



Université de Montpellier
Faculté des Sciences
Master Biologie-Agrosciences



Parcours : Biotechnologies et amélioration des plantes
tropicales

Année Universitaire 2022-2023

Exploration des potentialités de la prédiction génomique avec des données
multisites et multi-traits obtenues sur des core-collections de pêchers

INRAE PACA, UMR GAFL, équipe DADI

Co-encadré par : Morgane ROTH et Marie SERRIE

Soutenu le 09 Juin 2023

Octave CABEL



Table des matières

Parcours : Biotechnologies et amélioration des plantes tropicales.....	1
Abréviations et acronymes:.....	5
Remerciement.....	5
I. Introduction :.....	6
I.A. Optimiser l'amélioration des plantes par la génétique quantitative.....	6
I.A.1 : Augmenter la réponse à la sélection pour l'amélioration des plantes.....	6
I.A.2. Principe de la sélection génomique (GS).....	8
1.A.3. Etat des lieux de la sélection en arboriculture fruitière.....	8
I.B. Comment évaluer le potentiel de la sélection génomique sur une culture ?.....	12
I.B.1. Des modèles pour la sélection génomique : prédiction génomique.....	12
I.B.2 Modélisation statistiques de la prédiction génomique.....	14
I.B.3. Paramètres déterminants la capacité de prédiction (PA).....	16
I.C. Filière, biologie et caractéristiques du pêcher et des proches apparentés.....	18
I.C.1. Caractéristiques biologiques de l'espèce.....	18
I.C.2. La filière pêche.....	20
I.D. Objectifs et problématique du stage.....	20
I.D.1.Problématique.....	20
I.D.2. Objectifs.....	20
II. Matériel et méthodes :.....	22
II.A. Résumé de la méthode employée dans cette étude pour évaluer le potentiel de la sélection génomique sur le pêcher :.....	22
II.B. Le matériel végétal.....	22
II.B.1 Conception et description de la core-collection.....	22
II.B.2. Le dispositif expérimental, utilisation et entretien de la core-collection.....	24
II.C. Obtention des données phénotypiques et génotypiques.....	24
II.C.1. Phénotypage des caractères de l'études.....	24
II.C.2. Génotypage de la core-collection de pêchers.....	26
II.D. Traitement des données, modèles statistiques et méthodes de prédictions.....	26
II.D.1. Préparation préalable des données.....	26
II.D.2. Héritabilités.....	28
II.D.3. Décomposition de la variance (VD) :.....	28
II.D.4: Les méthodes de prédictions génomiques.....	30
III. Résultats.....	32

III.A Exploration préalable des données phénologiques, de dommage de cloque et leurs liens avec la consanguinité.	32
III.B. Héritabilités des caractères selon les environnements	34
III.C Décomposition de la variance, consanguinité et apparentement	34
III.D. Les capacités de prédictions : prediction accuracies (PAs).....	38
III.D.1 Conclusion et comparaison sur les PAs globales	38
III.D.2. PAs obtenues sur les résultats intra-environnementaux.....	38
IV. Discussion	40
IV.A. Pertinence de la méthode, du dispositif expérimental, et corrections statistiques	40
IV.B. Interprétation des différents résultats d’architectures génétiques et de prédictions pour chaque trait :	42
IV.B.1 Les héritabilités pour identifier la proportion du facteur génétique :.....	42
IV.B.2. Décomposition de la variance et hypothèse sur l’architecture des traits :	42
IV.B.3 Interprétations des données de PAs.....	44
IV.C.Intérêt en sélection pour la filière	46
IV.D. Conclusion et future application de la démarche.....	46
V. Références bibliographiques.....	49
VI. Annexe.....	52
Figure annexe 1 :	52
Figure annexe 2 :	52
Figure annexe 3 :	53
Figure annexe 4 :	56
Résumé :.....	57

Abréviations et acronymes:

- AA (Additivités par additivités)
- AD (Dominances par additivités)
- BLUP (Best Linear Unbiased Prediction)
- DD (Dominances par Dominances)
- FLO (Floraison)
- GBLUP (Genomic BLUP)
- GEBV (Genomic Estimated Breeding value, valeur croisement estimé pour la génétique d'un individu, \mathbf{y})
- GS (Genomic Selection)
- Inb (Inbreeding ; Consanguinité)
- PA (Prediction Accuracy ; prédictibilité, précision de prédiction)
- PV (Pointes-vertes)
- QTL (Quantitative traits locus)
- RR-BLUP (Ridge Regression BLUP)
- SNP (Single Nucleotide Polymorphism)
- VD (Variance decomposition ; decomposition de la variance)

Remerciement

Merci à Evelyne Joubert-Mazellier pour son soutien, à Sabrina, Patrice, Jean, Henri, et toute l'équipe DADI, REDD qui participèrent à bâtir ce cadre de travail formidable, bien sûr merci à mes encadrantes Marie SERRIE et Morgane ROTH qui furent humblement parfaite, dans ce rôle et qui m'ont beaucoup enseigné. Merci pour leurs patiences, leurs attentions, leurs discussions et leurs positivismes à toute épreuve. Merci aux autres stagiaires et collègues pour leurs soutiens, disponibilités, conseils et amitiés. Tous, nous avons amplement participé à ce que ce stage se passe à merveille dans un contexte de travail enrichissant.

Merci à tout le GAFL pour son cadre parfait, enrichissant, diversifiés, et ouvert d'esprit. Merci pour m'avoir accueilli et m'avoir permis de vivre cette expérience unique.

Et bien sûr, merci aux financeurs de CEP innovations et du GIS fruit, notamment David RAY pour avoir suivi mes travaux avec attention et sollicitude.

I. Introduction :

I.A. Optimiser l'amélioration des plantes par la génétique quantitative

I.A.1 : Augmenter la réponse à la sélection pour l'amélioration des plantes

I.A.1.a. Contexte d'amélioration variétale et recours aux marqueurs génétiques

Bien que les nouvelles méthodes d'amélioration des plantes nécessitent des expérimentations et du temps avant d'être adoptées en routine par les sélectionneurs, les méthodes actuelles de sélection reposent déjà sur les biotechnologies et la génétique. La MAS (Markers Assisted Selection) par exemple a déjà prouvé être une révolution (Ashraf et Foolad, 2013). Cette première permet dès les plus jeunes stades de vie d'une plante (graine ou plantule) d'obtenir des informations précises sur son phénotype adulte grâce à son génotype pour plusieurs marqueurs génétiques (fig. 1). Cela évite de recourir à la culture de l'ensemble de la descendance jusqu'à évaluer leurs phénotypes (Kumar et al., 2020). Cela peut représenter un gain considérable de temps et surtout d'argent notamment en ce qui concerne les plantes pérennes dont la période juvénile est plus longue.

La MAS, bien qu'assez communément adoptée en routine pour son efficacité et ses propriétés possède cependant des limitations telles que de ne prendre en compte qu'un nombre restreint de locus et de ne pas pouvoir efficacement sélectionner les traits à architecture quantitative tels que le rendement ou les résistances quantitatives. C'est pourquoi je cherche à évaluer les potentialités d'une méthode de sélection utilisant des marqueurs homogènement répartis sur la totalité du génome : la GS (Genomic Selection) (Kumar et al., 2012a; Kumar et al., 2020). Lors d'un programme de sélection, l'utilité et l'impact de ces méthodes sur un caractère peuvent être estimés en calculant le gain génétique ou bien la réponse à la sélection (Marulanda et al., 2016).

I.A.1.b Equation du sélectionneur : quantifier le progrès génétique

Le gain génétique peut-être défini comme les performances agronomiques des individus d'une variété en développement par rapport aux performances de leurs géniteurs (fig. 2). Celui-ci peut aussi se définir comme le changement de la moyenne et/ou de la variance du phénotype entre la génération parentale et la génération de leurs descendants (Longin et al., 2015; Marulanda et al., 2016; Marulanda et al., 2021). Le gain génétique se définit aussi ainsi :

$$\Delta G_a = ih\sigma_y / Y \quad (1)$$

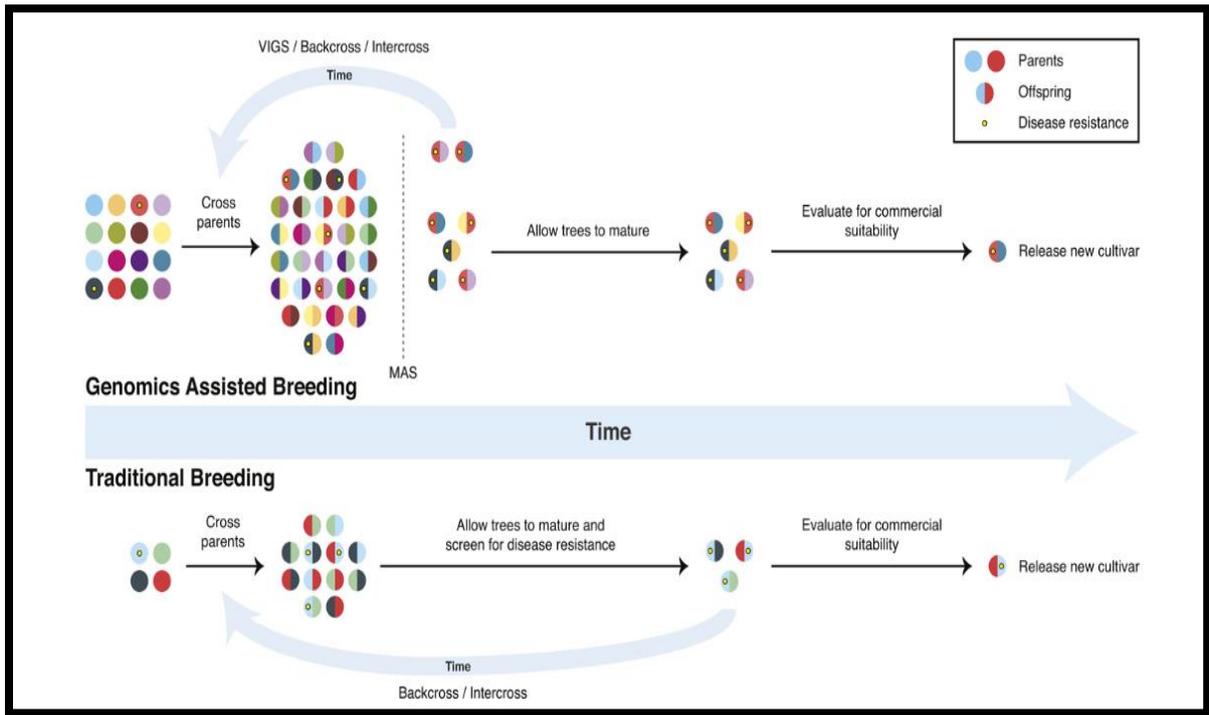


Figure 1 ; McClure., 2014, *Schéma comparatif de la sélection classique et de la sélection assistée par la génétique utilisant des informations génomiques à la place du phénotypage.*

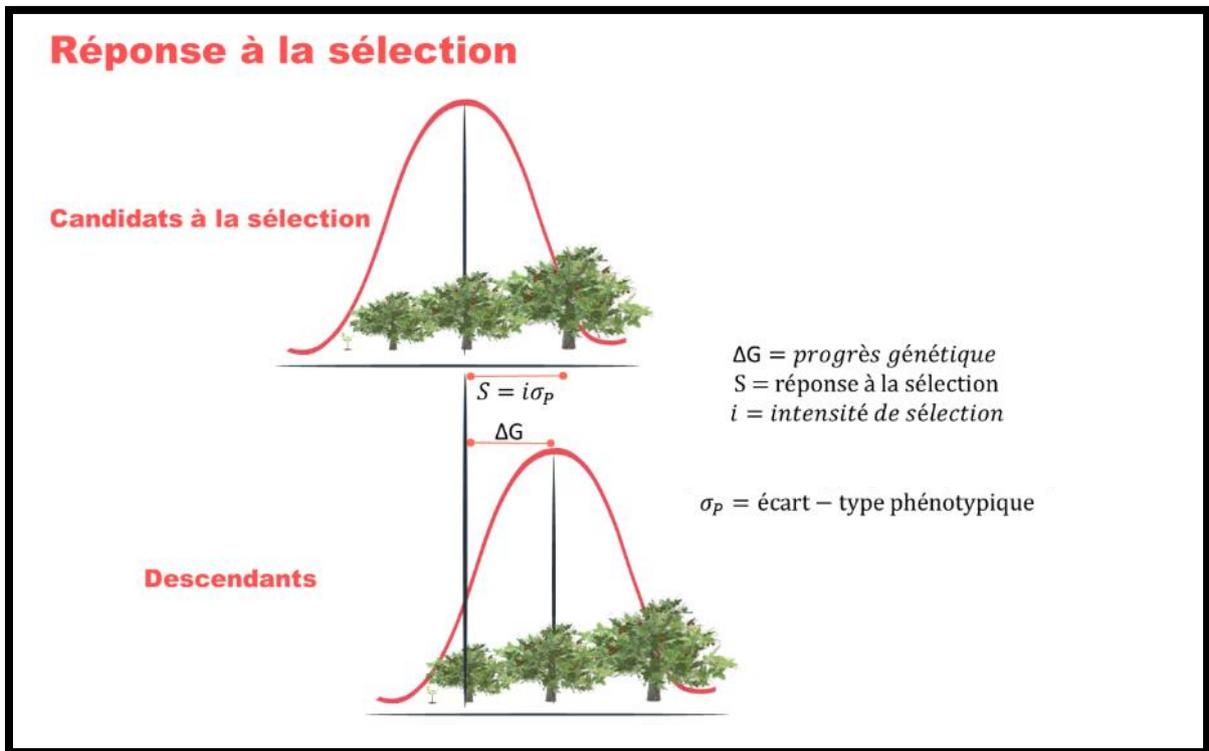


Figure 2 : extrait du manuscrit de Thèse de Nsibi et al., (2021); *réponse à la sélection en ce qui concerne un trait quantitatif continu d'une population d'un programme de sélection.* Les courbes rouges représentent la distribution du trait quantitatif continu. Le caractère à optimiser est la hauteur de la plante.

Tel que ΔG_a correspond aux gains génétiques entre deux générations, i est l'intensité de sélection, h la racine carrée de l'héritabilité du trait, σ_y l'écart-type génétique pour ce même trait, et Y est le nombre d'années nécessaires à l'obtention de la nouvelle variété.

I.A.2. Principe de la sélection génomique (GS)

La GS consiste à générer une valeur des performances GEBVs (Genomic Estimated Breeding Values) des individus en sélection vis-à-vis d'un caractère donné grâce à un modèle de prédiction se basant sur les variations alléliques de nombreux marqueurs répartis sur l'ensemble de leurs génomes (Elsen, 2022). Tout comme la MAS, cette méthode a la possibilité de réduire le temps total nécessaire aux programmes de sélection en optimisant le paramètre Y de l'équation (1) et d'augmenter le paramètre i en sélectionnant de manière plus stringente les individus à haut potentiel génétique (Marulanda et al., 2016). Elle consiste donc à estimer la valeur potentielle d'un candidat à la sélection grâce à sa combinaison d'allèles favorables sur des marqueurs répartis sur l'ensemble du génome. Mais contrairement à la MAS, elle est particulièrement appropriée pour des caractères au déterminisme génétique complexe (nombreux locus déterminant le trait) se manifestant par une distribution phénotypique de type quantitative continue (par opposition à discrète ou binaire) se rapprochant d'une loi normale (fig.1).

