



HAL
open science

Evaluation et mise en œuvre de la sélection phénotypique pour la sélection du blé tendre

Pauline Robert

► **To cite this version:**

Pauline Robert. Evaluation et mise en œuvre de la sélection phénotypique pour la sélection du blé tendre. Agronomie. Université Clermont Auvergne, 2022. Français. NNT : 2022UCFAC017 . tel-03870127

HAL Id: tel-03870127

<https://theses.hal.science/tel-03870127v1>

Submitted on 24 Nov 2022

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

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

Ecole doctorale des sciences de la vie, santé, agronomie et environnement

THESE

Présentée à l'Université Clermont-Auvergne

Pour l'obtention du grade de

DOCTEUR D'UNIVERSITE

Spécialité : Biologie Végétale

Evaluation et mise en œuvre de la sélection phénotypique pour la sélection du blé tendre

Soutenu le 10 mai 2022 par

Pauline Robert

Directeur de thèse : Jacques Le Gouis
Co-encadrement de la thèse : Renaud Rincant

Composition du jury :

Anne LAPERCHÉ , Maître de conférences (HDR) – Agrocampus Ouest	Rapportrice
Anne RICARD , Ingénieur en Chef des Ponts, des Eaux et des Forêts (HDR) – INRAE	Rapportrice
Christophe TATOUT , Professeur à l'université - UCA	Président du jury
Ellen GOUEMAND-DUGUE , Directrice de Recherche - Florimond Desprez	Examinatrice
Renaud RINCANT , Chargé de recherche - INRAE	Co-encadrant
Jacques LE GOUIS , Directeur de Recherche - INRAE	Directeur de thèse
Jérôme AUZANNEAU , Directeur recherche et développement – Agri-Obtentions	Invité
Sophie BOUCHET , Chargé de recherche – INRAE	Invitée

Travaux réalisés dans le cadre d'une convention CIFRE (2019/0060), avec l'aide de l'ANRT



Résumé

Evaluation et mise en œuvre de la sélection phénotypique pour la sélection du blé tendre

L'amélioration variétale connaît d'importants progrès depuis une vingtaine d'années grâce à l'apport de la génomique et des méthodes prédictives, comme la sélection génomique. D'autres types d'informations sont disponibles pour mieux modéliser la relation entre génotype et phénotype. L'objectif de cette thèse était d'évaluer et d'optimiser une nouvelle méthode de sélection, appelée sélection phénotypique, qui intègre la spectroscopie proche infrarouge dans les modèles prédictifs pour améliorer les programmes de sélection du blé tendre. Après avoir clairement défini la méthode, nous avons évalué, sur des données issues de programme de sélection, différents facteurs qui peuvent influencer les capacités de prédictions de cette méthode. Nous avons alors montré qu'à l'instar de la sélection génomique, la sélection phénotypique est aussi influencée par le modèle statistique, la taille et la composition du jeu de calibration. Les capacités de prédictions sont aussi influencées par l'origine d'acquisition des spectres et le nombre de spectres intégrés dans le modèle, et cela dépend du caractère à prédire. Nous avons ensuite évalué la sélection phénotypique dans différents scénarios de prédiction multi-environnementale et montré que l'interaction génotype-environnement est mieux modélisée par les spectres dans le proche infrarouge que par les marqueurs moléculaires. Sur l'ensemble de ces évaluations nous avons montré que la sélection phénotypique représente une bonne alternative ou un complément à la sélection génomique, et nous avons proposé différentes pistes d'intégration de la méthode dans les programmes de sélection.

Mots clés: amélioration des plantes, blé tendre, essais multi-environnementaux (MET), interaction génotype-environnement (GxE), sélection génomique, sélection phénotypique, spectroscopie proche infrarouge

Abstract

Evaluation and application of phenomic selection for bread wheat breeding

Plant breeding has made significant progress over the last twenty years with the contribution of genomics and predictive methods, such as genomic selection. Other types of information are available to better model the relationship between genotype and phenotype. The objective of this thesis was to evaluate a new selection method, named phenomic selection, which integrates near infrared spectroscopy in predictive models, to improve breeding programs for bread wheat. After having clearly defined the method, we evaluated on data from breeding programs, different factors that can influence the predictive ability of the method. We showed that like genomic selection, phenomic selection is also influenced by the statistical model, the size and the composition of the calibration set. The prediction abilities are also influenced by the origin of acquisition of the spectra and the number of spectra integrated in the model, and this depends on the trait to be predicted. We then evaluated the phenomic selection in different multi-environmental prediction scenarios and showed that the genotype-environment interaction is better modelled by NIR spectra than by molecular markers. On the whole, we showed that phenomic selection is an accurate and low-cost alternative or complement to genomic selection, and we proposed different ways of integrating the method in breeding programs.

Key words: Bread wheat, genomic selection, genotype-environment interaction (GxE), multi-environment trials (MET), near infrared spectroscopy (NIRS), phenomic selection, plant breeding.

Remerciements

Je tiens tout d'abord à remercier Anne Laperche et Anne Ricard, d'avoir accepté d'être les rapportrices de ce manuscrit, ainsi que l'ensemble des membres du jury, Christophe Tatout, Ellen Goudemand-Dugue, d'avoir accepté d'évaluer mon travail.

Je tiens à remercier sincèrement mon directeur de thèse Jacques Le Gouis et mon encadrant de thèse du quotidien, Renaud Rincet. Merci à vous qui m'avez fait confiance pour mener à bien cette thèse, et pour m'avoir soutenue et encouragée pendant ces trois années. Ces années ont été des plus particulières puisque marquées par de grands évènements : la pandémie mondiale et les multiples confinements mais aussi la migration de Renaud et moi-même au Moulon. Merci encore pour votre implication, votre bienveillance et votre suivi sur mon travail et pour nos discussions, aussi bien en présentiel qu'en distanciel.

Je tiens également à remercier les membres de mes comités de thèse, Vincent Segura, Martin Ecartot et Sophie Bouchet pour leur évaluation de l'avancement de mes travaux de thèse. Ces points étapes ont été source de riches discussions et de remarques très constructives. Merci encore pour l'expertise sur la spectroscopie proche infrarouge, celle-ci n'a plus de secret pour moi, ou presque.

Je remercie chaleureusement les équipes que j'ai rejoint au cours de ma thèse et qui m'ont toutes très bien accueillies. Pour ma première moitié de thèse au GDEC à Clermont-Ferrand, je remercie l'équipe QualiGrain pour les réunions que nous avons partagées et les pique-niques de fin d'année. De l'équipe Vista, que j'ai surtout côtoyée en pause-café, je remercie Vincent Allard, Boris Adam, David Cormier pour nos discussions sur la vie. Je n'oublie pas non plus Wa Ngo pour son aide lors des prises de spectres au champ et la remercie pour son sourire et sa bonne humeur inconditionnelle. Pour ma deuxième moitié de thèse au Moulon, merci à Laurence Moreau et Alain Charcosset de m'avoir adoptée dans l'équipe GQMS. Je remercie l'ensemble de l'équipe pour leur accueil bienveillant. Merci à Tristan Mary-Huard pour avoir répondu à mes questions en statistiques et pour les différents projets de codes sur R, j'ai beaucoup appris. Merci à Valérie Combes pour ces pâtisseries partagées aux pauses café et pour nos échanges dans le jardin du Moulon. Et oui ! j'ai eu la chance de connaître à la fois le site historique du Moulon, et le nouveau bâtiment resplendissant dans lequel GQE a déménagé. Il reste encore à faire quelques peintures...

Je remercie l'ensemble des personnes que j'ai rencontrées chez Agri-Obtentions et plus spécialement Jérôme Auzanneau, Laure Duchalais et Olivier Gardet qui ont toujours été à l'écoute, qui ont toujours pris le temps de répondre à mes questions et qui m'ont permis d'en connaître plus sur les programmes de sélection du blé tendre.

Je remercie également l'ensemble des personnes que j'ai rencontrées chez Florimond Desprez pour leur temps et l'intérêt qu'ils ont porté à mes travaux de thèse, et plus spécialement Philippe Lonnet et Thierry Demarquet. Finalement, je remercie chaleureusement Ellen Goudemand-Dugue, pour sa bienveillance, pour son entraide et son accueil dans le Ch'Nord toujours des plus agréable, avec une petite frite.

Je tiens à remercier aussi les sélectionneurs d'INRAE, François-Xavier Oury, Bernard Rolland, Emmanuel Heumez pour nos discussions autour de la sélection phénotypique et pour les explications de vos schémas de sélection. Je remercie encore Gilles Charmet pour avoir participé activement à nos réunions de brainstorming sur l'application de la sélection phénotypique dans les programmes de sélection du blé tendre. Merci encore une fois pour votre implication.

A tous ceux que j'ai croisé à INRAE, à tous les étudiants/doctorants/post-doctorants, un grand merci. Merci à Aurélien, Alizarine, Dimitri de m'avoir accueilli dans leur fine équipe, de m'avoir aidée sur le code et sur les accès au serveur. Merci pour nos petites soirées post confinement et d'ailleurs, j'attends toujours l'invitation pour la cabane d'ailleurs. Merci à Aurore, ma collègue de bureau et ma thésarde préférée, qui a toujours été là pour m'écouter, me soutenir et me transmettre sa bonne humeur.

Merci à ceux que j'ai rencontré tout au long de ce périple et qui sont devenus de vrais amis. Merci à Arthur et Simon avec qui la fête est plus folle ! Merci pour m'avoir fait partager vos aventures clermontoises ainsi que du bon rhum. Merci à mon amie de longue date Aurélie, pour être qui tu es, de toujours m'avoir soutenue depuis tellement d'années maintenant, pour tout ce que ton amitié m'apporte. J'ai hâte de voir ce qu'on va devenir !

Je souhaiterai finalement remercier ma famille pour m'avoir toujours donné les moyens de faire ce que je souhaitais, de m'avoir toujours soutenue même au cœur de la tempête. L'éclaircie n'est jamais loin. Merci tout spécialement à ma relectrice préférée, sans qui rien n'aurait été possible.

Enfin, merci à Pierre, tu sais tout déjà.

A mon père,

Liste des productions

Publications dans des revues à comité de lecture:

- Robert, P., Auzanneau, J., Goudemand, E., Oury, F. X., Rolland, B., Heumez, E., ... & Rincint, R. (2022). Phenomic selection in wheat breeding: identification and optimisation of factors influencing prediction accuracy and comparison to genomic selection. *Theoretical and Applied Genetics*, 1-20. <https://doi.org/10.1007/s00122-021-04005-8>
- Pauline Robert, Charlotte Brault, Renaud Rincint, Vincent Segura (2022). Phenomic selection: a new and efficient alternative to genomic selection (chapitre accepté pour faire partie de l'ouvrage « Complex Trait Prediction » de la série « Methods in Molecular Biology » édité par Springer) DOI: 10.1007/978-1-0716-2205-6
- Pauline Robert, Ellen Goudemand, Jérôme Auzanneau, François-Xavier Oury, Bernard Rolland, Emmanuel Heumez, Sophie Bouchet, Antoine Caillebotte, Tristan Mary-Huard, Jacques Le Gouis, Renaud Rincint. Phenomic selection in wheat breeding: prediction of the genotype-by-environment interaction in multi-environment breeding trials. (Soumis à la revue *Theoretical and Applied Genetics* le 18/02/2022)

Communications:

- Journées de l'Ecole Doctorale, 14-15 octobre 2020 (Présentation orale).
- Congrès international *EUCARPIA: Breeding and Biotechnologies* du 03 au 05 novembre 2021 (Présentation orale)

Listes des tableaux

Chapitre 1

Tableau 1-1. Caractères évalués pour la VATE lors de l'inscription de nouvelles variétés

Chapitre 2

Table 2-1. Selection of work using NIRS information in the context of breeding

Table 2-2. Comparative predictive ability reported for prediction based on G and H matrices

