

Grand Prix Sciences & Recherche 2022

Charlotte BRAULT

Optimisation de la sélection
de nouvelles variétés de vigne
à l'aide de la prédiction
génomique et phénotypique.

PRÉFACE

L'Académie Amorim est née, au début des années 90, d'une volonté des grands noms de l'œnologie de promouvoir leur discipline, leurs recherches, l'excellence des travaux conduits par leurs étudiants, en créant un concours scientifique international. Le groupe Amorim a immédiatement répondu à leur appel.

Contribuer à une meilleure connaissance du vin et de son environnement, tel était l'objectif poursuivi par ces passionnés, bientôt rejoints par leurs confrères universitaires, journalistes, artistes, juristes, économistes, tous animés par cette même passion.

Aujourd'hui, Jacques Puisais, Président-fondateur de l'Académie Amorim, nous a quitté. Tout comme Robert Tinlot, ancien directeur général de l'OIV, qui lui avait succédé à la présidence de notre institution avec talent. Et il en fallait pour donner un deuxième souffle à cette Académie et l'ouvrir sur l'économie, les marchés, le juridique, l'environnement... autant de disciplines indissociables du monde viti-vinicole, autant de sujets dont Jean-Marie Aurand s'est emparé avec enthousiasme en succédant à ses illustres prédécesseurs.

La modeste contribution du bouchon de liège, fleuron de l'activité du groupe Amorim, a été d'initier, de soutenir et promouvoir ce concours international visant à récompenser les travaux de jeunes chercheurs en œnologie et en sciences humaines dans le domaine du vin.

Cette collection, dédiée aux Lauréats du Grand Prix Sciences & Recherche de l'Académie Amorim, est devenue, au fil des ans, une référence et la mémoire vivante des efforts et des travaux engagés dans le monde entier pour servir la noble cause du Vin. Je ne peux que souhaiter la voir poursuivre dans la voie de l'excellence qu'elle s'est tracée durant ces dernières décades.

Antonio Rios de Amorim

Président directeur général Corticeira Amorim
Vice-Président de l'Académie Amorim

ÉDITO

L'adaptation au changement climatique sur la vigne et la réduction des produits phytosanitaires sont deux enjeux majeurs pour la filière viti-vinicole.

La thèse de Charlotte Brault y répond point par point.

La première partie développe des tests de phénotypage haut débit liés à la résistance à la sécheresse.

La deuxième partie, est une application de la sélection génomique à 2 étapes cruciales, la sélection des parents pour faire le choix de croisements et la sélection des variétés au sein des croisements. Les résultats démontrent que la prédiction génomique s'avère plus précise que la sélection phénotypique utilisée classiquement.

Enfin, la troisième partie, explore une nouvelle méthode de sélection, la sélection phénotypique qui combine l'approche phénotypique et génomique. Elle s'avère plus performante que la sélection phénotypique, plus rapide – cinq ans de gagnés ! – et bien moins coûteuse que la sélection génomique.

Notre Lauréate livre ici des travaux qui seront utilisés, dans un premier temps par des chercheurs pour permettre, ensuite, aux vignerons de bénéficier plus rapidement des nouvelles variétés adaptées aux attentes des producteurs et des consommateurs.

Encore une fois, les Membres de l'Académie se sont appliqués à récompenser des travaux d'une grande actualité et dont les recommandations pourront être mises en pratique à courte échéance pour permettre à la viticulture de s'adapter au changement climatique et à la filière de répondre aux attentes sociétales.

Jean-Marie Aurand
Président de l'Académie Amorim

Optimisation de la sélection de nouvelles variétés de vigne à l'aide de la prédiction génomique et phénotypique.

Charlotte BRAULT

Thèse soutenue le 6 décembre 2021.

Direction de thèse

Directeur de thèse :

Patrice This

Co-encadrants :

Agnès Doligez, Vincent Segura, Loïc Le Cunff, Timothée Flutre

Membres du jury :

Laurence Moreau, Jean-Luc Jannink, Jacques David, Patrice This, Laurent Bouffier, Christopher Sauvage, Agnès Doligez, Viviane Bécart (invitée).

1 - INTRODUCTION

1.1 A propos de la vigne

La vigne est cultivée sur 7,4 millions d'hectares dans le monde. La France arrive en cinquième position en termes de surface cultivée avec 789 000 hectares représentant 3% de la surface agricole cultivée (OIV 2017). Néanmoins, la viticulture est un secteur économiquement majeur en France qui représente 31,2% des exportations agricoles (FranceAgriMer 2019).

La vigne cultivée correspond à l'espèce *V. vinifera* du genre *Vitis* qui contient environ 80 espèces inter-fertiles. La vigne a subi une première domestication il y a environ 9 000 ans dans le Caucase, puis une domestication secondaire, avec une spécialisation selon l'usage du raisin (consommation fraîche, raisins secs ou transformation en vin). Cette domestication secondaire est à l'origine de la (faible) structuration génétique en trois groupes observée aujourd'hui chez la vigne cultivée : cuve est, cuve ouest et table est (Bacilieri et al. 2013; Péros et al. 2011).

La vigne cultivée est une espèce diploïde avec une grande diversité génétique et un haut niveau d'hétérozygotie (Laucou et al. 2011). Elle souffre de dépression de consanguinité, ce qui rend difficile l'obtention de descendants issus d'autofécondation d'une variété.

Avec la possibilité de multiplication végétative, certaines variétés existent depuis des centaines d'années, c'est le cas du Pinot Noir, du Chardonnay ou encore du Gamay qui existent depuis le Moyen-Age, voire avant (Bowers et al. 1999; Lacombe et al. 2013).

À la fin du 19^{ème} siècle, le mildiou, l'oïdium et le phylloxéra, ont été introduits en Europe. Quasiment toutes les variétés de vigne cultivées sont sensibles à ces ravageurs. Des traitements à base de soufre et de cuivre ont été développés pour lutter contre le mildiou et l'oïdium mais aucun traitement efficace n'existe contre le phylloxéra. Pour lutter contre le phylloxéra, deux solutions sont possibles : d'une part l'hybridation avec des sources de résistance génétique issues d'autres espèces du genre *Vitis* et d'autre part l'utilisation de variétés de porte-greffe résistantes, issues d'autres espèces de *Vitis*. La première option a conduit à la création des hybrides producteurs directs et leur abandon quasi complet pendant la seconde moitié du XX^{ème} siècle, alors que la seconde option est toujours en vigueur aujourd'hui dans une très grande majorité de vignobles de par le monde.

