Précision de la sélection génomique dans des populations constituées de matériel élite et de ressources génétiques chez le pommier

Babacar Diouf

Mémoire de fin d'études

Présenté pour l'obtention du diplôme d'Ingénieur systèmes agricoles et agroalimentaires durables au sud

Spécialité : Amélioration des Plantes et Ingénierie Végétale Méditerranéennes et Tropicales

Précision de la sélection génomique dans des populations constituées de matériel élite et de ressources génétiques chez le pommier

Par **Babacar DIOUF**

Année de soutenance : **2020**

Organisme d'accueil

Institut national de recherche pour l'agriculture, l'alimentation et l'environnement - INRAE
Institut de Recherche en Horticulture et Semences -IRHS

Mémoire de fin d'études

Présenté pour l'obtention du diplôme d'Ingénieur systèmes agricoles et agroalimentaires durables au sud

Spécialité : Amélioration des Plantes et Ingénierie Végétale Méditerranéennes et Tropicales

Précision de la sélection génomique dans des populations constituées de matériel élite et de ressources génétiques chez le pommier

Par **Babacar DIOUF**

Année de soutenance : **2020**

Mémoire préparé sous la direction de :

Hélène MURANTY

Xabi CAZENAVE

Jean Jaques KELNER

Organisme d'accueil :

INRAE- Pays de la Loire

IRHS

Présenté le : **17/09/2020**

Devant le jury :

Muriel TAVAUD

David CROS

Jacques DAVID

Remerciements :

J'adresse mes vifs remerciements à M. Jean Pierre RENOU directeur de l'IRHS, et à M. Charles-Eric DUREL responsable de l'équipe de recherche ResPom qui m'ont accueilli pour mon stage.

Je tiens à exprimer mes remerciements et ma reconnaissance à l'égard de mes deux maîtres de stage Mme Hélène Muranty et M. Xabi Cazenave pour leur disponibilité et leur soutien tout au long de ce stage.

J'associe à ces remerciements tout le personnel de l'IRHS pour leur accueil et pour tous les bons moments passés ensemble

Je remercie beaucoup mon tuteur pédagogique M. Jean Jaques Kelner pour son appui et sa contribution à l'élaboration de ce document. A travers lui tout le corps professoral et administratif de Montpellier SupAgro pour la qualité de la formation reçue.

Je remercie tous mes camarades de Montpellier SupAgro pour ces trois années passées ensemble.

Je ne saurais oublier ma famille, sans ce clin d'œil de reconnaissance pour tous vos soutiens plus particulières ceux de mon Papa Omar DIOUF et de mon Cousin Barthélemy FAYE.

Table des matières

Liste des figures

Liste des tableaux

Sigles et abréviations

ACP : Analyse en Composantes Principales

BLUP : Best Linear Unbiased Prediction

Cor-rang : corrélation des rangs

GBLUP : Genomic Best Linear Unbiased Prediction

GEBV : Genomic Estimated Breeding Values

HIDRAS : High Quality Disease Resistant Apples for a Sustainable Agriculture

MAF : Minor Allele Frequency

MSE : Erreur quadratique moyenne

RRBLUP : Ridge Regression Best Linear Unbiased Prediction

QTL : Quantitative Trait Locus

SAM : Sélection Assistée par Marqueurs

SNP : Single Nuclear Polymorphism

TBV : True Breeding Value (TBV)

Introduction

Le pommier (*Malus domestica* Borkh) est l'un des arbres fruitiers les plus cultivés dans le monde avec un volume de production de 86 millions de tonnes (FAOSTAT, 2018). Les régions tempérées enregistrent les plus importantes quantités de production. Parmi elles, la Chine seule, assure environ 45% de la production mondiale et se trouve au premier rang des pays producteurs de pomme. Avec 23% de la production mondiale, l'Europe est la deuxième plus grande région de production. La Pologne, l'Italie et la France représentent les principaux grands producteurs de l'Union Européenne. Pour ces dix dernières années, la production française de pomme est en moyenne 1,7 million de tonnes (Agreste, 2020). Ce niveau de production fait de la France, l'un des plus grands pays exportateurs de pomme. Cependant, la France voit sa part de marché diminuer lorsque la production européenne de pommes Gala (qui est la variété traditionnelle phare des exportations françaises), augmente par rapport à la moyenne (Agreste 2020). Selon World Apple and Pear Association (2018) environ la moitié de la culture européenne de pomme est basée sur quatre variétés : Golden Delicious, Gala, Idared, Red Delicious. Et pour l'année 2019, la variété Golden Delicious représente environ un quart des prévisions de récolte européenne.

L'origine de cette prépondérance de quelques variétés sur le marché, remonte aux années 90, où les investissements financiers dans la recherche pour l'amélioration variétale des pommes avaient généralement baissé (Noiton et Shelbourne, 1992). Cette situation a entrainé une utilisation répétée des variétés élites comme parents dans les programmes de sélection (Noiton et Alspach, 1996) . Par conséquent, au fil des années, la population élite s'est retrouvée avec une base génétique étroite. Cette faible diversité génétique induit une vulnérabilité importante des variétés face à la pression de facteurs biotiques (ravageurs, maladies) et abiotiques (sécheresse, salinité, gel) qui devient de plus en plus importante à cause des changements climatiques.

Ce problème pourrait être résolu en élargissant la base génétique de la population élite par transfert d'allèles favorables provenant des ressources génétiques grâce au croisement et à la sélection.

La sélection variétale du pommier est un long processus lié en partie à sa longue phase juvénile. De plus la qualité des fruits est une caractéristique complexe qui necessite l'évaluation de plusieurs traits et cela entraine des coûts onéreux de phénotypage. Dans ce contexte, l'emploi de plusieurs marqueurs génétiques pour prédire les phénotypes des individus est une stratégie prometteuse et permet aussi de diminuer la durée de la sélection (Heffner et al., 2009). Cette méthode appellée sélection génomique est apparue au début du XXI^{e} siècle et a été utilisée pour la première fois chez les bovins (Hayes et al., 2009b). Son application chez les végétaux s'est montrée aussi très efficace (Kumar et al., 2012a).

L'utilisation de la sélection génomique repose sur la construction d'un modèle de prédiction des valeurs génétiques d'individus génotypés mais non phénotypés à l'aide d'une population d'entrainement génotypée et phénotypée. La constitution de cette population d'entrainement est un des facteurs déterminants de la précision de la prédiction et donc de l'efficacité de la sélection génomique.

1 Synthèse bibliographique

1.1 Description de l'espèce

1.1.1 Taxonomie

Le pommier cultivé ou *Malus domestica* Borkh, appartient à la sous famille des *Maloideae,* de la famille des *Rosaceae* (Harris et al., 2002). Cette famille botanique regroupe de nombreuses autres cultures fruitières importantes sur le plan économique, telles que la poire, la pêche, la prune, l'abricot, l'amande, la cerise, la fraise et la framboise. Le genre *Malus* auquel appartient le pommier regroupe une trentaine d'espèces (Ma et al., 2017).

1.1.2 Origine

La domestication du pommier s'est faite il y a au moins 4000 ans (Cornille et al., 2014). De nombreuses études effectuées sur la domestication et l'origine de la pomme (Cornille et al., 2012 ; Cornille et al., 2014 ; Gross et al., 2014 ; Wani et al., 2015 ; Cornille et al., 2019) confirment qu'elle provient du continent asiatique. La région d'Asie centrale qui est identifiée comme le plus grand centre de diversité du pommier est considérée comme le lieu d'origine de la pomme domestiquée d'après Harris et al. (2002). Cette hypothèse est confortée par les travaux de Cornille et al. (2012), qui confirment que la pomme cultivée est originaire d'Asie centrale où l'on trouve encore aujourd'hui son ancêtre sauvage, *Malus sieversii*. Ces mêmes études ont permis d'identifier cette espèce comme le principal géniteur de *M. domestica* sur la base de la comparaison de séquences d'ADN. Par ailleurs, le génome du pommier montre des traces d'introgressions secondaires venant d'espèces sauvages dont les plus importantes sont *M. sylvestris, M. baccata* et *M.orientalis* (Wani et al., 2015). Cet échange de fragments d'ADN aurait eu lieu le long des routes des caravanes commerciales communément appelées "Route de la soie", et qui reliaient l'Asie et l'Europe par voie terrestre. Parmi ces espèces identifiées, *M. sylvestris* serait l'espèce d'Europe occidentale la plus proche de *M. domestica* d'après les résultats d'analyse des marqueurs microsatellites de Cornille et al. (2012).

1.1.3 Caractéristiques biologiques

Les pommiers en général sont des cultures pérennes qui présentent de longs cycles de graine à graine qui peuvent durer entre 3 et 8 ans. La maturité des fruits est atteinte généralement entre 70 et 180 jours après la floraison et dépend en grande partie du cultivar, qui peut être à maturation précoce, moyenne ou tardive (Kole, 2011). *M. domestica* est un arbre à fleurs hermaphrodites dont le régime de reproduction sexuée est assuré par une allogamie prédominante. Comme toutes les espèces du genre *Malus*, le pommier possède un système d'auto-incompatibilité gamétophytique entrainant une

faible réussite de l'autofécondation (Cornille et al., 2012). Le pommier domestiqué est caractérisé par un fort taux d'hétérozygotie alors qu'il présente une origine autopolyploïde. Au cours de son évolution, le génome du pommier est devenu diploïde, avec un nombre de chromosomes à 2n = 34 (Velasco et al., 2010; Ma et al., 2017), ce qui revient à un nombre de chromosome de base égal à 17, et un génome relativement petit de 650 Mb (Daccord et al., 2017).

1.1.4 Diversité génétique

Les caractéristiques de la qualité des fruits du pommier sont très variées. Les traits les plus évalués sont la couleur, la texture, le goût, la saveur, la taille et la forme des pommes, le stockage et la durée de conservation. Ainsi, la sélection variétale fait recours à l'exploitation de la diversité génétique pour enrichir ou résoudre certains problèmes identifiés chez les variétés commerciales. Les hybrides naturels sont d'une grande importance car ils possèdent généralement des événements de recombinaisons bénéfiques. Par exemple *M. sieversii,* qui est le principal ancêtre de la pomme cultivée est identifié comme une espèce avec un haut degré de diversité (Kole, 2011).

- **Gestion de la diversité**

La conservation de la diversité du pommier se fait sur des collections génétiques qui accueillent généralement des variétés locales ainsi que des variétés commerciales élites Elles peuvent contenir parfois des accessions sauvages. Beaucoup de pays producteurs de pommes ont recours à cette stratégie de conservation (Bramel et Volk, 2019). L'utilisation des marqueurs moléculaires pour la caractérisation et la gestion de ces collections devient de plus en plus répandue.

- **Diversité observée**

L'étude de Urrestarazu et al. (2016) sur la diversité des ressources génétiques de la pomme au niveau européen a identifié trois groupes et des sous-groupes qui illustrent une combinaison de processus historiques (migration, sélection) et des facteurs d'adaptation à divers environnements agricoles. Cette variation confirme en partie l'importance de la diversité génétique du germoplasme européen. En France l'étude de la diversité du pommier avec un nombre important de marqueurs microsatellites a montré une collection globale diverse malgré la redondance au sein des collections et entre celles-ci (Lassois et al., 2016).

1.1.5 Sélection phénotypique et SAM chez le pommier

La sélection phénotypique de création variétale chez le pommier est un processus long avec des coûts élevés de phénotypage. Cela est dû en partie à la longue phase juvénile du pommier qui retarde le phénotypage de certains caractères nécessaires à l'identification des génotypes les plus intéressants. C'est le cas de la productivité qui ne peut être évaluée qu'environ sept ans après les semis.
Pour pallier de telles contraintes, des marqueurs moléculaires ont été développés pour permettre une identification précoce des génotypes les plus intéressants, comme ceux présentant un allèle de

résistance à une maladie. Cependant, cette méthode de sélection assistée par marqueur est inefficace si le caractère d'intérêt est contrôlé par de nombreux gènes à effet faible (Muranty et al., 2014). Il est alors possible d'avoir recours à un grand nombre de marqueurs génétiques : cette approche a été baptisée sélection génomique.

1.2 La sélection génomique

1.2.1 Principe de la sélection génomique

Le phénotype de chaque individu dépend en partie des gènes qu'il porte. La somme des effets de ses gènes sur un caractère constitue sa valeur génétique. Elle peut être décomposée selon l'équation G = A + D + I avec A, la valeur génétique additive (ou breeding value), D, les effets de dominance résultant de l'interaction entre les allèles d'un même locus et I, les effets d'épistasies résultant des interactions entre les allèles de loci différents. La valeur génétique additive est la seule part de la valeur génétique qui est transmissible par reproduction sexuée à la descendance.

Le principe de la sélection génomique repose sur la prédiction de la valeur génétique additive de chaque candidat à la sélection, par l'utilisation d'un nombre important de marqueurs couvrant tout leur génome (Meuwissen et al., 2001). Les effets additifs des marqueurs génétiques sont estimés grâce aux déséquilibres de liaison existant entre ces marqueurs et les QTL causaux (Goddard et Hayes, 2007). Ensuite, la valeur génétique additive de chaque individu est estimée en calculant la somme des effets additifs de tous ses marqueurs.

Les estimations des effets des marqueurs sont effectuées à l'aide d'une population dite d'entrainement (ou référence) puis reportées sur une autre population dite candidate (ou validation). Cela est traduit concrètement par l'élaboration d'un modèle de prédiction établi sur la population d'entrainement dont les données génotypiques et phénotypiques ont été préalablement acquises. A partir de ce modèle, il devient possible de prédire les valeurs génétiques additives des individus de la population candidate dont seules les données génotypiques sont obtenues.