Chapitre 3

Table 3-1. Summary of the four datasets resulting from the trial evaluations

Table 3-2. Statistical models for computing adjusted means for GY, HD, and absorbance at each wavelength

Table 3-3. Description of the GS and PS models

Chapitre 4

Table 4-1. Summary of all genomic, phenomic and multi-omics models

Table 4-2. Predictive abilities of the different genomic, single NIRS, multi-NIRS and multi-omics models in scenario oGoE for GY and HD

Table 4-3. Predictive abilities of genomic, single NIRS, multi-NIRS and multi-omics models in scenario nGoE for GY and HD

Table 4-4. Predictive abilities of the genomic, single NIRS, multi-NIRS and multi-omics models in scenario oGoEref for GY

Table 4-5. Predictive abilities of genomic, single NIRS, multi-omics and multi-NIRS models in scenario nGoEref for GY

Chapitre 5

Tableau 5-1. Choix des modèles dont la capacité de prédiction a été prise en compte pour calculer la moyenne des capacités de prédiction de la SG, la SP et du Multiomique sur l'ensemble des données

Tableau 5-2. Comparaison des capacités de prédiction entre trois méthodes de prédictions différentes (SP, SG et Multiomique)

Tableau 5-3. Capacité de prédiction (PA) du rendement lorsqu'un modèle de prédiction phénomique (H-BLUP) est entraîné avec la génération F5 pour prédire la génération F6 et inversement

Listes des figures

Chapitre 1

Figure 0-1. Modèle de l'histoire phylogénique du blé tendre et de la structure de son génome

Figure 0-2. Morphologie du blé tendre

Figure 0-3. Anatomie du grain de blé tendre

Figure 0-4. Chronologie des différentes phases de développement du blé tendre, du semis à la récolte et de l'initiation des composantes du rendement

Figure 0-5. Production annuelle moyenne de blé dans le monde pour les dix premiers pays producteurs

Figure 0-6. Représentation des rendements du blé tendre par département et de la surface allouée à la culture du blé tendre par département

Figure 0-7. Représentation de la tendance des rendements dans les différents pays producteurs de blé tendre

Figure 1-8. Représentation des interactions qualitative et quantitative entre le génotype et l'environnement

Figure 1-9. Comparaison entre l'utilisation de la méthode Bulks, SSD et de l'haplodiploïdisation dans un schéma de sélection généalogique différée

Figure 1-10. Processus de sélection génomique utilisant une population d'entraînement pour estimer la valeur génétique d'une population de candidats

Figure 1-11. Évolution du coût de séquençage du génome humain depuis 2001

Figure 1-12. Comparaison d'un programme de sélection du blé tendre, basé sur la sélection assistée par marqueurs ou sur la sélection génomique

Chapitre 2

Figure 2-1. Prediction of a target trait for the selection candidates in different environments with GLOB selection

Figure 2-2. Visualization of different filters applied on a spectrum dataset

Chapitre 3

Figure 3-1. Scheme of NIR spectra acquisition at different stages on the same inbreds (at different generations of selfing)

Figure 3-2. Boxplot of predictive abilities of phenomic selection for grain yield (GY) and heading date (HD) based on four different wheat breeding datasets (cf. Table 3-1) with two different models (H-BLUP, PS-LASSO)

Figure 3-3. Boxplot of the predictive abilities of phenomic selection for grain yield (GY: **A,C**) and heading date (HD: **B,D**) in subgroup panels of wheat genotypes SubSet_SD (**A,B**) and SubSet_SDY (**C,D**) (cf. Table S5)

Figure 3-4. Predictive ability plotted as a function of the number of combined NIR spectra for grain yield (GY) and heading date (HD) for the panel-environment combinations of Set4 genotypes from year 2019

Figure 3-5. Boxplots of the predictive abilities of GS and PS for grain yield (GY) and heading date (HD) using different wheat breeding datasets (SubSet_SDM and SubSet_SYM)

Figure 3-6. Comparison of predictive abilities of GS and PS for GY as a function of the size of the training population for four predictive models

Figure 3-7. Comparison of predictive abilities for GY as a function of the size of the training population for four predictive models and three optimisation algorithms

Chapitre 4

Figure 4-1. Schematic overview of the six prediction scenarios representing the associated missing data to be predicted in the MET

Figure 4-2. Correlation coefficient matrices between trial environments for heading date adjusted means and grain yield adjusted means

Figure 4-3. Proportion of the genomic (G), the genomic x environment (GxE) and residual variances across the NIR spectra of winter wheat grains from 5 different MET sets

Chapitre 5

Figure 5-1. Boîte à moustache des prédictions phénotypiques pour le rendement (GY) et la date d'épiaison (HD sur l'ensemble des jeux de données pour les différents scénarios

Figure 5-2. Représentation des coefficients de corrélation de Pearson entre environnements et des moyennes de capacité de prédiction pour la teneur en protéine

Figure 5-3. Schéma des différents facteurs, influençant les capacités de prédiction de la sélection phénotypique, identifiés dans mes travaux de recherche (Chapitre III et Chapitre IV) mais aussi dans la littérature

Figure 5-4. Représentation des prédictions de rendements en fonction des moyennes ajustées pour 65 lignées du jeu de données Set2_2018_CF

Figure 5-5. Boîtes à moustaches de la capacité de prédiction phénotypique (SP), de la capacité de prédiction génomique (SG) et de la capacité de prédiction de la méthode combinant marqueurs moléculaires et spectres NIR (Multiomique) pour le rendement (GY) et la date d'épiaison (HD) basées sur l'ensemble des jeux de données et des scénarios de prédictions

Figure 5-6. Augmentation théorique du gain génétique attendu (%) en utilisant la SP plutôt que la SG dans le scénario SE, pour différentes valeurs de fiabilité de chacune des méthodes, pour différents coûts de génotypage (20 € ou 10 €) et d'acquisition des spectres NIR (10 €, 6 €, 2 €)

Figure 5-7. Augmentation théorique du gain génétique attendu (%) en utilisant la SP plutôt que la SG dans le scénario MET oGoE, pour différentes valeurs de fiabilité de chacune des méthodes, pour différents coûts de génotypage (20 € ou 10 €) et de coûts d'acquisition de spectres NIR (10 €, 6 €, 2 €)

Figure 5-8. Augmentation théorique du gain génétique attendu (%) en utilisant la méthode Multiomique plutôt que la SG dans le scénario SE, pour différentes valeurs de fiabilité de chacune des méthodes, pour différents coûts de génotypage (20 € ou 10 €) et de coûts d'acquisition de spectres NIR (10 €, 6 €, 2 €)

Figure 5-9. Augmentation théorique du gain génétique attendu (%) en utilisant la méthode Multiomique plutôt que la SG dans le scénario MET oGoE, pour différentes valeurs de fiabilité de chacune des méthodes et pour différents coûts de génotypage (20 € ou 10 €) et de coûts d'acquisition de spectres NIR (10 €, 6 €, 2 €)

Figure 5-10. Schéma de sélection alternatif intégrant la sélection phénotypique dans le cycle de sélection

Figure 5-11. Schéma de sélection alternatif intégrant la sélection phénotypique, et la sélection assistée par marqueurs dans le cycle de sélection

Figure 5-12. Schéma de sélection alternatif intégrant les prédictions phénotypiques, génomiques et multiomiques, ainsi que la sélection assistée par marqueurs dans le cycle de sélection

Figure 5-13. La sélection phénotypique en sélection du blé tendre

Listes des principales abréviations

BLUP: Best Linear Unbiased Predictor

CE: Covariables environnementales

CDmean: Moyenne du coefficient de détermination généralisé

DL: Déséquilibre de liaison

GEBV: Genomic Estimated Breeding Value

GLOB: Genomic-Like Omics-Based – Comme la génomique mais basé sur les omiques

GxE: Interaction génotype-environnement

GY: Grain Yield - Rendement

HD: Heading date

HTP: High-Throughput Phenotyping – Phenotypage Haut Débit

INRAE: Institut National de la Recherche en Agriculture, Alimentation et Environnement

MET: Essais multi environnementaux

NGS: Next Generation Sequencing – Séquençage de nouvelle génération

NIR: Near infra-red – Proche infrarouge

NIRS: Near infra-red spectroscopy – Spectroscopie proche infrarouge

QTL: Quantitative Trait Locus - Locus de caractères quantitatifs

PA: Predictive ability – Capacité de prédiction

RKHS: Reproducing Kernel Hilbert Space

SAM: Sélection assistée par marqueurs

SG (GS): Sélection génomique

SNP: Single Nucleotide Polymorphisms – Polymorphisme d'un seul nucléotide

SP (PS): Sélection phénotypique

SSD: Single Seed Descent – Descendance mono graine

T.: *Triticum*

TS: Training set -Population d'entraînement

UAV: Unoccupied Aerial Vehicle – Drone (aéronef sans équipage)

VI: Végétation index – Indice de végétation

Table des matières

Résumé	3
Abstract.....	5
Remerciements	7
Liste des productions	11
Listes des tableaux.....	13
Listes des figures	15
Listes des principales abréviations	17
Introduction générale	25
Chapitre 1	31
1. Histoire du blé tendre, une culture d'importance majeure	33
1.1. Le blé tendre (<i>Triticum aestivum</i> L.) et le genre <i>Triticum</i>	33
1.2. Histoire de la domestication du genre <i>Triticum</i>	33
1.3. Description botanique du blé tendre.....	35
1.4. Cycle de développement du blé tendre et rendement: de la graine aux grains	37
1.5. Une culture d'intérêt économique majeur	39
1.6. Les enjeux liés à la production du blé tendre	41
2. La sélection du blé tendre et son amélioration génétique.....	43
2.1. L'amélioration génétique pour la création de nouvelles variétés	43
2.2. La sélection phénotypique, une sélection basée sur l'observation des individus	44
2.3. Schéma de sélection pour la création de lignée pure de blé tendre	46
2.4. L'inscription des variétés de blé tendre.....	50
3. Utilisation des marqueurs moléculaires et autres omiques en sélection.....	52
3.1. La sélection assistée par marqueurs (SAM)	52
3.2. La sélection génomique.....	53
3.3. La sélection "genomic-like omics-based" (GLOB).....	64
4. Problématique et objectifs de la thèse	66
Chapitre 2	69
Préambule	71
Phenomic selection: a new and efficient alternative to genomic selection	73
1. Abstract.....	73
2. Introduction: the concept of phenomic selection and its relationship with other uses of spectra in breeding	74

3.	Literature review on the use of spectra in selection	79
3.1.	Types of technology.....	79
3.2.	Preprocessing NIR spectra.....	81
3.3.	Statistical models for phenotype prediction	83
3.4.	Relative performance of PS versus GS	86
3.5.	Factors affecting PS predictive ability.....	88
4.	Prospects	90
4.1.	Prebreeding: screening diversity collections at low cost	90
4.2.	Sparse testing: experimental design optimization in breeding programs	90
4.3.	Combining reduction of generation time (speed breeding) and performance prediction (PS) to increase genetic progress	91
4.4.	GxE prediction	92
4.5.	Making use of historical NIRS data in prediction.....	93
4.6.	The case of perennials	93
4.7.	Other applications	94
5.	Conclusion.....	95
	Chapitre 3.....	99
	Préambule	101
	Phenomic selection in wheat breeding: identification and optimisation of factors influencing prediction accuracy and comparison to genomic selection	103
1.	Abstract.....	103
2.	Introduction	105
3.	Materials and Methods	108
3.1.	Phenotyping, genotyping and NIRS data	108
3.2.	Statistical models for adjusting data and estimating heritability	112
3.3.	Calculation of the genetic and hyperspectral relationship matrices.....	113
3.4.	Reference prediction models and prediction scenario	114
3.5.	Evaluation of factors influencing predictive ability of PS.....	115
4.	Results	121
4.1.	Descriptive statistics, heritability, and correlation between traits and absorbances	121
4.2.	Phenomic predictions of grain yield and heading date in different bread wheat breeding programs	121
4.3.	Impact of NIR spectrum origin of acquisition on predictive ability	123
4.4.	Influence of the number of NIRS combined in multiple NIRS models on predictive ability.....	124
4.5.	Comparison of GS and PS predictive abilities	126
5.	Discussion.....	128
5.1.	Phenomic selection is an accurate complement or alternative to genomic selection in elite breeding material 128	
5.2.	Factors influencing PS predictive ability.....	129

