG	2.03E+00	11
Affx-113816797	7	34.73	2.38E-06	5.62	CC	6.98E+00	39		
					CT	0.00	87		
					TT	-5.44E+00	49		
Initial_leaf_area	MLM_K_Q	Affx-114114430	13	12.87	5.90E-06	5.23	CC	-1.09E+02	91
							CT	0.00E+00	65
							TT	-4.31E+01	21
	MLM_K	Affx-114211307	10	35.61	7.11E-06	5.15	AA	3.84872	16
							AG	0	71
							GG	0.26543	90
Number_of_leaves	MLM_K	Affx-114218179	13	9.54	1.02E-05	4.99	AA	2.35231	13
							AG	0	79
							GG	0.02344	85
		Affx-114078938	18	148.27	8.19E-06	5.09	CC	3.94782	5
							CT	0	66
							TT	0.47281	106
Maximum_internode_length	MLM_K	Affx-113764380	10	27.59	1.04E-05	4.98	AA	-2.8004	2
							AG	0	32
							GG	5.66061	143
		Affx-113916813	11	39.37	3.92E-06	5.41	AA	23.52636	1
							AG	0	33
							GG	4.70772	143
		Affx-114018746	11	39.54	2.74E-06	5.56	CC	23.70921	1
							CT	0	31
							TT	4.96439	145
		Affx-113940902	11	42.23	1.30E-06	5.89	AA	5.21	145
							AG	0	31
							GG	24.05078	1
Affx-113770453	18	77.29	2.61E-06	5.58	AA	4.87632	143		
					AG	0	33		
					GG	23.63879	1		
Internode_diameter	MLM_K	Affx-114130268	14	34.89	7.11E-06	5.15	AA	-2.85E-02	37
							AG	0	83
							GG	0.21354	55
		Affx-113976034	18	195.52	7.87E-06	5.10	CC	0.36562	12
							CT	0	65
							TT	0.14601	99

**Annexe 4a : tableau récapitulatif des différentes associations en condition de stress hydrique. La couleur orange représente les associations obtenues avec le modèle qui n'a pas été sélectionné par la minimisation du BIC**

Variable	modèle	Marqueur	Chromosome	Position (Mb)	p	-log(p)	Forme allélique	Effet allélique	Effectif	
Internode_dry_weight	MLM_K	Affx-113812487	5	32.06	3.48E-06	5.46	AA	-1.32E-02	143	
							AG	0	30	
							GG	0.12765	3	
		Affx-113948880	5	33.19	2.19E-06	5.66	5.66	AA	-1.50E-02	143
								AG	0	30
								GG	0.12722	3
		Affx-114025160	18	98.63	2.71E-06	5.57	5.57	CC	0.1255	3
								CT	0	34
								TT	-1.56E-02	137
Internode_density	MLM_K	Affx-113961588	18	62.75	5.39E-06	5.27	GG	0.01992	3	
							TG	0	60	
							TT	0.01382	113	
Leaf_dry_weight	MLM_K	Affx-114017122	2	11.93	8.81E-06	5.05	AA	0.06117	44	
							AC	0	83	
							CC	0.00998	49	
Leaf_area	MLM_K_Q	Affx-114130653	11	42.41	3.17E-06	5.50	GG	5.13961	109	
							GT	0	60	
							TT	-3.2071	7	
		Affx-114181357	11	42.49	7.76E-06	5.11	5.11	AA	5.03363	111
								AG	0	59
								GG	-3.1091	6
		Affx-114142881	13	44.71	2.30E-06	5.64	5.64	TT	7.72963	18
								CT	0	86
								CC	-2.6183	72
		Affx-113771065	18	105.58	2.71E-06	5.57	5.57	AA	7.28969	22
								AG	0	89
								GG	3.79535	63
Specific_leaf_weight	MLM_K_Q	Affx-113922200	13	5.57	4.75E-06	5.32	GG	4.46E-05	31	
							GT	0	75	
							TT	-3.43E-04	69	
Leaf_shoot_ratio	MLM_K	Affx-114217601	11	10.00	1.87E-06	5.73	AA	-2.28E-02	17	
							AG	0	70	
							GG	-1.73E-02	89	
		Affx-113987487	11	11.95	1.87E-06	5.73	5.73	AA	-2.28E-02	17
								AG	0	70
								GG	-1.73E-02	89
Leaf_transpiration	MLM_K	Affx-114122565	2	7.79	5.18E-06	5.29	AA	-5.43E-03	66	
							AG	0.00E+00	80	
							GG	9.19E-03	29	
		Affx-113828807	12	5.12	3.68E-06	5.43	5.43	CC	-1.06E-02	130
								CT	0	46
								TT	NA	0
		Affx-114093108	12	5.14	4.46E-06	5.35	5.35	CC	NA	0
								CT	0	45
								TT	-1.06E-02	131
		Affx-113750213	18	242.27	5.71E-06	5.24	5.24	AA	-1.29E-02	153
								AG	0	24
								GG	NA	0