La valeur génétique additive prédite chez un individu est appelée aussi Genomic Estimated Breeding Values (GEBV).

1.2.2 Apport de la sélection génomique chez le pommier

L'application de la sélection génomique chez le pommier permet d'augmenter considérablement le gain génétique qui représente la réponse à la sélection (Kumar et al., 2012a).

Le gain génétique par unité de temps ($\Delta G\iota$) est obtenu selon l'équation suivante (Falconer et Mackay, 2009) : $\Delta G\iota = \frac{\mathrm{i} \times \mathrm{r} \times \sigma}{\mathrm{L}}$

avec :

- i : l'intensité de la sélection
- r : la précision de la sélection
- σ : la variance génétique additive
- L : l'intervalle de générations

L'application de la sélection génomique permet de réduire l'intervalle entre les générations, ce qui a comme avantage une réduction considérable de la durée des programmes d'amélioration variétale chez le pommier. Elle permet aussi d'augmenter l'intensité de sélection, car les jeunes plants sont moins coûteux à produire que des arbres phénotypés en verger, même si on peut envisager de phénotyper des individus après une pré-sélection par la sélection génomique. Cela permet de réduire l'effectif des arbres à phénotyper, ce qui représente une réduction des surfaces occupées et par conséquence une diminution du coût du phénotypage. Le gain génétique peut aussi augmenter avec une meilleure précision de prédiction.

1.2.3 Modèles statistiques de prédiction génomique

Les modèles statistiques utilisés en sélection génomique ont pour objectif de calculer les GEBV d'une population candidate. Cela peut se faire directement en prédisant la GEBV de chaque individu ou par estimation des effets additifs associés à chaque marqueur pour ensuite faire la somme. La complexité méthodologique de la sélection génomique réside dans le fait que le nombre de variables explicatives (les marqueurs) est nettement supérieur au nombre de variables à expliquer (individus). Cette situation ne permet pas d'utiliser un modèle de régression linéaire multiple car le nombre de degrés de liberté est insuffisant et par conséquent il est impossible d'estimer les effets de tous les marqueurs. La solution à cette contrainte est d'appliquer dans les modèles, des méthodes statistiques de pénalisation comme le « shrinkage » qui est un rétrécissement vers 0 des effets estimés des marqueurs. En plus, selon les lois des modèles utilisés, différents types d'effets de marqueurs (fort faible ou nul) peuvent être autorisés. Les modèles les plus couramment utilisés en prédiction génomique, sont les modèles GBLUP, RRBLUP et les méthodes bayésiennes. Toutes ces méthodes sont basées sur des équations du modèle mixte d'évaluation de la valeur génétique. Le modèle mixte classique est le suivant :

$$\mathcal{Y} = \mathrm{X}\beta + \mathrm{Z}\mathcal{u} + e$$

Où $\mathcal{Y}$ est un vecteur de taille n contenant les observations où n est le nombre d'observations.

- $\mathrm{X}_{n,p}$ correspond à une matrice de design associée aux effets fixes ;
- β est un vecteur de taille p contenant les effets fixes où p est le nombre d'effets fixes;
- $\mathrm{Z}_{n,q}$ correspond à une matrice de design associée aux effets aléatoires ;
- $\mathcal{u}$ est un vecteur de taille q contenant les effets aléatoires où q le nombre d'effets aléatoires;
- e est un vecteur de taille n contenant les erreurs résiduelles.

Si $\mathcal{u}$ représente le vecteur des valeurs génétiques et $\mathcal{u} \sim \mathcal{N}(0, A\sigma_a^2)$, Henderson, (1975) a montré que les effets fixes et aléatoires peuvent être estimés grâce à l'équation suivante :

$$\begin{pmatrix} X'X & X'Z \\ Z'X & Z'Z + \lambda A^{-1} \end{pmatrix} \begin{pmatrix} \beta \\ \mathcal{u} \end{pmatrix} = \begin{pmatrix} X'y \\ Z'y \end{pmatrix}$$

Où $A_{q,q}$ est la matrice des corrélations génétiques additives basée sur le pédigrée

- $\lambda = \frac{\sigma_e^2}{\sigma_a^2}$ est le coefficient de pénalisation à estimer ;
- σ_e^2 est la variance résiduelle du modèle ;
- σ_a^2 est la variance génétique additive.

1.2.3.1 GBLUP

Le GBLUP ou BLUP génomique est une méthode utilisée en remplaçant la matrice d'apparentement généalogique $A_{q,q}$ par une matrice d'apparentement génomique $G_{q,q}$ (VanRaden, 2008). Cette dernière mesure l'apparentement entre individus sur la base de leur proportion d'allèles identiques pondérée par leur fréquence. Ce calcul de niveau d'apparentement basé sur l'information génétique est plus précis que celui basé sur l'apparentement généalogique qui ne prend pas en compte les différences entre individus dues à la ségrégation allélique. Avec la méthode GBLUP, le vecteur $\mathcal{u}$ de taille q contenant les valeurs génétiques des individus (ou BLUP) suit une loi normale de paramètre $(0, G\sigma_a^2)$. Et $\mathcal{u}$ peut être estimé à partir de l'équation du modèle mixte suivant :

$$\widehat{u} = \mathrm{GZ}'\,(\,\mathrm{ZGZ}' + \mathrm{I}\,\lambda)^{-1}\,(\mathrm{y} - 1\hat{\mu})$$

1.2.3.2 RR-BLUP

Le RR-BLUP ou ridge regression BLUP estime l'effet de chaque marqueur génétique pour en déduire la valeur génétique des individus génotypés. Dans ce cas la matrice de design associée aux effets aléatoires du modèle mixte correspond à une matrice de génotypage $W_{p,m}$. Et les effets des marqueurs sont contenus dans le vecteur g de taille m où m est le nombre de marqueurs. Le vecteur g suis une loi normale de paramètre $(0, \mathrm{I}\sigma_g^2)$. Avec la méthode RR-BLUP, la résolution des équations du modèle mixte ne nécessite pas l'utilisation de matrice d'apparentement. Ainsi les effets des marqueurs (ou g) peuvent être estimés à partir de l'équation suivante :

$$\hat{g} = \mathrm{W}'\,(\,\mathrm{WW}' + \mathrm{I}\,\lambda)^{-1}\,(\mathrm{y} - 1\hat{\mu})$$

1.2.3.3 Méthodes Bayésiennes

Les méthodes présentées estiment un effet non-nul pour chaque marqueur et supposent que les effets des marqueurs suivent tous une même loi normale centrée sur 0. Les méthodes bayésiennes proposent une distribution a priori des effets aléatoires attribués à chaque marqueur. En plus, les effets des marqueurs ne sont pas tous identiques, il est possible d'obtenir des marqueurs à effet faible voire nul et des marqueurs à fort effet (Gianola et al., 2009). Les méthodes bayesiennes peuvent donc être particulièrement utiles lorsque le caractère à prédire est gouverné par quelques QTL à effet fort (van den Berg et al., 2015).

1.2.4 Précision de la prédiction génomique et les facteurs qui l'influencent

La précision de prédiction (ou accuracy) est la corrélation entre les vraies valeurs génétiques appelées True Breeding Values (TBV) et celles estimées (GEBV)

En réalité, les vraies valeurs génétiques sont inconnues mais elles peuvent être estimées par BLUP, BLUE ou par calcul de phénotypes moyens obtenus à partir d'une évaluation multi-environnementale.

L'efficacité de la sélection génomique dépend de la précision des prédictions, qui est influencée par plusieurs facteur tels que :

- **l'héritabilité du caractère** : plusieurs études sur la sélection génomique ont montré que la précision de prédiction est fortement influencée par l'héritabilité du caractère étudié (Desta et Ortiz, 2014). Par exemple, les travaux de Muranty et al. (2015) ont montré que la précision de prédiction était plus élevée pour les caractères présentant les plus fortes héritabilités chez le pommier
- **la densité de marqueur** : Calus et al. (2008) ont montré que l'augmentation de la densité de marqueurs entraine une augmentation de la précision de prédiction. Mais la précision atteint un plateau à partir d'un seuil de densité de marqueurs.

- **la taille de la population d'entrainement** : la précision de prédiction des GEBV dépend aussi de la taille de la population d'entrainement car l'augmentation de la taille de la population entraine une augmentation du nombre d'haplotypes à estimer pour un caractère (Edwards et al., 2019). De même que pour la densité de marqueurs, passé un seuil, le nombre d'individus de la population d'entrainement n'améliore pas la qualité de la prédiction.
- **apparentement entre la population d'entrainement et la population candidate** : la prédiction génomique repose sur l'hypothèse selon laquelle les déséquilibres de liaison entre marqueurs et QTL identifiés dans la population d'entrainement sont les mêmes dans la population candidate. Pour que l'estimation des effets des marqueurs soit cohérente, les individus composant la population d'entrainement ne doivent pas être génétiquement éloignés de ceux de la population candidate (Clark et al., 2012; Pszczola et al., 2012). Ainsi, une faible précision de prédiction a été trouvée chez les bovins lorsque les GEBV de la race Holstein étaient calculées à partir de la race Jersey (Hayes et al., 2009a)

-

1.3 Objectifs

L'objectif principal de cette présente étude est d'évaluer les précisions de prédiction génomique sur deux populations différentes. La première est constituée uniquement de matériel élite et la deuxième uniquement de ressources génétiques. Nous caractériserons au préalable les deux populations afin de mieux expliquer les précisions de prédiction qui seront obtenues.

2 Matériels et Méthodes

2.1 Populations d'étude

2.1.1 Matériel élite

La majorité des génotypes du matériel élite proviennent du projet de recherche européen HIDRAS (High Quality Disease Resistant Apples for a Sustainable Agriculture). Ce projet a vu la participation de huit instituts de recherche situés dans six pays : Allemagne, Belgique, France, Italie, Pologne et Royaume-Uni. Des descendances produites dans le cadre des programmes d'amélioration de chaque institut ont été choisies pour constituer une population en vue d'une recherche de QTLs en pedigree. Le pedigree permettant de décrire les relations entre ces descendances compte aujourd'hui sept générations. Chaque descendant obtenu a été évalué dans l'institut où il avait été créé. Une trentaine de variétés, présentes dans tous les instituts, a été utilisée pour servir de témoin. Ce sont tous ces descendances appelées ici matériel élite qui ont été utilisées pour cette étude. D'autres descendances de même nature mais provenant d'autres projets ont été utilisées pour la caractérisation génomique du matériel élite mais pas lors de l'évaluation de la prédiction génomique, en l'absence de données phénotypiques pour les caractères retenus.

2.1.2 Ressources génétiques

Les données concernant les ressources génétiques ont été rassemblées ou produites dans le cadre du projet européen FruitBreedomics (Laurens et al., 2018). Ce matériel végétal est composé de variétés anciennes de pommes "à couteau" issues de six core-collections : CRA-W-Gembloux (Belgique), INRAE Angers (France), RBIP-Holovouzy (République tchèque), SLU-Balsgard (Suède), Université de Bologne (Italie) et NFC-Brogdale (Royaume-Uni).

2.2 Phénotypage des populations

2.2.1 Matériel élite

Le phénotypage du matériel élite a été standardisé et les méthodes de mesure ou d'évaluation harmonisées afin d'effectuer avec précision les mêmes évaluations dans tous les instituts. Les caractéristiques externes et sensorielles des fruits ont été évaluées sur tous les génotypes. Dans mon étude, seules les caractéristiques identiques à celles évaluées sur les ressources génétiques ont été retenues. Elles sont au nombre de sept et sont : la couleur superposée du fruit, la forme du fruit, l'acidité, le croquant, le goût, la jutosité et la date de récolte. Les BLUPs des valeurs phénotypiques

moyennes sur plusieurs années ont été calculés pour tous les génotypes lors du projet HiDRAS et ont été utilisés pour mon étude comme phénotypes du matériel élite.

2.2.2 Ressources génétiques

Le phénotypage des ressources génétiques a été effectué séparément dans chaque institut apportant une collection, suivant un protocole commun. Toutes les données d'évaluation des différents instituts ont été rassemblées pour calculer les moyennes ajustées des données phénotypiques. Comme indiqué précédemment seules les caractéristiques identiques à celles du matériel élite ont été utilisées lors de mon étude.

2.3 Génotypage des populations

2.3.1 Matériel élite

Le matériel élite a été génotypé avec une puce 20K (Bianco et al., 2014) dans le cadre du projet FruitBreedomics. Pour mon étude, un ensemble de 1509 génotypes provenant de 29 familles a été utilisé pour représenter le matériel élite. Les données génotypiques ont été imputées à l'aide du logiciel BEAGLE 4.0 (Browning et al., 2007) en utilisant la population de ressources génétiques phasée comme population de référence et en utilisant les relations de pedigree disponibles au moment de l'étude afin d'améliorer les résultats d'imputation. Cette imputation a permis d'augmenter le nombre de marqueurs à 303239.

2.3.2 Ressources génétiques

Les ressources génétiques ont été génotypées à l'aide de la puce à haute densité Axiom Apple 480K (Bianco et al., 2016) et 303239 marqueurs SNP ont été retenus. Un total de 1341 génotypes issus des ressources génétiques a été utilisé pour mon étude avec tous les 303239 SNPs.