6. Conclusion.....	133
7. Appendix 1	135
Chapitre 4.....	139
Préambule	141
Phenomic selection in wheat breeding: prediction of the genotype-by-environment interaction in multi-environment breeding trials	143
1. Abstract.....	143
2. Introduction	145
3. Materials and Methods	148
3.1. Plant material, genotyping and NIRS acquisition.....	148
3.2. Analysis of the phenotypes measured in the METs, genotypic data and NIRS data	149
3.3. Estimation of the covariance matrices describing the similarities between genotypes or environments.	151
3.4. Scenarios of prediction and corresponding cross-validation schemes.....	153
3.5. Models developed for the different multi-environment prediction scenarios.....	156
3.6. Imputation of missing spectra for unphenotyped varieties	159
4. Results	162
4.1. Characterisation of the METs based on phenotypes.....	162
4.2. Decomposition of the variance along NIR spectrum.....	162
4.3. Comparison of genomic, phenomic and multi-omics models in two original MET scenarios for the prediction of GY and HD.....	165
5. Discussion.....	169
5.1. Phenomic prediction and genomic prediction performed similarly and were outperformed by multi-omics prediction in MET designs	169
5.2. Modelling the GxE effect with NIR spectra improves the ability to predict grain yield	170
5.3. Original MET designs enable breeders to use PS at minimal cost	174
5.4. Phenomic selection is a promising tool to improve breeding programs	176
6. Conclusion.....	177
Chapitre 5.....	179
Préambule	181
1. La sélection phénomique: une nouvelle approche de prédiction pour l'amélioration des plantes.....	183
1.1. L'utilisation des spectres NIR et de la sélection phénomique pour la prédiction de caractères agronomiques	183
1.2. Des facteurs multiples qui influencent les qualités de prédiction et qui peuvent être optimisés	184
1.3. Intérêt de l'utilisation d'une variable telle que la spectroscopie proche infrarouge pour prédire les effets d'interactions génotype-environnement	192
2. La sélection phénomique peut-elle remplacer la sélection génomique ?	195

2.1.	Avantages et limites de l'utilisation des spectres NIR par rapport aux marqueurs moléculaires	195
2.2.	Sélection génomique et sélection phénotypique: des méthodes en concurrence... ..	199
2.3.	... ou qui peuvent être complémentaires	210
3.	Comment intégrer la sélection phénotypique dans les schémas de sélection du blé tendre ? ...	215
3.1.	Les éléments clés identifiés pour introduire la méthode dans des programmes de sélection	215
3.2.	Proposition de schéma alternatifs intégrant la sélection phénotypique à différentes étapes du programme de sélection en généalogie différée.	219
4.	Quelles autres perspectives d'applications de la sélection phénotypique en amélioration des plantes ?	227
	Références	233
	bibliographiques	233
	Annexes	245
	Annexe 1: données supplémentaires du chapitre 3.....	247
	Annexe 2 : données supplémentaires du chapitre 4.....	261

Introduction générale

Le blé tendre est la troisième céréale la plus produite dans le monde après le maïs et le riz (FAO 2022). Cultivée dans la plupart des régions du globe, elle fournit environ 20% des calories et protéines consommées par l'alimentation humaine (Shewry et Hey 2015). Un des enjeux pour la filière du blé tendre, est de pouvoir répondre à la demande croissante de la consommation humaine pour les prochaines décennies à venir. Depuis le milieu du 20^{ème} siècle, la croissance démographique augmente très rapidement. S'il a fallu des milliers d'années pour atteindre le milliard d'habitants sur Terre en 1800, les projections annoncent que la population mondiale va être multipliée par 10 en seulement 250 ans, pour atteindre 10 milliards d'habitants en 2050. Cette croissance de population ainsi que les changements de régimes alimentaires des pays à fort développement vont entraîner un doublement de la demande en céréales d'ici trente ans (Brisson *et al.* 2010).

Accroître la production mondiale est l'un des enjeux de la filière du blé tendre. Les possibilités d'accroissement des surfaces cultivées sont limitées (Pretty 2008). L'augmentation des rendements va donc être nécessaire et elle peut se faire par le développement de nouvelles variétés plus performantes, on parle alors de progrès génétique, ou par l'évolution des pratiques culturales (Oury *et al.* 2012). Malgré un progrès génétique constant en Europe, les rendements nationaux stagnent et cela depuis les années 90. Cette stagnation est le résultat de deux effets majeurs : le changement des pratiques agricoles et le changement climatique. D'un côté, les nouvelles pratiques agricoles tendent à limiter l'utilisation des produits pétro-chimiques (engrais, pesticides) et plus globalement à réduire les intrants dans les cultures (Brisson *et al.* 2010). D'un autre côté, le changement climatique amplifie l'intensité et la fréquence des stress (thermiques, hydriques) subis par les cultures durant leur cycle de développement (Semenov et Shewry 2011). L'amélioration variétale a donc pour défi de créer rapidement de nouvelles variétés qui répondront à ces différents enjeux.

L'amélioration variétale permet de créer de nouvelles variétés de blé tendre avec de meilleures performances pour le rendement, la qualité boulangère, les résistances aux maladies. Ce progrès génétique obtenu avec les programmes de sélection est globalement constant d'une année à l'autre mais insuffisant pour espérer atteindre les objectifs de production d'ici 2050. Pour gagner en efficacité dans le programme de sélection, une méthode de prédiction a été développée dans les années 2000 (Bernardo 1994; Meuwissen *et al.* 2001). Les prédictions génomiques permettent de prédire les performances d'individus non phénotypés pour le caractère d'intérêt, grâce au génotypage. Elle repose sur l'estimation simultanée de l'effet de tous les marqueurs moléculaires du génome. Cette méthode s'est rapidement popularisée dans le monde animal puis dans le monde

végétal, grâce notamment à une diminution du coût du génotypage. La sélection génomique est particulièrement efficace pour des caractères complexes, polygéniques. Elle permet donc en théorie d'améliorer le gain génétique annuel en réduisant le nombre d'années du cycle de sélection. Seulement, en réalité, très peu d'études ont encore montré l'efficacité de la sélection génomique dans de vrais programmes de sélection en création végétale (Crossa *et al.* 2017). Il existe en effet des freins à l'utilisation de la sélection génomique comme le coût du génotypage qui reste très coûteux pour des petites structures de sélection ou pour des espèces mineures. De plus, les capacités de la prédiction génomique sont aussi limitées par le fait que les marqueurs moléculaires modélisent difficilement les effets d'interactions entre gènes (épistasie) ou d'interactions entre le génotype et l'environnement.

Depuis quelques années, les modèles prédictifs intègrent de nouvelles sources de variabilité pour modéliser les différents effets responsables de la variance du caractère complexe. C'est le cas des endophénotypes qui correspondent à des caractères secondaires entre le génotype et le phénotype final. Ces endophénotypes se retrouvent aux différentes strates d'omiques comme la transcriptomique, la métabolomique. S'il a été montré que ces types d'endophénotypes sont des bons prédicteurs des performances du maïs (Riedelsheimer *et al.* 2012; Schrag *et al.* 2018; Azodi *et al.* 2020) ou encore du blé (Ward *et al.* 2015), ils restent tout de même difficiles à acquérir, très coûteux et donc difficiles à mettre en place en routine par les sélectionneurs, à l'échelle de leur programme. En 2018, Rincent *et al.* ont proposé une nouvelle méthode de sélection : la sélection phénotomique. Celle-ci repose sur l'utilisation de la spectroscopie proche infrarouge pour prédire différents caractères du blé tendre et du peuplier. De façon similaire à la sélection génomique, la sélection phénotomique repose sur l'estimation simultanée des effets des longueurs d'onde du spectre, qui remplacent les marqueurs moléculaires. Cette méthode a l'avantage d'être très accessible et peu coûteuse pour les sélectionneurs, puisque la prise de spectres est déjà faite en routine, chez les céréales, pour différentes applications, notamment la prédiction de la concentration en protéines du grain. Seulement, cette méthode étant très récente, elle a encore été très peu évaluée et nécessite donc une caractérisation plus fine sur des données de sélectionneurs.

Ainsi le présent travail de thèse s'inscrit dans l'évaluation d'une nouvelle méthode de sélection et de sa potentielle mise en application pour améliorer significativement le progrès génétique des programmes de sélection du blé tendre.

Chapitre 1

Synthèse bibliographique

1. Histoire du blé tendre, une culture d'importance majeure

1.1. Le blé tendre (*Triticum aestivum* L.) et le genre *Triticum*

Le blé tendre (*Triticum aestivum* L.) est une plante annuelle, monocotylédone, de la famille des *Poaceae* (anciennement les *Gramineae*), appartenant au genre *Triticum* (*T.*). Cette famille regroupe la plupart des plantes plus communément appelées herbes et céréales, et compte près de 12 000 espèces. Ainsi, les sous-familles des *Panicoideae* (sorgho, maïs), les *Ehrhartoideae* (riz) et les *Pooideae* (blé, orge) représentent les espèces les plus cultivées dans le monde. Le genre *Triticum* regroupe un certain nombre d'espèces désignées sous le terme de « blé » dont le blé tendre (*T. aestivum*) et le blé dur (*T. durum*), deux espèces de blés à grains nus d'intérêt économique majeur. Le blé tendre est principalement cultivé, en France, pour : l'alimentation humaine (meunerie, panification et biscuiterie), l'alimentation animale, l'amidonnerie, la glutennerie et enfin l'éthanolerie. Le blé dur, quant à lui, est cultivé en France uniquement pour l'alimentation humaine (confection de pâtes et de semoules). Dans ce genre *Triticum*, sont également cultivés de manière plus marginale, des blés à grain vêtus comme l'épeautre (*T. aestivum* ssp. *spelta*), l'engrain (*T. monococcum*) ou encore l'amidonnier (*T. turgidum* subsp. *dicoccon*).

Dans la suite du manuscrit, seul le blé tendre sera considéré et dénommé par le terme « blé ».

1.2. Histoire de la domestication du genre *Triticum*

La domestication du blé dans la région du Croissant Fertile remonte à il y a près de 10 000 ans, période pendant laquelle les espèces de blés sauvages (des genres *Triticum* et *Aegilops*) ont progressivement laissé place à des variétés diploïdes ou polyploïdes (tétra ou hexaploïdes) de blés domestiqués. Ceci s'est traduit par des changements majeurs dans leurs caractéristiques morphologiques, à savoir : l'augmentation de la taille du grain, un rachis (tige centrale de l'épi) plus rigide, permettant de maintenir les grains dans l'épi jusqu'à la récolte et enfin des glumes protégeant le grain et le libérant plus facilement (Salamini *et al.* 2002). Le blé tendre actuel est allohexaploïde ($2n = 6x = 42$ chromosomes), composé de trois sous-génomes (A, B et D) possédant chacun sept paires de chromosomes qui sont dit homéologues (numérotés de 1 à 7). Les trois génomes coexistent dans le même noyau de manière stable, car l'appariement entre chromosomes homéologues n'est pas possible lors de la méiose. L'appariement se fait exclusivement entre

chromosomes homologues, et cela grâce à un groupe de gènes, dont le gène Ph1 (*Pairing homeologous 1*) situé sur le chromosome 5B. Ainsi, à quelques exceptions près, le blé tendre se comporte comme une espèce diploïde lors de la méiose (Griffiths *et al.* 2006). *T. aestivum* est issu de deux hybridations interspécifiques successives (Figure 1-1). La première a eu lieu il y a environ 0.8 million d'années, entre deux espèces diploïdes ($2n=2x=14$) sauvages *T. urartu* (AA) et *Aegilops speltoides* (BB) (Marcussen *et al.* 2014). Après doublement chromosomique, une nouvelle espèce tétraploïde, *T. turgidum*, s'est alors répandue. Un second épisode d'hybridation a eu lieu entre une forme probablement domestiquée de *T. turgidum* (AABB ; $2n=4x=28$ chromosomes) et une espèce diploïde *Aegilops tauschii* (DD ; $2n = 2x = 14$ chromosomes) il y a environ 10 000 ans (Salamini *et al.* 2002). D'après Marcussen *et al.* (2014) l'apparition du blé tendre pourrait avoir eu lieu il y a 0.4 millions d'années et donc être bien plus âgé que l'estimation actuelle. Cependant, les traces archéologiques, et l'absence de blés hexaploïdes dans les populations sauvages corroborent l'estimation de 10 000 ans. C'est donc de cette seconde hybridation qu'est issu l'ancêtre hexaploïde du blé tendre.