Aujourd'hui, la vaste majorité des variétés cultivées sont toujours sensibles au mildiou et à l'oïdium et nécessitent donc de nombreux traitements phytosanitaires. Pour réduire l'utilisation de produits phytosanitaires, la sélection de variétés résistantes est une voie prometteuse. Dans ce contexte, le programme d'amélioration génétique français INRA-ResDur vise à combiner plusieurs sources de résistance au mildiou et à l'oïdium pour plus de durabilité (Schneider et al. 2019). Néanmoins, les étapes de sélection pour les autres caractères impliquent toujours le phénotypage des descendants issus de ces croisements, ce qui prend six à sept ans pour la vigne. La durée totale du cycle de sélection est actuellement de 15 ans (Figure 1).

Un idéotype correspond à une variété idéale, c'est-à-dire, à la variété qui remplit les objectifs de sélection pour tous les caractères.

Comme nous l'avons vu, la résistance au mildiou et à l'oïdium est un critère majeur chez la vigne pour diminuer l'utilisation de produits phytosanitaires. Toutefois, la vigne est également sensible à beaucoup d'autres maladies comme les maladies du

bois ou encore les insectes qui attaquent la vigne et transmettent des virus. L'autre menace qui pèse sur la vigne est le changement climatique. Même si la vigne est notamment cultivée dans des régions au climat relativement chaud comme le Sud du pourtour méditerranéen, les projections climatiques prévoient une augmentation des sécheresses, l'augmentation globale des températures et une plus grande probabilité des événements climatiques extrêmes (IPCC et al. 2021). Les conséquences du changement climatique sur la vigne sont une augmentation du stress hydrique, l'avancée des stades phénologiques, et plus particulièrement de la date de récolte, et le changement de la composition des baies avec notamment une augmentation du degré d'alcool et une baisse de l'acidité. Toutefois, les cépages cultivés actuellement ne représentent qu'une faible partie de la diversité génétique existante chez *V. vinifera* et la sélection génétique devrait pouvoir permettre d'atténuer les effets du changement climatique (Morales-Castilla et al. 2020; Myles 2013).

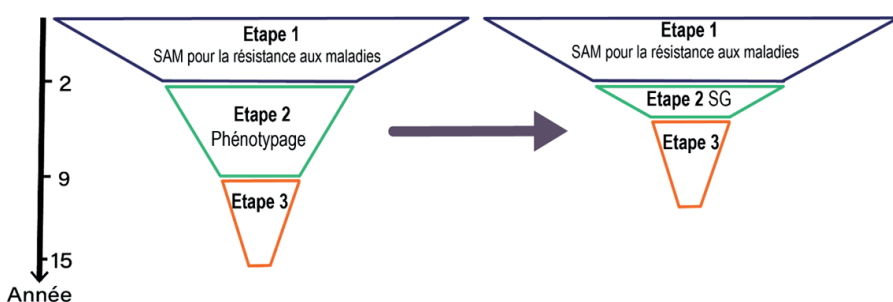


Figure 1 : Schéma de sélection actuel (à gauche) et futur (à droite) de nouvelles variétés de vignes résistantes. L'étape 1 correspond au premier croisement entre une variété résistante et une variété sensible et à la sélection des individus résistants par sélection assistée par marqueurs (SAM). L'étape 2 correspond à la sélection de variétés pour les caractères complexes, par phénotypage ou par sélection génomique (SG). L'étape 3 correspond à l'étape réglementaire d'inscription au catalogue.

1.2 Prédiction de la valeur génétique

Avant la prédiction génomique, la variation phénotypique était caractérisée au niveau moléculaire à l'aide de QTL (Quantitative Trait Loci), c'est-à-dire des zones du génome statistiquement associées à la variation d'un caractère. Une fois les QTL identifiés, le caractère d'intérêt pouvait être prédit à l'aide de marqueurs moléculaires associés aux QTLs via la sélection assistée par marqueurs (SAM). Cette méthode est par exemple utilisée actuellement pour suivre les gènes de résistance au mildiou et à l'oidium chez la vigne. Néanmoins, cette méthode devient rapidement inapplicable lorsque le nombre de QTL à suivre augmente. De plus, la majorité des caractères quantitatifs sont contrôlés par un grand nombre de QTL avec chacun de petits effets, ce qui rend l'application de cette méthodologie impossible (certains QTL ne sont pas détectés et il y a trop de QTL à suivre).

La sélection génomique (SG), permise par l'accès au génotypage haut-débit, a été proposée par (Meuwissen, Hayes, and Goddard 2001) pour répondre à ces limites. Le principe est d'utiliser une population d'entraînement phénotypée et génotypée pour estimer les effets statistiques associés aux marqueurs moléculaires et d'utiliser ces estimations pour prédire les valeurs génétiques de la population de sélection sur la base du génotypage uniquement.

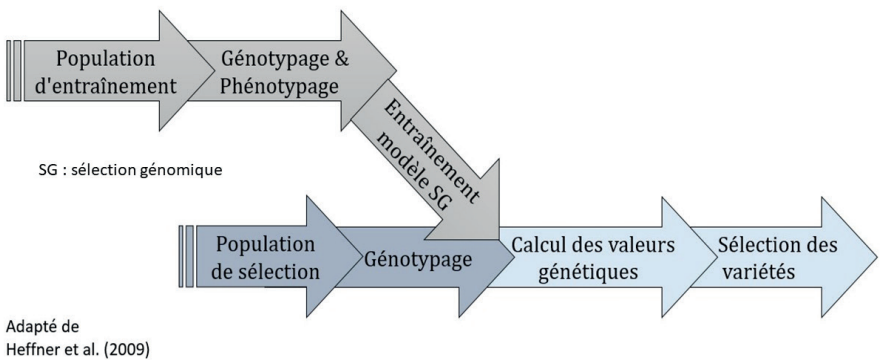


Figure 2 : Intégration de la sélection génomique dans les schémas de sélection.

Plusieurs méthodes statistiques existent pour l'estimation des effets des marqueurs, les plus populaires sont le RR-BLUP ou GBLUP, les méthodes à inférence bayésienne comme le BayesB, Bayes Cpi, BayesR, ou encore les méthodes non-linéaires comme le random forest, ou le RKHS (de los Campos et al. 2013). Les premières méthodes permettent uniquement d'estimer des effets alléliques additifs de substitution, donc une reformulation du modèle est nécessaire pour estimer la dominance ou l'épistasie. Les méthodes non-linéaires sont plus à même de capturer des effets non-linéaires (et donc non-additifs) mais ils demandent plus de temps de calcul et des jeux de données plus importants. Selon l'architecture génétique des caractères, le classement entre les méthodes peut varier mais les performances globales des différentes méthodes sont proches (Azodi et al. 2019).

Les principales variables qui affectent la précision de prédiction génomique sont : l'apparementement entre les populations d'entraînement et de validation, le type de marqueurs moléculaires associé à la densité de génotypage, et le phénotypage associé à la structure de la population. Ces paramètres interagissent entre eux et avec la méthode utilisée pour la SG.