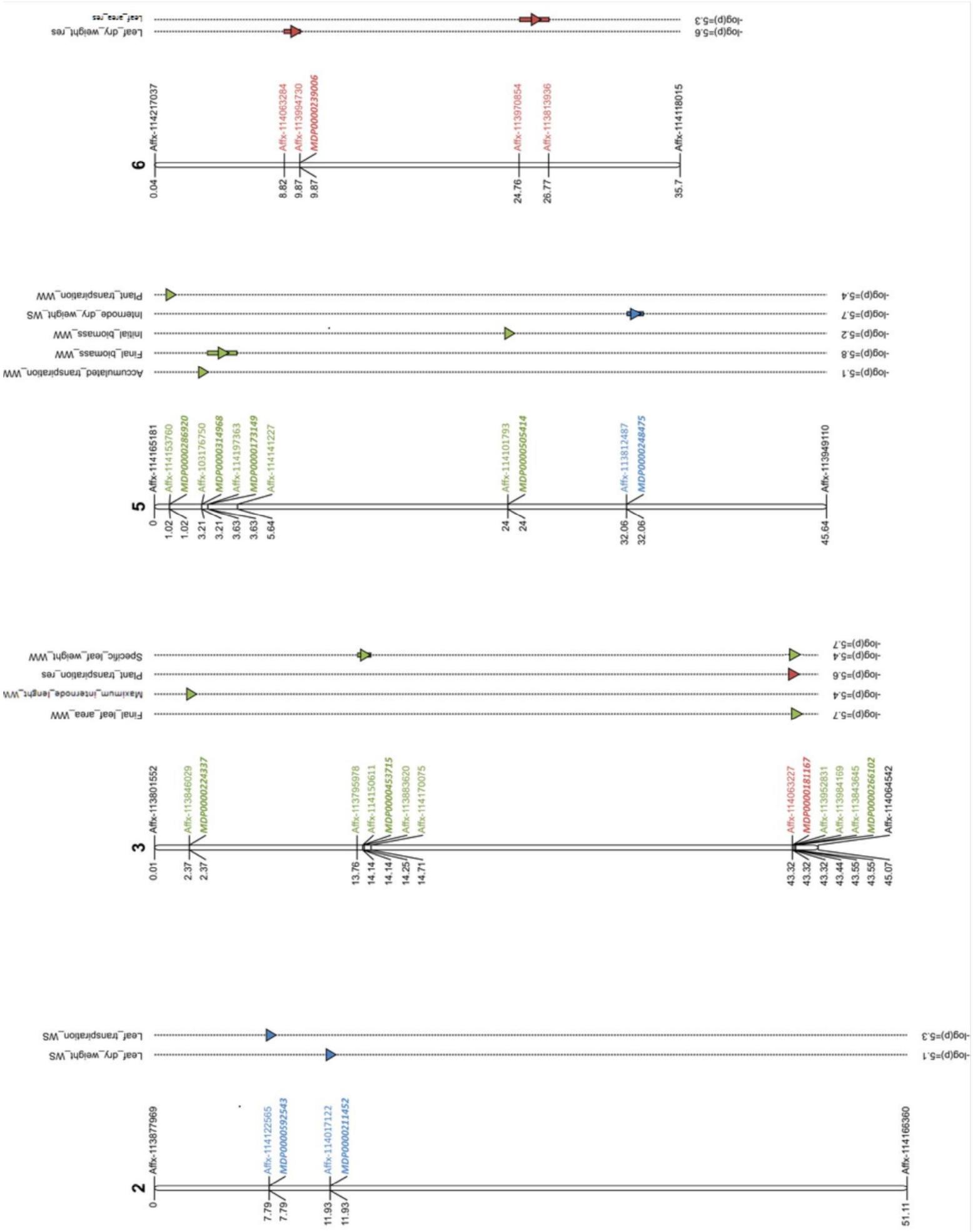
**Annexe 4b : tableau récapitulatif des différentes associations en condition de stress hydrique. La couleur orange représente les associations obtenues avec le modèle qui n'a pas été sélectionné par la minimisation du BIC**