Le blé tendre possède un génome de grande taille, de l'ordre de 17 000 mégabases (Mb), bien plus gros que celui du riz (400 Mb) ou du maïs (3 000 Mb), et qui est complexifié par la redondance des gènes et des séquences répétées portés par les trois sous génomes apparentés (Figure 1-1). Ainsi, 85% du génome est constitué de séquences répétées (The International Wheat Genome Sequencing Consortium (IWGSC) *et al.* 2018). Ces épisodes d'allopoléidisation, c'est-à-dire de duplication d'un génome hybride issu de deux espèces différentes, ont permis la création de nouvelles espèces aux génomes complexes, leur conférant des avantages évolutifs notamment en termes d'adaptabilité à diverses conditions climatiques, et ont contribué au succès de leur domestication (Dubcovsky et Dvorak 2007; Shewry 2009). Le blé tendre cultivé de nos jours, est donc le fruit de la domestication et de siècles de cultures.

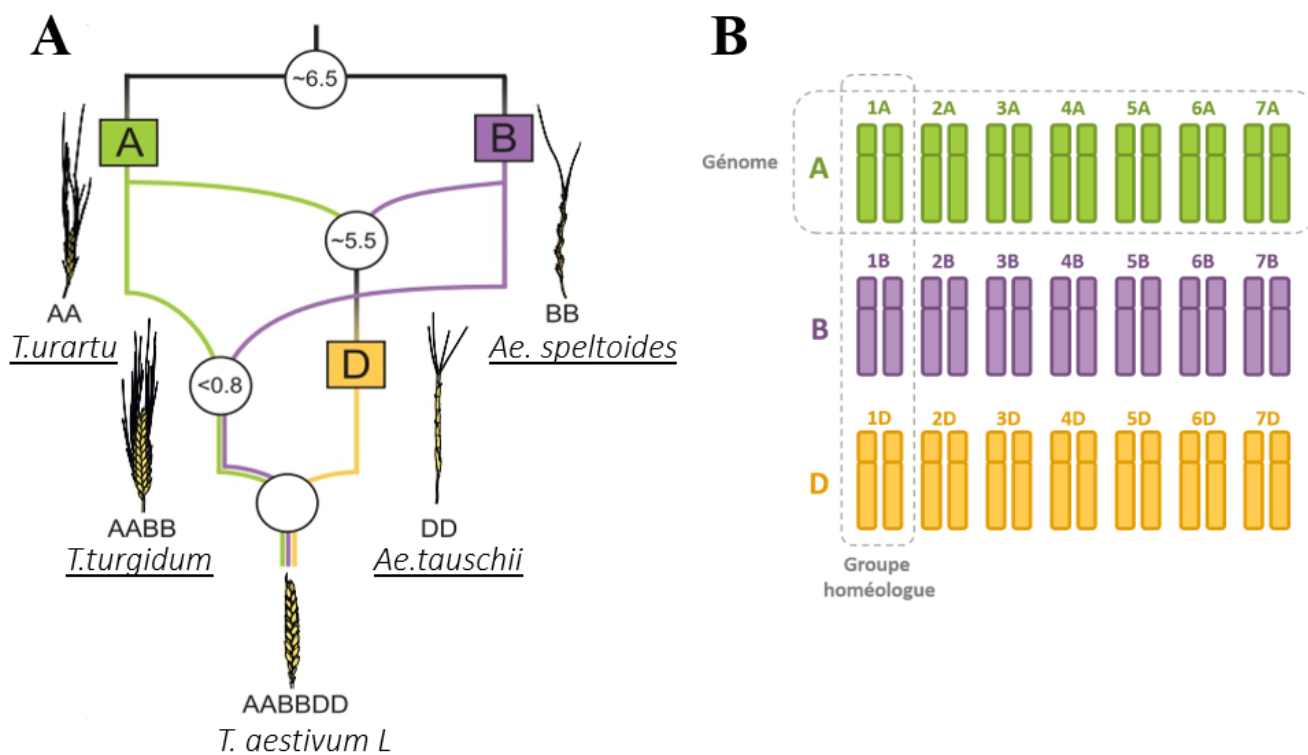
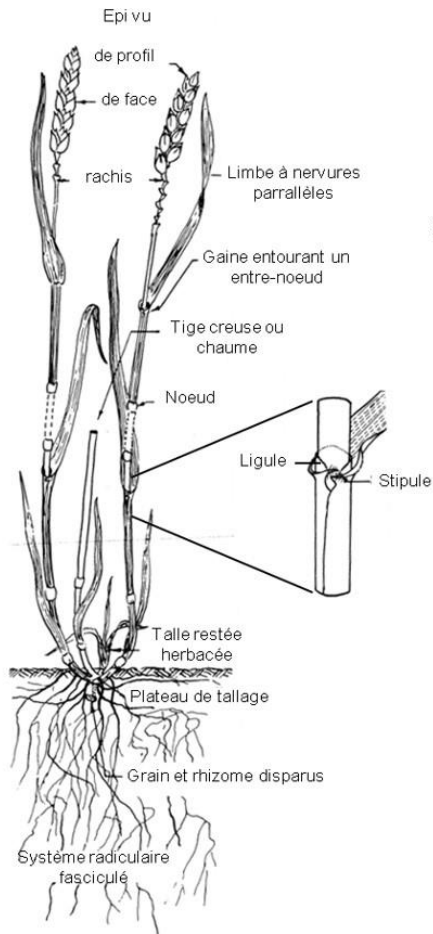


Figure 1-1. Modèle de l’histoire phylogénique du blé tendre et de la structure de son génome. (A) Evènements d’hybridation qui ont mené à la naissance de *Triticum aestivum* L. Dans les cercles sont précisées les dates approximatives (en millions d’années) des épisodes d’hybridation. Le cercle vide n’a pas de date d’hybridation car les preuves archéologiques désignent l’apparition du blé tendre il y a 10 000 ans et les travaux de Marcussen *et al.* (2014) désignent l’apparition il y a 0.4 million d’années. (B) Le génome du blé tendre est composé de trois sous-génomes A, B et D diploïdes contenant chacun 7 paires de chromosomes (adapté de Marcussen *et al.* 2014).

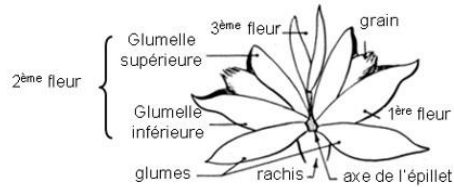
1.3. Description botanique du blé tendre

Le blé tendre est une plante herbacée annuelle à croissance définie. Elle est constituée d’un ensemble de brins appelés talles. Chaque talle est constituée d’une tige feuillée, encore appelée chaume, qui porte un épi à son extrémité (Figure 1-2, A). La première talle, issue du bourgeon apical, est appelée maître brin. Les talles suivantes, émises par les bourgeons axillaires, seront appelées “talles secondaires” et se développent au niveau du plateau de tallage. Chacune des talles se met en place de manière successive. L’épi porté par chacune des talles est constitué de deux rangées d’épillets (pour un total de 15 à 25 épillets) situés de part et d’autre de son axe principal, le rachis. Un épillet est un ensemble de fleurs, généralement trois, ordonnées à l’intérieur de deux glumes (Figure 1-2, B). Chaque fleur contient trois étamines et un ovaire surmonté de deux styles plumeux (Figure 1-2, C).

A – La plante entière



B – L'épillet



C – La fleur

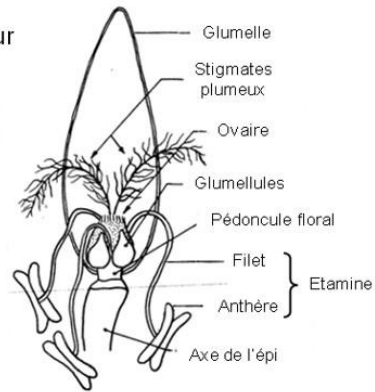


Figure 1-2. Morphologie du blé tendre (d'après Soltner, 1988 et tiré de Bednarek 2013). Sur ces schémas sont décrits les différents éléments constituant le système racinaire et le système aérien du blé tendre (A), et plus finement ce qui constitue un épillet (B).

La floraison ou anthèse, s'observe lorsque les étamines (ou anthères) sortent des glumelles de la fleur. C'est le point de référence pour la fécondation. La fleur du blé a la caractéristique d'être cléistogame, c'est-à-dire que la fécondation a lieu avant que la fleur ne s'ouvre et que les anthères n'apparaissent à l'extérieur. Le mode de reproduction le plus fréquent chez le blé est donc l'autofécondation. Finalement, le fruit de cette autofécondation est un grain sec indéhiscent, aussi appelé caryopse. Le grain est composé de trois principaux compartiments: l'embryon ou germe, l'albumen et les enveloppes (Figure 1-3). L'albumen constitue les réserves du grain et est le principal constituant de la farine ; il est composé entre 80% et 85% de matière sèche (Belderok *et al.* 2000). L'albumen amylicé contient à la fois la totalité de l'amidon présent dans le grain et une

partie des protéines du grain. L'embryon est, quant à lui, plutôt riche en lipides, huiles et globulines (Shewry *et al.* 2003).

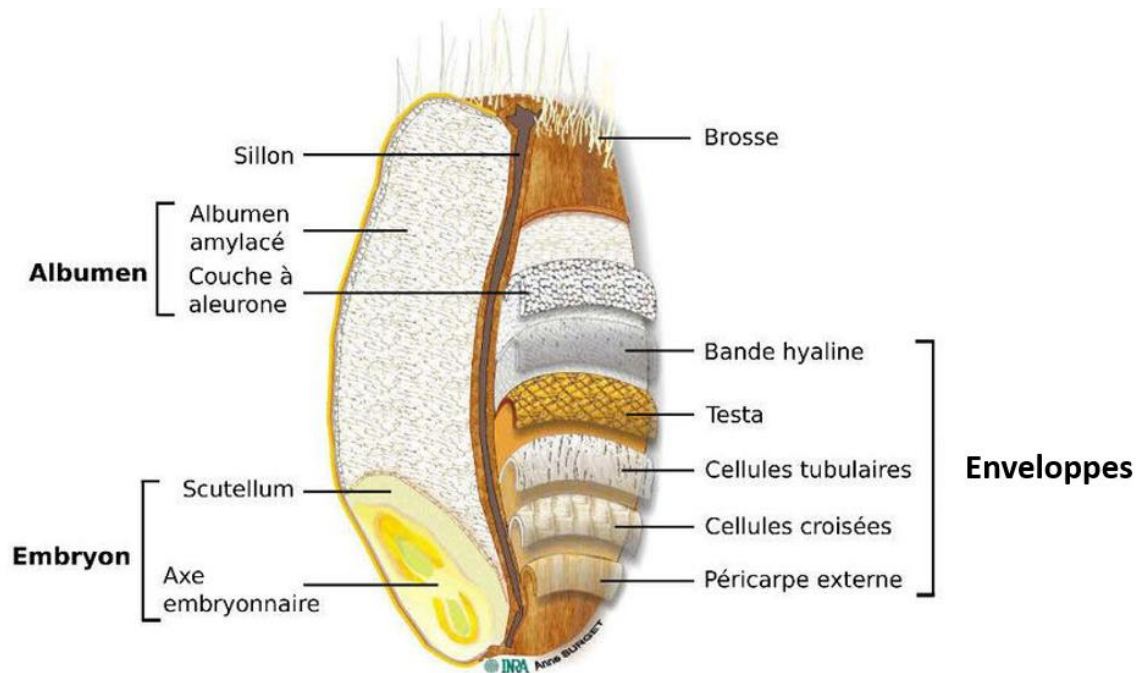


Figure 1-3. Anatomie du grain de blé tendre. Le grain de blé est constitué de trois principaux compartiments: l'embryon, l'albumen et les enveloppes (d'après (Surget et Barron 2005)).