Pour prédire, on « entraîne » donc un modèle capable à partir d'une population d'entraînement d'estimer les valeurs du trait y dans une population qui lui est apparentée. Cet entraînement sert à pondérer l'importance de chaque facteur génétique, ce qui s'illustre dans l'approche RR-BLUP par l'estimation des effets de chaque marqueur relativement aux autres et ainsi de sa contribution à la construction du trait quantitatif complexe continu. En cela, la prédiction génomique complète la MAS qui est adaptée à des traits au déterminisme plus simple (Monogéniques, ou Oligo-géniques).

I.A.3. Etat des lieux de la sélection en arboriculture fruitière

On considère que la vitesse d'amélioration des espèces pérennes fruitières est fortement en retard par rapport aux grandes cultures (McClure et al., 2014; Migicovsky and Myles, 2017) notamment à cause du fait que les infrastructures, le temps, et l'espace requis pour l'amélioration des plantes pérennes par croisement est drastiquement plus conséquent que pour les annuelles.

L'évaluation des traits commerciaux des pérennes est aussi souvent bien plus complexe (McClure et al., 2014). Les arbres fruitiers supportent aussi généralement très mal les autofécondations, ce qui rend compliqué le développement de lignées, bien qu'en l'occurrence le modèle pêcher soit autogame préférentiel (Byrne et al., 2012).

Concernant les cultures pérennes fruitières, le recours aux produits phytosanitaires a longtemps été la technique de choix pour faire face aux bioagresseurs, les activités de sélection se focalisant davantage sur les traits de rendement et qualité du fruit. McClure et al., (2014) titre son article : « Genomics : a potential panacea for the perennial problem » au vu des complémentarités de l'approche de la GS pour ces cultures. L'utilité de cette dernière est proportionnelle au temps et au coût requis pour l'acquisition du produit final de la culture depuis les graines (Longin et al., 2015), hors la période juvénile d'une pérenne fruitière est par essence plus longue. Le coût de l'amélioration de ces espèces ne favorise donc pas le développement de nouvelles variétés (McClure et al., 2014; Migicovsky and Myles, 2017). Les espèces pérennes fruitières ont besoin d'autres modes de sélection et de diversification pour faire face aux maladies et à l'évolution rapide des micro-organismes. La « course à l'armement » évolutive en phytopathologie se perd vite pour nombres de cultivars (Miller and Gross, 2011) notamment avec les méthodes actuelles d'introgessions qui ne permettent que d'introduire quelques QTLs majeurs à la fois rendant les résistances moins durables (Kumar et al., 2012a; Kumar et al., 2020). L'introduction de caractères de résistances et tolérances est encore plus difficile quand ceux-ci sont régis par un déterminisme génétique complexe et donc un grand nombre de locus déterminants avec de faibles effets sur le trait (inexistence de locus causatif majeur). Le coût élevé de la sélection et la propagation végétative majoritaire façonne le problème d'amélioration des pérennes, et alors qu'il y a de nombreux polymorphismes disponibles aux sélectionneurs, les contraintes agronomiques, biologiques, sociales et économiques empêchent l'exploitation de ces variations.

Les nouvelles méthodes de sélection assistées par des marqueurs génétiques sont lentement adoptées par les sélectionneurs (McClure et al., 2014). Ces derniers sont majoritairement utilisés pour ce qui est des tests de parentés et très peu pour les cribles de descendances. Des améliorations sont encore en cours et ces technologies relativement nouvelles sont encore en développement. Il est primordial de confirmer à l'aide d'études scientifiques de la précision de ces méthodes tout en investiguant l'optimisation des protocoles, designs et coûts de l'intégration du marquage génétique dans les programmes de sélections

des cultures fruitières pérennes (Kumar et al., 2012a; Grattapaglia, 2017; Roth et al., 2020; Jung et al., 2021). En ce qui concerne la GS : Une littérature scientifique abondante est présente pour les animaux et pour les grandes cultures comme les céréales, mais encore peu d'études existent chez ces pérennes et la méthode est loin d'être envisagée en sélection (Nakaya and Isobe, 2012; Grattapaglia, 2017).

I.B. Comment évaluer le potentiel de la sélection génomique sur une culture ?

I.B.1. Des modèles pour la sélection génomique : prédiction génomique

I.B.1.a. La précision de prédiction ou prédictibilité

Pour pouvoir sélectionner efficacement les meilleurs individus de nos programmes de sélection grâce à la génomique, il faut développer des modèles de prédiction capables d'attribuer des valeurs de performances (GEBV) fiables. Un paramètre essentiel à estimer pour se questionner sur la faisabilité de cette démarche se basant sur la modélisation, est la précision de prédiction ou « Prediction Accuracy » (PA) (r ou r^2), simple corrélation entre performances observées y , et performances prédites (GEBV). Le modèle est testé par validation simple ou croisée au sein de la population d'entraînement pour obtenir les précisions de prédiction. La validation croisée utilise des sous-ensembles d'un même jeu de données pour entraîner et valider le modèle. L'apparement entre population de validation (ou de sélection) et population d'entraînement est nécessaire pour pouvoir prédire efficacement.

I.B.1.b Les limites des modèles additifs concernant l'amélioration des plantes

Classiquement, la sélection repose sur l'exploitation des effets additifs car hérités de manière stable d'une génération sur l'autre, et ce sont également, à juste titre, les effets majeurs pris en compte en sélection génomique (Meuwissen et al., 2001). Cela permet d'améliorer d'année en année les performances de nos populations par rapport à celle de leurs parents vis-à-vis de traits quantitatifs continus (fig 2). Cependant dans le souci d'une prédiction génomique prenant en compte l'architecture des traits de manière plus complète, il peut être pertinent de tenir compte d'autres types d'effets génétiques liés aux interactions entre allèles intra-locus (dominance) et aux interactions entre allèles inter-locus (épistasie) (Vitezica et al., 2017; Roth et al., 2022). Si on ne prend pas en compte ces effets liés aux interactions, il y a un risque de sélectionner des individus n'étant pas réellement les plus performants (Kumar et al., 2015).

I.B.2 Modélisation statistiques de la prédiction génomique

Les méthodes de prédiction utilisées ici sont appelées GBLUP (eq. 2) et RR-BLUP (eq. 3). Elles s'opposent par la nature des effets estimés/pris en compte. Dans le cadre du GBLUP on exploite soit la covariance génétique entre individus estimée à l'aide des marqueurs génétiques via des matrices d'apparentement calculées selon une méthode choisie (en l'occurrence Vitezica et al., 2017). Dans le cadre du rrBLUP (Endelman, 2011), les effets des marqueurs sont directement estimés et les GEBVs reflètent la somme des effets de ces marqueurs pour chaque individu, connaissant son génotype à ces mêmes marqueurs. Ces deux méthodes reposent sur une même hypothèse que tous les effets aux marqueurs génomiques sont quantitatifs et normalement distribués. Ce n'est pas le cas de méthodes probabilistes Bayésienne comme le LASSO ou certain Bayes Alphabet (Habier et al., 2011). Ces dernières peuvent permettre de mieux prédire des caractères sous-tendus par des gènes à la fois à effets mineurs et majeurs comme des traits simples.

Les différents types d'effets génétiques nous renseignent sur le déterminisme du caractère et il est possible de modéliser la variance apportée par chacun de ces facteurs génétiques (Additif (A), Dominance (D), et Épistasie (I)) dans la variation du caractère y .

$$y_i = \mu + \beta Inb_i + A_i + D_i + AA_i + AD_i + DD_i + \varepsilon_i \quad (2)$$

Où y est le phénotype observé ou prédit, Inb_i est une variable quantitative résumant le taux d'Homozygotie de l'individu i , le phénotype observé, μ l'intercepte, β est le coefficient de régression associée à la consanguinité. A_i et D_i sont les effets génétiques (aléatoires) de l'additivité et de la dominance chez l'individu i alors que AA_i , AD_i et DD_i sont les effets composant l'épistasie. On révèle ainsi les individus plus performants pour chaque caractère. La GEBV peut être prédite par différents types de modèle comme (2) ou bien :

$$y = x\beta + \sum_{j=1}^m z_j a_j + e \quad (3)$$

Où X une matrice environnementale ; β les coefficients attribués aux environnements ; z_j le $j^{\text{ème}}$ marqueurs appartenant à Z une matrice génotypique de dimension $N * M$, N étant le

nombre d'individus total dans les matrices et M le nombre de marqueurs total m ; a un vecteur de coefficient comportant les effets des allèles à chaque locus dont a_j est le coefficient associée à l'allèle « positive » du $j^{\text{ème}}$ locus. Les variables explicatives génétiques du modèle linéaire de GS sont donc exprimées sous la forme de matrice N^*M de SNP (Single Nucleotide Polymorphism), marqueurs di-alléliques codé en tertiaire $\{-1, 0, 1\}$ ou $\{0,1,2\}$ pour les génotypes $\{AA ; AB ; BB\}$ respectivement, représentant le nombre d'allèles « positives » à chaque locus m pour chaque individu n .

I.B.3. Paramètres déterminants la capacité de prédiction (PA)

Un des premiers paramètres à considérer est l'héritabilité. Celle-ci est définie comme la proportion de variance génétique expliquant la variance du phénotype (Schmidt et al., 2019). Elle est, au moins en partie, transmise d'une génération à l'autre. Avec une faible héritabilité, c'est l'environnement qui va affecter majoritairement le trait observé et il sera vain de vouloir de sélectionner sur une base génétique. La prédictibilité est fortement liée à l'héritabilité du trait.

La densité de SNPs criblant le génome est aussi importante, une trop grande couverture s'avère souvent inutile, mais il existe un juste-milieu propre à l'architecture du trait (nombre de marqueurs impliqués, leur localisation, étendue du déséquilibre de liaison autour des gènes causaux), à la biologie de l'espèce traitée, et au type de populations choisies qui vont impacter l'étendue du Linkage disequilibrium (LD).

De plus pour qu'un modèle soit à même de prédire convenablement les valeurs phénotypiques, il faut que ce premier ai été calibré sur une population suffisamment apparentée à la population d'entraînement. Si les populations d'entraînement et de validation ou de sélection sont trop éloignées, le modèle peut être peu efficace et la prédictibilité plus faible. Cet apparentement tend à décroître au fur et à mesure des générations notamment à cause des événements de recombinaison pouvant affecter le LD (Linkage Disequilibrium) entre marqueurs génétiques et SNPs causatifs.

I.C. Filière, biologie et caractéristiques du pêcher et des proches apparentés.

I.C.1. Caractéristiques biologiques de l'espèce

I.C.1.a. Description de la Famille botanique

Le pêcher *Prunus persica* (L.) Batsch est désigné comme une espèce modèle parmi les Rosacées parmi lesquelles existent de nombreux genres d'intérêt agronomique comme les *Pyrus*, les *Malus*, ou les *Prunus*. Ce dernier genre comprend des cultures comme : l'amandier (*P. dulcis*), l'abricotier (*P. armeniaca*) ou le cerisier (*P. avium* ou *P. cerasus*) (Shulaev et al., 2008). Le pêcher appartient à la sous-famille des *Prunoideae* (*Amygdaloideae* ; ou encore *Drupacées* en référence au nom du fruit caractéristique) et donc plus étroitement apparenté à l'amandier. Les *Prunus*, comprenant environ 250 espèces, partagent un petit génome (220–300 Mbp) hautement synténique, majoritairement diploïdes sauf quelques exceptions comme chez les pruniers (Aranzana et al., 2019). Les espèces membres du genre *Prunus* ont de faibles barrières reproductives à l'intérieur de leurs sous-genres respectifs. Ces barrières peuvent dans certains cas être contournées grâce à des méthodes telles que le sauvetage d'embryon.

I.C.1.d. Le pêcher

La faible taille de son génome (0.23 Gb) est un plus pour les généticiens sachant aussi le haut degré de caractérisation fonctionnelle aujourd'hui atteint (International Peach Genome Initiative et al., 2013). Son importance économique majeure (cf **I.C.2.**), ainsi que sa large distribution selon les climats et latitudes, lui confèrent un potentiel futur proche de la pomme (première culture fruitière mondiale). Un génome de référence de haute qualité est disponible pour la pêche depuis 2010 et d'autres principaux *Prunus* furent aussi séquencés les années suivantes (International Peach Genome Initiative et al., 2013; Aranzana et al., 2019). La banque de données sur les Rosaceae « GDR » (Genome Database for Rosaceae) est en pleine expansion (Jung et al., 2019). Les pêcheurs ont un régime de reproduction autogame préférentiel (5% allogamie) en partie responsable d'une diversité génétique relativement faible dans l'espèce, mais conférant cependant des facilités pour la création de cultivars notamment par sa tolérance à la consanguinité (Byrne et al., 2012). L'ensemble des pêcheurs cultivés à des fins agro-alimentaires et ornementales seraient issus d'un seul et unique événement de domestication dans l'actuelle région géographique de la Chine (Cao et al., 2014).

I.C.1.e. Proches apparentées

Dans le sous-genre *Amygdalus*, les espèces les plus apparentées aux pêcheurs (*P. persica*) sont *P. mira* Koehne ; *P. kansuensis* Rehd. ; et *P. davidiana* (Carr.) Franch . Ces espèces

peuvent être utilisées pour élargir la diversité génétique dans les variétés de pêcheurs et servent de sources pour des résistances aux insectes, nématodes et pathogènes dans des croisements pour améliorer à la fois les variétés de porte-greffes et de scions (*i. e.* qui produisent les fruits) (Martinez-Gomez et al., 2003).

I.C.2. La filière pêche

La production mondiale de pêches, nectarines et pavies s'élève à 23 millions de tonnes par an (FAOSTAT, 2023). L'Asie participerait à plus de 60% du marché, avec la Chine produisant jusqu'à presque 9 millions de tonnes par an contre 1.5 million pour l'Italie ou 1.25 million pour l'Espagne, respectivement 2^{ème} et 3^{ème} leaders mondiaux. La France serait 8^{ème} au classement avec 332,000 tonnes produites à l'année (FAOSTAT, 2023).

En France, les terroirs produisant le plus de pêches sont concentrés dans le Sud-Est entre les régions PACA, Occitanie et Rhône-Alpes principalement. On voit notamment les plus grandes productions dans la vallée du Rhône (Gard ; Drome ; Bouches-du-Rhône) ainsi que dans les Pyrénées-Orientales et l'Ardèche. Les Pyrénées-Orientales à elles seules contribuent à la production de pêches à hauteur de 100,000 tonnes par an (Ctifl, 2012).

I.D. Objectifs et problématique du stage

I.D.1. Problématique

La pertinence de la sélection génomique chez les arbres fruitiers pose un certain nombre de questions auxquelles nous nous proposons de répondre avec l'exemple du pêcheur. Est-il possible de prédire de manière précise les valeurs phénotypiques du pêcheur grâce à la prédiction génomique ? Sur quels traits et avec quelle population d'entraînement et quel modèle les prédictions sont-elles les plus précises ? Peut-on appliquer la sélection génomique à l'amélioration variétale chez le pêcheur ?