Variable	modèle	Marqueur	Chromosome	Position (Mb)	p	-log(p)	Forme allélique	Effet allélique	Effectif	
Initial_biomass	MLM_K	Affx-113795365	17	8.54	1.23E-05	4.91	GG	NA	0	
							AG	0.00	45	
							AA	-4.43E+00	130	
Initial_leaf_area	MLM_K	Affx-113791654	17	8.74	7.73E-06	5.11	GG	NA	0	
							GT	0	45	
							TT	-8.03E+01	131	
		Affx-113795365	17	8.54	9.09E-06	5.04	5.04	GG	NA	0
								AG	0	45
								AA	-8.02E+01	130
	Affx-113983729	17	8.51	6.32E-07	6.20	6.20	GG	-9.47E+01	137	
							AG	0	40	
							AA	NA	0	
	MLM_K_Q	Affx-113983729	17	8.51	1.77E-06	5.75	5.75	GG	-9.69E+01	137
								AG	0	40
								AA	NA	0
Number_of_leaves	MLM_K	Affx-114004122	11	8.55	8.81E-06	5.05	CC	0.80866	90	
							CT	0	71	
							TT	1.62892	14	
		Affx-113765943	14	31.52	4.99E-06	5.30	5.30	AA	3.11771	4
								AG	0	59
								GG	0.66903	112
Maximum_internode_length	MLM_K	Affx-114179844	8	0.13	5.87E-06	5.23	CC	-4.2217	13	
							CT	0	88	
							TT	0.85832	75	
		Affx-114133571	8	0.13	5.87E-06	5.23	5.23	CC	0.85832	75
								CT	0	88
								TT	-4.2217	13
Internode_dry_weight	MLM_K	Affx-114218923	11	42.57	9.34E-06	5.03	AA	-1.47E-02	84	
							AC	0	73	
							CC	0.03498	19	
Internode_volume	MLM_K	Affx-114196772	18	241.35	9.76E-06	5.01	AA	0.02122	89	
							AG	0	74	
							GG	0.14502	12	
	MLM_K_Q	Affx-114218923	11	42.57	8.85E-06	5.05	5.05	AA	-2.66E-02	84
								AC	0	73
								CC	0.09197	19
Leaf_dry_weight	MLM_K	Affx-114063284	6	8.82	2.63E-06	5.58	CC	-3.93E-02	141	
							CT	0	33	
							TT	0.09783	2	
		Affx-113994730	6	9.87	8.97E-06	5.05	5.05	CC	0.06949	15
								CT	0	101
								TT	7.61E-04	60
		Affx-113867538	15	25.97	5.66E-06	5.25	5.25	AA	0.07406	13
								AG	0	66
								GG	0.02512	96
		Affx-113839674	18	201.70	1.53E-06	5.82	5.82	AA	0.0779	14
								AG	0	70
								GG	0.02129	90
Leaf_area	MLM_K	Affx-113994730	6	9.87	1.08E-05	4.97	CC	7.06083	15	
							CT	0	101	
							TT	0.72658	60	
		Affx-113970854	6	24.76	9.17E-06	5.04	5.04	CC	2.09646	105
								CT	0	60
								TT	7.62318	11
		Affx-113813936	6	26.77	4.68E-06	5.33	5.33	CC	0.37731	9
								CT	0	73
								TT	3.99582	94
		Affx-114029287	7	31.36	2.58E-06	5.59	5.59	CC	3.38274	39
								CT	0	84
								TT	-2.4091	51
		Affx-113898375	10	9.66	3.32E-06	5.48	5.48	CC	0.74998	74
								CT	0	83
								TT	-5.9558	19
		Affx-113867538	15	25.97	1.26E-06	5.90	5.90	AA	7.017	13
								AG	0	66
								GG	3.19674	96
		Affx-113755966	15	25.98	6.65E-07	6.18	6.18	AA	2.97626	102
								AG	0	60
								GG	7.6993	13
		Affx-114080754	16	12.45	5.71E-06	5.24	5.24	AA	1.55127	9
								AC	0	74
								CC	4.11175	92
		Affx-114178769	18	36.31	5.27E-06	5.28	5.28	AA	0.86655	127
								AG	0	41
								GG	-7.3262	8
		Affx-114005513	18	96.70	4.29E-06	5.37	5.37	AA	-1.12E+01	4
								AC	0	51
								CC	-9.02E-01	121
Affx-114221177	18	112.71	7.32E-06	5.14	5.14	CC	2.77427	149		
						CT	0.00E+00	22		
						TT	-1.05E+01	4		
Affx-113826154	18	191.81	9.50E-06	5.02	5.02	CC	-5.79E+00	17		
						CT	0.00E+00	76		
						TT	-2.79E+00	83		
Affx-113839674	18	201.70	5.87E-08	7.23	7.23	AA	7.88E+00	14		
						AG	0.00E+00	70		
						GG	2.75E+00	90		