Globalement, la SG permet d'accélérer le cycle de sélection, avec une prédiction des valeurs génétiques à un stade précoce. Ce point est particulièrement important chez la vigne où le temps de génération est très long et les essais phénotypiques coûteux. De plus, la SG permet d'augmenter la précision de sélection et de gérer la diversité génétique sur le long terme (Consortium et al. 2021).

Malgré des avancées considérables, le prix du génotypage peut constituer un frein à l'application de la SG dans les programmes d'amélioration, où des milliers d'individus doivent être génotypés. C'est pourquoi, (Rincint et al. 2018) ont proposé la sélection phénotypique (SP), comme alternative à la SG. Le principe repose sur l'hypothèse que les spectres mesurés sur les tissus des plantes dépendent de leur composition en différentes molécules, qui est déterminée génétiquement. Les spectres reflètent donc le génome des individus, en interaction avec l'environnement. Cette méthode a été testée avec succès sur le blé et le peuplier et quelques autres espèces mais reste nouvelle et de nombreuses questions demeurent.

1.3 Problématique de la thèse

Cette thèse résulte d'un partenariat entre l'interprofession viticole Inter-Rhône, l'IFV (Institut Français de la Vigne et du Vin) et INRAE. Elle s'est déroulée au sein de l'UMT GéoVigne et de l'équipe DAAV de l'UMR AGAP Institut. L'objectif était de répondre

aux problématiques d'Inter-Rhône en termes de création variétale pour accélérer la sélection de nouvelles variétés de vigne résistantes au mildiou et à l'oïdium.

Si la SG a démontré son utilité dans les programmes d'amélioration des animaux et de diverses espèces végétales (Consortium et al. 2021), son application chez la vigne reste limitée à quelques études préliminaires restreintes. L'objectif de ma thèse était donc d'implémenter la SG et la SP dans des configurations variées et sur un grand nombre de caractères pour étudier leurs intérêts respectifs en vue d'optimiser la sélection de nouvelles variétés de vigne. Pour cela, j'ai utilisé trois populations de vigne déjà génotypées et phénotypées.

Dans un premier temps, j'ai comparé plusieurs méthodes de prédiction génomique dans une population bi-parentale, en testant notamment si des méthodes multivariées (multi-caractères) pouvaient avoir un intérêt. Dans un second temps, j'ai appliqué la SG dans un contexte plus proche de celui des programmes de sélection, c'est-à-dire en inter-population, avec une population d'entraînement plus ou moins apparentée avec la population de validation. Enfin, j'ai testé l'implémentation de la SP, en comparant sa précision à celle de la SG. Enfin, une discussion générale a permis de discuter certains points de mes travaux, de proposer quelques pistes d'application en sélection, et des perspectives de recherche.

2 PRÉDICTION GÉNOMIQUE UNIVARIÉE ET MULTIVARIÉE APPLIQUÉES À UNE POPULATION BI-PARENTALE

La modélisation statistique multivariée permet d'analyser conjointement plusieurs caractères. Jusqu'à présent, il a été montré que cette modélisation permettait d'obtenir une meilleure précision de prédiction des phénotypes en prenant en compte les corrélations génétiques et non génétiques entre les caractères e.g., (Jia and Jannink 2012). Par ailleurs, certaines méthodes de prédiction génomique font de la sélection de variables (effets alléliques) ; on peut donc s'en servir pour l'étude du déterminisme génétique des caractères (détection de QTL). J'ai comparé plusieurs méthodes de SG (ridge regression, elastic net, LASSO, gradient boosting, et régression pénalisée structurée) avec des méthodes classiques de détection de QTL par cartographie d'intervalle (SIM pour Simple Interval Mapping et MIM pour Multiple Interval Mapping).

Dans ce contexte, j'ai utilisé une population bi-parentale de 188 individus, issue d'un croisement réciproque entre les variétés Syrah et Grenache. Cette population a été phénotypée en conditions semi-contrôlées (plateforme PhenoArch) pour 7 caractères liés à la tolérance au déficit hydrique (la transpiration diurne et nocturne spécifique, le potentiel hydrique foliaire, la différence de potentiel hydrique entre sol et feuille, la conductance hydraulique, la différence de production de biomasse et l'efficacité d'utilisation de l'eau) (Coupel-Ledru et al. 2014; 2016). Cette même population a été plantée au champ, un stress hydrique a été appliqué sur un des deux blocs, et 7 caractères agronomiques ou liés à la tolérance à la sécheresse ont été étudiés (le poids de bois de taille, le poids de baie, la date de véraison, la fertilité, le rendement, le nombre de grappes et le qui reflète le stress hydrique subit par la plante au cours de la saison).

Pour tester le potentiel intérêt de la modélisation multivariée, et des méthodes

de SG pour la détection de QTL, j'ai construit une nouvelle carte génétique dense avec des marqueurs SNP (Single Nucleotide Polymorphism), et j'ai simulé deux caractères avec une corrélation génétique, en faisant varier l'architecture génétique et l'héritabilité. Par ailleurs, j'ai utilisé les données expérimentales obtenues en champ et en plateforme de phénotypage pour valider les résultats de prédiction et détecter de nouveaux QTL.

Sur les données simulées, les résultats de prédiction génomique ont montré que certaines méthodes étaient particulièrement sensibles à l'architecture génétique avec une baisse de précision lorsque l'architecture génétique était complexe (50 QTL), plus particulièrement pour SIM et MIM. D'une manière générale, une meilleure précision de prédiction était associée à une plus grande héritabilité. Pour les méthodes multivariées, les précisions de prédiction n'étaient pas supérieures aux meilleures méthodes univariées. Pour la détection de QTL, les méthodes de SG faisant de la sélection de variable surpassaient SIM et MIM en termes de précision de sélection lorsque 50 QTL étaient simulés. Toutefois, l'utilisation de méthodes de SG pour la détection de QTL requiert que la valeur du paramètre de pénalisation soit adaptée à la sélection de variable et non à la prédiction. Ceci a été rendu possible grâce à des extensions de ces méthodes comme la Stability Selection ou le marginal False Discovery Rate.

Sur les données expérimentales obtenues en plateforme de phénotypage, les méthodes de SG avaient les meilleurs résultats pour la prédiction, avec une très légère supériorité globale des méthodes multivariées. Les précisions de prédiction obtenues variaient de -0.1 à 0.68 selon les caractères et les méthodes, avec une moyenne de 0.38 pour la meilleure méthode (elastic net multivarié). Ces valeurs sont encourageantes pour l'application de la SG sur des caractères liés à la tolérance à la sécheresse. (Coupel-Ledru et al. 2014; 2016) avaient déjà appliqué une détection de QTL sur ces données. En comparaison, j'ai utilisé de nouvelles méthodes pour la détection de QTL, qui ont prouvé leur utilité sur des données simulées ; de plus, j'ai utilisé une carte génétique plus dense. Ces nouveautés m'ont permis de réduire les intervalles de confiance des QTL et de trouver de nouveaux QTL pour ces caractères. Sur les données expérimentales obtenues au champ, j'ai appliqué les meilleures méthodes précédemment définies pour la SG. Le classement entre les méthodes était similaire à celui observé sur les données de la plateforme de phénotypage. Pour chaque caractère, j'ai calculé la différence normalisée entre modalités (présence vs absence de stress hydrique) pour chaque génotype, considérant cette différence comme un nouveau caractère reflétant la réponse au stress. La précision de prédiction de la réponse au stress hydrique était faible pour tous les caractères et pour toutes les méthodes testées.