I.D.2. Objectifs

Les objectifs de ce stage se répartissent en 4 axes :

- Caractériser la diversité phénotypique d'une core-collection de pêcheur en se basant sur des caractères à l'architecture génétique contrastée de phénologie et de tolérance à la cloque
- Disséquer l'architecture génétique par l'étude des effets additifs et non-additifs ainsi que de la consanguinité sur les traits considérés

- Évaluer la qualité des prédictions génomiques pour l'ensemble de ces traits à l'aide de différents modèles prenant en compte ou non les effets génétiques non-additifs
- Analyser le potentiel de la GS le pêcher et poursuivre la discussion auprès des sélectionneurs sur l'intérêt de cette méthode pour les fruits à noyaux du genre *Prunus*.

II. Matériel et méthodes :

II.A. Résumé de la méthode employée dans cette étude pour évaluer le potentiel de la sélection génomique sur le pêcher :

La méthode appliquée ici peut être synthétisée en 3 étapes et pouvant être elles-mêmes subdivisées : (1-i) Choix des caractères à analyser ; (1-ii) Estimation des héritabilités ; (1-iii) Analyse des corrélations entre homozygotie et caractères phénotypiques ; (2-i) Calculs des matrices d'apparentement « Kinships » codant pour les effets génétiques (apparentement additif ou non-additif) ; (2-ii) Représenter les proportions de ces effets dans la décomposition de la variance (VD); (3-i) Valider et entraîner les modèles pour obtenir les PAs avec les modèles GBLUP ; (3-ii) Développer des modèles RR-BLUP et obtentions des PAs

II.B. Le matériel végétal

II.B.1 Conception et description de la core-collection

La core-collection pêchers est composée de 206 accessions toutes greffées sur des porte-greffes Montclar® Chanturque et provenant des collections de ressources génétiques de INRAE conservées par le centre de ressources biologique (CRB) *Prunus* à Bordeaux et à Avignon. Celle-ci fut constituée pour maximiser la diversité génétique du germplasma international des pêchers dans une population de taille optimale pour les expérimentations. On y retrouve de nombreux cultivars de références, des variétés élites, ainsi que les plus proches apparentés sauvages de l'espèce (*P.kansuensis* , *P.davidiana*, *P.mira*). Des analyses en composantes principales, DAPC ainsi que des analyses de type des sNMF ont été effectuées par (Serrie, 2021) et ont permis de conclure à une structuration en 3 grands groupes génétiques des individus de la core-collection. Les groupes analysés sont très liés par leurs provenances géographiques (majoritairement Européennes, Nord-Américaine et Asiatique) et montre un faible taux d'admixture.

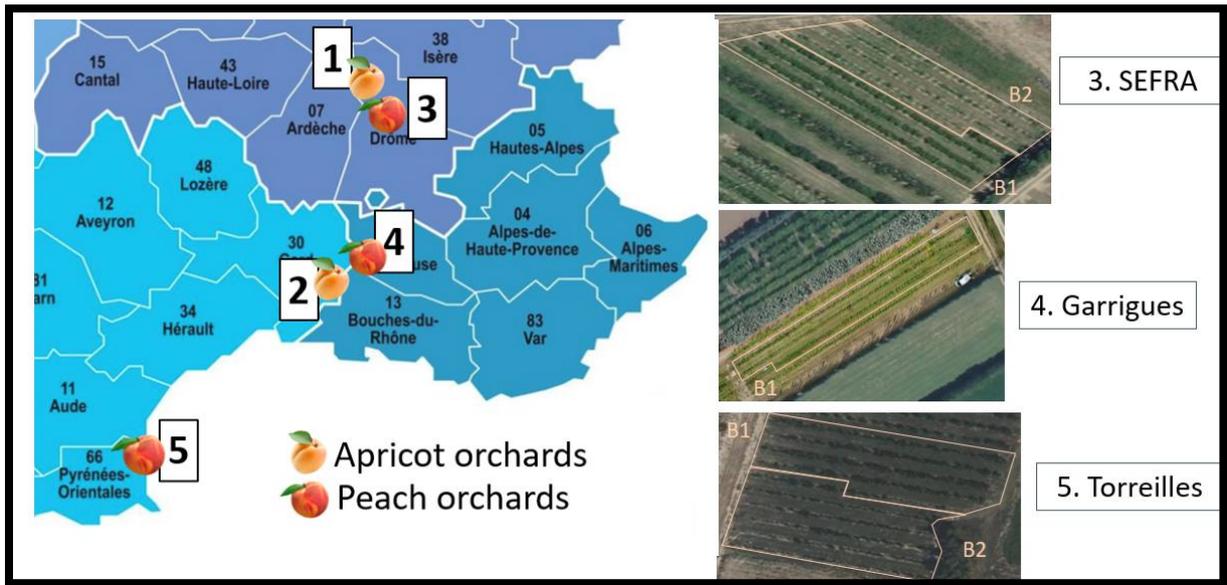


Figure 3 : Carte de la répartition géographique des dispositifs de culture des core-collections abricotiers et pêchers de l'INRAE PACA, projet RésiDiv, avec vues satellites sur la topologie des vergers de pêchers (Serrie, résultats pas encore publiés).

II.B.2. Le dispositif expérimental, utilisation et entretien de la core-collection

II.B.2.a. Design expérimental

La core-collection de pêchers a été plantée en janvier 2019 en suivant un plan randomisé avec deux blocs complets comportant chacun 1 réplication par génotype. Les premiers blocs ont été plantés directement avec des arbres déjà greffés (plantation dite en scion) alors que les arbres des seconds blocs, eux, furent greffés sur place dans les vergers. Les plants n'ayant pas pu se développer furent ensuite replantés les hivers suivants. Sur le site des Garrigues cependant, la collection est avant tout plantée pour la conservation in-situ, en cela la randomisation est partielle et il n'y a qu'un seul bloc au lieu de deux (fig 3).

II.B.2.b Conditions de culture

Ce dispositif expérimental d'où proviennent nos mesures est cultivé sur 3 sites dans le quart sud-est de la France (fig.3). Les 3 sites sont (i) les Garrigues (GA) à Avignon sous la responsabilité de l'INRAE PACA ; (ii) Torreilles (TO) avec la SICA-CENTREX dans les Pyrénées orientales ; et (iii) Étoile sur Rhône dans la Drôme où la plantation est gérée une station d'étude régionale, la SEFRA (SE). Les abréviations citées sont les nomenclatures permettant d'identifier les différents sites et sont les diminutifs utilisés quotidiennement par notre équipe.

Au vu des différences de latitudes entre les sites, il y a une hétérogénéité environnementale entre les sites (méditerranéens ou plus continentaux). Notamment à cause des microclimats des sites (Aude, Vallée du Rhone, Drôme) ainsi que des itinéraires techniques des différentes stations gérant la culture du dispositif. Les entretiens des vergers sont gérés par les stations locales, mais l'INRAE effectue les notations. Les pêchers ont été maintenus en protection intégrée pendant les 2 premières années après leurs plantations. À partir de 2022, des traitements ponctuels sont appliqués si la survie du verger est en jeu, car ceux-ci sont conduits en conditions de faible intrants-phytosanitaire.

II.C. Obtention des données phénotypiques et génotypiques

II.C.1. Phénotypage des caractères de l'études

Les suivis saisonniers permettent l'observation phénotypique du développement des maladies, de la vigueur et des stades phénologiques de nos collections. Dans cette étude nous baserons sur des traits quantitatifs de phénologie annuel ainsi que sur les indices de dégâts de cloques.



Figure 4 : Photographies illustrant les étapes du stade des pointes-vertes ; (A) Pointes-vertes considérées comme sorties ; (B) pointes-vertes déjà passées, le plumeau du bourgeon foliaire est déjà trop ouvert, feuilles en débouurements ; les écailles du bourgeon s'affaissent : on peut apercevoir les différentes feuilles.

Note	% de feuilles atteintes
0	0 %
1	1 à 10 %
2	11 à 20 %
3	21 à 30 %
4	31 à 40 %
5	41 à 50 %
6	51 à 60 %
7	61 à 70 %
8	71 à 80 %
9	81 à 90 %
10	> 90 %



Figure 5 : Echelle de phénotypage des symptômes de la cloque du pêcher ; (A) échelle de notation permettant l'évaluation de l'intensité des dégâts de cloque ; (B) Photographies des rameaux d'arbres peu atteints notés 2 ; (C) Photographies des rameaux d'arbres très atteints notés 9 ou 10.

Les dates des stades phénologiques sont reportées au rythme de 1 à 2 fois par semaine selon le trait ciblé. L'avancement du cycle de débourrement des boutons floraux ou des bourgeons foliaires est suivies en notant lorsque certains seuils (10 ; 50 et 100%) sont atteints. On note les seuils d'évolutions selon la proportion du débourrement sur les rameaux de l'individu mesuré, c.à.d quand 50% des rameaux possèdent 50% de bourgeons ayant déjà débourrés leurs fleurs, alors le stade 50% est atteint. Seul ce seuil de 50% sera utilisé dans notre étude, respectivement « flo_50 » (FLO) (code BBCH : 65) pour la floraison et « pv_visible_50 » (PV) (BBCH : 09) pour les pointes vertes (prédébourrement foliaire) (fig. 4).

Pour estimer l'intensité de l'attaque de cloque, on utilise une grille de notation allant de 0 à 10 (fig.5) où 0 correspond à une absence totale de symptômes et une note au-dessus de zéro correspond au pourcentage de feuilles attaquées, légèrement ou totalement, sur le nombre de feuilles totales de l'arbre.

II.C.2. Génotypage de la core-collection de pêchers

192 individus de la core-collection pêcher ont été génotypés par la société Neogen en utilisant la puce 18 K SNPs issue du consortium IRSC. L'ADN a été extrait au laboratoire d'analyses moléculaires du GAFL. Les SNPs de la puce ont été choisis pour mettre en valeur le polymorphisme à l'échelle de l'espèce entière et certains SNPs d'intérêt de la littérature scientifique y sont aussi incorporés. Dans le présent travail, j'ai également appliqué des filtres sur les données de génotypage pour (i) retenir les SNPs avec moins de 10% de NA, (ii) retenir les SNPs correspondant aux critères de MAF > 5%, ce qui laisse une matrice avec 13975 marqueurs. L'imputation des données manquantes a été effectuée en remplaçant les valeurs manquantes par la moyenne pour chaque locus.

II.D. Traitement des données, modèles statistiques et méthodes de prédictions

II.D.1. Préparation préalable des données.

Le jeu de données entier est composé de 2651 individus appareillés répartis en 3 sites sur 3 années, dont le produit fait 9 environnements avec 1 à 2 répétitions du dispositif core-collection pêchers chacun (fig. 3). Un individu est un génotype, cloné des accessions de la core-collection, mesuré une année donnée sur un site donné. Une correction des données phénotypique est effectuée pour s'affranchir des effets spatiaux dans les différents blocs. Cette correction est appliquée grâce à la formule :

$$Y_{irc} = \mu + G_i + R_r + C_c + E_{irc} \quad (5)$$

Où y_{irc} est le phénotype de l'individu i situé à la $i^{\text{ème}}$ ligne de la $c^{\text{ème}}$ colonne, G_i est l'effet génétique du $i^{\text{ème}}$ individu, R_i est l'effet lié à la ligne de l'individu dans le design, et C_c est l'effet lié à sa colonne. E_{irc} est l'erreur liée à l'individu i , $r^{\text{ème}}$ ligne et $c^{\text{ème}}$ colonne. Tous les facteurs en lettres latines sont aléatoires. Les définitions de ces facteurs sont fournies annexe 1.

Les analyses GBLUP globales, c.à.d. basées sur la totalité des individus du dispositif (répétitions environnementales considérées comme blocs), sont réalisées sur des moyennes ajustées obtenues grâce à la méthode des moindres carrés avec le package « doBy ». Plusieurs ANOVAs sont réalisées avec le test non-paramétrique de Kruskal-Wallis suivies d'une analyse 2 à 2 avec corrections de Bonferroni face à la multiplicité des tests.

II.D.2. Héritabilités

Pour obtenir l'héritabilité des caractères quantitatifs étudiés, il faut d'abord bâtir un modèle linéaire mixte d'explication :

$$y = Z + \beta + \lambda + \beta\lambda + e \quad (6)$$

Où Z est l'effet du génotype, β l'effet du site, λ l'effet du lieu, et leurs produits est l'interaction soit le facteur environnementale, e est l'erreur. On obtient ainsi des variances grâce aux facteurs aléatoires. Cette équation sera utilisée pour obtenir des BLUPs des classements des individus en valeurs normées BLUP (Best Linear Unbiased Prediction) vis-à-vis du trait y . Il est aussi possible de calculer précisément l'héritabilité du trait grâce à la formule :

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2/n_{\text{Rep}}} \quad (7)$$

On fait ici le rapport de la variance liée au facteur génétique (σ_g^2) sur la variance totale, c.à.d. liée à l'environnement, à la génétique, à leurs interactions et à l'erreur résiduelle (ici σ_e^2) divisée par n_{Rep} le nombre moyen de répétitions des génotypes par environnements.

II.D.3. Décomposition de la variance (VD) :

Pour observer les différentes proportions de variances expliquées et par quelles facteurs génétiques, on compte utiliser 5 modèles : A ; A_Inb ; AD ; AD_Inb ; ADI_Inb , dont les définitions précises des paramètres les composants sont fournies annexe 2. Ces modèles sont exposés comme tel, du plus au moins complexe, avec ajouts des facteurs génétiques au fur et à mesure pour comparer l'effet de l'implémentation des différentes distinctions d'effets non-additifs.

$$\mathbf{A} ; y_{ij} = A_i + e_{ij} \quad (8-a)$$

Ici dans le modèle **A**, seul l'effet additif **a** de l'individu est pris en compte avec l'erreur **e** pour expliquer **y** de l'individu **i** pour l'année **j**.

$$\mathbf{A_Inb} ; y_{ij} = A_i + Inb_i + e_{ij} \quad (8-b)$$

Dans cette version on ajoute « l'Inbreeding », le facteur de consanguinité, ou degré d'apparentement entre les individus représentés par le facteur « **Inb** ».

$$\mathbf{AD} ; y_{ij} = A_i + D_i + e_{ij} \quad (8-c)$$

$$\mathbf{AD_Inb} ; y_{ij} = A_i + D_i + Inb_i + e_{ij} \quad (8-d)$$

Dans ces 2 versions on rajoute avant tout la dominance, puis l'Inbreeding peut être pris en compte ou non.

$$\mathbf{ADI_Inb} ; y_{ij} = A_j + D_i + AA_i + AD_i + DD_i + Inb_i + e_{ij} \quad (8-e)$$

Ci-dessus le modèle le plus complexe où l'on ajoute les 3 matrices liées aux effets épistatiques (**I** ; décomposé en 3 facteurs différents, **aa**, **ad**, et **dd** ; cf **eq. (2)** chp. **I.B.2.**). Ces implémentations issues de la méthode de (Roth et al., 2022) de décomposition de la variance (VD) est effectuée grâce à une adaptation du script de l'article mis en ligne publiquement **sur INRAE dataverse**. Ces modèles y sont testés grâce au package R MM4LMM et notamment via la fonction principal « MMEst », adaptée pour les modèles de décomposition de la variance, capable d'inférer à la proportion de responsabilité de chacun des facteurs aléatoire grâce à un algorithme de Min-Max. Ces modèles seront ensuite utilisés pour prédire et seront comparés aux résultats observés pour obtenir des PAs via la méthode GBLUP (cf. **eq.2**) (utilisant les matrices d'effets génétiques bâties selon la méthode de (Vitezica et al., 2017)).