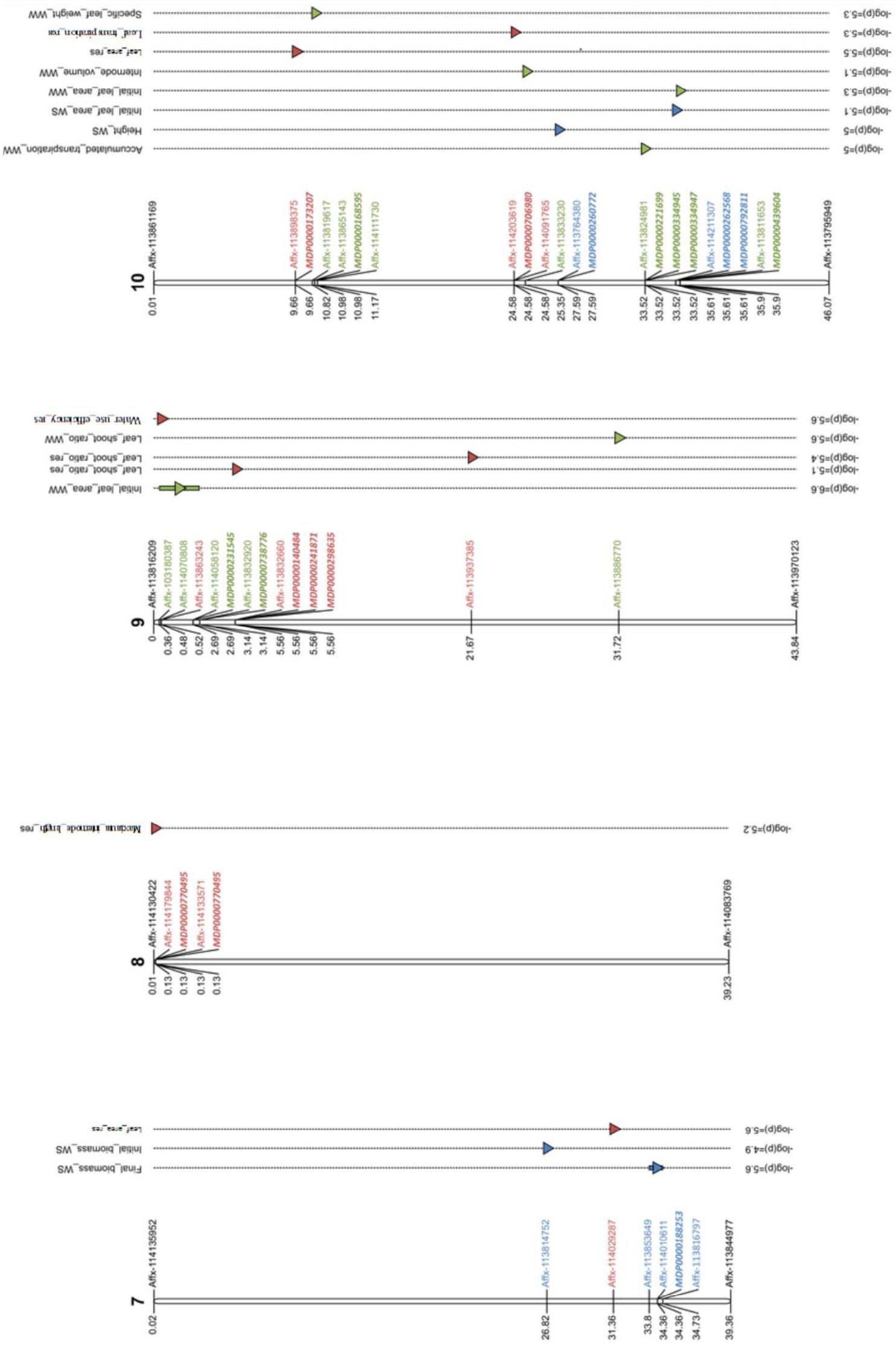
Annexe 5a : Tableau de l'ensemble des associations significatives trouvées pour les variables de la modalité résiduelle. La couleur orange représente les associations qui ne correspondent pas au modèle sélectionné par le BIC.