3 PRÉDICTION GÉNOMIQUE INTER-POPULATION

La SG est souvent appliquée au sein d'une population génétiquement homogène à la fois pour entraîner et valider le modèle. Cependant, cette configuration correspond rarement à celle d'un programme d'amélioration, où la population de sélection correspond à des croisements qui ne sont pas encore réalisés et qui n'ont pas servi à entraîner le modèle. Dans cette partie, j'ai donc testé la SG dans des configurations plus réalistes en création variétale. Pour cela, j'ai testé l'utilisation d'un panel de diversité, composé de 279 variétés représentant l'ensemble de la diversité génétique de *Vitis vinifera*, comme potentielle population d'entraînement universelle. La population de validation était composée de 622 individus répartis en 10 populations bi-parentales équilibrées issues des croisements entre 5 variétés, chacune de ces 5 variétés étant impliquée dans 4 croisements (plan de croisement appelé demi-diallèle). Ces deux populations ont été phénotypées différentes années et génotypées avec des SNP, dont 32 894 sont en commun entre ces deux populations. Les quinze caractères communs étudiés sont classés en quatre catégories : caractères morphologiques pour le poids, la largeur, la longueur, la compacité de la grappe, le nombre de grappes et le poids des baies ; caractères de composition des baies pour le malate, le tartrate et le shikimate au stade mûr et des ratios associés ; caractères de phénologie avec la date de véraison (début de la maturation), la date de récolte et l'intervalle entre les deux, et le caractère de vigueur.

J'ai décomposé mon étude en deux sous-objectifs. Dans un premier temps, j'ai prédit la moyenne des descendants d'un croisement, ce qui correspond à la première étape de sélection des parents pour réaliser les meilleurs croisements ; puis j'ai prédit les valeurs des individus au sein de chaque croisement, ce qui correspond à la seconde étape de sélection une fois les croisements réalisés. Pour chaque étape, j'ai mesuré la précision de prédiction avec deux méthodes de SG : ridge regression (RR) et LASSO. La performance relative de ces deux méthodes dépendant de l'architecture génétique du caractère, j'ai gardé la valeur de la prédiction de la meilleure méthode afin de s'abstraire de cet effet. J'ai conçu trois scénarios de prédiction : dans le scénario 1a, le modèle est entraîné et validé au sein du demi-diallèle par validation croisée ; dans le scénario 1b, trois populations issues de croisement avec un parent en commun sont utilisées pour entraîner le modèle, la prédiction se faisant dans le quatrième croisement du demi-diallèle impliquant ce même parent. Enfin, dans le scénario 2, le panel de diversité est utilisé pour entraîner le modèle et les croisements demi-diallèle servent à sa validation. Le scénario 1a correspond à la précision de prédiction maximale, le scénario 2 correspond à ce qui pourrait être implémenté dans un programme de sélection et le scénario 1b est un intermédiaire entre ces deux scénarios. Enfin, j'ai testé différents paramètres pouvant expliquer la précision de prédiction obtenue à chaque étape.

L'analyse des données phénotypiques montre que les populations du demi-diallèle affichent des moyennes et des variances différentes selon les caractères. Au niveau génétique, l'apparement entre le demi-diallèle et le panel de diversité est partiel, avec 1/3 des individus du panel (de la sous-population cuve ouest) étant plus apparementés avec le demi-diallèle que les autres (des sous-populations cuve est et table est).

Pour la prédiction de la moyenne des descendants des croisements, la distribution des précisions de prédiction étaient différentes selon que l'on considérait les

moyennes par croisement tous caractères confondus ou les moyennes par caractère tous croisements confondus. Par croisement, les précisions de prédiction étaient autour de 0.6 pour les scénarios 1a et 1b et beaucoup plus variables (de -0.3 à 0.72) pour le scénario 2, avec une forte corrélation négative avec la distance génétique entre les parents des croisements.

Pour la prédiction des performances des individus au sein de chaque croisement, le scénario 1a permettait d'obtenir la meilleure précision de prédiction, cette dernière diminuant avec le scénario 1b et le scénario 2, ce qui était attendu. Les précisions de prédiction demeuraient toutefois très variables selon le croisement et le caractère considérés. Pour les scénarios 1a et 1b, le paramètre le plus déterminant pour la précision de prédiction était l'héritabilité, alors que pour le scénario 2, les paramètres les plus déterminants étaient l'héritabilité, la différenciation entre les croisements pour chaque caractère, et le modèle de prédiction (RR ou LASSO).

Pour améliorer la précision de prédiction et étant donné le faible apparentement global entre le panel de diversité et les croisements du demi-diallèle, nous avons échantillonné dans le panel les individus les plus apparentés avec chaque croisement pour constituer une nouvelle population d'entraînement optimisée. Cette optimisation a permis d'améliorer la prédiction de la moyenne des croisements, lorsque qu'ils étaient mal prédits. Pour la prédiction de la valeur des individus au sein d'un croisement, l'optimisation de la population d'entraînement n'a permis qu'une légère amélioration de la précision de prédiction en comparaison avec un échantillonnage aléatoire.

Par ailleurs, nous avons également implémenté la prédiction de la variance des populations issues des croisements du demi-diallèle, dans le scénario 1a. En effet, les croisements les plus prometteurs pour un caractère donné sont ceux produisant des descendants avec à la fois une moyenne et une variance élevées, ce qui permet d'augmenter la probabilité de générer les meilleurs individus. La prédiction de la variance s'est avérée moins précise que celle de la moyenne, avec une corrélation de 0.58 entre la variance prédite et la variance observée.

En conclusion, il s'agit de la première étude sur la prédiction génomique en inter-population chez la vigne utilisant des données expérimentales et impliquant plusieurs croisements. Les précisions de prédiction obtenues pour la moyenne des croisements et les valeurs des individus dans chaque croisement sont prometteuses pour une implémentation pratique de la SG chez la vigne.