2.4 Caractérisation des populations

2.4.1 Évaluation de la diversité génétique et de la structure des populations

2.4.2 Diversité génétique

Pour évaluer le niveau de diversité génétique de chacune des deux populations, les indices d'hétérozygotie moyenne observée (H_o) et attendue (H_e) ont été calculés. De plus, l'indice de différenciation moyenne (F_{st}) entre les deux populations a été évalué (voir tableau 2). Par ailleurs, les répartitions de fréquences d'allèles minoritaires (MAF) de tous les SNPs ont été comparées entre les deux populations à l'aide de représentations graphiques par fenêtre glissante sur tout le long du

génome. La taille des fenêtres glissantes a été fixée à 2 Mb avec un pas de 400 kb. La donnée n'a pas été calculée quand la fenêtre comportait moins que 10% du nombre de SNPs attendus. Les intervalles de confiance à 95% ont été calculés par bootstrap avec 2000 répétitions. Afin de prendre en compte la distribution asymétrique des fréquence alléliques, la méthode "BCA" (Bias-Corrected and Accelerated bootstrap) du package "boot" de R a été utilisée (Canty et al., 2020).

2.4.3 Structure génétique

Les données génétiques des deux populations étaient trop volumineuses pour effectuer directement les analyses de structure. Par conséquent, un élagage a été réalisé sur la base du déséquilibre de liaison entre marqueurs. Le logiciel PLINK 1.9 a été utilisé pour comparer de façon successive des paires de marqueurs sur des fenêtres glissantes afin de retenir uniquement les marqueurs présentant une corrélation inférieure à 0,1. L'élagage a été fait en utilisant des fenêtres glissantes de 50 SNPs avec un pas de 10 SNPs.

Après élagage des données génotypiques, les structures des populations ont été évaluées par deux méthodes différentes :

- L'analyse en composante principale (ACP) : les ACP ont été calculées avec la fonction prcomp() de R (R Core Team 2020) et les données ont été représentées avec ggplot2. L'ACP consiste à représenter tous les génotypes dans un espace de dimension réduite avec au minimum deux axes principaux. Cette schématisation de l'ACP est basée sur les distances euclidiennes entre individus en se référant à leur niveau de ressemblance. Ainsi les individus groupés en un endroit du graphe témoignent un fort de degré de ressemblance entre eux par rapport aux autres individus distants.
- La structure a été aussi évaluée à l'aide du logiciel Admixture 1.3 (Alexander et al. 2015). Son principe de fonctionnement repose sur l'approche de la validation croisée (expliquée au 2.5.2) pour calculer les erreurs de prédictions des génotypes d'appartenir à un groupe génétique ancestral (K) dont le nombre est renseigné apriori. Le logiciel établit en parallèle une matrice Q évaluant les probabilités de chaque génotype d'appartenir à un groupe. Le nombre de K pour lequel l'erreur de la validation croisée est plus faible est celui qui est retenu comme décrivant mieux la structure la population. Un nombre de K allant de 1 à 20 a été appliqué pour évaluer la structure du matériel élite et des ressources génétiques.

2.5 Prédiction génomique

2.5.1 Modèle de prédiction

Le modèle GBLUP (expliqué au 1.2.3.1) a été choisi pour effectuer les prédictions génomiques du matériel élite et des ressources génétiques. Les matrices d'apparentement des populations ont été calculées grâce au package « rrblup » (Endelman, 2011). Le modèle appliqué pour prédire les données phénotypiques des différents caractères est le suivant :

$$\mathcal{Y} = \mathrm{X}\beta + \mathrm{Z}u + e$$

Où $\mathcal{Y}$ est un vecteur de taille n contenant les valeurs phénotypiques avec n le nombre d'individus.

- X_n est la matrice d'incidence des effets fixes;
- β_n représente le vecteur contenant la moyenne des valeurs phénotypiques des individus de la population,
- Z_n correspond à la matrice d'incidence des effets aléatoires;
- u_n est le vecteur contenant les valeurs additives des individus de la population, $u \sim \mathcal{N}(0, G\sigma_a^2)$ où G est la matrice contenant les apparentements génomiques estimés et σ_a^2 est la variance génétique additive. ;
- e est le vecteur des erreurs résiduelles de taille n.

2.5.2 Évaluation de la précision de prédiction

La précision de prédiction du modèle a été déterminée par la méthode de la validation croisée qui consiste à prédire les valeurs phénotypiques d'une partie de la population (validation) à l'aide de l'autre partie restante (entrainement). La précision de la prédiction du modèle est obtenue en calculant la corrélation de Pearson entre les valeurs phénotypiques observées et celles prédites.

Pour faire la validation croisée, les individus de la population totale ont été répartis en 5 groupes et les phénotypes d'un des groupes ont été masqués pour servir de population de validation. Les 80% des individus restants ont été utilisés comme population d'entrainement devant servir à prédire les valeurs phénotypiques de la population de validation. La constitution des populations d'entrainement et de validation se faisait au hasard.

Au total 100 répétitions ont été réalisées et 5 précisions de prédiction ont été obtenues à chaque répétition en faisant passer chacun des 5 groupes comme population de validation. La précision de prédiction du modèle a été déterminée en faisant la moyenne des 500 précisions obtenues.

En parallèle des calculs de précision de prédiction de trois autres estimateurs ont été calculés pour mieux évaluer le modèle :

- le biais de prédiction est obtenu en calculant le coefficient de régression entre les vraies valeurs phénotypiques et celle prédites : Un coefficient de régression de 1 indique une

- absence de biais, tandis qu'un coefficient de régression inférieur à 1 indique une « surestimation », c'est-à-dire une variance plus importante des valeurs prédites par rapport aux valeurs observées. Un coefficient de régression supérieur à 1 indique une « sous-estimation », c'est-à-dire une variance plus faible des valeurs prédites par rapport aux valeurs observées.
- l'erreur quadratique moyenne (ou MSE) : en considérant l'erreur de prédiction d'un génotype comme étant la différence entre sa valeur phénotypiques observée et celle prédite, la MSE de la population est obtenue en calculant la moyenne des carrés des erreurs.
- La corrélation des rangs (cor-rang) est déterminée en calculant la corrélation des rangs des individus avant et après prédiction. Les rangs sont établis par rapport aux valeurs phénotypiques des individus. Les rangs sont moins différents lorsque leur corrélation est proche de 1.

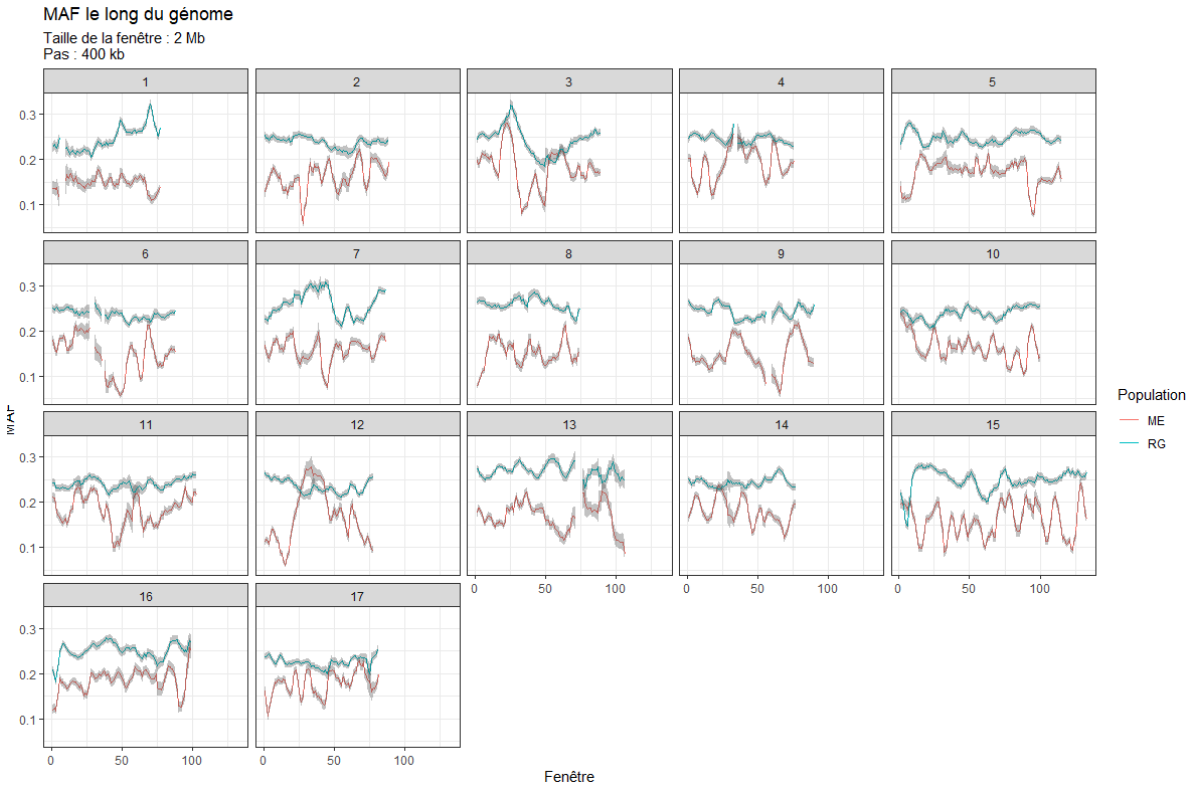


Figure 1 : MAF (Minor allele frequency) des SNPs dans le matériel élite et les ressources génétiques le long du génome. Les courbes représentent la moyenne des MAF dans des fenêtres de 2 Mb avec un pas de 400 kb.

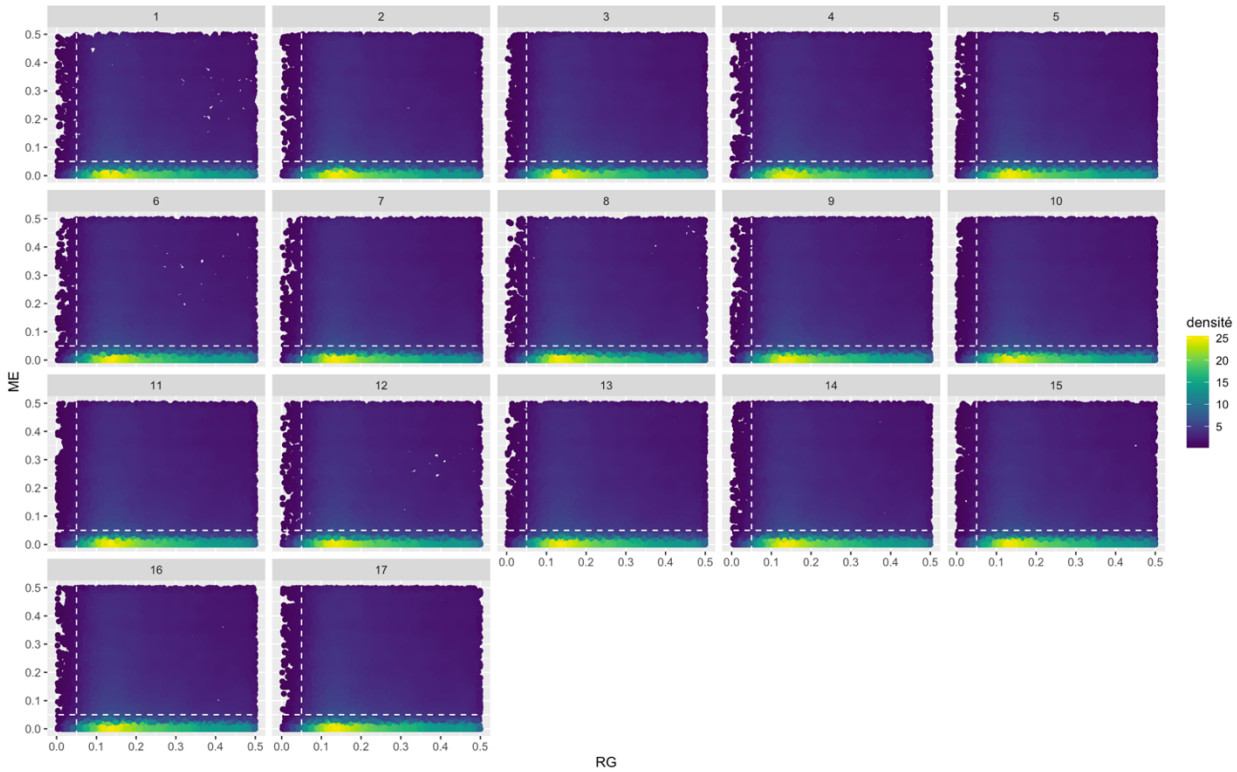


Figure 2 : Densités de SNPs selon leur MAF dans le matériel élite et dans les ressources génétiques sur tout le génome, par chromosome. Les pointillés horizontaux et verticaux représentent les limites des MAF inférieures à 0,05 pour le matériel élite et les ressources génétiques, respectivement.

3 Résultats

3.1 Diversité génétique

3.1.1 Indices de diversité génétique

Les fréquences des allèles de référence calculés sont de 0,80 pour le matériel élite par rapport à 0,75 pour les ressources génétiques. Les taux moyens d'hétérozygoties observés et attendus sont plus élevés sur les ressources génétique que sur le matériel élite (Tableau 1). L'indicateur de différenciation génétique moyenne entre populations est égal à 0,004.