II.D.4: Les méthodes de prédictions génomiques

II.D.4.a. Points communs des deux démarches.

Pour obtenir les PAs : les corrélations entrent valeurs prédites et valeurs observées sont effectuées 200 fois pour obtenir une moyenne robuste. Elles sont calculées selon une validation croisée 5-fold (4/5 ou 80/100), où 80% des individus sont compris dans le groupe d'entraînement et 20% dans la population de validation. Ce système d'entraînement et de validation du modèle sera dans un premier temps sera appliqué de façon globale c.à.d au jeu de données entier puis elles seront ensuite appliquées sur des sous-ensembles du dataset comprenant les valeurs des individus pour un environnement spécifique.

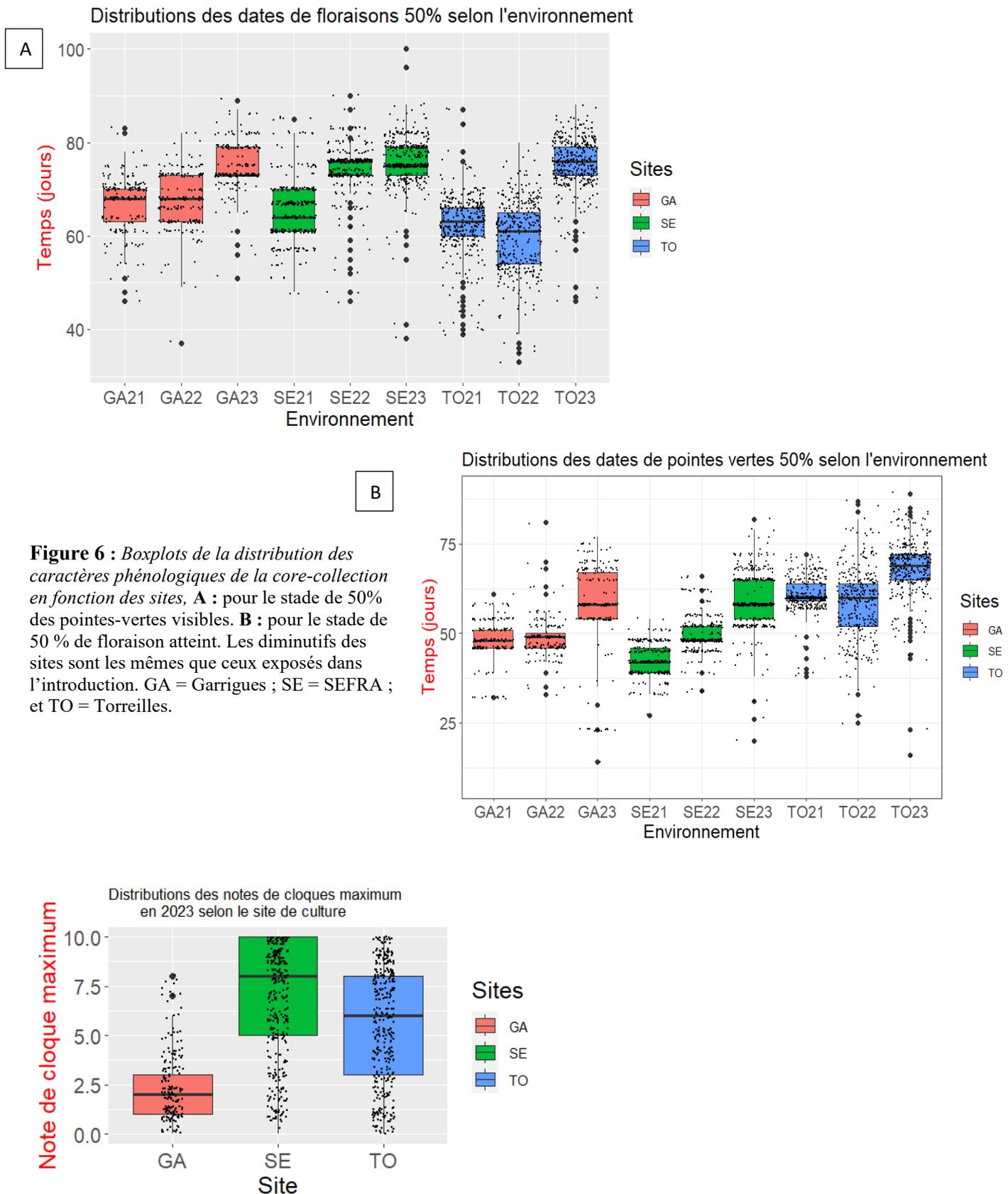


Figure 6 : *Boxplots de la distribution des caractères phénologiques de la core-collection en fonction des sites, A : pour le stade de 50% des pointes-vertes visibles. B : pour le stade de 50 % de floraison atteint. Les diminutifs des sites sont les mêmes que ceux exposés dans l'introduction. GA = Garrigues ; SE = SEFRA ; et TO = Torreilles.*

Figure 7 : *distribution des valeurs maximales de dommages de cloque de la core-collection pêcher en fonction des sites pour l'année 2023, les abréviations des cites sont respectivement SEFRA(SE), les Garrigues(GA) et Torreilles(TO).*

II.D.4.a. LE RR-BLUP

Les analyses sont conduites sur R 4.3.0 où les modèles (cf. **eq. (6)**) sont implémentés à l'aide du package RR-BLUP. En entrant ensuite les BLUPs obtenus et la matrice génotypique dans la fonction « mixed.solve » du package RR-BLUP, l'algorithme est capable de trouver les coefficients attribués à chaque marqueur. Les coefficients estimés des marqueurs sont ensuite utilisés pour la prédiction (cf. **eq. (3)** chp. **I.B.2**) sur les populations de tests choisies aléatoirement par le principe de validations croisées.

II.D.4.b. LE G-BLUP

Nous testons les 5 modèles (cf. **eq.8**) utilisés dans la décomposition de la variance pour leurs capacités à prédire. On ajoute au fur et à mesure les différentes variables explicatives **x** correspondant à des effets génétiques codées grâce à Vitezica et al., (2017) et utilisées en facteur sous formes de matrice. Le taux de consanguinité, ou Inbreeding, est le taux de marqueurs homozygote d'un individu par rapport au nombre total de marqueurs.

III. Résultats

III.A Exploration préalable des données phénologiques, de dommage de cloque et leurs liens avec la consanguinité.

On aperçoit sur la figure 6 les boxplots de la distribution des données collectées sur les différentes années. On peut constater graphiquement une variabilité intra-environnementale. L'année 2023 présente des dates plus tardives sur les 3 sites pour les 2 traits. Les données sur le site TO ont des moyennes d'apparitions de pointes vertes toujours plus élevées que celles de toutes les années sur les 2 autres sites ($p < 2^{E-16}$). Ces dernières plafonnent à environ 60 jours avant débourrement alors que les moyennes de Torreilles sont de minimum 60 jours soit débutent fin février. Pour ce qui est des dates de 50 % de floraisons, on retrouve des hétérogénéités intra- et inter-sites. La moyenne du temps nécessaires avant d'atteindre le stade de 50 % de floraison est d'environ 70 jours (soit début mars). Les valeurs extrêmes sont clairement affichées et on peut constater une large variabilité phénotypique dans la core-collection et dans chaque environnement. En ce qui concerne les dégâts de cloque (fig.7), leurs distributions sont elles aussi bien variables d'un lieu à l'autre, en lien avec les différents itinéraires technique qui changent le contexte épidémiologique.

En effet le début des traitements contre la cloque en 2023 montre aux Garrigues une diminution drastique des infections à *Taphrina deformans*. En revanche, le site le plus touché

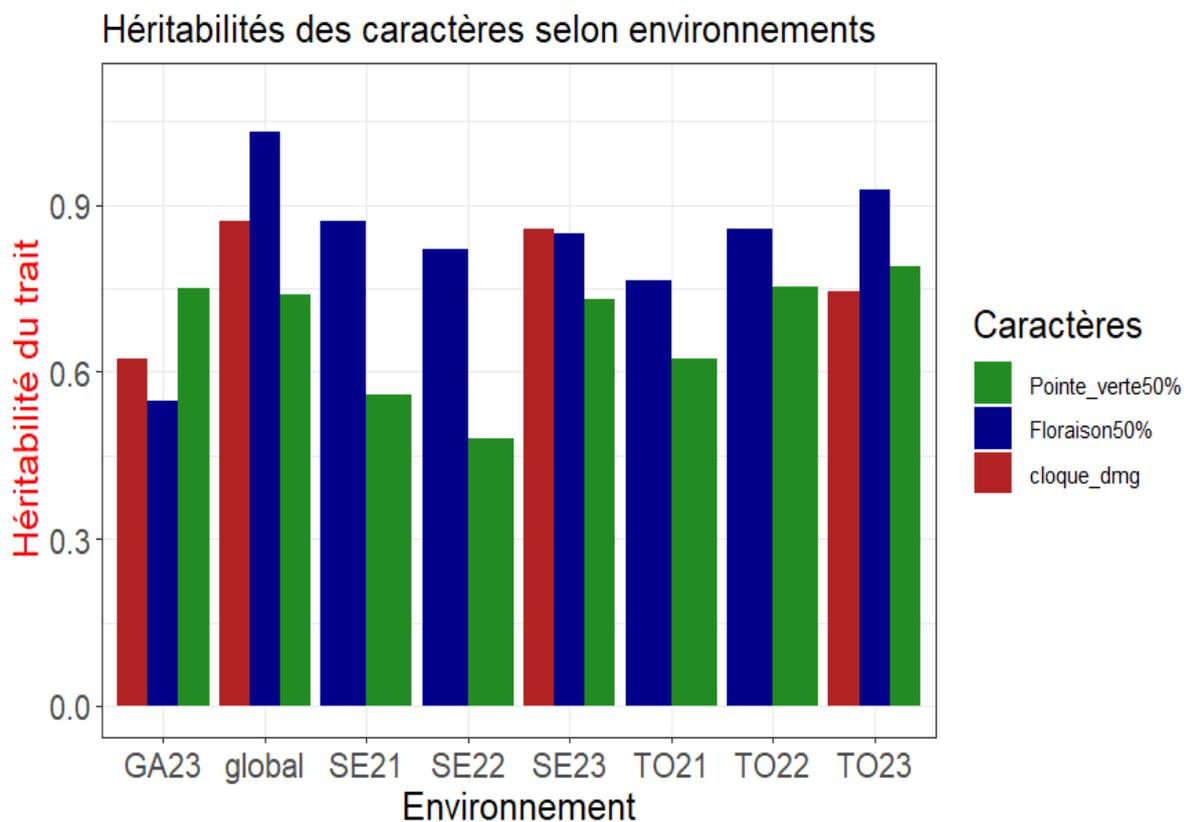


Figure 8 : Histogramme des niveaux d'héritabilités obtenues dans les différents environnements où la prédiction a été testée. Chaque couleur correspond à un trait. Les sites se lisent de gauches à droites, dans l'ordre Garrigues (GA), SEFRA (SE), et Torreilles (TO), les années sont dans l'ordre chronologique de gauche à droite, seuls les derniers chiffres sont indiqués 21, 22 ou 23 pour les années 2021, 2022 à 2023 respectivement. L'héritabilité n'est pas calculable en GA. L'environnement « global » comprend la totalité des répétitions inter-sites.

est de loin la SEFRA avec une moyenne centrée sur 7.5 et tout le contenu du 4^e quartile de sa distribution noté 10 en dégâts de cloque (fig.7).

III.B. Héritabilités des caractères selon les environnements

Dans un premier temps, la juxtaposition des héritabilités (fig.8) par environnements nous permet de nous rendre compte des caractères ayant la variance génétique la plus importante. Les héritabilités pour la date de floraison ne descendent jamais en dessous de 0.75 et sont centrées autour de 0.80 ce qui est très élevé, alors que les héritabilités du trait « stade d'apparition de 50% des pointes vertes sur l'individu » plafonnent à 0.75 et semblent plus centrées autour d'environ 0.60. On conclut donc graphiquement à une différence de variance génétique impliquée dans le polymorphisme des différents traits. Si la quantité de variance génétique impliquée est moindre pour ce trait c'est que son déterminisme génétique total (facteur G, du $P=G \times E$) est plus limité et qu'il (P) est plus dépendant de l'environnement (E). « L'environnement global » résume l'héritabilité à l'échelle du jeu de données entier. L'héritabilité globale de la tolérance à la cloque est de 0.84.

III.C Décomposition de la variance, consanguinité et apparentement

Sur la figure annexe 3, on peut lire la distribution des proportions d'homozygoties (sur la totalité des locus marqueurs) des individus de la core-collection. On voit ici une distribution de l'homozygotie entre 0.5 et 0.95. Les autres graphiques nous renseignent sur la distribution de l'apparentement au sein de la population (matrices en annexes). On aperçoit notamment sur la matrice d'additivité, par effet de regroupement (ou 'clustering'), les 3 groupes structuraux représentés dans la core-collections et connus à l'échelle de la diversité du pêcher. Les matrices d'apparentement sont employées dans la décomposition des effets génétiques grâce à une analyse de variance-covariance, puis en prédiction génomique avec le modèle GBLUP.

Nous avons effectué des tests de corrélations de Pearson entre le taux de consanguinité et les différents caractères. La floraison est ici corrélée à la consanguinité avec un r de 0.13 ($p < 0.001$) ; les pointes vertes à 0.08 ($p < 0.005$) ; et la tolérance à la cloque, cependant, serait inversement corrélée avec un r de -0.24 ($p < 9 \times 10^{-10}$). Ceci indiquerait que plus la consanguinité augmente, plus les pêchers ont tendance à débousser tardivement (au niveau floral et végétatif, tendance faible), et moins ils seraient sensibles à la cloque.