4 INTÉRÊT DE LA PRÉDICTION PHÉNOMIQUE COMME ALTERNATIVE À LA PRÉDICTION GÉNOMIQUE CHEZ LA VIGNE

Dans ce chapitre, j'ai testé l'utilisation de spectres au lieu des marqueurs moléculaires pour la prédiction des phénotypes. Cette méthodologie a été originellement proposée par (Rincint et al. 2018) et je l'ai appliquée pour la première fois chez la vigne. L'intérêt de l'utilisation de spectres est qu'ils sont moins chers et plus rapides à acquérir que les données de génotypage. L'hypothèse sous-jacente est que la réflectance à une longueur d'onde donnée dépend de la composition biochimique dans le tissu considéré et que cette composition est déterminée génétiquement. Cette hypothèse a été testée avec succès sur le blé et le peuplier dans la première étude de (Rincint et al. 2018). Ces résultats ont été confirmés pour plusieurs autres espèces céréalières mais la prédiction phénomique est une nouvelle méthodologie et beaucoup d'interrogations demeurent. Pour mieux comprendre l'intérêt de la prédiction phénomique comparativement à la prédiction génomique, nous avons mesuré en 2020 et 2021 des spectres en proche infra-rouge sur des échantillons séchés de feuilles et de bois, dans le demi-diallèle et le panel de diversité précédemment étudiés. J'ai ainsi pu tester plusieurs modalités d'utilisation des spectres. D'abord, j'ai ajusté un modèle linéaire mixte pour la réflectance à chaque longueur d'onde, afin d'en extraire la composante génétique. Ce modèle était modifié selon que les deux années et/ou les deux tissus étaient inclus ou non. A partir des valeurs génétiques extraites à chaque longueur d'onde, j'ai construit une matrice d'apparement phénomique, et je l'ai comparée à la matrice d'apparement génomique. Les résultats montrent que l'année ou le tissu ne modifie presque pas la corrélation (ou co-inertie) entre ces matrices. L'analyse des composantes de la variance par longueur d'onde montre que l'ajustement d'un modèle unique pour les deux tissus conduit à une variance génétique très faible. J'ai donc écarté ce modèle pour la prédiction phénomique. La prédiction phénomique a été implémentée au sein du demi-diallèle et du panel de diversité pour les 15 caractères précédemment étudiés. Deux méthodes de prédiction ont été comparées : RR-BLUP et GBLUP. La première se base sur une matrice de marqueurs SNP pour la SG et une matrice de réflectance pour la SP. La seconde se base sur la matrice d'apparement génomique ou phénomique précédemment calculée. J'ai comparé l'utilisation de la moyenne des spectres bruts et l'utilisation des BLUP de réflectance issus des modèles linéaires mixtes décrits ci-dessus. L'utilisation des BLUP de réflectance a conduit à une augmentation drastique de la précision de prédiction. Cette conclusion n'avait jamais été mise en évidence auparavant. En revanche, de manière générale, l'année, le tissu et le pré-traitement des spectres avaient un effet négligeable sur la précision de prédiction. Toutefois, utiliser des spectres issus de deux années et combiner les deux tissus dans un modèle multi-matriciel donnait globalement la meilleure précision de prédiction. J'ai comparé les meilleures précisions de prédiction ainsi obtenues pour la prédiction phénomique, avec celles de la prédiction génomique. Pour quelques caractères dans quelques croisements, la précision de prédiction phénomique était supérieure à celle de la génomique mais en moyenne elle était de -0.2. De façon intéressante, j'ai observé dans les deux populations une corrélation significative entre les deux précisions de prédiction, avec une pente de régression proche de 1. Cela suggère que le rang des précisions de prédiction est conservé entre la prédiction génomique et phénomique, quel que soit le caractère.

Enfin, j'ai combiné dans un même modèle de prédiction les matrices d'apparement génomique et phénotomique, en testant les différents tissus (bois, feuille ou bois + feuille). Globalement, l'addition des matrices phénotomiques conduisait à un léger gain de précision de prédiction dans le panel de diversité et à une légère diminution dans le demi-diallèle, comparé à la prédiction génomique seule. Ceci peut s'expliquer par le fait que j'ai cherché à maximiser la variance génotypique des spectres et à exclure les autres composantes de la variance. De plus, comme les phénotypes ont été mesurés plusieurs années avant les spectres, la précision de prédiction ne peut reposer que sur la composante génétique seule, et non sur le lien entre la nature du spectre et le caractère à prédire, comme l'utilisation de spectres pris sur bois pour prédire la vigueur.

5 DISCUSSION GÉNÉRALE

5.1 Données et populations étudiées

Cette thèse avait pour objectif de proposer des optimisations de la sélection de nouvelles variétés de vigne en faisant appel à la prédiction génomique ou phénotomique. Bien que la prédiction génomique ne soit pas une méthodologie nouvelle (la publication de référence date de 2001), les premières applications chez la vigne datent de 2014 et il y a eu assez peu de travaux publiés. Pourtant, la SG gagnerait d'autant plus à être appliquée que le temps de génération est long et que la taille des essais phénotypiques chez la vigne devient vite un problème. Pour mesurer l'intérêt de l'implémentation de la SG dans les programmes d'amélioration vigne, j'ai d'abord comparé plusieurs méthodes de SG, dans une population bi-parentale homogène génétiquement. Les avantages de cette population sont : 1) que l'apparement théorique est identique et maximal pour tous les individus, 2) que cette descendance a été phénotypée pour de nombreux caractères relatifs à la réponse au stress hydrique, à la fois en conditions semi-contrôlées et au champ. Ensuite, je me suis placée dans un cadre plus proche de celui rencontré dans les programmes d'amélioration, avec une population d'entraînement génétiquement plus éloignée de la population de validation. Pour cela, j'ai utilisé un panel de diversité comme population d'entraînement et des populations bi-parentales apparentées en demi-diallèle comme populations de validation. Ces populations ont été phénotypées pour 15 caractères en commun. L'avantage de ces populations est que les résultats obtenus sont plus facilement extrapolables à d'autres croisements puisque la diversité génétique explorée est plus grande, même si les précisions de prédiction obtenues sont moins élevées. Enfin, dans une optique de baisse de coût et d'augmentation du débit, j'ai testé la sélection phénotomique comme alternative à la sélection génomique. Pour simplifier l'analyse, j'ai implémenté la SP seulement en intra-population, en utilisant les populations et les caractères étudiés dans la partie précédente.

5.2 Décomposition du modèle de prédiction

Nous pouvons discuter les résultats obtenus en analysant pour chacune des composantes du modèle de prédiction, ses caractéristiques et modifications

possibles pour augmenter la précision de prédiction. Le modèle de prédiction global est : $y = X\beta + \epsilon$ avec y le vecteur des phénotypes ajustés pour un caractère donné, X la matrice génotypique des SNP ou la matrice phénotypique des réflectances pour chaque longueur d'onde, pour chaque individu, β les effets estimés des SNP ou des réflectances, et ϵ le vecteur de la résiduelle.