Tableau 1: Indices de diversité génétique du matériel élite et des ressources génétiques

Pop	**Effectif**	**Freq_ref**	**Freq_alt**	$\mathbf{H_o}$	$\mathbf{H_e}$	$\mathbf{H_s}$	$\mathbf{H_t}$	$\mathbf{H_i}$	$\mathbf{F_{st}}$	$\mathbf{F_{is}}$	$\mathbf{F_{it}}$
ME	1517	0,80	0,20	0,25	0,32	0,343	0,35	0,30	**0,004**	0,14	0,14
RG	1341	0,75	0,25	0,34	0,37						

* **Freq_alt** : fréquence des allèles alternatifs, **freq_ref** : fréquence des allèles de référence, $\mathbf{H_o}$: Taux moyen d'hétérozygotie observée, $\mathbf{H_e}$: Taux moyen d'hétérozygotie attendue, $\mathbf{H_s}$: Moyenne des hétérozygoties attendues, $\mathbf{H_t}$: hétérozygotie attendue moyenne de la population totale, $\mathbf{H_i}$: Moyenne des Hétérozygoties observées, $\mathbf{F_{st}}$: différenciation génétique moyenne entre les deux populations, $\mathbf{F_{is}}$: déviance de l'hétérozygotie globale observée par rapport à l'hétérozygotie globale attendue, $\mathbf{F_{it}}$: déficit global en hétérozygotes

3.1.2 Évaluations des fréquences alléliques minoritaires (MAF)

Les fréquences alléliques minoritaires déterminés sur l'ensemble des marqueurs montrent des niveaux de fréquences différents sur tout le long du génome entre le matériel élite et les ressources génétiques (Figure 1). En comparaison avec les ressources génétiques, le matériel élite présente généralement des niveaux de MAF plus faibles. Cela est confirmé par l'ampleur des variations et leur répétition sur les courbes des MAF du matériel élite. De plus la densité de MAF du matériel élite est supérieure à celle des ressources génétiques sur tout le génome. Le pourcentage de SNPs de MAF inférieure à 0,05 est de 36,5% pour le matériel élite contre 3,4% pour les ressources génétiques. Ainsi, les plus fortes densités de SNPs se trouvent dans parmi ceux ayant une MAF inférieure à 0,05 dans le matériel élite et comprise entre 0,1 et 0,2 dans les ressources génétiques (Figure 2). Les deux populations ne montrent pas de différence de densité de SNP pour des MAF supérieures à 0,05.

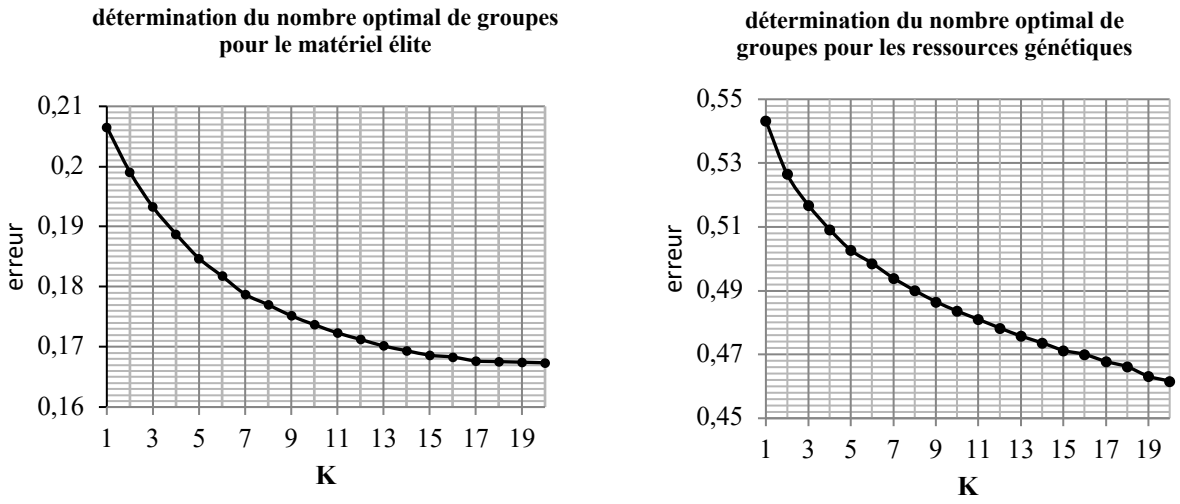


Figure 3 : Résultats des analyses d'Admixture avec K représentant le nombre de groupes génétiques ancestraux. L'erreur représente l'erreur de la validation croisée.

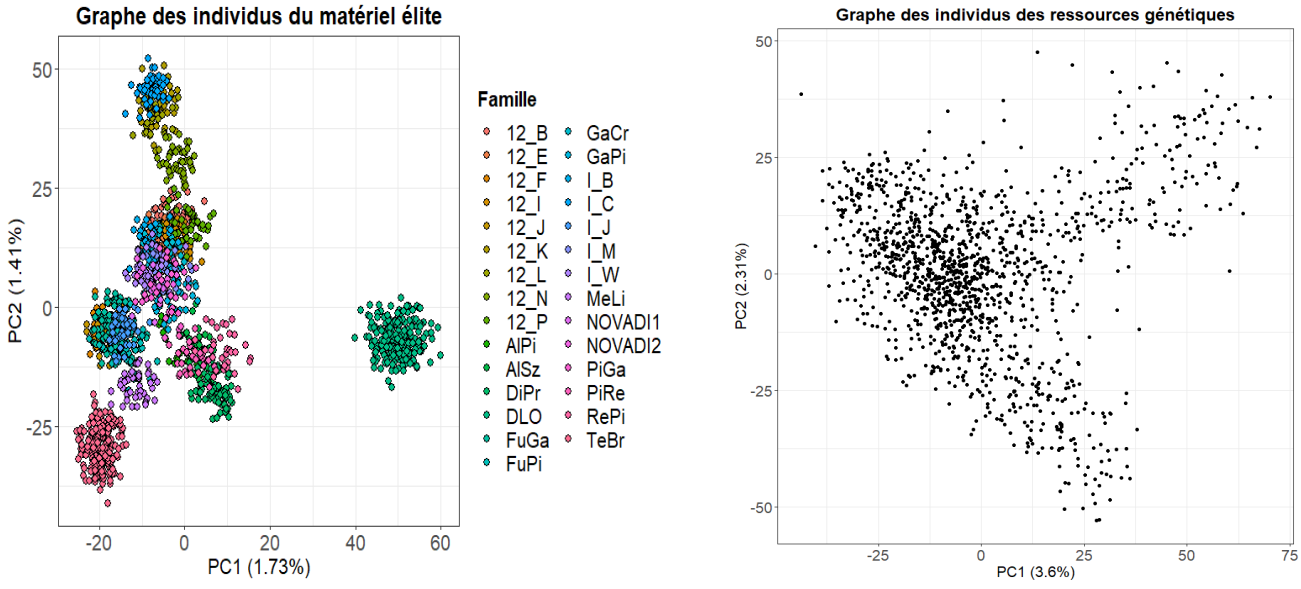


Figure 4 : ACP des individus du matériel élite et des ressources génétiques.

3.2 Structure des populations

- **Élagage des données génotypiques**

Pour le matériel élite le nombre de marqueurs SNP obtenu après élagage est de 27324 ce qui représente 9% du nombre total de marqueurs avant élagage. Par contre, un effectif de 12429 marqueurs SNP a été retenu après élagage sur les ressources génétiques ce qui correspond à 4% du nombre initial de marqueurs (voir Tableau 1).

Deux méthodes d'évaluation de la structure ont été appliquées:

- **Admixture**

L'analyse Admixture montre une décroissance continue de l'erreur de la validation croisée sur les deux populations pour le nombre de groupes K allant de 1 à 20 (figure 3) Par contre, les niveaux d'erreur de validation croisée sont plus faible pour le matériel élite (entre 0,16 et 0,21) que pour les ressources génétiques (entre 0,45 et 0,55). En plus à la différence des ressources génétiques, la diminution de l'erreur sur le matériel élite reste très faible à partir de K = 17. Ces résultats ne révèlent aucune structure dans les deux populations.

- **Analyse en composantes principales (ACP)**

Les deux principaux axes de l'ACP du matériel élite ont une inertie totale de 3,14%. L'axe 1 qui a une inertie de 1,73% fait apparaitre une structuration visible entre la famille DLO, qui comporte 210 individus; et les 28 autres familles restantes, qui comportent entre 9 et 192 individus. Ces dernières familles forment un sous-groupe peu homogène où certaines familles sont superposées avec d'autres (Figure 4). L'ACP réalisée sur les ressources génétiques a deux axes principaux d'une inertie totale 5,97% et le graphe obtenu ne montre aucune structuration de la population (Figure 4).

Tableau 2 : Résumé des estimations moyennes permettant d'apprécier la qualité de prédiction du modèle utilisé.

population	estimation	couleur		forme		acidité		croquant		jutosité		goût		date de récolte	
		moyenne	écart-type	moyenne	écart-type	moyenne	écart-type	moyenne	écart-type	moyenne	écart-type	moyenne	écart-type	moyenne	écart-type
ME	précision de prédiction	**0,70**	0,03	**0,42**	0,06	**0,46**	0,06	**0,34**	0,06	**0,50**	0,05	**0,23**	0,05	**0,73**	0,03
	Biais	1,00	0,07	1,00	0,20	1,00	0,17	1,02	0,25	1,00	0,14	1,01	0,31	1,01	0,06
	MSE	2,83	0,29	2,13	0,19	2,25	0,22	1,38	0,14	1,35	0,14	0,22	0,07	155,60	16,67
	Cor Rang	0,70	0,04	0,39	0,06	0,45	0,06	0,31	0,06	0,45	0,06	0,56	0,06	0,66	0,04
	Effectif	960		963		851		851		849		847		1017	
RG	précision de prédiction	**0,60**	0,04	**0,24**	0,06	**0,44**	0,05	**0,35**	0,05	**0,40**	0,05	**0,41**	0,06	**0,82**	0,02
	Biais	1,00	0,10	1,07	0,38	0,96	0,15	1,00	0,20	0,98	0,17	1,03	0,21	1,01	0,05
	MSE	3,16	0,26	9,01	0,79	1,65	0,15	1,16	0,11	1,02	0,08	1,08	0,12	242,90	24,14
	Cor Rang	0,61	0,04	0,23	0,06	0,41	0,05	0,34	0,05	0,38	0,05	0,42	0,06	0,81	0,02
	Effectif	878		766		1178		1159		1183		912		1025	

3.3 Précision de prédiction :

Le tableau 3 présente les résultats des précisions de prédiction moyennes obtenus par validation croisée sur les sept traits étudiés. Ces précisions sont différentes d'une population à une autre et d'un caractère à un autre.

L'amplitude de variation pour le matériel élite et pour les ressources génétiques, des 7 caractères évalués est 0,23 à 0,82.

La couleur, la jutosité et la forme sont les 3 caractères pour lesquels la précision est plus grande pour le matériel élite que pour les ressources génétiques, avec un écart variant entre 0,1 et 0,18.

L'acidité et le croquant sont les 2 caractères pour lesquels les précisions sont presque de même niveau pour les deux types de populations.

La date de récolte et le goût sont les 2 caractères pour lesquels la précision est plus grande pour les ressources génétiques que pour le matériel élite, avec un écart de 0,09 à 0,18.

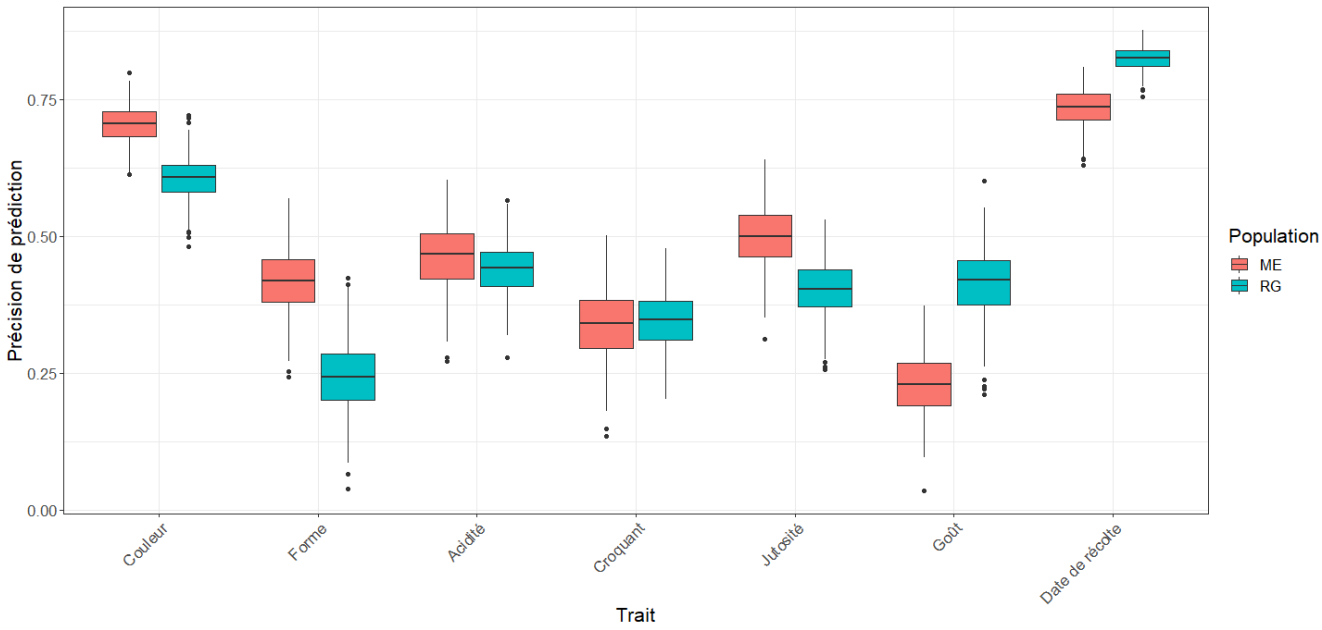


Figure 5 : Distribution des 500 précisions de prédiction de chaque caractère du matériel élite (ME) et des ressources génétiques (RG).