A

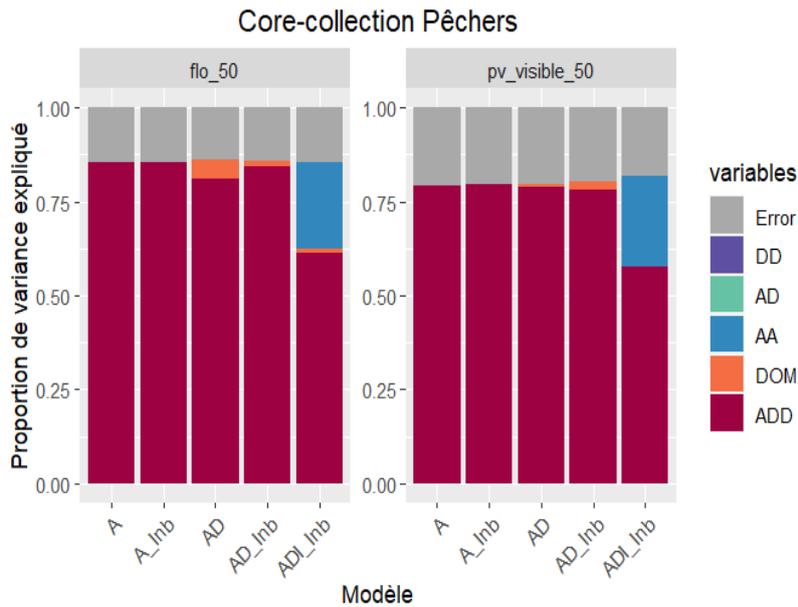
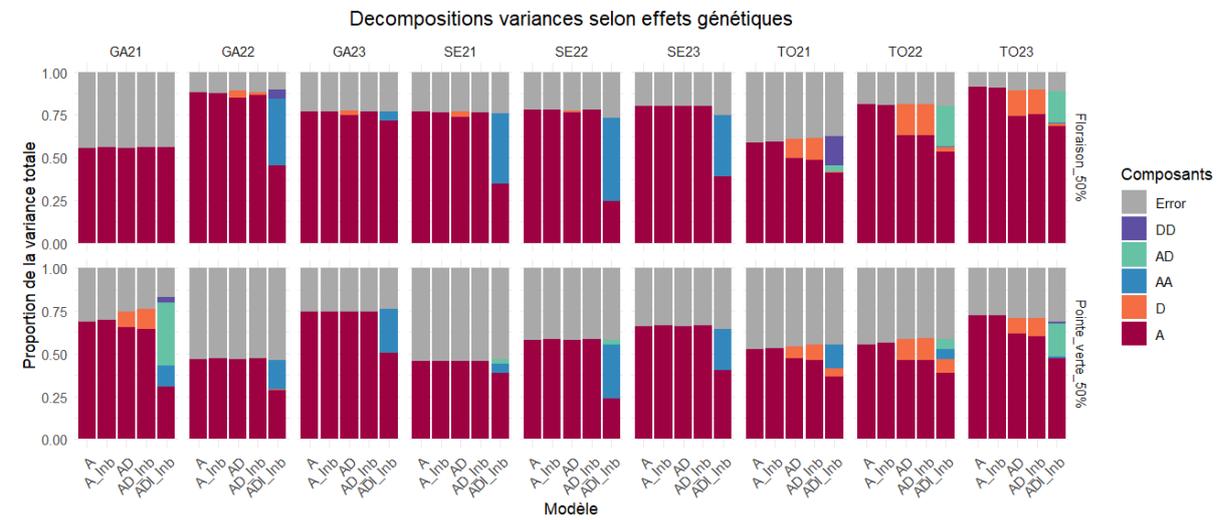


Figure 9 : Décomposition de la variance totale des caractères de phénologies par les 5 modèles utilisés (A) sur les moyennes ajustées par clone pour totalité des différents environnements, ou (B) sur des BLUP intra-environnementales. On montre ici les différentes proportions d'effets génétique impactant la variabilité de traits quantitatifs complexes (variance totale = variance phénotypique = 1)

B



A

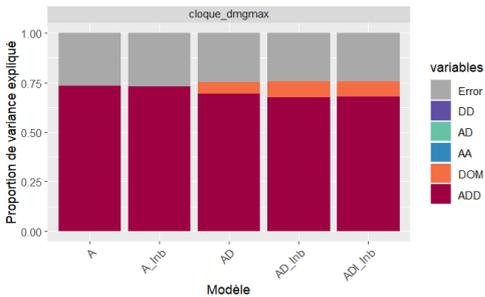
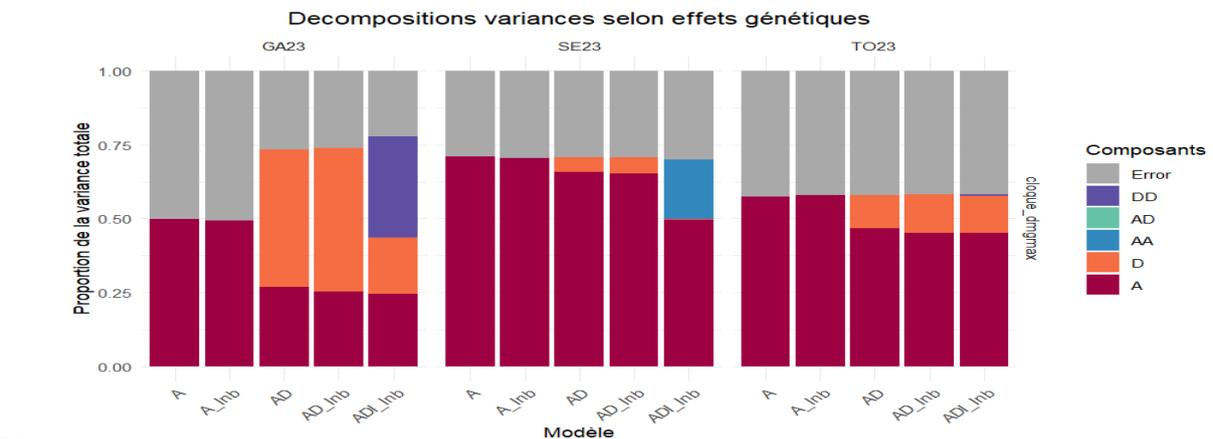


Figure 10 : Décomposition de la variance totale du caractère « dégâts maximum de cloque » par les 5 modèles utilisées (A) sur les moyennes ajustées par clone pour la totalité des différents environnements, ou (B) sur des BLUP intra-environnementales.

B



Si on prend les décompositions de la variance réalisées sur les moyennes ajustées (fig 9, A et fig 10, A), on peut constater ensuite l'importance du facteur lié à la dominance dès que les modèles le prennent en compte aussi bien pour les traits phénologiques que pour le trait lié aux dégâts de cloque. La proportion de dominance diminue ensuite dans le modèle ajoutant la consanguinité (**Inb**) avec la dominance (modèle **AD_Inb**) en ce qui concerne le trait de phénologie florale (sauf TO21), alors qu'elle augmente dans ce 4^e modèle pour les traits liés aux pointes vertes et aux dommages de cloque (fig. 9, A et fig. 10, A). Pour les deux traits phénologiques, le modèle le plus complexe (**ADI_Inb**) semble mettre en évidence un effet conséquent d'épistasie, lié à l'effet d'additivité par additivité (AA).

Afin d'analyser plus finement la variabilité intra-environnementale, les modèles ont été appliqués aux sous-sets environnementaux de notre jeu de données (fig 9 et 10, B). Pour les dates de phénologies, en général, on peut constater une forte hétérogénéité des résultats selon les environnements (site/année). Si la variance liée à la dominance est bien toujours présente, on voit ici apparaître des variances d'épistasie de type AA et AD voir plus rarement de DD. Ces variances semblent site-dépendantes, en effet l'AA semble revenir à la SEFRA alors que l'AD semble plus spécifique de Torreilles. La dominance bien qu'inférée, comme statistiquement significative par le modèle, au site GA est aussi est très importante à TO mais quasi-absente de SE. En ce qui concerne la sensibilité à la cloque (fig.10) l'impact majeur de la dominance pour le trait cloque observée en fig. 10 A, est bien confirmé ici en B malgré une apparition d'effet d'épistasie DD et AD. Pour ce dernier trait, l'épistasie n'intervient que lors des VD par environnements (fig.10). La proportion de variance génétique totale peut être augmentée, et la variance d'erreur réduite comme en GA23 de la figure 10, B.

On rappelle que les résultats d'analyse de variance pour le site des Garrigues, du fait de l'unique répétition du dispositif de chaque génotype, doivent être interprétés avec précaution.

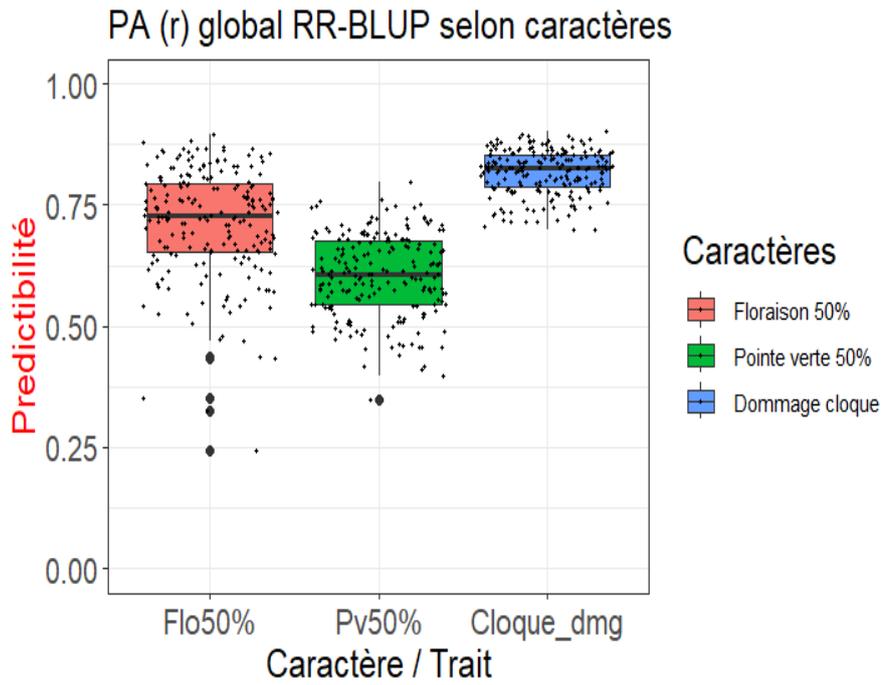


Figure 11 : Précision de prédiction avec la méthode RR-BLUP pour les différents traits grâce à un modèle entraîné sur l'ensemble des 3600 individus du jeu de données. 200 corrélations (PAs) par boxplot.

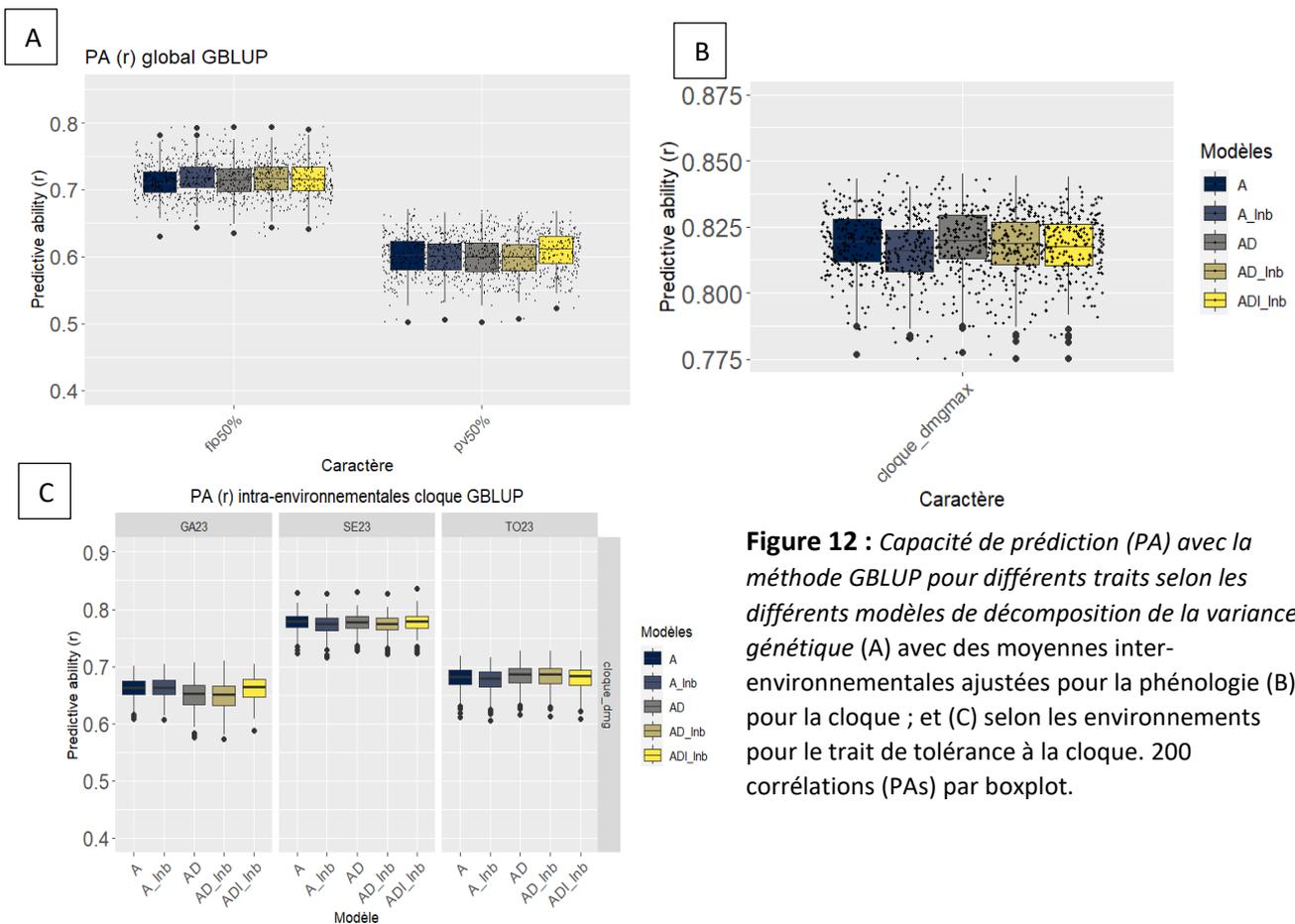


Figure 12 : Capacité de prédiction (PA) avec la méthode GBLUP pour différents traits selon les différents modèles de décomposition de la variance génétique (A) avec des moyennes inter-environnementales ajustées pour la phénologie (B) pour la cloque ; et (C) selon les environnements pour le trait de tolérance à la cloque. 200 corrélations (PAs) par boxplot.

III.D. Les capacités de prédictions : prediction accuracies (PAs)

III.D.1 Conclusion et comparaison sur les PAs globales

RR-BLUP

En ce qui concerne les PAs globales obtenues via la méthode RR-BLUP avec un entraînement sur les 2651 observations (fig. .11), on constate sur les graphiques une meilleure capacité à prédire le trait de date de floraison (r moyen de 0.75, écart-type de 0.11) par rapport à celui de date de sortie de pointes vertes (r moyen de 0.6, écart-type de 0.09). Les meilleures performances sont obtenues pour les dégâts de cloque avec des PAs centrées autour de 0.8. L'écart-type est aussi le plus faible (0.05) pour la distribution des valeurs de précision de prédiction obtenues.