1.4. Cycle de développement du blé tendre et rendement: de la graine aux grains

Le cycle de développement du blé est une succession d'étapes allant de la germination à la récolte de la plante. L'observation de ces différentes étapes a permis l'identification de stades clés et la construction d'échelles de développement, dont celle de (Zadoks *et al.* 1974), qui est largement utilisée en agronomie. Cette échelle annotée de 00 à 99, décrit finement les différents stades. Le premier chiffre définit les 10 stades les plus importants du cycle comme par exemple: (0) la germination, (1) la plantule, (2) le tallage, (5) l'épiaison, (6) la floraison, (8) la maturation des graines et (9) la sénescence. Le deuxième chiffre définit une sous étape comme par exemple: 55 qui correspond à l'étape de mi-épiation pour laquelle 50% de l'épi est sorti de la gaine. Plus simplement, le cycle se décompose en trois grandes phases: la phase végétative, la phase

reproductive et la phase de remplissage des grains (Figure 1-4). Durant la phase végétative, qui se déroule de la germination au début de la montaison (allongement de la tige principale), la plante va mettre en place son système racinaire et son système foliaire (feuilles et tiges). Lors de la phase reproductive, l'inflorescence va se développer: l'épi va sortir progressivement de la gaine, suivi de la sortie des fleurs. La dernière phase est celle du remplissage des grains, phase pendant laquelle les grains formés après fécondation vont s'enrichir en glucides et protéines. Pour garantir une bonne conservation des grains au cours du stockage, la récolte se fait généralement lorsque la teneur en eau des grains est inférieure à 15%. Le passage de la phase végétative à la phase reproductive est remarquable chez le blé, car elle permet de distinguer deux classes de blés: les blés d'hiver et les blés de printemps. Pour les blés d'hiver, une vernalisation est nécessaire pour initier ensuite l'appareil reproducteur, c'est-à-dire une période de 30 à 60 jours pendant laquelle la plante est soumise à des températures froides. Les blés d'hiver sont généralement semés en octobre, en France. Les blés de printemps eux n'ont pas besoin de cette période de vernalisation et peuvent être semés après l'hiver (Acevedo *et al.* 2002). Ce critère de durée du cycle végétatif permet une adaptabilité du blé tendre à différentes conditions climatiques. Par exemple, au Canada où les hivers sont trop froids pour cultiver du blé d'hiver, seul le blé de printemps peut y être cultivé. En France, la culture d'un blé d'hiver qui n'aurait pas supporté des températures trop basses durant l'hiver peut être rattrapé par une culture de blé de printemps à la sortie de l'hiver. Cependant le cycle de croissance d'un blé de printemps étant plus court que celui d'un blé d'hiver, les rendements sont généralement moins élevés.

C'est durant les trois phases de développement que se fait l'élaboration du rendement. Celle-ci peut se décomposer en différentes composantes principales: le nombre d'épis par m² (NE/m²), le nombre de grains par épi (NG/E), et la masse moyenne des grains exprimée en Poids de Mille Grains (PMG). La première composante peut être elle-même décomposée en sous composantes: le nombre de plantes par m² (NP/m²) et le nombre d'épis par plante (NE/P). L'élaboration du rendement est donc la multiplication de chacune de ces composantes, qui sont influencées à la fois par la génétique de la plante mais aussi par les conditions environnementales. Ces composantes se mettent en place en cascade durant le cycle de développement de la plante (Figure 1-4).

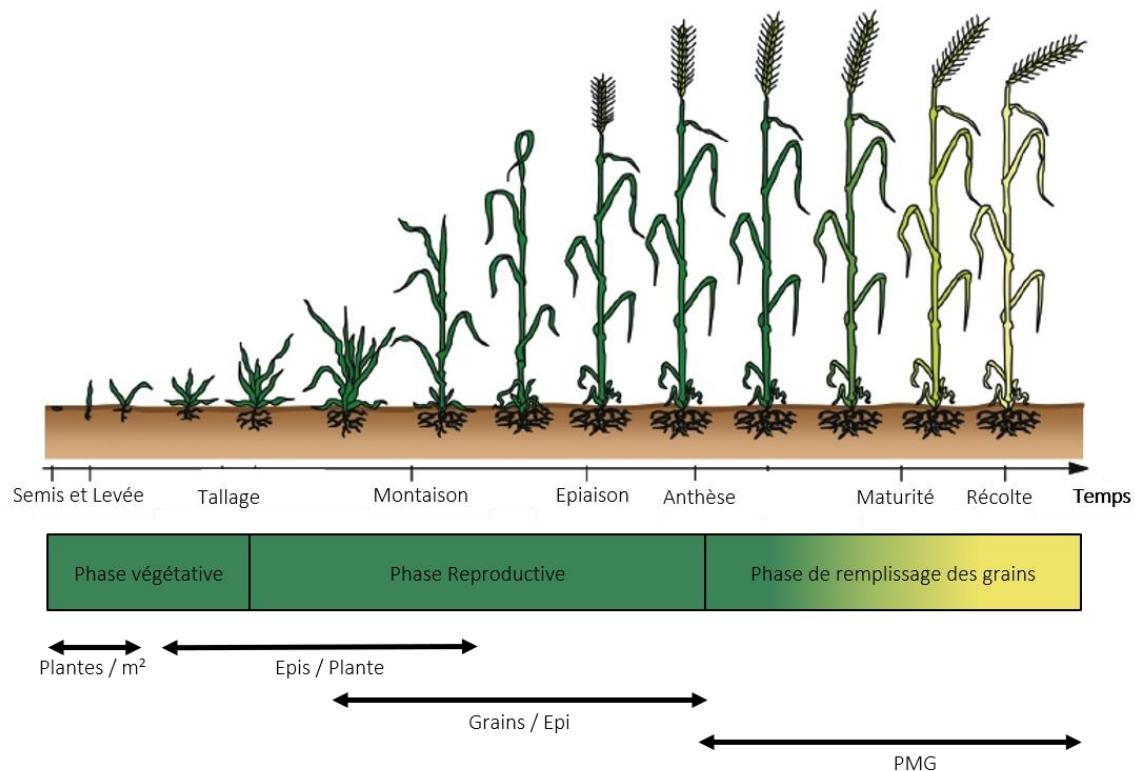


Figure 1-4. Chronologie des différentes phases de développement du blé tendre, du semis à la récolte et de l'initiation des composantes du rendement (adapté de (Guo *et al.* 2015)).

1.5. Une culture d'intérêt économique majeur

Le blé est l'une des cultures majeures mondiales : étant largement utilisé comme aliment de base dans de nombreux pays, il fournit un cinquième des calories consommées par la population mondiale (Reynolds *et al.* 2011). Actuellement, 95% des blés cultivés dans le monde sont du blé tendre et les 5% restants sont majoritairement du blé dur. La culture du blé tendre est très répandue en termes de latitude puisqu'on le trouve de la Scandinavie à l'Argentine. A l'inverse, le blé dur est principalement cultivé dans des zones méditerranéennes car il est plus adapté à un climat sec (Shewry 2009).

La production de blé dans le monde s'élève à 760 millions de tonnes en 2020 pour une surface de production de 219 Mha et une consommation d'environ 700 millions de tonnes (FAO 2022). Bien que ce soit la culture la plus cultivée en termes de surfaces, c'est la troisième céréale la plus produite derrière le maïs et le riz. Les trois premiers pays producteurs de blé sont la Chine avec 133 millions de tonnes, suivie de l'Inde (100 millions de tonnes) et de la Russie (78 millions de

tonnes) (Figure 1-5), et à eux trois, ils représentent presque la moitié de la production mondiale.

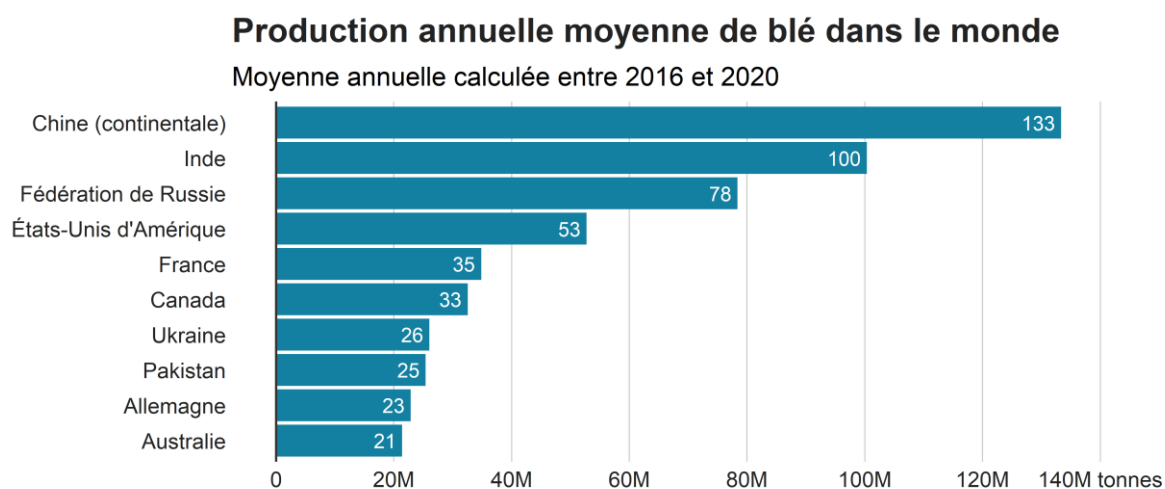


Figure 1-5. Production annuelle moyenne de blé dans le monde pour les dix premiers pays producteurs (FAO 2022).

Sur une période de 5 ans, entre 2016 et 2020, la France occupe toujours la 5^{ème} position du classement mondial, avec une production moyenne de 35 millions de tonnes et est le premier producteur de l'Union européenne. Sur les 9,4 millions d'hectares alloués à la culture des céréales en France, le blé tendre est cultivé sur 5 millions d'hectares. Elle est la première céréale produite, devant le maïs (20%), les orges (18%) et le blé dur (3%) (FranceAgriMer 2021). Globalement, le blé tendre est cultivé sur l'ensemble des régions métropolitaines mais principalement dans le nord de la France, avec le plus gros bassin de production dans les Hauts-de-France. Cette région représente à elle seule 20% de la production nationale. Globalement les surfaces cultivées en blé sont beaucoup plus importantes dans la moitié nord de la France que la moitié sud, et les rendements obtenus y sont également plus élevés. Pour la saison 2019-2020, les rendements varient entre 18 et 92 q/ha, la moitié nord de la France présentant des rendements allant de 60 à 92 q/ha (Figure 1-6), avec un rendement moyen sur la France de 88,9 q/ha (chiffres provisoires au 28/10/2021, (Agreste 2021).