Les meilleures précisions de prédiction ont été observées pour les caractères couleur de fruit et date de récolte dans les deux populations. Les plus faibles précisions de prédiction ont été obtenues sur le caractère goût dans le matériel élite et le caractère forme de fruit dans les ressources génétiques.

On observe une faible variation (avec des écarts-types inférieurs à 0,07) des 500 précisions de prédiction calculées sur chaque caractère et ayant permis d'obtenir les précisions de prédiction moyennes. Les biais de prédiction des valeurs phénotypiques observées par les valeurs prédites sont très faibles car ils sont tous compris entre 0,96 et 1,07.

Le coefficient de détermination R^2 des précisions de prédiction moyennes entre le matériel élite et les ressources génétiques est 0,62 (figure 5).

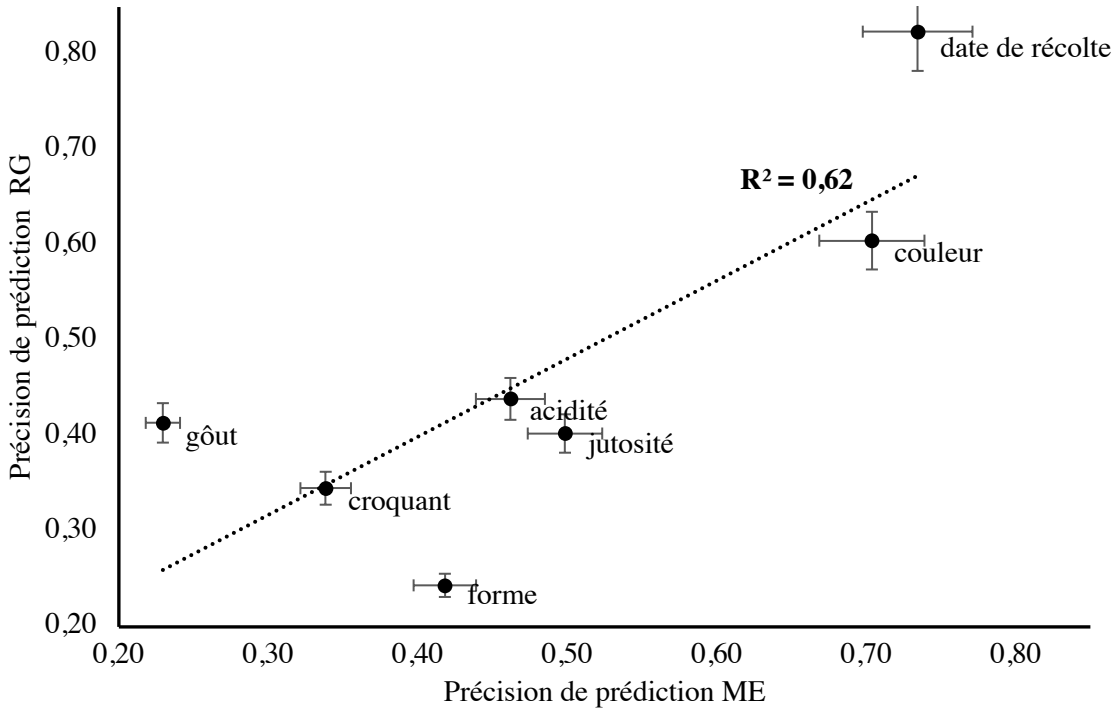


Figure 6 : Corrélation entre les précisions de prédiction moyennes du matériel élite et des ressources génétiques. Les barres d'erreurs des précisions de prédiction moyennes du matériel élite sont représentées horizontalement et celles des ressources génétiques sont représentées verticalement.

3.4 Corrélation entre précision de prédiction et effectif d'individus

Les effectifs des individus ayant permis d'effectuer les prédictions sont différents entre populations et entre caractères. Tous les effectifs utilisés sont compris entre 766 et 1025 génotypes. Le R^2 entre les précisions de prédiction moyennes et les effectifs utilisés par caractère est de 0,56 pour le matériel élite alors qu'elle est de 0,005 pour les ressources génétiques.

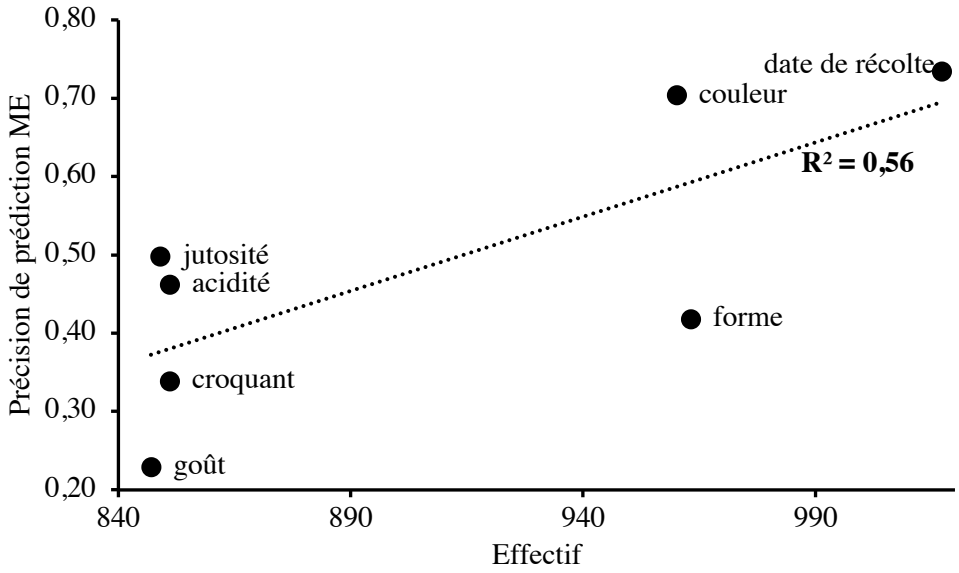


Figure 7 : Corrélation entre les précisions de prédiction moyennes du matériel élite et les effectifs d'individus ayant servi à la prédiction des valeurs phénotypiques du caractère.

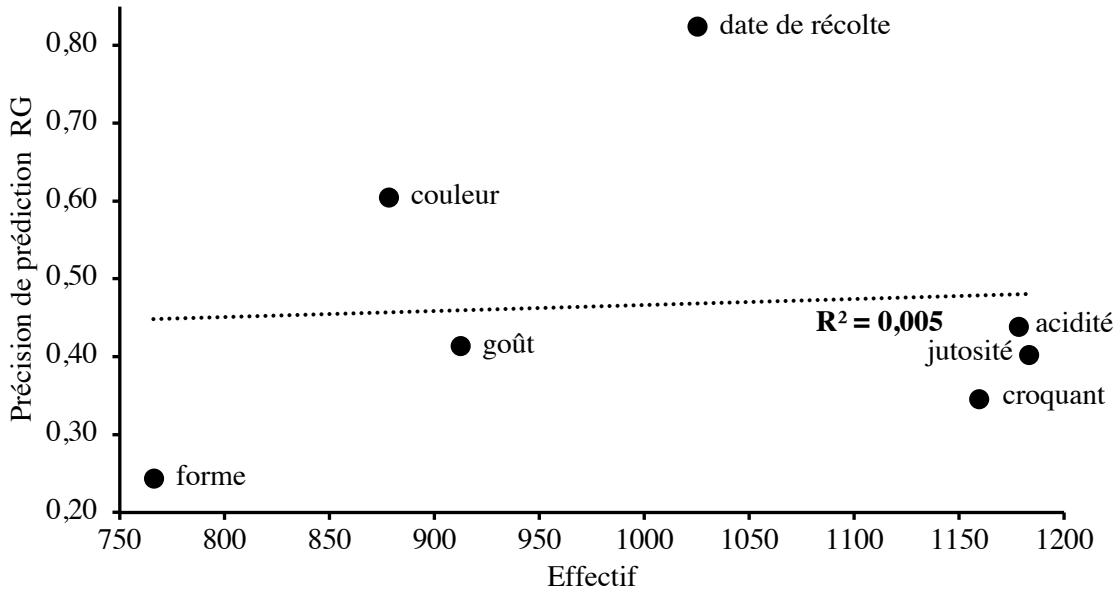


Figure 8 : Corrélation entre les précisions de prédiction moyennes des ressources génétiques et les effectifs d'individus ayant servi à la prédiction des valeurs phénotypiques du caractère.

4 Discussion

L'évaluation de la précision de prédiction du matériel élite et des ressources génétiques a été menée en deux étapes : la première a été consacrée à l'étude de la diversité et de la différenciation entre les deux populations pour mieux analyser et interpréter les résultats de prédictions obtenus en seconde étape.

4.1 Diversité et différenciation génétique

L'élagage sur le déséquilibre de liaison, visant à ne retenir que les marqueurs en faible déséquilibre de liaison pour les analyses de structure, a abouti à retenir environ deux fois plus de marqueurs sur le matériel élite que sur les ressources génétiques. On peut en conclure que les déséquilibres de liaison entre marqueurs sont moins importants sur le matériel élite que sur les ressources génétiques. Ceci pourrait être dû aux quelques générations (environ sept) qui séparent le matériel élite des ressources génétiques car à chaque génération d'intercroisement le niveau de déséquilibre de liaison est réduit. D'après Mackay et Powell, (2007) une série de recombinaisons diminue le déséquilibre de liaison des allèles non étroitement liés, même si la sélection directionnelle modifie les fréquences des allèles étroitement liés et génère des déséquilibres de liaison entre les marqueurs autour des loci sélectionnés dans le matériel élite.

Le taux d'hétérozygotie observée est plus faible pour le matériel élite ($H_o = 0,25$) par rapport à celui des ressources génétiques (Ho = 0,34) peut être le résultat de la sélection. Ces taux d'hétérozygotie observée obtenus au sein des deux populations sont similaires au taux d'hétérozygotie observée (H_o=0,35) dans une collection de ressources génétiques par Kumar et al. (2014) en utilisant une puce 20K. Malgré l'existence d'un écart de fréquences génotypiques entre le matériel élite et les ressources génétiques, les deux populations sont presque similaires ($F_{st} = 0,004$).

Malgré la provenance diverse des ressources génétiques, aucune structuration basée sur leurs environnements d'origine n'a été trouvée, ce qui confirme les résultats d'autres travaux. Urrestarazu et al. (2016) ont montré que les ressources génétiques de la pomme au niveau européen présentent une variation génétique importante mais une structure de population limitée. Par contre le matériel élite présenterait une structuration par famille avec des différences faibles entre elles à l'exception de la famille DLO (Figure 4).

Par ailleurs l'écart entre les intervalles de MAF les plus fréquentes existant entre les deux populations illustrerait que la sélection effectuée chez le matériel élite à partir d'un nombre limité de fondateurs (environ 40) a causé une perte de diversité par dérive génétique. Et si cette différence de MAF entre

Tableau 3 : Facteurs pouvant faire varier nos précisions de prédiction

Facteurs pouvant faire varier la précision de prédiction des caractères phénotypiques du matériel élite et des ressources génétiques
• **l'héritabilité du caractère ;** • **caractère polygénique ou non du trait phénotypique ;** • le degré d'apparentement des individus ; • **la distribution phénotypique** ; • **l'interaction entre génotype et environnement pour le caractère phénotypique**; • **l'effectif des individus** ; • patrimoines génétiques et culturels du notateur ; • l'imputation des données génotypique ; • les fréquences alléliques ; • le modèle de prédiction génomique.

***en gras :** les facteurs confirmés par nos résultats

population est confirmée, il serait pertinent de rechercher les gènes candidats par rapport aux QTLs connus dans chaque région. En plus on suspecte qu'il y a des allèles (et/ou haplotypes) présents dans le matériel élite et pas dans les ressources génétiques par exemple le gène *Rvi6/Vf* de résistance à la tavelure qui se trouve sur le chromosome 1 (Vinatzer et al., 2004). Il serait donc intéressant de chercher ces allèles pour expliquer les différences entre populations. Il peut aussi être intéressant de réintroduire de la diversité à partir des ressources génétiques à certains endroits du génome du matériel élite surtout là où de nombreux allèles sont fixés ou quasi-fixés.

4.2 Analyse et interprétation des précisions de prédiction

Les résultats de précision de prédiction obtenus varient de 0,23 à 0,82 et ceci peut être expliqué par différents facteurs liés à chaque caractère. Les plus importantes précisions de prédiction ont été obtenues sur le trait date de récolte dans les deux populations (ME = 0,73, RG = 0,82). Ces résultats sont similaires à ceux de Jung et al. (submitted), qui ont trouvé une précision de prédiction de 0,75 sur le même caractère avec une population de 534 génotypes de pommiers comprenant à la fois des ressources génétiques et des génotypes issus de familles de programmes d'amélioration du pommier européens. Les auteurs justifient leur résultat par la forte l'héritabilité du caractère qui est de 0,96 et du faible effet de l'interaction entre le génotype et l'environnement pour la date de récolte.

Les caractères phénotypiques peuvent être sous le contrôle de plusieurs loci ou sous l'influence d'un QTL majeur, alors que la prédiction génomique est plus efficace pour les caractères polygéniques (Heffner et al., 2009). La revue de Kumar et al. (2012) sur la sélection génomique du pommier indique que des QTL influencent l'expression des caractères croquant, jutosité, goût et acidité. Mais seule la dépendance de l'acidité à un QTL majeur a été confirmée sur plusieurs environnement et par plusieurs études (Nybom, 1959; Maliepaard et al., 1998; Liebhard et al., 2003). Nos résultats de prédiction sur ces caractères vont dans le sens de confirmer cette hypothèse car les précisions de ces caractères sont compris entre 0,23 et 0,50. Comme le niveau d'expression des QTLs majeurs peut varier en fonction des environnements, le fait que les génotypes ont été phénotypés dans différents environnements peut expliquer les quelques niveaux de prédictions moyens de ces caractères. Il aurait été possible d'obtenir de meilleures précisions sur ces caractères avec des modèles Bayesiens qui attribuent aux marqueurs des effets différents.