Variable	modèle	Marqueur	Chromosome	Position (Mb)	p	-log(p)	Forme allélique	Effet allélique	Effectif	
Specific_leaf_weight	MLM_K	Affx-114186775	13	37.09	8.27E-06	5.08	CC	2.85E-04	118	
							CT	0.00E+00	54	
							TT	-2.41E-04	3	
Leaf_shoot_ratio	MLM_K	Affx-113832660	9	5.56	8.16E-06	5.09	CC	-4.77E-03	113	
							CT	0.00E+00	54	
							TT	-3.12E-02	9	
		Affx-113937385	9	21.67	3.98E-06	5.40	5.40	CC	-1.59E-02	14
								CT	0.00E+00	80
								TT	1.01E-02	82
		Affx-114217601	11	10.00	1.73E-06	5.76	5.76	AA	-1.55E-02	17
								AG	0.00E+00	70
								GG	-1.55E-02	89
		Affx-113987487	11	11.95	1.73E-06	5.76	5.76	AA	-1.55E-02	17
								AG	0.00E+00	70
								GG	-1.55E-02	89
		Affx-113801202	13	33.54	4.10E-06	5.39	5.39	AA	-1.64E-02	14
								AG	0.00E+00	79
								GG	9.77E-03	82
		Affx-113810847	13	33.54	3.98E-06	5.40	5.40	CC	1.01E-02	82
								CT	0.00E+00	80
								TT	-1.59E-02	14
		Affx-114204873	16	29.19	1.61E-06	5.79	5.79	CC	-1.67E-02	63
								CT	0.00E+00	75
								TT	-1.16E-02	36
		Affx-113945924	16	29.22	1.61E-06	5.79	5.79	AA	-1.16E-02	36
								AG	0.00E+00	75
								GG	-1.67E-02	63
Affx-114134235	18	36.94	3.98E-06	5.40	5.40	CC	-1.59E-02	14		
						CT	0.00E+00	80		
						TT	1.01E-02	82		
Plant_transpiration	MLM_K	Affx-114063227	3	43.32	2.49E-06	5.60	AA	5.96E+00	32	
							AG	0.00E+00	93	
							GG	1.67E+00	51	
Leaf_transpiration	MLM_K	Affx-114203619	10	24.58	5.81E-06	5.24	CC	5.86E-03	40	
							CT	0.00E+00	96	
							TT	-5.13E-03	41	
	MLM_K_Q	Affx-114091765	10	24.58	5.32E-06	5.27	5.27	AA	3.78E-03	16
								AG	0.00E+00	68
								GG	-6.49E-03	93
		Affx-114176443	11	38.84	1.04E-06	5.98	5.98	AA	2.90E-03	104
								AG	0.00E+00	64
								GG	1.79E-02	9
Water_use_efficiency	MLM_K	Affx-113863243	9	0.52	2.75E-06	5.56	GG	1.08E-03	10	
							GT	0.00E+00	57	
							TT	-5.78E-04	110	
Accumulated_transpiration	MLM_K_Q	Affx-113788331	18	192.30	4.26E-06	5.37	CC	-1.76E+02	127	
							CT	0.00E+00	46	
							TT	-6.48E+01	3	