GBLUP

Pour la méthode GBLUP cette fois, on n'observe pas de différences notables entre les prédictions des différents modèles pour chacun des caractères : les capacités de prédictions semblent en général équivalentes pour les 5 modèles (fig.12). Les précisions de prédictions restent conséquentes et comprises entre 0.6 et 0.85, ce qui démontre que les modèles arrivent efficacement à prédire les valeurs phénotypiques d'individus non-observés. Le modèle le plus complexe (couleur jaune, **ADI_Inb**) n'est pas forcément le modèle à même de prédire le mieux, excepté pour la date de débourrement foliaire (trait « pv50% », fig. 12. A).

III.D.2. PAs obtenues sur les résultats intra-environnementaux

Pour le trait lié aux pointes vertes, on constate que la totalité des moyennes de PA sont comprises entre 0.4 et 0.55. Si on regarde pour la cloque, on aperçoit que les PAs moyennes sont comprises entre 0.65 et 0.75. Les meilleures précisions de prédiction sont observées dans les environnements SE22 et TO23 pour la date de floraison et à T023 en ce qui concerne la date de sortie de pointes vertes (fig 13). Il semble que les dégâts de cloque, notés uniquement en 2023, soient mieux à même d'être prédits dans le site de la SEFRA (« SE23 », Fig. 12C). Pour la floraison, les moyennes sont comprises entre 0.6 et 0.7 sauf en ce qui concerne l'environnement TO21 dont la moyenne se dégrade par rapport aux autres. Malgré l'absence de répétition des clones aux Garrigues, les précisions de prédiction ne semblent pas fortement dégradées, à l'exception du trait de sorties de pointes vertes (fig. 14). Le modèle le plus complexe de prédiction GBLUP (**ADI_Inb**) ne prédit pas souvent mieux que le reste (fig. 12, C ; fig. 13 ; fig 14). La cinétique des PAs en fonction de la complexification des modèles semble parfois inversé d'une année à l'autre (TO21 s'oppose à TO22 et TO23 pour le trait floraison50% et SE21 s'oppose à SE22 et SE23 pour le trait pointe_verte50%).

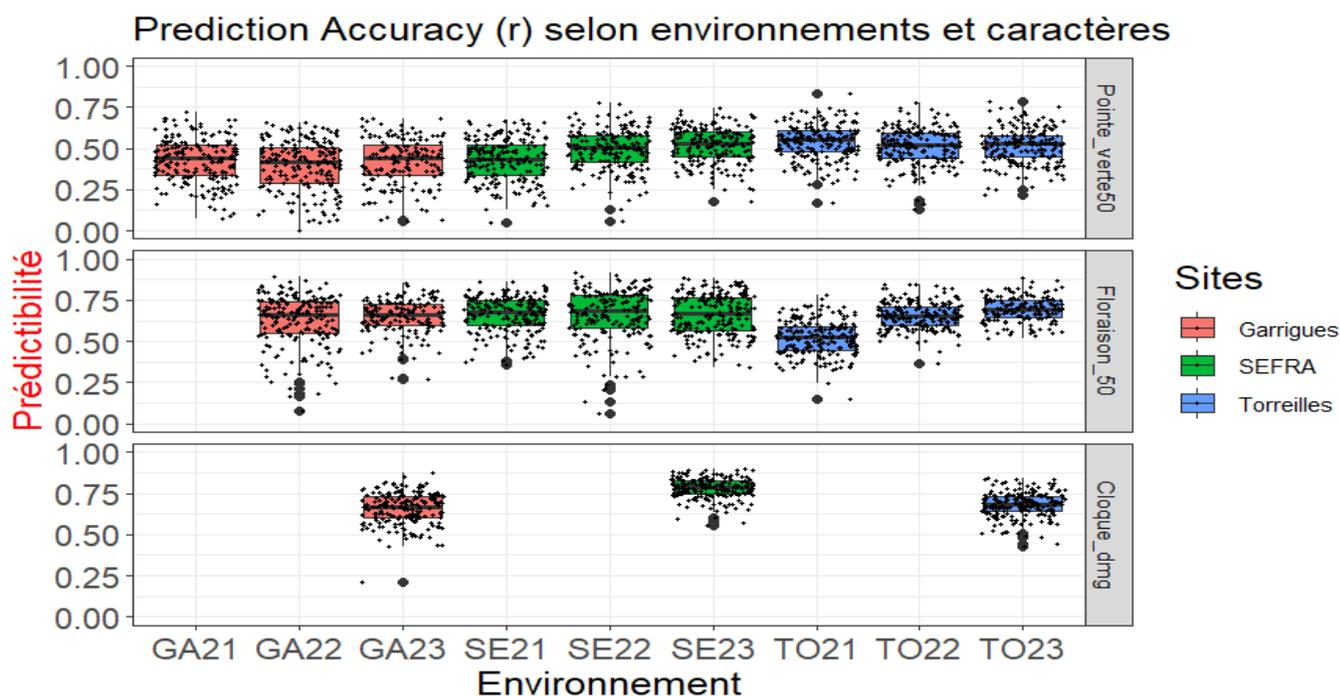


Figure 13 : Prediction accuracies (coefficient r de Pearson, répétées 200 fois) obtenue par la méthode RR-BLUP pour les différents caractères phénotypiques selon les environnements du dispositif core-collection. Les sites sont ceux cité en II.B. ; pointe_verte50 = date de phénologie code BBCH 09 ; floraison_50 = date de phénologie BBCH 65 ; Cloque_dmg = intensité des dommages liés à la cloque

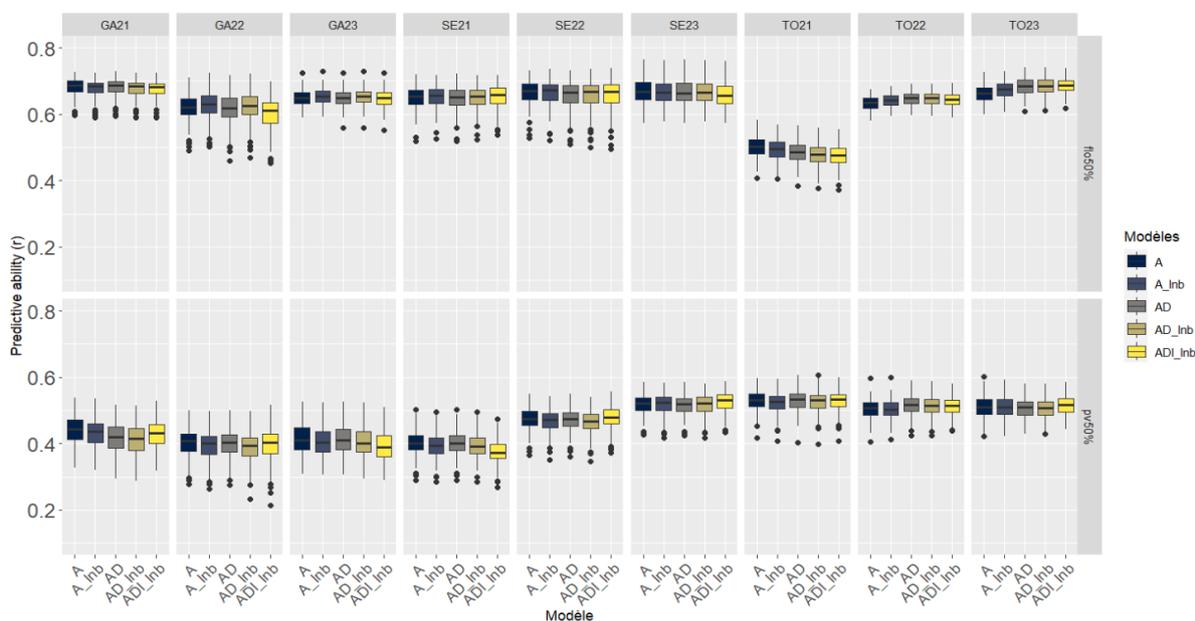


Figure 14 : Prediction Accuracies (r , répétées 200 fois) obtenues par la méthode GBLUP pour les différents caractères de phénologies selon les environnements du dispositif core-collection pêcheurs. Les sites sont ceux cité en II.B. ; et les modèles cités en II.D.3 ; pv = date de sortie des pointes vertes ; fio = date de floraison.

IV. Discussion

IV.A. Pertinence de la méthode, du dispositif expérimental, et corrections statistiques

Pour investiguer les potentielles efficacités de la GS, il nous a donc fallu tester par empirisme la qualité des prédictions sur des caractères spécifiques. Pour cela, une première approche est d'identifier les caractères les plus fastidieux à sélectionner pour compenser les problématiques de la filière de sélection de l'espèce travaillée. C'est pourquoi ces travaux sur la prédiction génomique sont les prémices d'une plus large investigation sur les données accumulées lors des années d'observations de la core-collection.

La culture de la collection est établie en faibles nombres de répétitions intra-environnementales (1 à 2). L'analyse des données comme un ensemble de blocs d'un même site augmente le nombre de répétitions globales et permet une plus grande robustesse statistique. Malgré les problèmes d'estimations de la variance génétique aux Garrigues liés à la présence d'un seul arbre par génotype, intégrer ce site aux données y est possible et contribue à améliorer l'estimation des effets génétiques à travers le dispositif. En effet, dans un dispositif multisite, on peut considérer que les données enregistrées sur ce site font office d'une répétition supplémentaire, tel un bloc. La VD peut donc être effectuée en considérant les différentes répétitions inter-environnementales, toutefois, il ne sera pas possible d'estimer de variance d'interaction $G \times E$ en incluant les garrigues. En tout cas, les deux méthodes de prédictions se révèle très efficaces même au Garrigues, ce qui confirme l'hypothèse comme quoi le dispositif expérimental a un bon potentiel de prédiction. La validation croisée semble très efficace grâce à la variabilité de la core-collection et du génotypage. Il est légitime de se demander si, dans un programme d'amélioration, un modèle entraîné sur la core-collection serait applicable sur une population de sélection biparentale. Sachant la diversité génétique de cette première, l'apparementement serait plus faible que lors de la validation croisée appliquée ici.

Les perdus de vue sont aléatoires : certains génotypes sont donc absents de certains blocs. Ici, deux designs sont pris en compte pour les analyses de « prediction accuracy » : Soit (i) tous les environnements sont considérés par le biais de valeurs moyennées ; soit (ii) les prédictions sont faites à l'intérieur d'un environnement ce qui réduit drastiquement

le nombre de répétitions par géotypes. L'héritabilité est estimable sur tous les sites sauf en GA. Cela influence les prédictions intra-environnementales RRBLUP car celles-ci prennent en compte le nombre de répétitions des individus. Le modèle de prédiction subit un entraînement moins robuste en GA et la part génotypique expliquée est réduite tout comme en GA21 (flo), GA22 (pv) (fig. 9 B) et GA23 (cloque) (fig. 10 B). Pour pouvoir corriger l'héritabilité par l'hétérogénéité du nombre de répétitions, il faut ajuster l'équation par le nombre moyen de répétitions par environnement (cf **eq.7**).

IV.B. Interprétation des différents résultats d'architectures génétiques et de prédictions pour chaque trait :

IV.B.1 Les héritabilités pour identifier la proportion du facteur génétique :

Si une héritabilité moindre peut indiquer une situation où la GS sera limitée, l'inverse est aussi possible, si l'héritabilité est très élevée la MAS peut suffire et le caractère peut-être trop facilement héritable (Oligogénique). L'héritabilité est un paramètre qui sert à identifier au préalable les caractères régis par un déterminisme génétique complexe. Cela peut aider à identifier les caractères sur lesquels la GS pourrait être appliquée. Les traits sélectionnés ici contrastent en termes d'héritabilités par site, nous nous attendons donc à observer des PAs contrastantes. Les environnements peuvent aussi faire fluctuer les valeurs : il y a effectivement des différences de PAs entre les années de cultures sur un même site. On conclut ici que ces traits semblent prometteurs pour la GS, car les fluctuations de ces héritabilités souvent comprises entre 0.5 et 0.7 nous indiquent de potentiels caractères au déterminisme génétique complexe mais aussi assez conséquent pour une sélection efficace.

IV.B.2. Décomposition de la variance et hypothèse sur l'architecture des traits :

Les modèles mixtes expliquant les différentes proportions d'effets génétiques (VD) représentent aussi la part de variation génétique par rapport à la part de variation phénotypique totale. On peut donc avoir une approximation de l'héritabilité graphiquement sur la plupart des figures de décomposition de la variance. Il faut bien garder en tête que les effets d'épistasie sont directement liés aux effets d'additivité et de dominance (Roth et al., 2022). Une proportion de l'effet additif est redondante avec l'effet épistatique d'additivité par additivité, ce qui se traduit par un gain très faible ou nul d'héritabilité lorsqu'on ajoute l'épistasie dans le modèle (**ADI_Inb**). On voit que la dominance est impliquées dans la tolérance à la cloque (fig 10), bien que le site des Garrigues réduise sa part pour l'inférer au facteur DD. L'architecture des traits de phénologie varie en fonction des environnements (fig. 9, B). Ils semblent à moindre mesure

impactés par la dominance. Les sites SE et TO contrastent totalement dans l'estimation des effets d'épistasie.

IV.B.3 Interprétations des données de PAs

Ces modèles linéaires mixtes ont ensuite été utilisés pour réaliser des prédictions génomiques. Il semblerait que les 5 modèles de GBLUP (fig.12 C ; fig 14) n'apportent pas davantage de capacité à prédire comparé au modèle RR-BLUP (fig 13), et ce dans les prédictions intra- comme en inter-environnementales. La méthode GBLUP n'étant pas à même de prédire d'avantage que le RRBLUP il n'y a pas de raison de privilégier cette première méthode au vu des faibles différences graphiques de PA. En effet, la prise en compte des matrices d'effets génétiques ne confère pas une meilleure précision de prédiction. Peut-être peut-on optimiser les modèles linéaires mixtes en utilisant des répétitions et pas des moyennes ajustées ?

Il semble efficace, via la sélection génomique, de sélectionner des pêcheurs aux différentes phénologies de floraisons ou de débourrements végétatifs. La phénologie de floraison et la tolérance à la cloque sont ici les caractères qui semblent les plus à même d'être sélectionnés par la GS et les PAs obtenues sur la cloque sont les plus élevées (atteignant 0.8 avec modèle global, fig. 11 et 12, B). Cela nous donne bon espoir de pouvoir réussir à développer efficacement des génotypes plus tolérants grâce à la GS. On pourrait sélectionner les dates de pointes vertes assez efficacement aussi et on sait que ces dates sont importantes car elles influent sur le moment de la contamination par *Taphrina deformans*, la cloque du pêcheur, un des pathogènes majeurs de l'espèce pêcheur.

L'étude de Jung et al., (2022) nous montre les PAs pour 30 traits quantitatifs chez la pomme. Ils sembleraient que ceux-ci fluctuent entre 0.18 et 0.88, mais très peu de traits atteignent une PA moyenne supérieure à 0.8. Les traits liés à la date de floraison sont prédits avec moins de précision que dans le travail présenté, mais les méthodes RR-BLUP ou G-BLUP qui sont connues comme une référence en termes de capacité de prédiction n'y ont pas été testées. Kumar et al., (2012b) nous montre que le RR-BLUP permet d'estimer les GEBV pour différents caractères de la pomme avec une précision souvent supérieure à 0.8. Dans l'article de référence de (Roth et al., 2022) la moyenne des PAs pour la phénologies de floraison chez le maïs est de 0.75 ce qui est plus élevé qu'ici, peut-être est-ce liée à l'utilisation d'unité degré-jour qui confère une plus forte comparabilité entre environnements ? Nsibi (2021) sur une population biparentale, nous montre une précision de prédiction de 0.7 pour ce qui est de la date de floraisons en GS.