Les données phénotypiques (BLUPs, moyennes ajustées, une seule valeur par génotype et par caractère) utilisées lors de cette étude n'ont pas permis de calculer les héritabilités des différents caractères évalués. L'étude de Kouassi et al. (2009) sur l'estimation des paramètres génétiques et la prédiction des valeurs additives était basée sur une grande partie du matériel élite utilisé et avait permis de calculer les héritabilités au sens strict de différents caractères : le croquant ($h^2 = 0,16$), le goût ($h^2 = 0,31$), la jutosité

($h^2 = 0,35$), la couleur ($h^2 = 0,55$) et l'acidité ($h^2 = 0,63$). Les caractères les moins héritables comme le croquant et le goût ont obtenu les plus faibles précisions de prédiction sur le matériel élite. A cela s'ajoute une distribution non symétrique des valeurs phénotypiques du caractère goût dans le matériel élite contrairement à la distribution observée dans les ressources génétiques (Annexe 1a/b). Ceci permet de confirmer que les précisons de prédiction des caractères phénotypiques du pommier sont fortement influencées par leur héritabilité et leur distribution phénotypique (Muranty et al., 2015b). Les précisions de prédiction des ressources génétiques ont été parfois meilleures que celles du matériel élite dont l'apparentement entre génotypes est plus fort. Cette situation pourrait être due aux différences de distributions phénotypiques entre les deux populations (Annexe 1a/b).

Chez les ressources génétiques on observe une absence de corrélation ($R^2 = 0,005$) entre les précisions de prédiction et les effectifs des individus utilisés. Par contre cette corrélation ($R^2 = 0,56$) est plus forte avec le matériel élite. Ces résultats pourraient illustrer, qu'avec le degré d'apparentement des individus du matériel élite, l'augmentation de l'effectif des individus permettrait d'améliorer la précision de prédiction. Nos résultats confirment les hypothèses d'autres études de sélection génomique. Par exemple l'évaluation de la prédiction génomique avec une population de 1120 génotypes de pommier provenant seulement de 6 parents a permis d'obtenir des précisions variant de 0,70 à 0,90 pour 6 caractères de qualité du fruit (Kumar et al., 2012b). Les individus ayant permis d'obtenir ces niveaux élevés de prédiction ont été seulement génotypés avec une puce 8K. Ainsi dans le cas de notre étude la densité de marqueurs à utiliser pouvait être appréciée en fonction du niveau d'apparentement des individus qui constituent notre population. C'est-à-dire on aurait pu utiliser moins de marqueurs pour la prédiction du matériel élite qui présente des individus plus apparentés et une densité importante de marqueurs (303239 SNPs) pour la prédiction des ressources génétiques. Par ailleurs une prédiction du matériel élite avec les données génotypiques non imputées (celles obtenues avec la puce 20K) permettrait d'évaluer l'apport de l'augmentation du nombre de marqueurs sur la précision.

Conclusion

Cette étude a permis de décrire la diversité et la différenciation du matériel élite et des ressources génétiques et d'évaluer les niveaux de précision de prédiction des caractères phénotypiques de ces populations. Le matériel élite et les ressources génétiques présentent une très faible différenciation (F_{st}= 0,004). Une structuration faible des familles a été obtenue chez le matériel élite alors qu'avec les ressources génétiques aucune structuration n'a été observée. Une perte de diversité par dérive génétique et par sélection a été constatée chez le matériel élite par rapport aux ressources génétiques.
L'amplitude de variation des précisions de prédiction de tous les caractères est de 0,23 à 0,82. Sur les sept caractères phénotypiques évalués sur chaque population, trois des meilleures précisions de prédiction ont été enregistré avec le matériel élite et deux meilleures précisions ont été obtenues avec les ressources génétiques. Les précisons des deux caractères restants était presque similaires pour les deux populations. Les faibles à moyennes précisions de prédiction obtenues sur cinq caractères fait qu'il s'avère intéressant d'améliorer la précision en augmentant par exemple le nombre d'individus.
Pour bien estimer l'effet des marqueurs sur une population d'entrainement, il faut un effectif suffisant d'individus, et donc un grand nombre d'haplotypes, afin d'obtenir des précisions de prédiction élevées. Si nos résultats se confirment, et que le nombre d'individus phénotypés du matériel élite est faible au point de ne pas pouvoir assurer une évaluation intra-population correcte, la combinaison de populations pourrait être une stratégie intéressante à appliquer. Par ailleurs, elle est considérée être une solution pour calculer efficacement les GEBV des candidats issus des croisements entre races chez les bovins, car le déséquilibre de liaison entre marqueurs et QTL de la population d'entrainement serait encore conservé dans la population de validation (Karoui et al., 2012; Zhou et al., 2014). Cette méthode pourrait être utilisée chez les végétaux pour prédire efficacement les GEBV des génotypes issus de croisement entre populations ou entre groupes hétérotiques différents. C'est le cas des génotypes qui seront issus du croisement entre le matériel élite et les ressources génétiques comme prévu dans le cadre du projet dans lequel est inscrite cette étude.

D'après de Roos et al. (2009) la combinaison de populations serait aussi un bon recours pour améliorer la précision de la prédiction des caractères à faible héritabilité si elle est utilisée avec une densité relativement importante de marqueurs. Une augmentation du nombre d'individus constituant la population d'entrainement du matériel élite pourrait donc améliorer la précision de prédiction des caractères comme le croquant, le goût, la jutosité qui présentent des héritabilités faibles.

Cette méthode de combinaison de populations peut nécessiter d'adapter les modèles de prédiction utilisés pour qu'ils prennent en compte d'éventuelles différences qui pourraient exister entre

populations, comme les différences de déséquilibre de liaison et des fréquences alléliques entre populations. Dans le cas où la combinaison de populations serait appliquée entre le matériel élite et les ressources génétique il serait nécessaire d'étudier de façon comparative les déséquilibres de liaison des deux populations.

Contribution de l'étudiant

Au début du stage, j'ai effectué la description d'une partie des données phénotypiques et génotypiques mises à ma disposition. Par la suite j'ai évalué la diversité et la différenciation des deux populations étudiées. J'ai eu à faire toutes les analyses d'élagage et d'évaluation de la structure en me servant des logiciels Plink 1.9 et Admixture1.3 installés sur serveur. J'ai écrit les scripts R. J'ai utilisé ceux qui étaient destinées à faire la prédiction en faisant varier les données d'entrées et les nombres de répétitions. Par la suite j'ai lancé toutes les analyses de prédiction génomique sur serveur. Au final j'ai eu à organiser et représenter graphiquement les résultats obtenus. Toutes ces étapes ayant permis à l'aboutissement de l'étude ont été encadrées par mes maîtres de stage qui ont contribué à l'amélioration et parfois à l'orientation des analyses.

Références bibliographiques

Alexander, D.H. & Novembre, J., 2015. Admixture 1 . 3 Software Manual.

Canty Angelo and Brian Ripley (2020). boot: Bootstrap R (S-Plus) Functions. R package version 1.3 25.

Bianco, L., Cestaro, A., Linsmith, G., Muranty, H., Denancé, C., Théron, A., Poncet, C., Micheletti, D., Kerschbamer, E., Di Pierro, E.A., Larger, S., Pindo, M., Van de Weg, E., Davassi, A., Laurens, F., Velasco, R., Durel, C.-E., Troggio, M., 2016. Development and validation of the Axiom ® Apple480K SNP genotyping array. Plant J 86, 62–74.

Bianco, L., Cestaro, A., Sargent, D.J., Banchi, E., Derdak, S., Di Guardo, M., Salvi, S., Jansen, J., Viola, R., Gut, I., Laurens, F., Chagné, D., Velasco, R., van de Weg, E., Troggio, M., 2014. Development and Validation of a 20K Single Nucleotide Polymorphism (SNP) Whole Genome Genotyping Array for Apple (Malus × domestica Borkh). PLoS ONE 9, e110377.

Bramel, P., Volk, G.M., n.d. A global strategy for the conservation and use of apple genetic resources 52.

Calus, M.P.L., Meuwissen, T.H.E., de Roos, A.P.W., Veerkamp, R.F., 2008. Accuracy of Genomic Selection Using Different Methods to Define Haplotypes. Genetics 178, 553–561.

Clark, S.A., Hickey, J.M., Daetwyler, H.D., van der Werf, J.H., 2012. The importance of information on relatives for the prediction of genomic breeding values and the implications for the makeup of reference data sets in livestock breeding schemes. Genet Sel Evol 44, 4.

Cornille, A., Antolín, F., Garcia, E., Vernesi, C., Fietta, A., Brinkkemper, O., Kirleis, W., Schlumbaum, A., Roldán-Ruiz, I., 2019. A Multifaceted Overview of Apple Tree Domestication. Trends in Plant Science 24, 770–782.

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

Cornille, A., Gladieux, P., Smulders, M.J.M., Roldán-Ruiz, I., Laurens, F., Le Cam, B., Nersesyan, A., Clavel, J., Olonova, M., Feugey, L., Gabrielyan, I., Zhang, X.-G., Tenaillon, M.I., Giraud, T., 2012. New Insight into the History of Domesticated Apple: Secondary Contribution of the European Wild Apple to the Genome of Cultivated Varieties. PLoS Genet 8, e1002703.

Daccord, N., Celton, J.-M., Linsmith, G., Becker, C., Choisne, N., Schijlen, E., van de Geest, H., Bianco, L., Micheletti, D., Velasco, R., Di Pierro, E.A., Gouzy, J., Rees, D.J.G., Guérif, P., Muranty, H., Durel, C.-E., Laurens, F., Lespinasse, Y., Gaillard, S., Aubourg, S., Quesneville, H., Weigel, D., van de Weg, E., Troggio, M., Bucher, E., 2017. High-quality de novo assembly of the apple genome and methylome dynamics of early fruit development. Nat Genet 49, 1099–1106.

de Roos, A.P.W., Hayes, B.J., Goddard, M.E., 2009. Reliability of Genomic Predictions Across Multiple Populations. Genetics 183, 1545–1553.

Desta, Z.A., Ortiz, R., 2014. Genomic selection: genome-wide prediction in plant improvement. Trends in Plant Science 19, 592–601.

Edwards, S.M., Buntjer, J.B., Jackson, R., Bentley, A.R., Lage, J., Byrne, E., Burt, C., Jack, P., Berry, S., Flatman, E., Poupard, B., Smith, S., Hayes, C., Gaynor, R.C., Gorjanc, G., Howell, P., Ober, E., Mackay, I.J., Hickey, J.M., 2019. The effects of training population design on genomic prediction accuracy in wheat. Theor Appl Genet.

Endelman, J.B., 2011. Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. The Plant Genome 4, 250–255.

Falconer, D.S., Mackay, T.F.C., 2009. Introduction to quantitative genetics, 4. ed., [16. print.]. ed. Pearson, Prentice Hall, Harlow.

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

Goddard, M.E., Hayes, B.J., 2007. Genomic selection: Genomic selection. Journal of Animal Breeding and Genetics 124, 323–330.

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, 1770–1779.

H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

Harris, S.A., Robinson, J.P., Juniper, B.E., 2002. Genetic clues to the origin of the apple. Trends in Genetics 18, 426–430.

Hayes, Ben J, Bowman, P.J., Chamberlain, A.C., Verbyla, K., Goddard, M.E., 2009a. Accuracy of genomic breeding values in multi-breed dairy cattle populations. Genet Sel Evol 41, 51.

Hayes, B.J., Bowman, P.J., Chamberlain, A.J., Goddard, M.E., 2009b. Invited review: Genomic selection in dairy cattle: Progress and challenges. Journal of Dairy Science 92, 433–443.

Heffner, E.L., Sorrells, M.E., Jannink, J.-L., 2009. Genomic Selection for Crop Improvement. Crop Sci. 49, 1–12.

Henderson, C.R., 1975. Best linear unbiased estimation and prediction under a selection model. Biometrics 31, 423–447.

Jung M., Roth M., Aranzana M. J., Auwerkerken A, Bink M., Denancé C., Dujak C., Durel C-E., Font i F. C., Cantin C M., Guerra W, Howard N. P., Lewandowski M., Ordidge M., Rymenants M., Sanin N., Studer B., Zurawicz E., Laurens F., Patocchi A., Muranty H., "The apple REFPOP a reference population for genomics-assisted breeding in apple" (submitted)

Karoui, S., Carabaño, M.J., Díaz, C., Legarra, A., 2012. Joint genomic evaluation of French dairy cattle breeds using multiple-trait models. Genet Sel Evol 44, 39.

Kole, C. (Ed.), 2011. Wild Crop Relatives: Genomic and Breeding Resources. Springer Berlin Heidelberg, Berlin, Heidelberg.

Kouassi, A.B., Durel, C.-E., Costa, F., Tartarini, S., van de Weg, E., Evans, K., Fernandez-Fernandez, F., Govan, C., Boudichevskaja, A., Dunemann, F., Antofie, A., Lateur, M., Stankiewicz-Kosyl, M., Soska, A., Tomala, K., Lewandowski, M., Rutkovski, K., Zurawicz, E., Guerra, W., Laurens, F., 2009. Estimation of genetic parameters and prediction of breeding values for apple fruit-quality traits using pedigreed plant material in Europe. Tree Genetics & Genomes 5, 659–672.