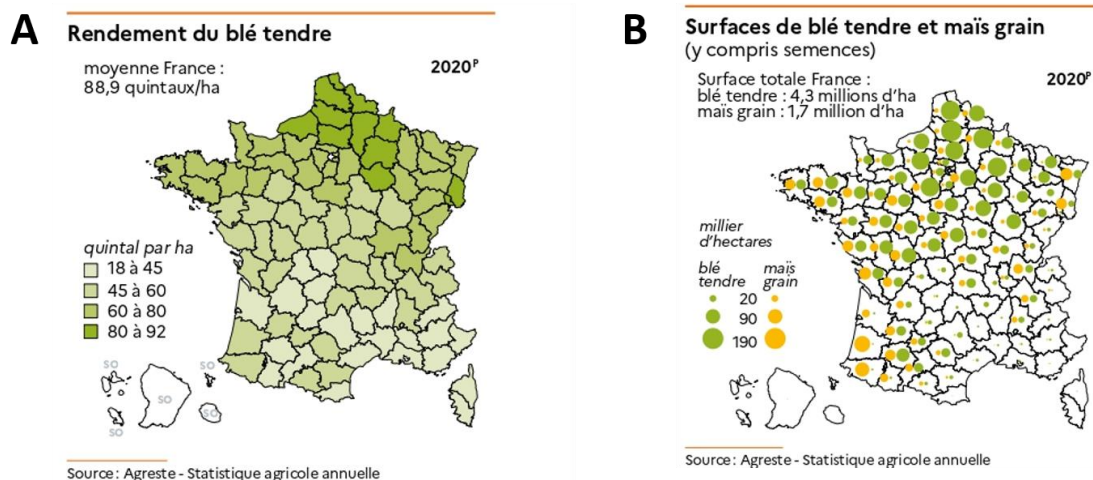


Figure 1-6. Représentation des rendements du blé tendre par département (A) et de la surface allouée à la culture du blé tendre par département (B) (adapté d'Agreste 2021, P: chiffres provisoires).

Sur la totalité de la production de blé tendre française, la moitié est destinée au marché intérieur, c'est-à-dire, à la consommation humaine, l'alimentation animale et les usages industriels et l'autre moitié est destinée à l'exportation. En moyenne entre 2016 et 2020, 17,4 millions de tonnes ont été exportées, vers des pays de l'Union Européenne (7,6 millions de tonnes) et hors Union Européenne (9,8 millions de tonnes). Les principaux marchés d'export hors Union Européenne sont le Maghreb et l'Afrique Subsaharienne (FranceAgriMer 2021).

1.6. Les enjeux liés à la production du blé tendre

L'agriculture et la filière du blé tendre doivent répondre à de nouveaux enjeux pour les décennies à venir. La croissance démographique mondiale qui explose depuis les années 1950, associée à des changements de régimes alimentaires des pays à forte croissance comme la Chine ou l'Inde (Brisson *et al.* 2010) sont deux facteurs qui montrent que la demande en céréales va doubler d'ici trente ans. Il est donc nécessaire d'augmenter la production mondiale pour répondre à cette demande croissante.

Malgré une amélioration constante des variétés de blé, la production mondiale ne s'accroît que de 1% chaque année (Ray *et al.* 2013). Au XX^{ème} siècle, les rendements ont progressé rapidement pour stagner à la fin des années 90, notamment en France. Cependant, cette tendance ne s'observe pas dans tous les pays (Figure 1-7), ce qui explique la légère augmentation de la production mondiale (Brisson *et al.* 2010).

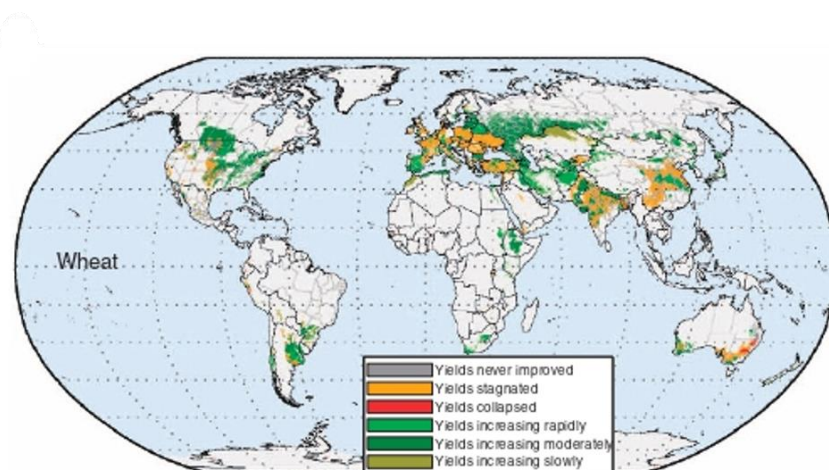


Figure 1-7. Représentation de la tendance des rendements dans les différents pays producteurs de blé tendre (d'après (Ray *et al.* 2012)).

L'augmentation des rendements peut se faire par le développement de nouvelles variétés plus performantes, on parle alors de progrès génétique, ou par l'évolution des pratiques culturales (Oury *et al.* 2012).

Le progrès génétique est toujours stable et n'a pas décliné depuis les années 90 mais son effet sur la production est contrebalancé par les effets négatifs du changement climatique (Brisson *et al.* 2010). Ceci est vérifié notamment par une perte de stabilité des rendements, due à des fluctuations climatiques interannuelles et inter-régionales de plus en plus importantes (FAO 2016). Le changement climatique est caractérisé notamment par une augmentation de la température moyenne, des températures extrêmes en été, mais aussi par des différences de pluviométrie interrégionale plus prononcées. Ces caractéristiques auront un effet direct sur les performances des cultures. Pour le blé, les risques de sécheresse et d'échaudage pendant la phase de remplissage des grains ou encore une prolifération accrue des bio-agresseurs et des maladies impacteront le rendement final (Gallais 2015). Associées à cela, les politiques de diminution d'apport de fertilisants, pour réduire l'effet négatif des pratiques agricoles sur l'environnement, contribuent aussi à diminuer le potentiel de rendement (Brisson *et al.* 2010).

Ainsi, des programmes sont lancés aux échelles nationale et internationale pour répondre à ces enjeux. En France, le programme BreedWheat a démarré en 2011 et a pour objectif d'aider la filière française de sélection du blé à relever le défi d'une production plus durable et de qualité. Ce projet, qui s'est terminé fin 2020, réunissait de nombreux acteurs français, aussi bien publics que privés dans un but commun d'une sélection française plus efficace (Paux *et al.* 2022). Un effort

international du séquençage du génome du blé tendre a aussi été lancé en 2005 par l'International Wheat Genome Sequencing Consortium. Ce travail a permis de séquencer plus de 94% du génome et d'identifier plus de 100 000 gènes présents sur les 21 chromosomes du cultivar Chinese Spring (IWGSC *et al.* 2018). La séquence de référence est un outil fondamental pour mieux comprendre la biologie du blé et pour aider les sélectionneurs dans leur travail de création variétale, assistés de la génomique.

2. La sélection du blé tendre et son amélioration génétique

2.1. L'amélioration génétique pour la création de nouvelles variétés

L'amélioration des plantes a pour objectif de créer de nouvelles variétés qui apportent du progrès génétique. (Gallais 2011) définit une variété comme une population artificielle à base génétique étroite, reproductible à l'identique, ayant des caractéristiques agronomiques bien définies et conçus pour répondre à différents besoins des utilisateurs. D'un point de vue génétique, la création de nouvelle variété consiste à réunir le plus de gènes favorables pour des caractères d'intérêt agronomique, dans une population artificielle voire dans un seul génotype. Seulement, réunir tous les gènes favorables dans une même variété en une seule étape n'est pas possible. Un processus de sélection sur plusieurs générations va donc être mis en place par le sélectionneur, pour maximiser les chances de combiner les allèles favorables.

L'objectif de la sélection est de choisir les meilleurs individus issus d'un croisement initial. Le groupe d'individus supérieurs à la moyenne pour le caractère à améliorer, constituera les parents de la génération suivante. Le processus de sélection se fait alors de génération en génération. Avec l'effet de la sélection, on peut alors déterminer un progrès génétique s'exprimant comme $\Delta G = i r \sigma A / T$, où i est l'intensité de sélection, r la précision de sélection, σA l'écart-type génétique du caractère et T l'intervalle de génération (Falconer 1996). Cette équation est surnommée l'équation du sélectionneur car elle permet de rendre compte de l'efficacité du programme de sélection. Tous les termes de l'équation sont des composantes importantes pour améliorer les méthodes de sélection, pour les rendre plus efficaces tout en minimisant le coût pour les entreprises de sélection.

2.2. La sélection phénotypique, une sélection basée sur l'observation des individus

Pour sélectionner les meilleurs individus, les plus transgressifs, le sélectionneur se base sur l'observation ou sur la mesure des caractères visibles des individus. Globalement, le caractère étudié ou phénotype (P), est le résultat de l'effet du génotype (G), de l'effet de l'environnement (E) et de l'effet de l'interaction entre le génotype et l'environnement (GxE).

$$P = G + E + GxE$$

L'effet du génotype et de l'environnement sur le phénotype ne sont pas indépendants, la réponse phénotypique à un changement de condition environnementale ne sera pas la même pour tous les génotypes. C'est ce qu'on appelle l'interaction GxE (Comstock et Moll 1963). Les plantes ont une capacité à s'adapter aux différents environnements ce qui fait qu'un simple génotype peut produire de multiples phénotypes. Cette adaptation passe par des changements de physiologie, métabolisme ou encore de développement au sein de la plante, on parle alors de plasticité phénologique. La valeur phénotypique des génotypes dans différents environnements peut être représentée par une courbe de réponse. Sont alors observés différents types de réponse entraînant différentes interactions GxE: quantitative ou qualitative (Figure 1-8). La présence d'interaction quantitative ne change pas le classement des variétés dans les différents environnements: la variété la plus performante reste la plus performante d'un environnement à l'autre. A l'inverse, l'interaction qualitative entraîne une inversion du classement des variétés (Allard et Bradshaw 1964).

L'efficacité de la sélection repose sur la concordance entre valeur phénotypique et valeur génétique (l'héritabilité), qui varie selon le type de caractère étudié. L'effet du génotype (G) est défini comme la valeur génétique (g) qui est, elle-même, décomposable en différentes valeurs: la valeur additive (a), la valeur de dominance (d) et la valeur d'épistasie (ξ) de telle sorte que la valeur génétique est la somme des différentes valeurs:

$$g = a + d + \xi$$

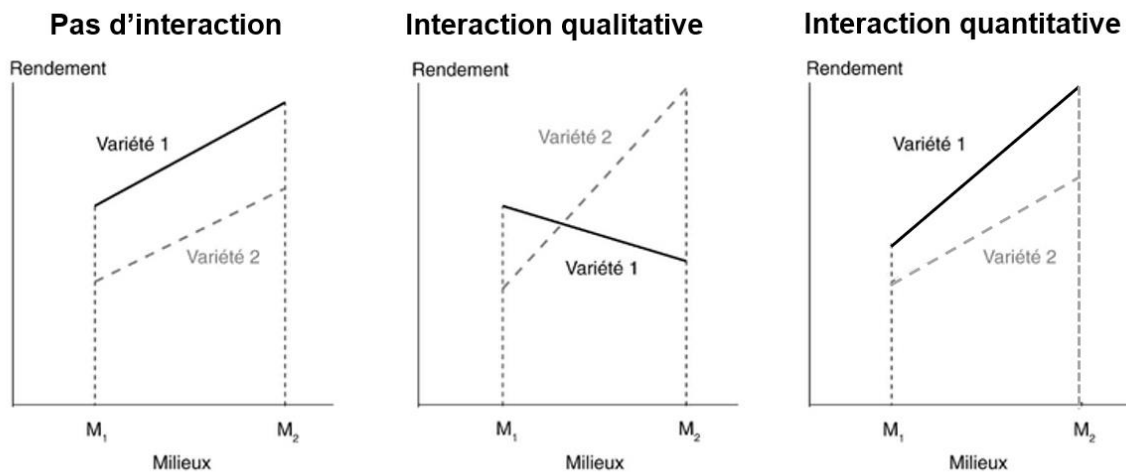


Figure 1-8. Représentation des interactions qualitative et quantitative entre le génotype et l'environnement.