Les données phénotypiques sont une composante majeure de l'équation de prédiction. Notamment, nous avons démontré que l'héritabilité du caractère était un paramètre déterminant de la précision de prédiction en intra-population. L'héritabilité peut être améliorée en augmentant le nombre de répétitions, ou en décomposant le caractère en ses composantes physiologiques pour diminuer la complexité du caractère. Néanmoins, le coût de phénotypage constitue encore un goulot d'étranglement dans de nombreux programmes d'amélioration.

La matrice d'entrée du modèle peut prendre des formes différentes. En SG, il s'agit, pour chaque individu et SNP, du nombre d'allèles en commun avec la séquence de référence, ainsi les valeurs possibles sont 0, 1 et 2. En SP, il s'agit de la réflectance pour un individu à une longueur d'onde donnée, soit une valeur numérique comprise entre 0 et 1. D'un point de vue génétique, les populations issues des croisements du demi-diallèle et les sous-populations du panel sont plus différenciées dans la matrice d'apparentement génomique que phénotypique.

La méthode d'estimation des effets des variables (estimation du vecteur β) est susceptible d'impacter la précision de prédiction, en interaction avec les autres facteurs mentionnés. A titre d'exemple, j'ai comparé plusieurs méthodes univariées et multivariées, et montré par la suite la supériorité de GBLUP sur RR-BLUP dans le cas d'une population plus structurée.

5.3 Recommandations pour améliorer les programmes d'amélioration vigne

Dans le cadre de la sélection de variétés résistantes à Inter-Rhône, les cépages Syrah et Grenache ont été croisés avec 4 génotypes résistants au mildiou et à l'oïdium. Parmi les descendants, 155 individus ont une résistance polygénique à ces deux maladies et sont génotypés par la technologie GBS (SNP). Pour ce programme spécifiquement, je conseille de prédire précocement les valeurs génétiques des individus avec la SG, en utilisant comme populations d'entraînement les croisements du demi-diallèle apparentés à Syrah ou à Grenache. Ainsi, on peut s'attendre à ce que les précisions de prédiction soient proches de celles obtenues dans le scénario 1b. Pour les futurs programmes d'amélioration vigne, un nouveau panel créé dans le cadre du projet ANR SelGenVit pourra être utilisé. Il est composé de 132 variétés et inclut des individus avec des gènes de résistance au mildiou, à l'oïdium mais aussi à la maladie de Pierce ou au black-rot, provenant d'autres espèces de *Vitis*.

Prédiction de la moyenne des croisements

Nous avons vu que la moyenne observée des descendants ne correspondait que rarement à la moyenne observée des parents. Néanmoins, la moyenne des descendants d'un croisement peut être facilement prédite par SG si les parents sont génotypés. La précision de prédiction est variable selon la proximité génétique des parents mais de bonnes précisions peuvent être atteintes si celle-ci est grande. Cette étape requiert qu'un panel de référence soit génotypé et phénotypé pour estimer les effets associés aux marqueurs. La prédiction peut se faire sur tous les

croisements possibles entre les variétés de ce panel. En pratique, les programmes auront probablement comme objectif d'utiliser au moins un parent résistant aux maladies. Dans ce cas, lorsque les croisements sont réalisés, une sélection assistée par marqueurs fait un premier tri parmi les descendants pour les caractères avec un gène majeur (Figure 3).

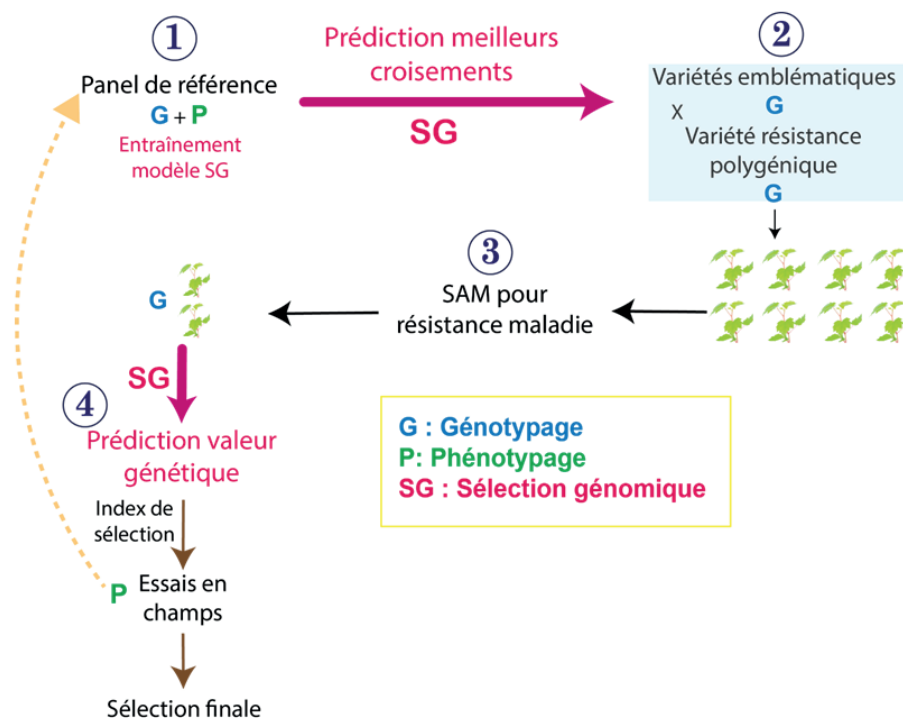


Figure 3 : Proposition d'un nouveau schéma de sélection de la vigne.

- 1 : conception d'un panel de référence pour calibrer le modèle de sélection génomique,
- 2 : sélection des meilleurs croisements à réaliser, intégrant des variétés emblématiques et résistantes,
- 3 : sélection assistée par marqueurs pour les caractères à gènes majeurs,
- 4 : sélection des variétés pour les caractères complexes, à l'aide de la sélection génomique

Option 1 : prédiction génomique

Une fois que les individus résistants des populations ont été sélectionnés, une option est d'appliquer la SG. Pour cela, les descendants résistants doivent être génotypés. Si le panel de référence est utilisé, les précisions de prédiction attendues sont de l'ordre de celle obtenues dans le scénario 2.

Option 2 : prédiction phénotypique

Pour cette seconde option, les descendants ne sont plus génotypés mais des spectres seront mesurés. Ces mesures devront également être faites dans le panel de référence. Jusqu'ici chez la vigne, la SP n'a pas été testée en inter-population. Il faudra d'abord s'assurer que la diminution de précision de prédiction attendue en inter-population ne soit pas trop forte pour la SP par rapport à celle de la SG.