Kumar, S., Chagné, D., Bink, M.C.A.M., Volz, R.K., Whitworth, C., Carlisle, C., 2012a. Genomic Selection for Fruit Quality Traits in Apple (Malus×domestica Borkh.). PLoS ONE 7, e36674.

Kumar, S., Bink, M.C.A.M., Volz, R.K., Bus, V.G.M., Chagné, D., 2012b. Towards genomic selection in apple (Malus × domestica Borkh.) breeding programmes: Prospects, challenges and strategies. Tree Genetics & Genomes 8, 1–14.

Kumar, S., Raulier, P., Chagné, D., Whitworth, C., 2014. Molecular-level and trait-level differentiation between the cultivated apple (*Malus* × *domestica* Borkh.) and its main progenitor *Malus sieversii*. Plant Genet. Res. 12, 330–340.

Laurens, F., Aranzana, M.J., Arus, P., Bassi, D., Bink, M., Bonany, J., Caprera, A., Corelli-Grappadelli, L., Costes, E., Durel, C.-E., Mauroux, J.-B., Muranty, H., Nazzicari, N., Pascal, T., Patocchi, A., Peil, A., Quilot-Turion, B., Rossini, L., Stella, A., Troggio, M., Velasco, R., van de Weg, E., 2018. An integrated approach for increasing breeding efficiency in apple and peach in Europe. Hortic Res 5, 11.

Lassois, L., Denancé, C., Ravon, E., Guyader, A., Guisnel, R., Hibrand-Saint-Oyant, L., Poncet, C., Lasserre-Zuber, P., Feugey, L., Durel, C.-E., 2016. Genetic Diversity, Population Structure, Parentage Analysis, and Construction of Core Collections in the French Apple Germplasm Based on SSR Markers. Plant Mol Biol Rep 34, 827–844.

Liebhard, R., Kellerhals, M., Pfammatter, W., Jertmini, M., Gessler, C., 2003. Mapping quantitative physiological traits in apple (Malus × domestica Borkh.). Plant Mol Biol 52, 511–526.

Ma, B., Liao, L., Peng, Q., Fang, T., Zhou, H., Korban, S.S., Han, Y., 2017. Reduced representation genome sequencing reveals patterns of genetic diversity and selection in apple: Genetic diversity and selection in apple. J. Integr. Plant Biol. 59, 190–204.

Mackay, I., Powell, W., 2007. Methods for linkage disequilibrium mapping in crops. Trends in Plant Science 12, 57–63.

Maliepaard, C., Alston, F.H., van Arkel, G., Brown, L.M., Chevreau, E., Dunemann, F., Evans, K.M., Gardiner, S., Guilford, P., van Heusden, A.W., Janse, J., Laurens, F., Lynn, J.R., Manganaris, A.G., den Nijs, A.P.M., Periam, N., Rikkerink, E., Roche, P., Ryder, C., Sansavini, S., Schmidt, H., Tartarini, S., Verhaegh, J.J., Vrielink-van Ginkel, M., King, G.J., 1998. Aligning male and female linkage maps of apple (Malus pumila Mill.) using multi-allelic markers: Theor Appl Genet 97, 60–73.

Meuwissen, T.H.E., Hayes, B.J., Goddard, M.E., 2001. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps 11.

Muranty, H., Jorge, V., Bastien, C., Lepoittevin, C., Bouffier, L., Sanchez, L., 2014. Potential for marker-assisted selection for forest tree breeding: lessons from 20 years of MAS in crops. Tree Genetics & Genomes 10, 1491–1510.

Muranty, H., Troggio, M., Sadok, I.B., Rifaï, M.A., Auwerkerken, A., Banchi, E., Velasco, R., Stevanato, P., van de Weg, W.E., Di Guardo, M., Kumar, S., Laurens, F., Bink, M.C.A.M., 2015b. Accuracy and responses of genomic selection on key traits in apple breeding. Hortic Res 2, 15060.

Noiton, D., Shelbourne, C.J.A., 1992. Quantitative genetics in an apple breeding strategy Euphytica 60:213-219.

Noiton, D., Alspach P., 1996. Founding Clones Imbreending, Coanscestry, and Status Number of Modern Apple Cultivars. J. Amer. Soc. Hort. Sci 121(5):773-782. 1996

Nybom, N., 1959. On the Inheritance of Acidity in Cultivated Apples. Hereditas 45, 332–350.

Pszczola, M., Strabel, T., Mulder, H.A., Calus, M.P.L., 2012. Reliability of direct genomic values for animals with different relationships within and to the reference population. Journal of Dairy Science 95, 389–400.

Rapport Agreste « synthèses conjoncturelle » N° 356 du mai 2020

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Urrestarazu, J., Denancé, C., Ravon, E., Guyader, A., Guisnel, R., Feugey, L., Poncet, C., Lateur, M., Houben, P., Ordidge, M., Fernandez-Fernandez, F., Evans, K.M., Paprstein, F., Sedlak, J.,

Nybom, H., Garkava-Gustavsson, L., Miranda, C., Gassmann, J., Kellerhals, M., Suprun, I., Pikunova, A.V., Krasova, N.G., Torutaeva, E., Dondini, L., Tartarini, S., Laurens, F., Durel, C.-E., 2016. Analysis of the genetic diversity and structure across a wide range of germplasm reveals prominent gene flow in apple at the European level. BMC Plant Biol 16, 130.

van den Berg, S., Calus, M.P.L., Meuwissen, T.H.E., Wientjes, Y.C.J., 2015. Across population genomic prediction scenarios in which Bayesian variable selection outperforms GBLUP. BMC Genet 16, 146.

VanRaden, P.M., 2008. Efficient Methods to Compute Genomic Predictions. Journal of Dairy Science 91, 4414–4423.

Velasco, R., Zharkikh, A., Affourtit, J., Dhingra, A., Cestaro, A., Kalyanaraman, A., Fontana, P., Bhatnagar, S.K., Troggio, M., Pruss, D., Salvi, S., Pindo, M., Baldi, P., Castelletti, S., Cavaiuolo, M., Coppola, G., Costa, F., Cova, V., Dal Ri, A., Goremykin, V., Komjanc, M., Longhi, S., Magnago, P., Malacarne, G., Malnoy, M., Micheletti, D., Moretto, M., Perazzolli, M., Si-Ammour, A., Vezzulli, S., Zini, E., Eldredge, G., Fitzgerald, L.M., Gutin, N., Lanchbury, J., Macalma, T., Mitchell, J.T., Reid, J., Wardell, B., Kodira, C., Chen, Z., Desany, B., Niazi, F., Palmer, M., Koepke, T., Jiwan, D., Schaeffer, S., Krishnan, V., Wu, C., Chu, V.T., King, S.T., Vick, J., Tao, Q., Mraz, A., Stormo, A., Stormo, K., Bogden, R., Ederle, D., Stella, A., Vecchietti, A., Kater, M.M., Masiero, S., Lasserre, P., Lespinasse, Y., Allan, A.C., Bus, V., Chagné, D., Crowhurst, R.N., Gleave, A.P., Lavezzo, E., Fawcett, J.A., Proost, S., Rouzé, P., Sterck, L., Toppo, S., Lazzari, B., Hellens, R.P., Durel, C.-E., Gutin, A., Bumgarner, R.E., Gardiner, S.E., Skolnick, M., Egholm, M., Van de Peer, Y., Salamini, F., Viola, R., 2010. The genome of the domesticated apple (Malus × domestica Borkh.). Nat Genet 42, 833–839.

Vinatzer, B.A., Patocchi, A., Tartarini, S., Gianfranceschi, L., Sansavini, S., Gessler, C., 2004. Isolation of two microsatellite markers from BAC clones of the Vf scab resistance region and molecular characterization of scab-resistant accessions in Malus germplasm*. Plant Breeding 123, 321–326.

Wani, A.A., Dar, J.A., Bhat, T.A., 2015. Malus × domestica Borkh. - from wild resources to present day cultivated apple 13.

Whittaker, J.C., Thompson, R., Denham, M.C., 2000. Marker-assisted selection using ridge regression. Genet. Res. 75, 249–252.

Zhou, L., Heringstad, B., Su, G., Guldbrandtsen, B., Meuwissen, T.H.E., Svendsen, M., Grove, H., Nielsen, U.S., Lund, M.S., 2014. Genomic predictions based on a joint reference population for the Nordic Red cattle breeds. Journal of Dairy Science 97, 4485–4496.