En théorie pour le sélectionneur de plantes autogames, seule la valeur additive des gènes est importante car c'est la plus grande part de variance qui est transmise à la descendance, c'est la part héritable. La part de dominance est liée à l'interaction entre allèles d'un même gène et la part d'épistasie est liée à l'interaction entre gènes. En pratique, pour certains caractères ou populations, les valeurs génétiques dues aux effets additifs et non additifs peuvent être confondues et une grande partie de la variance due aux interactions des allèles (dominance et d'épistasie) peut se manifester sous forme de variance additive (Hill *et al.* 2008). La proportion de variance additive attribuable aux interactions des allèles dépend en grande partie de la distribution des fréquences alléliques aux loci causaux (Muñoz *et al.* 2014).

Il est à noter que dans le cas de la génétique quantitative, c'est l'épistasie statistique qui est considérée et non l'épistasie physiologique (Sackton et Hartl 2016). L'épistasie statistique correspond à la variance épistatique déterminée par la fréquence des allèles dans une population. On parle donc d'effet épistatique sur la variance génétique à l'échelle de la population (Cheverud et Routman 1995). L'épistasie physiologique correspond aux effets génotypiques déterminant le phénotype de l'individu. C'est-à-dire, que le phénotype obtenu est le résultat d'un gène dominant qui interagit avec d'autres gènes, et qui masque l'effet de ces derniers.

Le rendement est un caractère influencé par de nombreux gènes (caractère quantitatif) moyennement héritables, c'est-à-dire qu'une part significative de la variance phénotypique n'est pas génétique et ne peut donc pas être soumise à sélection. La sélection de ce caractère nécessite donc de répéter le phénotypage d'une même variété un grand nombre de fois pour estimer au mieux les valeurs génétiques. De plus, ce caractère est très influencé par le milieu et est soumis à de fortes interactions GxE. Les variétés les plus performantes ne seront donc pas forcément les mêmes d'un environnement à l'autre. Dans le cas d'interaction quantitative, des inversions de classements entre les variétés peuvent apparaître d'un environnement à l'autre, compliquant le choix des meilleures variétés à sélectionner.

Pour prendre en compte l'effet interaction GxE sur l'expression du caractère, il est donc essentiel de produire des essais multi-environnementaux (MET) ce qui permet d'évaluer la performance des génotypes dans différentes conditions (Burgueño *et al.* 2012). Ces derniers doivent donc être testés sur plusieurs années et dans différents lieux pour déterminer leur potentiel de rendement sur une large gamme d'environnements caractérisés par différentes combinaisons de stress. Ces MET sont coûteux puisqu'ils nécessitent de tester un grand nombre d'individus dans différents environnements. Pour ne pas perdre de temps, les sélectionneurs se contentent souvent d'observer les variétés candidates sur une ou deux années. Cela peut avoir pour conséquence d'orienter la sélection des candidats sur des contextes climatiques particuliers, ne correspondant pas forcément aux besoins du marché. Les mauvais génotypes sont éliminés rapidement alors qu'ils pourraient présenter de meilleures performances dans d'autres environnements (Jarquín *et al.* 2017).

2.3. Schéma de sélection pour la création de lignée pure de blé tendre

Le blé tendre est une espèce autogame et aura donc tendance à privilégier l'autofécondation plutôt que la fécondation croisée. Dans ce cas, il y a de fortes chances pour que la descendance soit entièrement homozygote et génétiquement identique à la plante mère initiale (Gallais 2011). Il est donc nécessaire d'apporter de la variabilité génétique pour espérer sélectionner de la descendance transgressive. Vilmorin en 1856, expliquait que pour recréer de la variabilité génétique, il fallait croiser deux lignées parentales puis sélectionner et fixer une nouvelle lignée qui surpasse les lignées parentales. La sélection généalogique, c'est-à-dire une sélection génération après génération a donc été proposée et adoptée.

2.3.1. La sélection généalogique directe

La méthode de la sélection généalogique consiste à sélectionner, au fur et à mesure de la fixation par autofécondation, la descendance la plus intéressante issue du croisement initial. En pratique, le schéma se compose donc d'un croisement initial de lignées parentales entièrement homozygotes (lignée pure). A l'issue de ce croisement est obtenue la génération F1 entièrement hétérozygote, si les deux parents ont des allèles différents aux loci. Les générations suivantes seront issues de l'autofécondation de la plante mère. A la génération F2, la sélection peut commencer puisque les différents loci hétérozygotes ségrègent. Il est donc possible de sélectionner sur cette nouvelle base de variabilité génétique. Aux générations F2 et F3, les individus sont semés aux champs pour être sélectionnés sur des caractères visuels de conformité de la plante. A partir de la F4, les individus vont être semés sur plusieurs lignes dans un dispositif de pépinière. Les critères de sélection vont être plus élaborés (date d'épiaison, hauteur de plante, résistance aux maladies) et des critères de qualité vont être mesurés (taux de protéine et poids spécifique par exemple). En génération F5, les premiers essais vont être réalisés avec des dispositifs en micro-parcelle, sur un ou deux lieux. En génération F6, les essais vont être mis en place dans le réseau d'essais, c'est-à-dire dans plusieurs environnements. Ces évaluations pour le rendement et la fixation du matériel vont continuer jusqu'à la génération F8 voire F9 pour potentiellement être déposés à l'inscription au catalogue officiel.

Il est nécessaire d'avoir toutes ces générations de fixation pour obtenir des lignées pures puisque la fréquence des génotypes hétérozygotes à un locus donné, à la génération N, est calculée comme $F_n = (1/2)^{n-1}$. A la génération F8, la proportion de loci hétérozygotes est inférieure à 1%.

Cette sélection généalogique directe, c'est-à-dire avec de la sélection à chaque génération du programme de sélection, a l'avantage d'améliorer efficacement les caractères fortement héritable. Le sélectionneur connaît aussi très bien la lignée qu'il va déposer à l'inscription puisqu'il évalue son comportement dans différents sites et sur plusieurs années, dès la F2. Par contre, cette sélection a aussi l'inconvénient d'être très longue (une dizaine d'années jusqu'à l'inscription) et elle induit une réduction de la variabilité génétique très rapidement dans le programme. En effet, la sélection se faisant dès la F2, des individus présentant un bon potentiel pour le rendement, caractère qui est évalué très tardivement dans le programme, pourraient être écartés très tôt.

2.3.2. *La sélection généalogique différée*

Pour pallier la perte de diversité génétique rapide liée à la sélection généalogique directe, d'autres méthodes de sélection généalogique indirecte ont été proposées: la méthode des bulks et la méthode de la SSD.

La méthode des bulks consiste à sélectionner au hasard les individus qui continueront le programme de sélection, aux générations précoces (F2-F4). Pour cela, les grains sont choisis en vrac dans un mélange, ce qui permet d'éviter la sélection artificielle par l'homme. Ces grains sont ensuite ressemés en mélanges pour constituer la génération suivante. Dans les programmes de sélection, les générations de bulks se feront sur les générations F2 et F3. En F4, une pépinière est mise en place pour une évaluation plus poussée des lignées. Cette méthode a l'avantage de conserver plus longtemps de la diversité génétique dans le programme de sélection, de faciliter la phase de fixation en F2 et F3 avec un dispositif simple et peu coûteux. Par contre elle a aussi le désavantage de laisser faire la sélection naturelle, qui va privilégier des plantes hautes et précoces qui seront de meilleurs compétiteurs que les plantes courtes et tardives. Il est possible d'intervenir en coupant les plantes les plus hautes pour limiter ces effets de compétition.

La méthode de la descendance sur un grain (single seed descent, SSD) a aussi pour objectif de maintenir la diversité génétique jusqu'à des générations plus avancées dans le cycle de sélection. La sélection commencera donc plus tard dans le schéma. Ainsi la méthode des bulks peut être remplacée par de la SSD un seul grain sera récolté par plante et semé pour obtenir la génération suivante, limitant ainsi l'apparement élevé entre lignée. Il y a une probabilité non nulle pour que le grain choisi ne donne pas de descendance. Dans ce cas, pour limiter des pertes, un épi peut être choisi à la place d'un seul grain. L'ensemble des grains de l'épi seront semés pour ensuite sélectionner à nouveau un seul épi par plante. L'avantage de la SSD, par rapport aux bulks, est qu'elle permet de limiter l'effet de la sélection naturelle tout en maintenant une variabilité génétique. Elle a aussi l'avantage majeur d'accélérer le cycle de sélection si celle-ci est pratiquée en serre. Pour le blé tendre d'hiver, il est possible de faire deux générations sous serre en une seule année. En contrepartie, la production de lignée aura un coût plus élevé que les bulks si celle-ci se fait en conditions artificielles (serre, chambre de culture).

2.3.3. *L'haplodiploïdisation en sélection généalogique*

L'haplodiploïdisation est une technique qui permet d'obtenir des lignées entièrement homozygotes en une génération. Elle va donc servir à accélérer le cycle de sélection en réduisant la phase de fixation des lignées. Plusieurs techniques existent pour obtenir le doublement chromosomique. *In vitro*, la régénération des plantes entières à partir de cellules haploïdes, peut être obtenue sur ovules ou microspores. Le doublement chromosomique peut être alors spontané ou induit par un agent chimique, le plus répandu étant la colchicine. Chez certaines espèces, l'obtention des embryons haploïdes peut se faire *via* des croisements interspécifiques. Le principe se base sur le rejet des chromosomes provenant du gamète mâle d'une autre espèce. Il faut alors mettre en place un sauvetage d'embryons pour qu'ils ne dégèrent pas. Appliquée dans le schéma de sélection sur la F1, l'haplodiploïdisation permet d'obtenir des lignées totalement homozygotes et observables au champs une ou deux années après le croisement initial. Les étapes de bulks et SSD peuvent donc être remplacées par la production d'haploïde doublé.

Par contre, cette méthode est très coûteuse, puisqu'elle nécessite l'investissement dans un laboratoire spécialisé. Le taux de réussite d'obtention des haploïdes doublés peut être variable d'une espèce à l'autre et d'un génotype à l'autre.

De toutes les méthodes présentées dans la sélection généalogique, ce sont la SSD en serre et l'utilisation d'haploïde doublé qui apparaissent les plus efficaces pour améliorer le gain génétique puisque le cycle de sélection est accéléré (Figure 1-9). Leur utilisation n'est pas généralisée dans tous les programmes de sélection à cause de l'investissement économique qu'elle représente.