Gestion du cycle de sélection

Les individus prédits au sein de chaque croisement par la SP ou la SG feront l'objet d'un index de sélection combinant les objectifs de sélection pour les différents caractères cibles, pour ne conserver que les individus les plus prometteurs. Ainsi, des essais en champ seront ensuite réalisés pour un nombre réduit d'individus en vue de l'inscription au catalogue. Ces individus pourront être incorporés au modèle de prédiction dans le panel de référence pour les prochains cycles de sélection. Au fur et à mesure des cycles, la diversité génétique va diminuer, ce qui va impacter le progrès génétique sur le long terme. Il faut être vigilant lors de l'étape de sélection des croisements, pour conserver des génotypes peu apparentés, malgré la plus faible précision de prédiction associée. Des contraintes peuvent être ajoutées au modèle de prédiction pour gérer la consanguinité.

5.4 Perspectives

La prise en compte de l'interaction GxE est capitale pour les futurs programmes d'amélioration. Pour cela, il est nécessaire d'avoir une population phénotypée plusieurs années dans plusieurs environnements. Certains modèles statistiques sont spécifiquement adaptés pour prendre en compte ces effets, en estimant des covariances entre environnements, en étudiant des normes de réaction à une variable environnementale ou en utilisant un modèle multivarié, où chaque caractère correspond à un phénotype dans un environnement donné.

Dans ma thèse, j'ai utilisé des spectres mesurés dans le proche infra-rouge sur des échantillons séchés en laboratoire mais il existe d'autres dispositifs pour mesurer des spectres. Dans la même gamme de longueur d'onde, il existe des spectromètres portables qui peuvent être utilisés au champ. Sinon, des spectromètres peuvent être embarqués sur des drones pour mesurer des images hyper-spectrales. Toutefois, les spectres issus de ces modalités sont probablement plus bruités que ceux que nous avons utilisés, il faut donc trouver un compromis entre le débit et la qualité des données. Plusieurs interrogations demeurent quant à l'impact de l'utilisation de la SP dans les programmes d'amélioration. La SP est-elle adaptée à tous les types d'architecture génétique ? Est-ce qu'il y a des zones du génome qui ne sont pas couvertes par les spectres ? Est-ce que la SP permet de mieux prendre en compte les effets génétiques non-additifs ? L'utilisation de simulations permettrait de répondre à certaines de ces questions.

BIBLIOGRAPHIE

- Azodi, Christina B., Emily Bolger, Andrew McCarren, Mark Roantree, Gustavo de los Campos, and Shin-Han Shiu.** 2019. "Benchmarking Parametric and Machine Learning Models for Genomic Prediction of Complex Traits." *G3: Genes, Genomes, Genetics* 9 (11): 3691–3702. <https://doi.org/10.1534/g3.119.400498>.
- Bacilieri, Roberto, Thierry Lacombe, Loïc Le Cunff, Manuel Di Vecchi-Staraz, Valérie Laucou, Blaise Genna, Jean-Pierre Péros, Patrice This, and Jean-Michel Boursiquot.** 2013. "Genetic Structure in Cultivated Grapevines Is Linked to Geography and Human Selection." *BMC Plant Biology* 13 (1): 25. <https://doi.org/10.1186/1471-2229-13-25>.
- Bowers, John, Jean-Michel Boursiquot, Patrice This, Kieu Chu, Henrik Johansson, and Carole Meredith.** 1999. "Historical Genetics: The Parentage of Chardonnay, Gamay, and Other Wine Grapes of Northeastern France." *Science* 285 (5433): 1562–65. <https://doi.org/10.1126/science.285.5433.1562>.
- Campos, G. de los, J. M. Hickey, R. Pong-Wong, H. D. Daetwyler, and M. P. L. Calus.** 2013. "Whole-Genome Regression and Prediction Methods Applied to Plant and Animal Breeding." *Genetics* 193 (2): 327–45. <https://doi.org/10.1534/genetics.112.143313>.
- Consortium, R2d2, Aline Fugerey-Scarbel, Catherine Bastien, Mathilde Dupont-Nivet, and Stéphane Lemarié.** 2021. "Why and How to Switch to Genomic Selection: Lessons From Plant and Animal Breeding Experience." *Frontiers in Genetics* 0. <https://doi.org/10.3389/fgene.2021.629737>.
- CoupeL-Ledru, Aude, Éric Lebon, Angélique Christophe, Agnès Doligez, Llorenç Cabrera-Bosquet, Philippe Péchier, Philippe Hamard, Patrice This, and Thierry Simonneau.** 2014. "Genetic Variation in a Grapevine Progeny (*Vitis Vinifera* L. Cvs Grenache×Syrah) Reveals Inconsistencies between Maintenance of Daytime Leaf Water Potential and Response of Transpiration Rate under Drought." *Journal of Experimental Botany* 65 (21): 6205–18. <https://doi.org/10.1093/jxb/eru228>.
- CoupeL-Ledru, Aude, Eric Lebon, Angélique Christophe, Agustina Gallo, Pilar Gago, Florent Pantin, Agnès Doligez, and Thierry Simonneau.** 2016. "Reduced Nighttime Transpiration Is a Relevant Breeding Target for High Water-Use Efficiency in Grapevine." *Proceedings of the National Academy of Sciences* 113 (32): 8963–68. <https://doi.org/10.1073/pnas.1600826113>.
- FranceAgriMer.** 2019. "La Filière Vin | FranceAgriMer - Établissement National Des Produits de l'Agriculture et de La Mer." La Filière Vin | FranceAgriMer - Établissement National Des Produits de l'Agriculture et de La Mer. 2019. <https://www.franceagrimer.fr/filieres-Vin-et-cidre/Vin/La-filiere-Vin>.
- IPCC, P. R. Shukla, J. Skea, E. Calvo Buendia, V. Masson-Delmotte, H.-O. Pörtner, D. C. Roberts, et al.** 2021. "IPCC, 2019: Summary for Policymakers. In: Climate Change and Land: An IPCC Special Report on Climate Change, Desertification, Land Degradation, Sustainable Land Management, Food Security, and Greenhouse Gas Fluxes in Terrestrial Ecosystems."
- Jia, Yi, and Jean-Luc Jannink.** 2012. "Multiple-Trait Genomic Selection Methods Increase Genetic Value Prediction Accuracy." *Genetics* 192 (4): 1513–22. <https://doi.org/10.1534/genetics.112.144246>.
- Lacombe, Thierry, Jean-Michel Boursiquot, Valérie Laucou, Manuel Di Vecchi-Staraz, Jean-Pierre Péros, and Patrice This.** 2013. "Large-Scale Parentage Analysis in an Extended Set of Grapevine Cultivars (*Vitis Vinifera* L.)." *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik* 126 (2): 401–14. <https://doi.org/10.1007/s00122-012-1988-2>.
- Laucou, V., T. Lacombe, F. Dechesne, R. Siret, J.-P. Bruno, M. Dessup, T. Dessup, et al.** 2011. "High Throughput Analysis of Grape Genetic Diversity as a Tool for Germplasm Collection Management." *Theoretical and Applied Genetics* 122 (6): 1233–45. <https://doi.org/10.1007/s00122-010-1527-y>.
- Meuwissen, Theo, Ben Hayes, and Mike Goddard.** 2001. "Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps." *Genetics*, 11.
- Morales-Castilla, Ignacio, Iñaki García de Cortázar-Atauri, Benjamin I. Cook, Thierry Lacombe, Amber K. Parker, Cornelis van Leeuwen, Kimberly A. Nicholas, and Elizabeth M. Wolkovich.** 2020. "Diversity Buffers Winegrowing Regions from Climate Change Losses." *Proceedings of the National Academy of Sciences* 117 (6): 2864–69. <https://doi.org/10.1073/pnas.1906731117>.
- Myles, Sean.** 2013. "Improving Fruit and Wine: What Does Genomics Have to Offer?" *Trends in Genetics* 29 (4): 190–96. <https://doi.org/10.1016/j.tig.2013.01.006>.
- OIV.** 2017. "Distribution of the World's Grapevine Varieties." *Focus OIV* 2017.
- Péros, Jean-Pierre, Gilles Berger, Aurélien Portemont, Jean-Michel Boursiquot, and Thierry Lacombe.** 2011. "Genetic Variation and Biogeography of the Disjunct *Vitis* Subg. *Vitis* (Vitaceae)." *Journal of Biogeography* 38 (3): 471–86. <https://doi.org/10.1111/j.1365-2699.2010.02410.x>.
- Rincent, Renaud, Jean-Paul Charpentier, Patricia Faivre-Rampant, Etienne Paux, Jacques Le Gouis, Catherine Bastien, and Vincent Segura.** 2018. "Phenomic Selection Is a Low-Cost and High-Throughput Method Based on Indirect Predictions: Proof of Concept on Wheat and Poplar." *G3 Genes | Genomes | Genetics*, October, g3.200760.2018. <https://doi.org/10.1534/g3.118.200760>.
- Schneider, C., C. Onimus, E. Prado, V. Dumas, S. Wiedemann-Merdinoglu, M.A. Dorne, M.C. Lacombe, et al.** 2019. "INRA-ResDur: The French Grapevine Breeding Programme for Durable Resistance to Downy and Powdery Mildew." *Acta Horticulturae*, no. 1248 (August): 207–14. <https://doi.org/10.17660/ActaHortic.2019.1248.30>