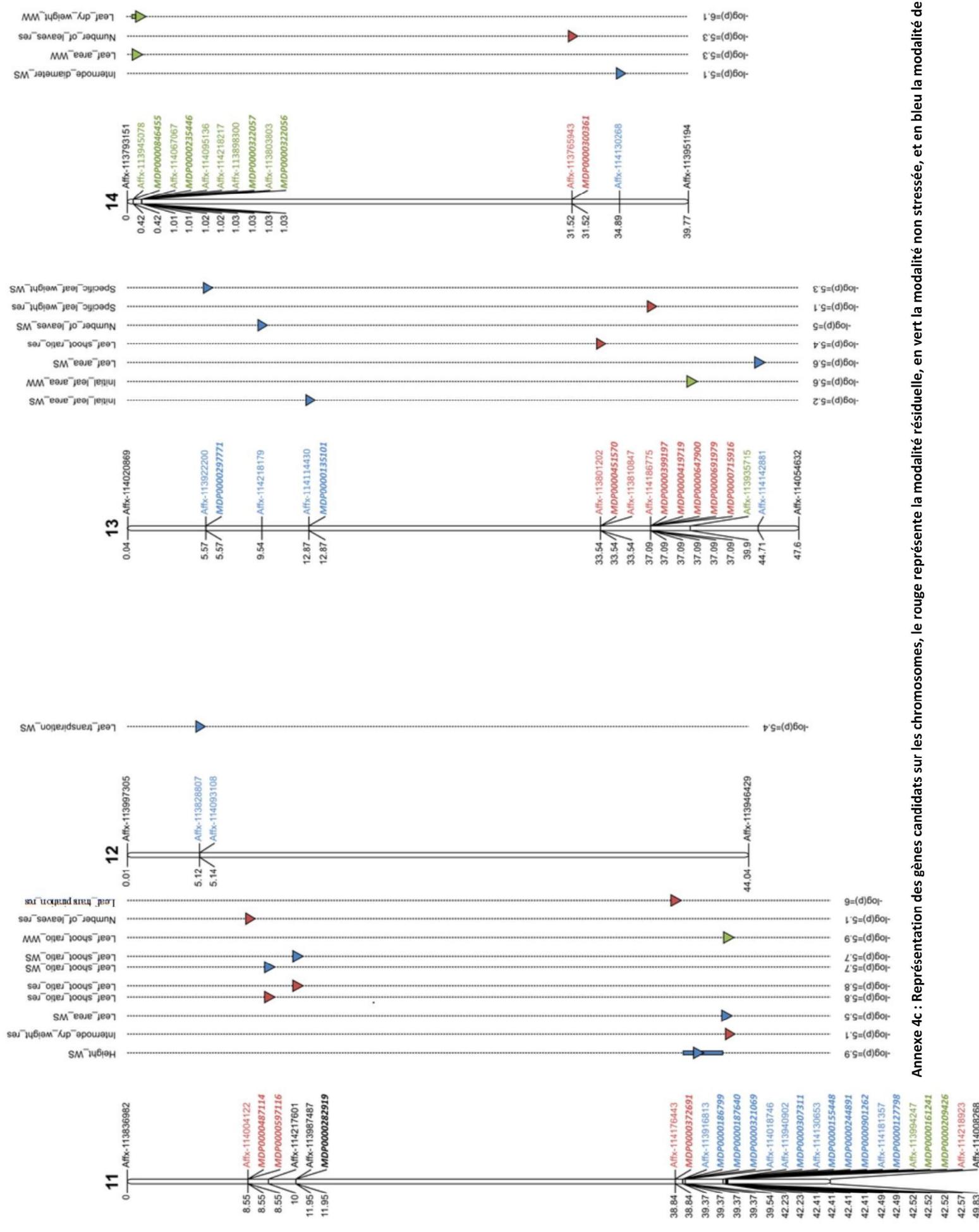
Annexe 5b : Tableau de l'ensemble des associations significatives trouvées pour les variables de la modalité résiduelle. La couleur orange représente les associations qui ne correspondent pas au modèle sélectionné par le BIC.