IV.C. Intérêt en sélection pour la filière

Les résultats sur le déterminisme des dégâts de cloque et sur la capacité à les prédire sont très encourageants pour l'avenir de la GS contre cette maladie voire pour des problématiques plus larges de sélection de tolérance ou résilience. Fu et al., (2022) ont par exemple montré que malgré de nombreuses méthodes de GS expérimentées, les dégâts liés à *Monilina spp.* chez la pêche sont difficiles à prédire (PA à 0.3 maximum). C'est aussi le cas pour différentes maladies chez l'abricotier investiguées dans la thèse de Nsibi (2021) tel que le Mildiou, Monilina, la rouille, ou le chancre bactérien, avec des PAs moyennes de 0.4 , 0.2 , 0.6 et 0.1 respectivement.

Jusqu'alors en plus d'être souvent contournée quand introduite par des méthodes de sélection par phénotypage ou même MAS sur gènes majeurs uniques (Kumar et al., 2012a; McClure et al., 2014; Migicovsky and Myles, 2017; Kumar et al., 2020), les améliorations des résistances sont très coûteuses à obtenir, comme toute sélection chez une pérenne (McClure et al., 2014). Avec des marqueurs répartis dans tout le génome, la GS a le potentiel de travailler sur des formes de résiliences et de résistances les plus complexes (et ainsi plus ou moins héritable) grâce à la sélection de nombreux locus influençant le phénotype. Les traits doivent certes être évalués 1 à 1 pour leurs habilités à être sélectionnés par GS dans un programme d'amélioration, mais l'automatisation des prédictions génomiques par des pipelines bio-informatiques régulièrement mis à jour pourrait permettre de mettre en place une stratégie de GS en partenariat avec les filières de sélection.

IV.D. Conclusion et future application de la démarche

Nous avons pu étudier le déterminisme génétique lié aux traits de phénologie et de sensibilité à la cloque. Cela peut nous apporter des indices fondamentaux sur le fonctionnement général de ces traits quantitatifs complexes et nous indiquer des modèles pour renforcer les méthodes de prédictions chez cette espèce dans notre contexte expérimental voire à plus large échelle (autre pérennes, autres variables à expliquer...).

Le dispositif de cette étude nous confère une robustesse statistique conséquente grâce à un grand nombre de répétitions de génotypes à travers les environnements, avec une haute variabilité des phénotypes exprimés en verger, reflet de la diversité génétique mobilisée. Nous sommes optimistes quant à l'application de la présente méthode à d'autres caractères, dont des bioindicateurs de résilience.

À long terme, le dispositif pourra contribuer à sélectionner efficacement des variétés résilientes face aux maladies pour réduire les besoins en intrants phytosanitaires des cultures fruitières pérennes. Les résultats de prédictions des dégâts de cloque doivent être consolidés mais sont encourageants quant à l'application de la GS pour l'amélioration d'un trait quantitatif complexe d'importance majeure pour la filière.

La core-collection pêcher étudiée ici est gérée par INRAE en parallèle d'une core-collection abricotier (présentée en figure 3). Une prochaine étape serait d'appliquer la méthode présentée dans mon travail à l'abricotier grâce à ce dispositif pour pouvoir comparer les différents résultats entre ces espèces très apparentées, mais aux régimes de reproductions bien différents.

v. Références bibliographiques

- Aranzana MJ, Decroocq V, Dirlewanger E, Eduardo I, Gao ZS, Gasic K, Iezzoni A, Jung S, Peace C, Prieto H, et al** (2019) Prunus Genetics and Applications after de Novo Genome Sequencing: Achievements and Prospects. *Horticulture Research* **6**: 58
- Ashraf M, Foolad MR** (2013) Crop breeding for salt tolerance in the era of molecular markers and marker-assisted selection. *Plant Breeding* **132**: 10–20
- Byrne DH, Raseira MB, Bassi D, Piagnani MC, Gasic K, Reighard GL, Moreno MA, Pérez S** (2012) Peach. *Fruit Breeding*. Springer, Boston, MA, pp 505–569
- Cao K, Zheng Z, Wang L, Liu X, Zhu G, Fang W, Cheng S, Zeng P, Chen C, Wang X, et al** (2014) Comparative Population Genomics Reveals the Domestication History of the Peach, *Prunus Persica*, and Human Influences on Perennial Fruit Crops. *Genome Biology* **15**: 415
- Ctifl** (2012) Monographie Ctifl sur l’abricot. <https://www.ctifl.fr/l-abricot-monographie>
- Elsen J-M** (2022) Genomic Prediction of Complex Traits, Principles, Overview of Factors Affecting the Reliability of Genomic Prediction, and Algebra of the Reliability. *Methods Mol Biol* **2467**: 45–76
- Endelman JB** (2011) Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. *The Plant Genome*. doi: 10.3835/plantgenome2011.08.0024
- FAOSTAT** (2023) FAO, statistiques agricoles : pêches, nectarines et abricots dans le monde jusqu’à aujourd’hui.
- Fu W, da Silva Linge C, Lawton JM, Gasic K, Department of Plant and Environmental Sciences, Clemson University, Clemson, SC 29634-0002, USA** (2022) Feasibility of Genomic Prediction for Brown Rot (*Monilinia Spp.*) Resistance in Peach. *Fruit Research* **2**: 1–12
- Grattapaglia D** (2017) Status and Perspectives of Genomic Selection in Forest Tree Breeding. *Genomic Selection for Crop Improvement: New Molecular Breeding Strategies for Crop Improvement*. pp 199–249
- Habier D, Fernando RL, Kizilkaya K, Garrick DJ** (2011) Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics* **12**: 186
- International Peach Genome Initiative, Verde I, Abbott AG, Scalabrin S, Jung S, Shu S, Marroni F, Zhebentyayeva T, Dettori MT, Grimwood J, et al** (2013) The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nat Genet* **45**: 487–494
- Jung M, Keller B, Roth M, Aranzana MJ, Auwerkerken A, Guerra W, Al-Rifaï M, Lewandowski M, Sanin N, Rymenants M, et al** (2021) Genetic Architecture and Genomic Prediction Accuracy of Apple Quantitative Traits across Environments. doi: 10.1101/2021.11.29.470309
- Jung M, Keller B, Roth M, Aranzana MJ, Auwerkerken A, Guerra W, Al-Rifaï M, Lewandowski M, Sanin N, Rymenants M, et al** (2022) Genetic architecture and genomic predictive ability of apple quantitative traits across environments. *Horticulture Research* **9**: uhac028

- Jung S, Lee T, Cheng C-H, Buble K, Zheng P, Yu J, Humann J, Ficklin SP, Gasic K, Scott K, et al** (2019) 15 years of GDR: New data and functionality in the Genome Database for Rosaceae. *Nucleic Acids Res* **47**: D1137–D1145
- Kumar S, Bink MCAM, Volz RK, Bus VGM, Chagné D** (2012a) Towards Genomic Selection in Apple (*Malus × domestica* Borkh.) Breeding Programmes: Prospects, Challenges and Strategies. *Tree Genetics & Genomes* **8**: 1–14
- Kumar S, Chagné D, Bink MCAM, Volz RK, Whitworth C, Carlisle C** (2012b) Genomic Selection for Fruit Quality Traits in Apple (*Malus × domestica* Borkh.). *PLOS ONE* **7**: e36674
- Kumar S, Hilario E, Deng CH, Molloy C** (2020) Turbocharging introgression breeding of perennial fruit crops: a case study on apple. *Horticulture Research* **7**: 47
- Kumar S, Molloy C, Muñoz P, Daetwyler H, Chagné D, Volz R** (2015) Genome-Enabled Estimates of Additive and Nonadditive Genetic Variances and Prediction of Apple Phenotypes Across Environments. *G3 (Bethesda)* **5**: 2711–2718
- Longin CFH, Mi X, Würschum T** (2015) Genomic selection in wheat: optimum allocation of test resources and comparison of breeding strategies for line and hybrid breeding. *Theor Appl Genet* **128**: 1297–1306
- Martinez-Gomez P, Arulsekhar S, Potter D, Gradziel TM** (2003) Relationships among Peach, Almond, and Related Species as Detected by Simple Sequence Repeat Markers. *Journal of the American Society for Horticultural Science*
- Marulanda J, Mi X, Melchinger A, Xu J-L, Würschum T, Longin F** (2016) Optimum breeding strategies using genomic selection for hybrid breeding in wheat, maize, rye, barley, rice and triticale. *Theoretical and Applied Genetics*. doi: 10.1007/s00122-016-2748-5
- Marulanda J, Mi X, Utz H, Melchinger A, Würschum T, Longin F** (2021) Optimum breeding strategies using genomic and phenotypic selection for the simultaneous improvement of two traits. *Theoretical and Applied Genetics*. doi: 10.1007/s00122-021-03945-5
- McClure KA, Sawler J, Gardner KM, Money D, Myles S** (2014) Genomics: a potential panacea for the perennial problem. *Am J Bot* **101**: 1780–1790
- Meuwissen TH, Hayes BJ, Goddard ME** (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**: 1819–1829
- Migicovsky Z, Myles S** (2017) Exploiting Wild Relatives for Genomics-assisted Breeding of Perennial Crops. *Frontiers in Plant Science* **8**:
- Miller AJ, Gross BL** (2011) From forest to field: perennial fruit crop domestication. *Am J Bot* **98**: 1389–1414
- Nakaya A, Isobe SN** (2012) Will genomic selection be a practical method for plant breeding? *Annals of Botany* **110**: 1303–1316
- Nsibi M, Duchêne E, David J, Dirlewanger E, Moreau L, Bouffier L, Audergon J-M, Regnard J-L, Pot D, Sauvage C** (2021) Thèse Mariem Nsibi. PhD Thesis

- Roth M, Beugnot A, Mary-Huard T, Moreau L, Charcosset A, Fiévet JB** (2022) Improving genomic predictions with inbreeding and nonadditive effects in two admixed maize hybrid populations in single and multienvironment contexts. *Genetics* **220**: iyac018
- Roth M, Muranty H, Di Guardo M, Guerra W, Patocchi A, Costa F** (2020) Genomic prediction of fruit texture and training population optimization towards the application of genomic selection in apple. *Horticulture Research* **7**: 148
- Schmidt P, Hartung J, Rath J, Piepho H-P** (2019) Estimating Broad-Sense Heritability with Unbalanced Data from Agricultural Cultivar Trials. *Crop Science* **59**: 525–536
- Serrie M** (2021) Identification and characterization of phenotypic and genetic components of resilience in apricot and peach by using genome wide association studies. other. Institut Agro - Montpellier SupAgro
- Shulaev V, Korban SS, Sosinski B, Abbott AG, Aldwinckle HS, Folta KM, Iezzoni A, Main D, Arús P, Dandekar AM, et al** (2008) Multiple Models for Rosaceae Genomics. *Plant Physiology* **147**: 985–1003
- Vitezica ZG, Legarra A, Toro MA, Varona L** (2017) Orthogonal Estimates of Variances for Additive, Dominance, and Epistatic Effects in Populations. *Genetics* **206**: 1297–1307

VI. Annexe

$$\begin{aligned} A &\sim N(0, \sigma_A^2 K_A), \quad D \sim N(0, \sigma_D^2 K_D), \quad \varepsilon \sim N(0, \sigma_\varepsilon^2 I) \\ AA &\sim N(0, \sigma_{AA}^2 K_{AA}), \quad AD \sim N(0, \sigma_{AD}^2 K_{AD}), \quad DD \sim N(0, \sigma_{DD}^2 K_{DD}) \\ A &\perp D \perp AA \perp AD \perp DD \perp \varepsilon \end{aligned}$$

Figure annexe 1 :

Définitions des paramètres des équations de décomposition de la variance.

$$\begin{aligned} G &\sim N(0, \sigma_g^2 I) \\ R &\sim N(0, \sigma_r^2 I) \\ C &\sim N(0, \sigma_c^2 I) \\ E &\sim N(0, \sigma_e^2 I) \\ G &\perp R \perp C \perp E \end{aligned}$$

Figure annexe 2 :

Définitions des paramètres du modèle de correction des valeurs phénotypiques par les effets spatiaux (R et C)