Grand Prix 2021 - Delphine Winstel

Recherches sur les composés du bois de chêne modulant la saveur des vins et des eaux-de-vie

Grand Prix 2019 - Yann Raineau

Défis environnementaux de la viticulture : une analyse comportementale des blocages et des leviers d'action.

Grand Prix 2017 - Liming Zeng

Étude de la composition macromoléculaire du raisin et des vins : impact sur la qualité sensorielle

Grand Prix 2014 - Olivier Serra

Le législateur et le marché vinicole sous la Troisième République

Grand Prix 2013 – Dorothée Boyer-Paillard

Essai sur les notions d'origine et de provenance en droit du commerce, la reconnaissance juridique des « territoires des productions typiques »

Grand Prix 2012 – Guillaume Antalick

Bilan biochimique et sensoriel des modifications de la note fruitée des vins rouges lors de la fermentation malolactique, le rôle particulier des esters

Grand Prix 2011 – Axel Marchal

Recherches sur les bases moléculaires de la saveur sucrée des vins secs

Grand Prix 2010 – Caroline Le Goffic

La protection des indications géographiques en France, dans la Communauté européenne et aux Etats-Unis

Grand Prix 2009 - Katharina Zott

Les levures non-Saccharomyces : dynamique, caractérisation et interaction avec Saccharomyces durant les étapes pré-fermentaires et la fermentation alcoolique

Grand Prix 2008 – Elise Sarrazin

*Recherches sur l'arôme des vins liquoreux de pourriture noble issus des cépages Sémillon et Sauvignon blanc
Caractérisation de composés clés et étude de leur genèse*

Grand Prix 2007 – Alexandre Pons

Recherche sur l'arôme de vieillissement prématuré des vins blancs secs

Grand Prix 2006 - Paulo Lopes

Etude des phénomènes oxydatifs pendant le vieillissement des vins en bouteille. Rôle de l'obturateur

Grand Prix 2005 - Stéphane La Guerche

Les défauts moisissés terreux des vins : la géosmine identifiée comme principale responsable

Grand Prix 2004 - Michael Jourdes

Réactivité, sythèse, couleur et activité biologique d'Ellagitannins C-Glycosidique et Flavano-Ellagi tannins

Grand Prix 2003 Dominique Roujou de Boubée

*Recherches sur la 2-méthoxy-3-isobutylpyrazine dans les raisin
et dans les vins*

Grand Prix 2002 -- Catherine Peyrot des Gachons

*Recherches sur le potentiel aromatique des raisins
de Vitis vinifera L.cv Sauvignon*

Grand Prix 2001 - René Siret

*Etude du polymorphisme génétique de la vigne cultivée (Vitis vinifera
L.) à l'aide de marqueurs microsatellites : application à la caractérisa-
tion des cépages dans les vins*

Grand Prix 2000 - Takatoshi Tominaga

*Recherches sur l'arôme variétal des vins de Vitis vinifera
L. cv. sauvignon blanc et sa genèse à partir de précurseurs inodores
du raisin*

Grand Prix 1999 - Isabelle Cutzach-Billard

*Etude sur l'arôme des vins doux naturels non muscatés au cours de leur
élevage et de leur vieillissement. Son origine. Sa formation*

Grand Prix 1998 - Virginie Moine-Ledoux

*Recherches sur le rôle des Mannoprotéines de levure
vis à vis de la stabilité protéique et tartrique des vins*

Grand Prix 1997 - Valérie Lavigne-Cruege

*Recherche sur les composés soufrés formés par la levure au cours de la
vinification et l'élevage des vins blancs secs*

Grand Prix 1996 - Sylvie Biau

Etude de la matière colorante des vins blancs de Bordeaux

Grand Prix 1995 - Samuel Lubbers

*Etude des interactions entre les macromolécules d'origine levurienne
du vin et les composés d'arôme*

Grand Prix 1994 - Ziya Günata

*Etude et exploitation par voie enzymatique
des précurseurs d'arôme du raisin, de nature glycosidique*

Grand Prix 1993 - Pierre-Louis Teissedre

Le plomb, du raisin au vin

Grand Prix 1992 - Pascal Chatonnet

*Incidence du bois de chêne sur la composition chimique et les qualités
organoleptiques des vins, applications technologiques*

Secrétariat de l'Académie Amorim

Emilie Loubié : +33 (0) 6 62 20 91 97

Email : emilie@academie-amorim.com