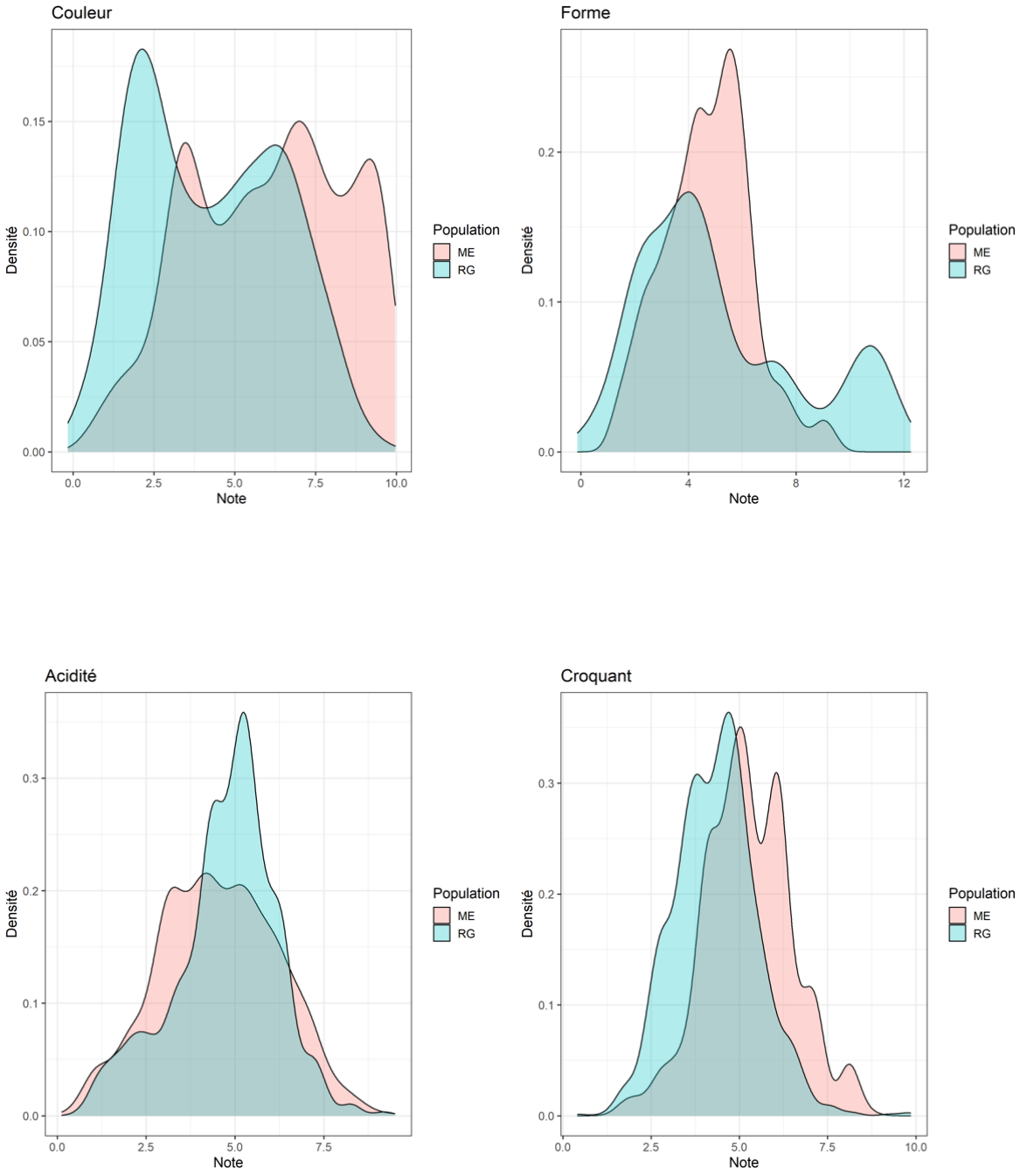
Sitographie

http://www.fao.org/faostat/fr/#data consulté le 23 mai 2020

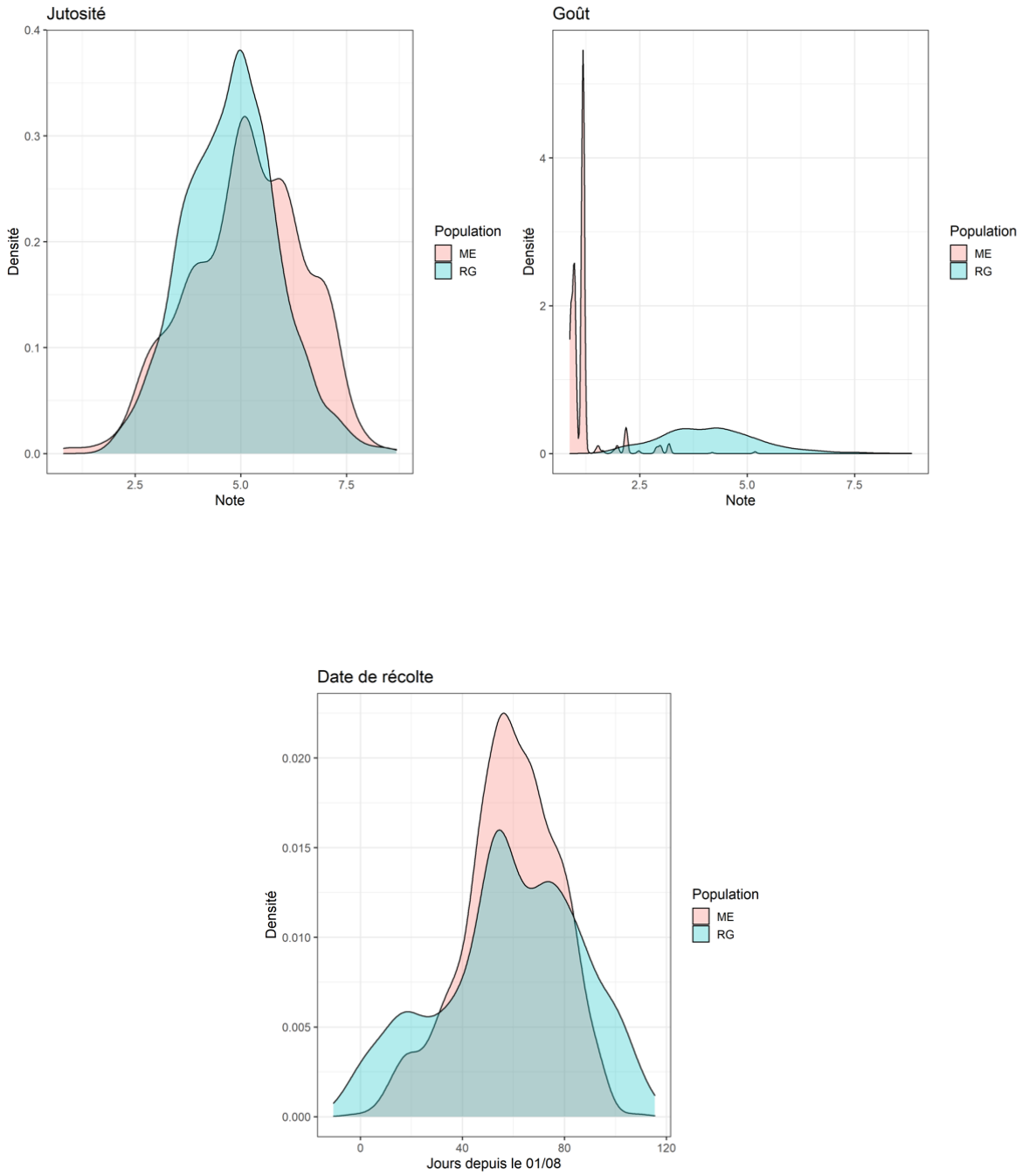
http://www.wapa-association.org/docs/2019/European_summary_reduced.pdf consulté le 25 mai 2020

Annexe

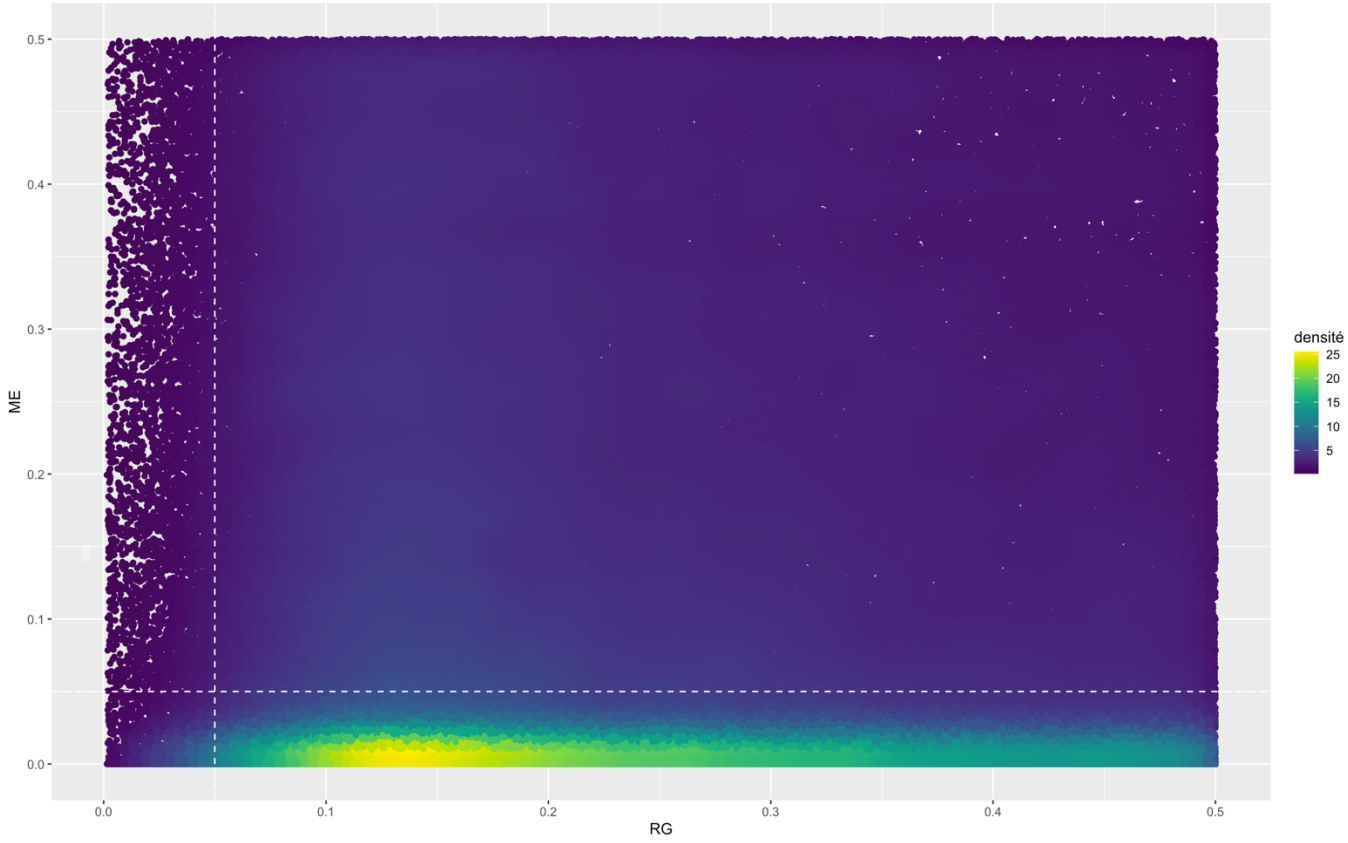
Annexe 1a : Distribution des valeurs phénotypiques des caractères



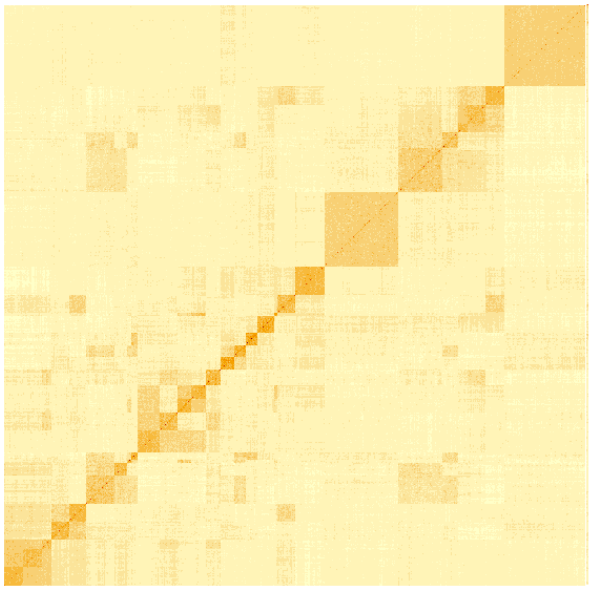
Annexe 1b: Distribution des valeurs phénotypiques des caractères



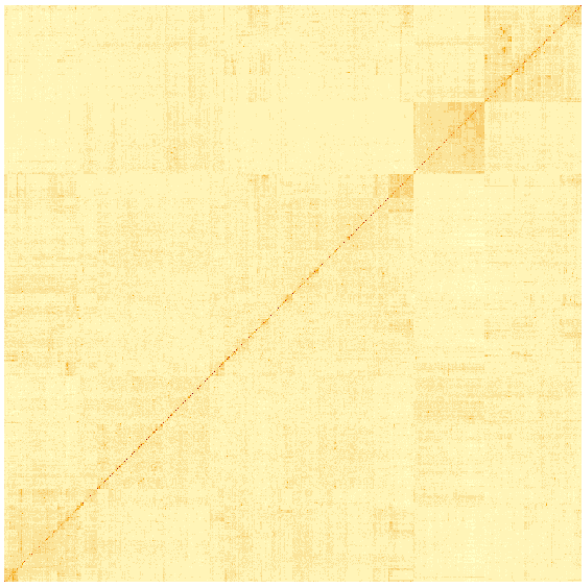
Annexe 2: Densités de SNPs selon leur MAF dans le matériel élite (ME) et dans les ressources génétiques (RG) sur tout le génome



Annexe 3: heatmaps des matrices d'apparentement génomique des deux populations



Matériel élite



Ressources génétiques

Annexe 4: Effectifs des individus du matériel élite par famille

	Famille	Effectif
1	12_B	47
2	12_E	57
3	12_F	48
4	12_I	46
5	12_J	23
6	12_K	47
7	12_L	25
8	12_N	45
9	12_P	46
10	AlPi	18
11	AlSz	32
12	DiPr	75
13	DLO	210
14	FuGa	115
15	FuPi	42

	Famille	Effectif
16	GaCr	32
17	GaPi	43
18	I_B	39
19	I_C	49
20	I_J	48
21	I_M	36
22	I_W	36
23	MeLi	43
24	NOVADI1	9
25	NOVADI2	10
26	PiGa	31
27	PiRe	30
28	RePi	35
29	TeBr	192

Résumé

La culture européenne de pomme est basée sur un petit nombre de variétés dites élites. Ces variétés deviennent de plus en plus vulnérables dans le contexte actuel de changement climatique car elles présentent peu de diversité et une base génétique étroite. Le transfert d'allèles à partir des ressources génétiques vers le matériel élite par la sélection génomique pourrait permettre de résoudre ce problème. Cette étude vise à évaluer la précision de la prédiction génomique du matériel élite et des ressources génétiques. Nous disposons de données génotypiques à haute densité (303239 SNPs) pour des individus représentant du matériel élite et des ressources génétiques en effectifs variant de 847 à 1025. La caractérisation des deux populations par les données génotypiques a montré une différentiation très faible (F_{st} = 0,004) entre le matériel élite et les ressources génétiques. Par contre des différences de fréquence d'allèles minoritaires ont été observées sur tout le long du génome entre les deux populations. Le matériel élite présente une faible structuration en familles à la différence des ressources génétiques qui ne montrent aucune structuration. Le modèle GBLUP et l'approche de validation croisée ont été appliqués pour estimer la précision de prédiction du modèle pour sept caractères phénotypiques évalués dans les deux populations. L'amplitude de variation de toutes les précisions calculées va de 0,23 à 0,82. Les précisions de prédiction étaient plus élevées pour le matériel élite que pour les ressources génétiques sur trois caractères, avec un écart variant entre 0,1 et 0,18, et plus élevées pour les ressources génétiques que pour le matériel élite pour deux caractères, avec un écart de 0,09 à 0,18. Les deux caractères restant ont obtenu des précisions presque similaires sur les deux populations. Les faibles précisions de prédiction obtenues sur cinq des sept caractères indiquent la nécessité d'améliorer la précision de prédiction en augmentant le nombre d'individus.
Mots clés : Prédiction génomique, Précision de prédiction, Matériel élite, Ressources génomiques, Marqueur SNP, Pomme

Abstract

European apple cultivation is based on a small number of elite cultivars. These cultivars are becoming increasingly vulnerable in the current climate change context because of their reduced diversity and narrow genetic base. Genetic resources may be a source of favorable alleles that are absent from elite material and thus the transfer of such alleles from genetic resources to elite material could help solve this problem. Genomic selection has been proposed to achieve this goal. This study aims to assess the accuracy of genomic prediction of elite material and genetic resources. We have used high density genotypic data (303239 SNPs) and populations of 847 to 1025 individuals. The characterization of the two populations using genotypic data showed a very weak differentiation (F_{st} = 0.004) between elite material and genetic resources. On the other hand, differences in minor allele frequencies (MAF) were observed throughout the genome of the two populations. Elite material has a weak family structure, while genetic resources show no structure. The GBLUP model and the cross-validation approach were applied to estimate the accuracy of predictions for seven phenotypic traits assessed in the two populations. The range of variation of all calculated accuracies ranged from 0.23 to 0.82. Accuracies were higher for elite material than for genetic resources for three traits, with a mean difference ranging from 0.1 to 0.18, and higher for genetic resources than for elite material for two traits, with a mean difference ranging from 0.09 to 0.18. The two remaining traits obtained almost similar accuracies for the two populations. The low accuracies obtained for five of the seven traits indicates the need to improve accuracy by increasing the number of individuals.
Key words: Genomic prediction, Accuracy, Elite material, Genomic resources, SNP marker, Apple

Pour citer cet ouvrage : DIOUF, Babacar (2020). Précision de la sélection génomique dans des populations constituées de matériel élite et de ressources génétiques chez le pommier. d'Ingénieur systèmes agricoles et agroalimentaires durables au sud, Amélioration des Plantes et Ingénierie Végétale Méditerranéennes et Tropicales, Montpellier SupAgro, 87 pages.