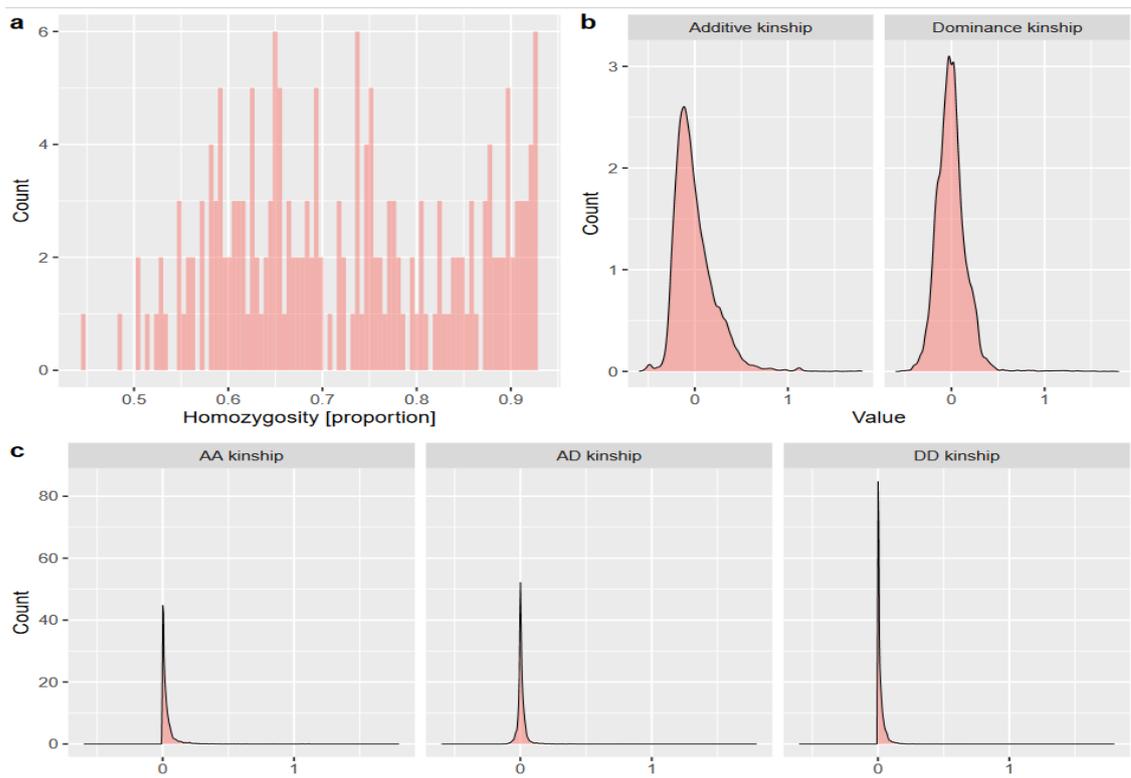
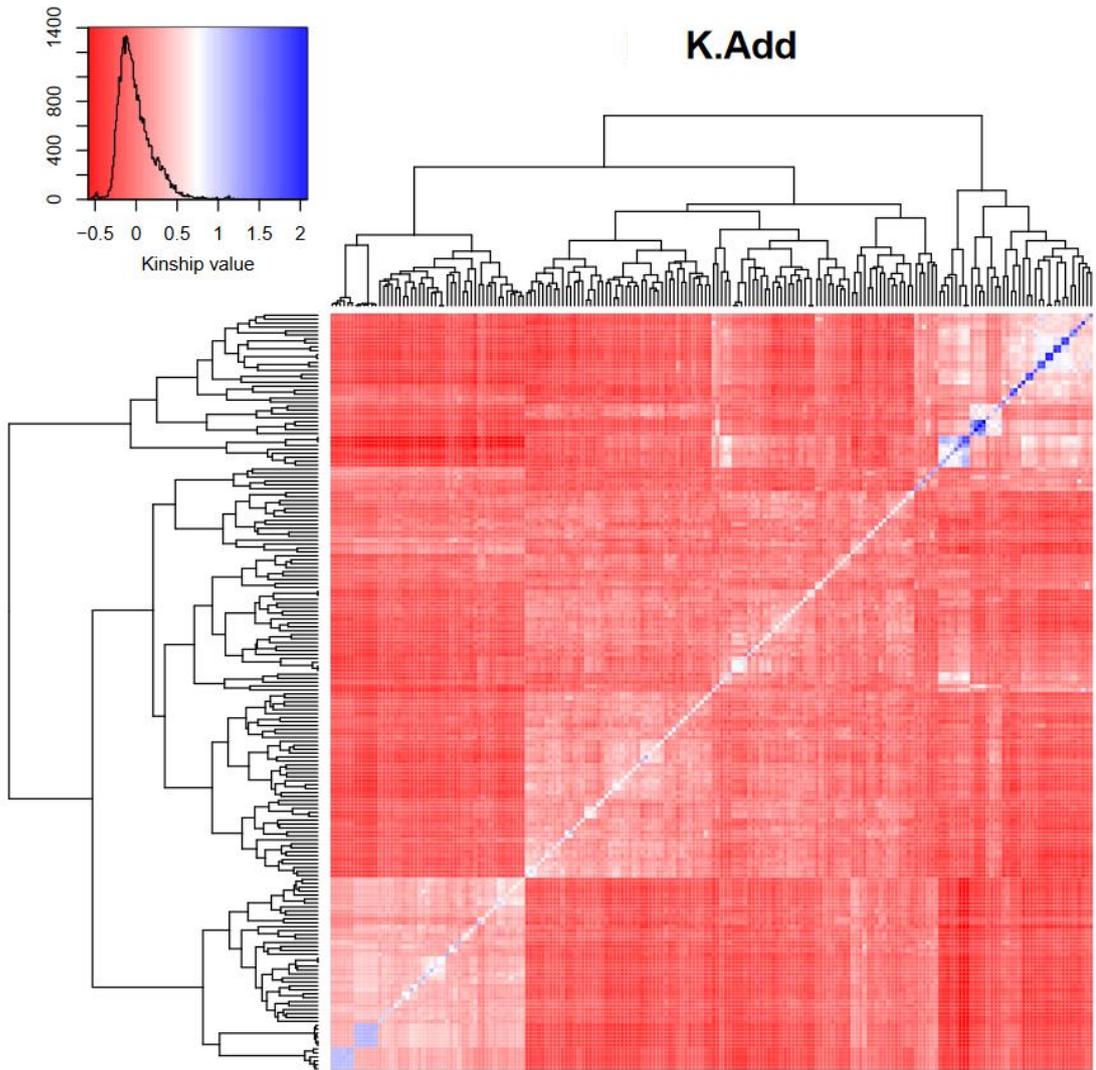
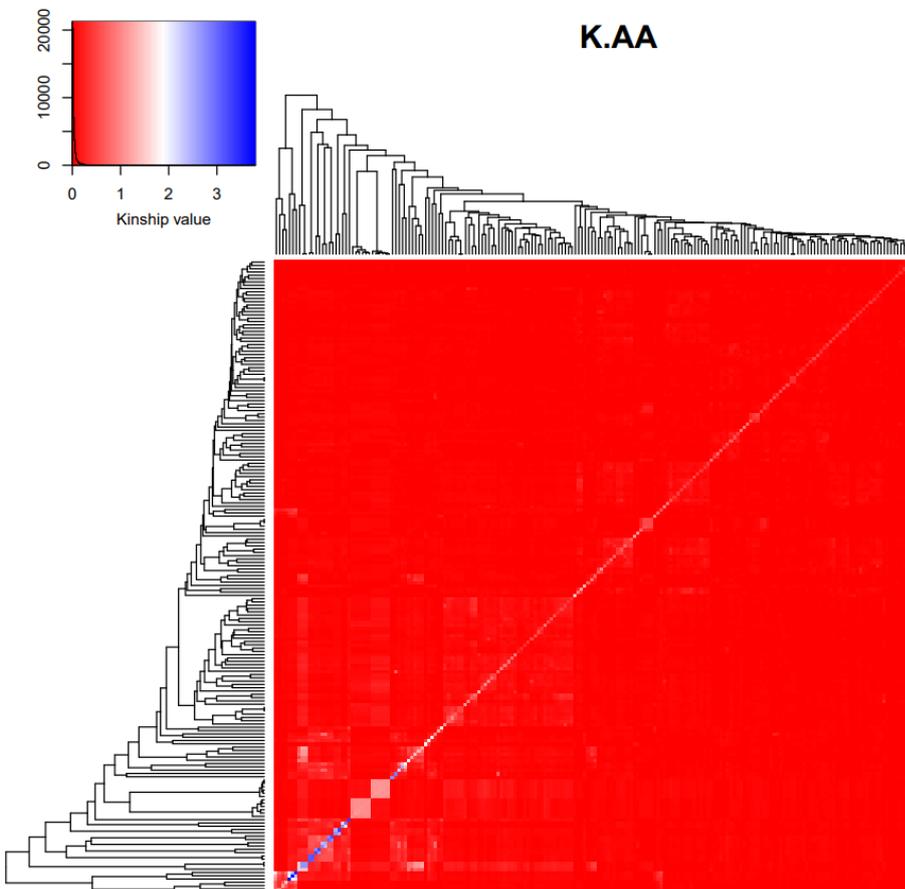
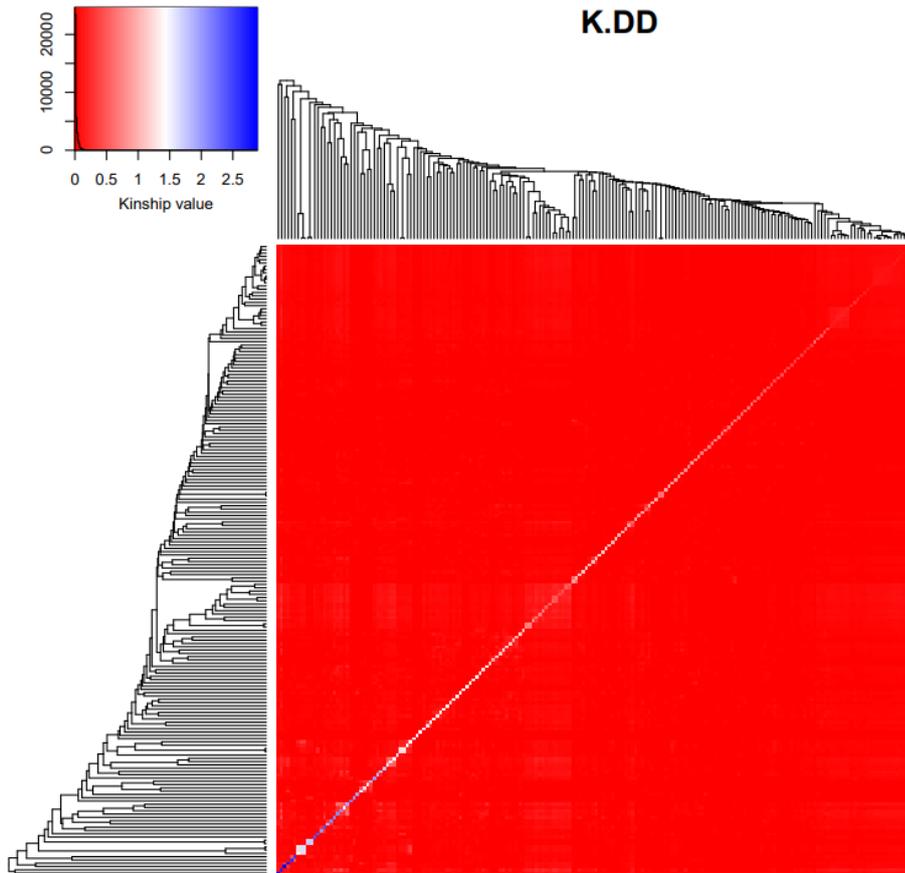


Figure annexe 3 : Distribution des valeurs du vecteur de consanguinité et des valeurs d'apparementement deux à deux pour les différents effets génétiques employés dans les modèles linéaires mixtes ; (a) décompte des proportions d'homozygotie dans la population ; (b) Valeurs des matrices d'effets additif et de dominance ; (c) valeurs des matrices d'effets épistatiques (AA , AD et DD)





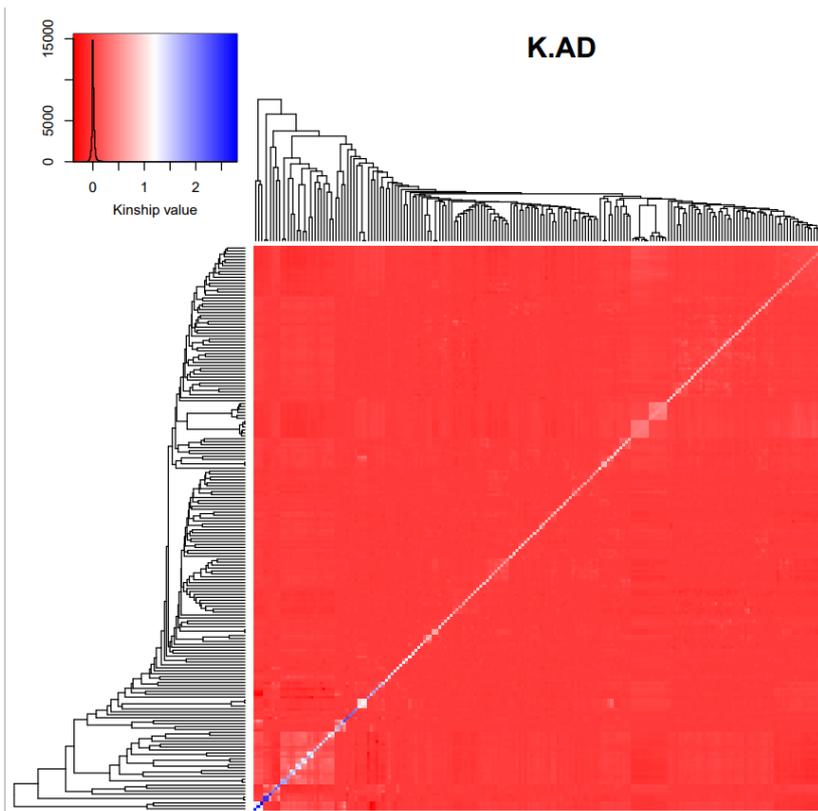


Figure annexe 4 : Ensemble de « heatmap » représentant les degrés d'apparentements des individus de la core-collection de pêcher pour leurs valeurs d'effets additif ou non-additifs. Ces cartes sont utilisées comme cofacteurs dans les modèles linéaires de décomposition de la variance. Dans ces travaux le calcul des effets liés aux matrices de génotypes est effectué selon Vitezica et al., (2017). Le K signifie Kinship, ou apparentement. A signifie Additivité et D Dominance.

Résumé :

Dans l'optique d'optimiser les programmes de sélection sur plantes pérennes agricoles majeurs nous investiguons les potentialités d'applications de la sélection génomique sur trois traits grâce à une core-collection de pêchers implantée et maintenue dans des conditions de bas intrant phytosanitaire sur trois sites, à travers trois années, dans le sud-est de la France. Des prédictions génomiques sont conduites sur deux traits de phénologies (date de floraison et date d'apparition des pointes vertes, BBCH 65 et 09) et sur la tolérance à la cloque (*Taphrina deformans*). La core-collection étudiée est composée de 206 accessions, dont 192 ont été génotypées avec la puce 16K IRSC. Les niveaux d'apparements d'additivité, de dominance et d'épistasie sont estimés selon Vitezica et al 2017. Ces apparements ont été utilisés pour obtenir des décompositions de la variance génotypique sur laquelle nous analyserons l'influence de la consanguinité. Des prédictions sont ensuite conduites avec cinq modèles de GBLUP et un de RRBLUP. Elles sont effectuées à l'intérieur des environnements ou à travers les répétitions du dispositif core-collection. Les effets de dominance sont importants dans l'explication des variances. On obtient de bonnes précisions de prédictions pour tous les traits : date de floraison (0.75), la tolérance à la cloque (0.8), pointes-vertes (0.5) mais il y a une grande disparité des résultats entre sites. Les prédictibilités de nos modèles sont des arguments en faveur du déterminisme génétique complexe des caractères analysés. Ces résultats confirment nos hypothèses comme quoi nous avons bon espoir d'un jour pouvoir appliquer la GS pour optimiser le temps de génération de nouvelles variétés chez les arbres fruitiers à noyaux ou la sélection génomique est encore marginales dans les programmes de sélections.

In the purpose of optimising selection programs on major agricultural perennial crops we are investigating potentiality of genomic selection applications on 3 traits from a peach core-collections growing in low phytosanitary inputs conditions in 3 sites, through 3 years, in south-east of France. Prediction analyses are conducted on 2 phenological traits (half flowering time and green leaf-spike blushing, BBCH 65 and 09) and on *Taphrina deformans* tolerance. The core collection is composed of 206 accessions, for whom 192 has been genotyped with 16K IRSC. Additivity, dominance, and epistatic kinships have been built, according to the NOIA used in Vitezica and al 2017, to get genotypic variances decompositions. Predictions are then carried out with 5 GBLUP's models and 1 of RRBLUP on global number of repetitions or on environmental subsets. Dominance effect seems to matter in much variance's explanations. Much traits are efficiently predictable: flowering time (0.75), *T.deformans* tolerance (0.8), and the predictive ability means for green leaf-spike blushing is (0.5). We can conclude to strong potentiality of genomic selection with the model used. It's an argument for a genetical complex determinism of the analysed traits. Those results are confirming our hypothesis: we can have good expectations of using genomic selection to optimize the amelioration of perennial fruits crops for whom this method isn't already applied.