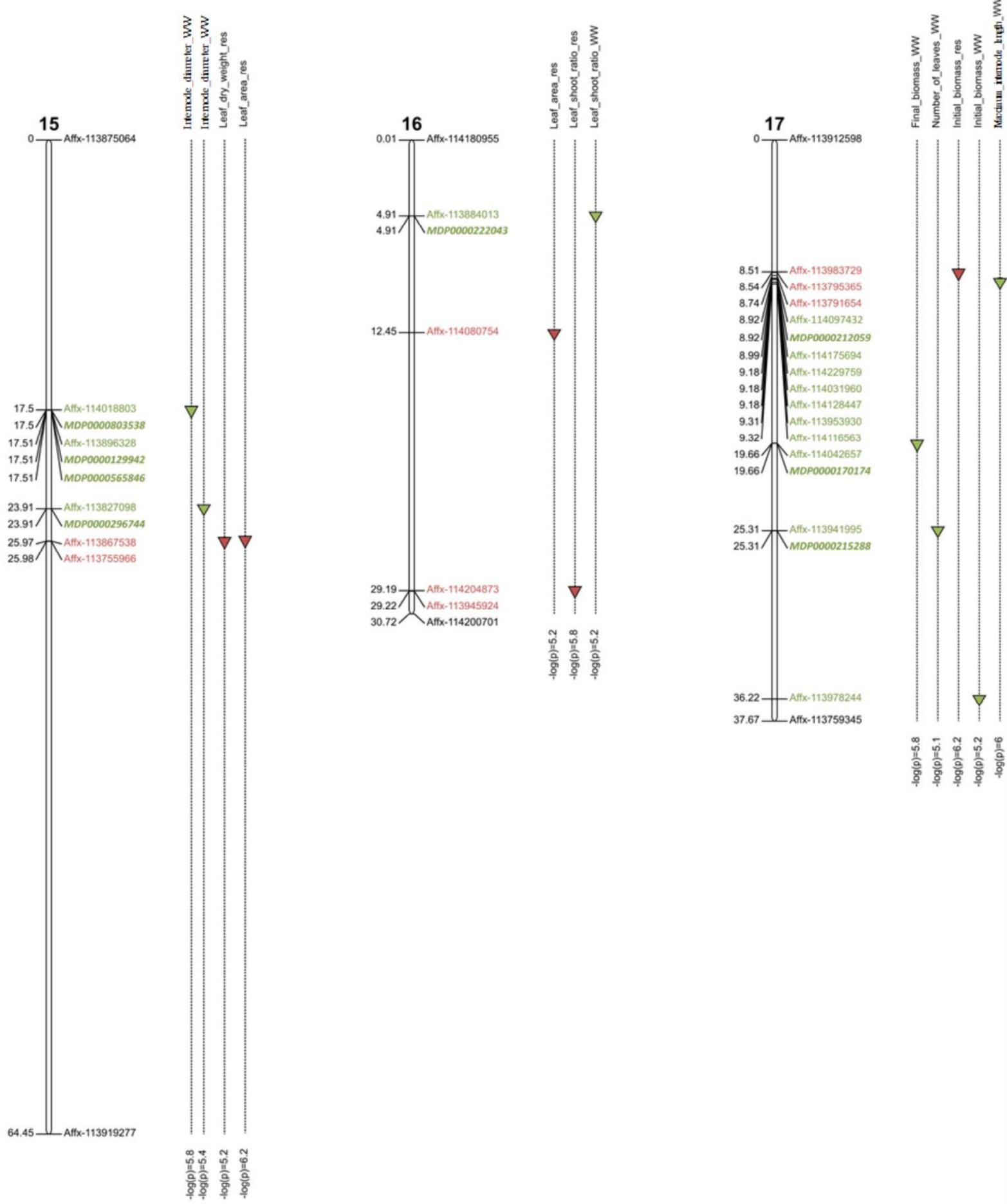
Annexe 4a : Représentation des gènes candidats sur les chromosomes, le rouge représente la modalité résiduelle, en vert la modalité non stressée, et en bleu la modalité de stress hydrique



Annexe 4b : Représentation des gènes candidats sur les chromosomes, le rouge représente la modalité résiduelle, en vert la modalité non stressée, et en bleu la modalité de stress hydrique



Annexe 4c : Représentation des gènes candidats sur les chromosomes, le rouge représente la modalité résiduelle, et en bleu la modalité de stress hydrique



Annexe 6d : Représentation des gènes candidats sur les chromosomes, le rouge représente la modalité résiduelle, en vert la modalité non stressée, et en bleu la modalité de stress hydrique



Diplôme / Mention : Master 2 Sciences Technologie Santé  
Mention Biologie Agronomie Santé  
Spécialité : Production et Technologie du Végétal (ProTeV)  
Parcours II. Sélection, Production et Valorisation des Végétaux (SPVV)  
Option : Innovation variétale  
Responsable d'option : Maria Manzanares-Dauleux

Auteur : Victor Debernardi

Organisme d'accueil : CIRAD UMR AGAP

Date de naissance : 03/10/1991

Adresse : CIRAD complexe Lavalette

Nb de pages : 19 Annexes : 6

avenue agropolis

Année de soutenance : 2016

34090 Montpellier cedex

Maître de stage : Evelyne Costes

### **Recherche des bases génétiques de caractères de croissance et de réponse au stress hydrique d'une core-collection de pommiers, par génétique d'association.**

Le pommier (*Malus domestica*) est l'arbre fruitier le plus cultivé des régions tempérées. La recherche de génotype plus tolérant à la sécheresse est dans le contexte actuel du réchauffement global une priorité en amélioration variétal afin de maintenir une bonne production dans les années à venir.

Le but de cette étude est de rechercher au sein d'une core collection de pommier français des gènes impliqué dans la tolérance à la sécheresse. La puce à ADN 487K (487K SNPs Affymetrix Axiom®) a été utilisée dans le cadre d'une étude de génétique d'association sur génome entier. Deux modèles linéaires mixtes ont été utilisés, l'un prenant structure et apparentement en compte, l'autre ne prenant en compte que l'apparentement. A l'issu des associations trouvés, une approche gène candidat a été réalisé et plusieurs gènes d'intérêts impliqués dans la tolérance à la sécheresse ont été révélés. Une étude transcriptomique sur les gènes candidats permettra par la suite d'en savoir plus sur l'importance de ces gènes vis-à-vis du stress hydrique.

Mots-clés: *Malus domestica*, stress hydrique, core collection, analyse d'association, SNP

### **Searching the genetics basis of growth traits and response to water stress in an apple core-collection, by association genetics**

The apple tree (*Malus domestica*) is the most cultivated fruit tree in temperate regions. The search for more tolerant genotype to drought is, in the context of global warming, a priority in varietal improvement in order to maintain a good production in the coming years.

The goal of this study is to search for genes involved in drought within a french apple core collection The 487K SNPs Affymetrix Axiom® was used as part of a genome wide association study. Two linear mixed models were used, one taking in account the structure and the kinship, the other only taking in account the kinship. When all the association were found, a candidate gene approach was conducted and several genes of interest were revealed. A transcriptomic study on candidate genes will allow us to know more about the importance of those genes regarding water stress.

Keywords: *Malus domestica*, water stress, core-collection, association analysis, SNP