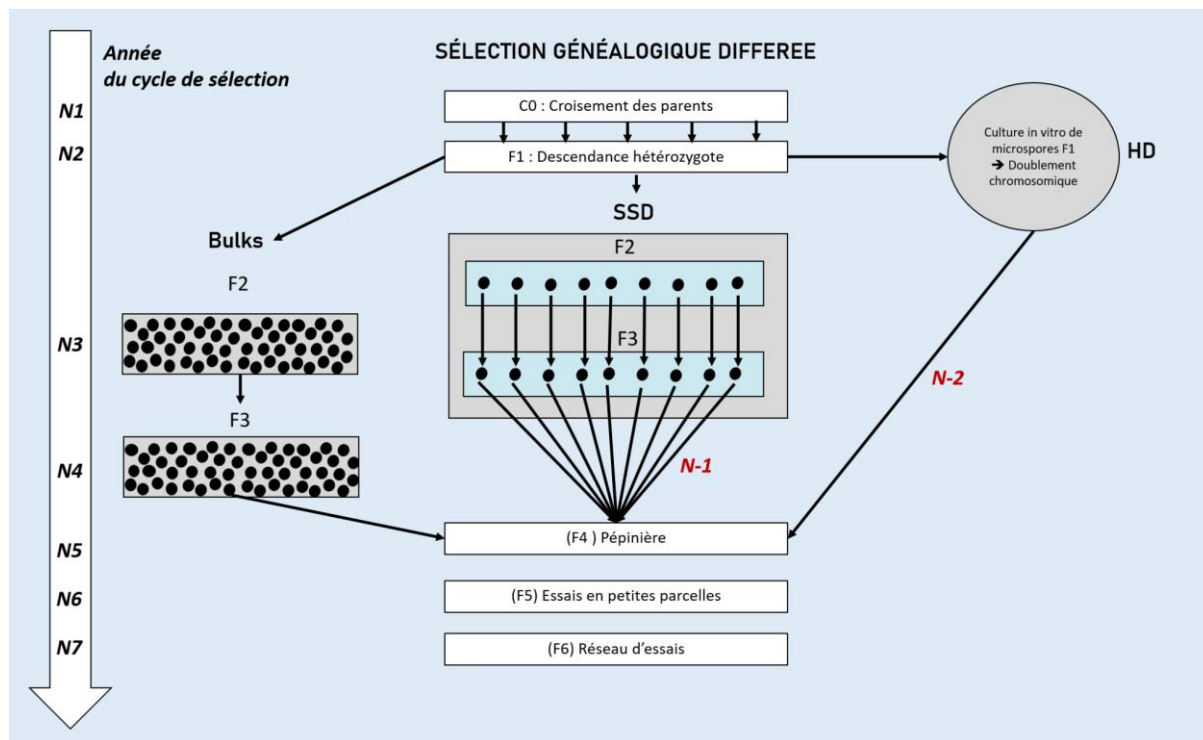


Figure 1-9. Comparaison entre l'utilisation de la méthode Bulks, SSD et de l'haplodiploïdisation dans un schéma de sélection généalogique différée. En blanc est présentée les étapes communes de la sélection généalogique différée. La SSD en bleu est réalisée en serre pour permettre d'obtenir plusieurs générations en une année. Les générations entre parenthèses sont précisées pour les générations issues des Bulks ou de la SSD, le matériel étant fixé pour les haploïdes doublés (HD).

2.4. L'inscription des variétés de blé tendre

Pour qu'une variété soit commercialisée en France, elle doit être inscrite au catalogue officiel français et par extension au catalogue officiel européen en répondant à un certain nombre de critères réglementaires.

La variété présentée par l'obteneur doit répondre aux critères de distinction, d'homogénéité et de stabilité (DHS). Plus précisément, il faut que la variété puisse être distinguée des autres par des caractères morphologiques, être homogène et être identique à elle-même d'une année à l'autre. Pour le blé tendre, la production de variétés se fait en lignée pure, où l'homozygotie totale n'est jamais réellement atteinte en réalité. En effet, des résidus d'hétérozygotie seront retrouvés à certains loci, ce qui sera sans conséquence pour l'obtention si les caractères agronomiques et morphologiques ne sont pas affectés (Gallais 2011). Les épreuves de DHS sont conduites par le Groupe d'Etude et de contrôle des Variétés Et des Semences (GEVES) et consiste en l'évaluation

26 caractères morphologiques et physiologiques sur une durée minimale de 2 ans (Doré et Varoquaux 2006). A la suite de ces épreuves, l'obteneur obtient une certification des semences et un titre de protection de la variété, qui lui confère l'exclusivité de son exploitation.

Pour être commercialisée, la variété doit également répondre aux critères de valeur agronomique, technologique et environnementale (VATE). La variété va être évaluée pour son rendement, sa valeur technologique, ses caractéristiques physiologiques et sa résistance aux maladies, pendant 2 ans dans un réseau national d'essais agronomiques (Tableau 1-1). Ces réseaux d'essais sont spécifiques à la précocité de la variété (tardif, semi précoce et précoce) et les variétés sont évaluées en conduite classique (conduite optimale avec traitements fongicides) et en conduite non traitée (conduite optimale sans traitement fongicide). Cette évaluation se fait en comparant la variété testée à des témoins de référence qui sont les variétés de référence sur le marché. La variété proposée à l'obtention doit apporter un progrès génétique par rapport aux variétés actuelles.

Tableau 1-1. Caractères évalués pour la VATE lors de l'inscription de nouvelles variétés (GEVES 2019)

Le rendement	Valeur technologique	Caractéristiques physiologiques et autres	Les résistances aux bioagresseurs
Rendement dans les essais traités et non traités fongicides.	<ul style="list-style-type: none"> - Teneur en protéines - Poids Spécifique - W et P/L - Dureté - Indice de Hagberg - Valeur en panification française (Test biscuitier) - Viscosité pour alimentation animale (volailles) 	<ul style="list-style-type: none"> - Grain Protein Deviation (écart à la régression négative existant entre le rendement et la teneur en protéines) - Alternativité - Précocité d'épiaison - Précocité à montaison - Hauteur - Résistance à la verse - Résistance au froid - Résistance à la germination sur pied 	<ul style="list-style-type: none"> - Rouille jaune - Rouille Brune - Piétin verse - Oïdium - Septorioses (<i>Z. tritici</i> et <i>P. nodorum</i>) - Fusarioses (<i>F. graminearum</i> et autres spp.) - Tolérance globale aux maladies (écart T-NT) - Mosaïques (SBCMV et WSSMV) - Cécidomyie orange
Le rendement est exprimé en % des variétés témoins.	A partir des résultats des analyses, la classe technologique de la variété est définie	Les notations dans le réseau (contamination naturelle) et dans les essais spécifiques (infection artificielle avec des souches choisies en collaboration avec l'INRAE) sont traduites en cotations de résistance indépendantes de l'année et du lieu (1 = très sensible ; 9 = résistant).	

Pour le blé tendre, les analyses de la valeur technologique permettent de classer les variétés dans différentes catégories: Blé améliorant (BA), Blé panifiable (BP) et Blé Panifiable Supérieur (BPS), Blé Biscuitier (BB) et Blé Autres Usages (BAU, BAU imp). Ces catégories définissent les utilisations industrielles et débouchés des variétés. Des seuils de rendement sont requis pour chaque classe technologique. Par exemple, une variété de classe BPS devra avoir un rendement supérieur à 102% du rendement moyen des témoins, quand une variété de classe BA devra avoir un rendement supérieur à 80% du rendement moyen des témoins (GEVES 2019).

Les critères pris en compte, pour orienter la sélection du blé en fonction des attentes des utilisateurs, de la société et des avancées de la sélection sont définis par le Comité Technique Permanent de la Sélection des Plantes Cultivées (CTPS) (GEVES 2022). Par exemple, depuis 2017, une commission a été créée pour favoriser l'inscription au catalogue français de variétés liées à l'agriculture biologique.

Comme décrite précédemment, la création variétale a pour objectif d'amener du progrès génétique pour des caractères agronomiques d'intérêt. Seulement, la méthode de sélection phénotypique est longue, malgré les différentes méthodes existantes, pour accélérer le cycle de sélection et coûteuse avec les dispositifs de phénotypage de centaines de lignées. Ces limites tendent à réduire l'efficacité de la sélection. Une des possibilités, pour améliorer les méthodes de sélection, est alors d'utiliser des biotechnologies pour assister le sélectionneur dans ses décisions.

3. Utilisation des marqueurs moléculaires et autres omiques en sélection

3.1. La sélection assistée par marqueurs (SAM)

Depuis les années 80, les généticiens et sélectionneurs ont travaillé pour inclure des technologies basées sur le marquage moléculaire en amélioration variétale. L'essor de différents systèmes de marquage moléculaire a permis d'augmenter significativement le nombre de marqueurs polymorphes disponibles pour les sélectionneurs (Crossa *et al.* 2017). L'intérêt de la SAM réside dans le fait que cette sélection ne se fait pas directement sur le caractère d'intérêt mais sur le marqueur moléculaire associé à ce caractère.

Des techniques comme la cartographie et le séquençage du génome des plantes avaient pour objectif d'identifier la fonction, la régulation et l'expression des gènes contrôlant certains caractères agronomiques. C'est ainsi que des cartes de liaisons ont été produites et que des marqueurs moléculaires ont été utilisés pour identifier un gène souhaité. Les marqueurs moléculaires sont des séquences d'ADN qui présentent un polymorphisme permettant de différencier plusieurs allèles. Ces marqueurs moléculaires sont intéressants lorsqu'ils sont en déséquilibre de liaison (associés non aléatoirement) avec un gène d'intérêt ou un locus de caractères quantitatifs (QTL: région du génome où un ou plusieurs gènes sont associés à un caractère quantitatif). Grâce à la cartographie de liaison, il a alors été possible de noter sur les

chromosomes la position des marqueurs et leur distance génétique relative et ainsi d'identifier l'emplacement des QTLs associés à des caractères d'intérêt.

Ces outils de marquage moléculaire, de détection de QTLs, sont utiles aux sélectionneurs pour différencier de manière plus efficace les individus avec la meilleure combinaison allélique possible pour les caractères d'intérêt. C'est sur ces outils que se base la SAM, consistant à sélectionner les variétés en fonction des allèles aux marqueurs.

La SAM est une méthode qui présente certains avantages par rapport à la sélection traditionnelle notamment celui d'augmenter l'efficacité de la sélection. Elle permet de gagner du temps en substituant des essais aux champs par des tests moléculaires. En effet, une fois que les QTLs associés à un caractère d'intérêt ont été identifiés par le biais des marqueurs, un test moléculaire peut être effectué pour identifier les individus portant la bonne combinaison allélique et ainsi diminuer l'effectif à évaluer au champ (Collard *et al.* 2005 ; Bernardo 2008). Cette méthode de sélection présente cependant des limites pour son application dans les programmes de sélection. Une de ces limites est la nécessité pour le sélectionneur de connaître les allèles favorables et leurs effets moyens sur le phénotype (Bernardo et Charcosset 2006). Cela est donc principalement applicable pour les caractères qualitatifs à effets majeurs mais pas pour les caractères quantitatifs qui sont influencés par de nombreux loci à effet faible et souvent dépendants de l'environnement. En effet, quand l'interaction GxE est importante, les effets des QTLs peuvent varier en fonction de l'environnement (Moreau *et al.* 2004; Xu et Crouch 2008). De plus, la part de variance génétique expliquée par les QTLs détectés est souvent faible car les tests statistiques sont très conservatifs afin de diminuer le risque de détecter des faux positifs.

3.2. La sélection génomique

3.2.1. Principes de la méthode

En 1990, Lande et Thompson proposent une méthode pour améliorer la SAM. Celle-ci se fait en deux étapes. Dans un premier temps, les marqueurs significatifs d'un grand jeu de marqueurs sont sélectionnés. Puis dans un deuxième temps, les informations phénotypiques sont couplées avec les marqueurs significatifs, dans un index de sélection pour expliquer alors une proportion significative de la part de variance génétique additive (Heffner *et al.* 2009). L'estimation simultanément de l'effet de tous les marqueurs de manière à capturer une plus grande

part de variance génétique appelée sélection génomique (SG), est une méthode proposée dès les années 90 avec l'utilisation des modèles mixtes linéaires (Bernardo 1994) puis dans les années 2000 avec l'inférence bayésienne (Meuwissen *et al.* 2001).

La SG est une méthode qui va utiliser l'information portée par des marqueurs moléculaires couvrant l'ensemble du génome dans le but de prédire la valeur d'un caractère agronomique avec une précision assez bonne pour permettre de sélectionner les individus sur la base seule de cette prédiction (Lorenz *et al.* 2011). C'est une forme de SAM excepté le fait que tous les effets des haplotypes (ensembles de gènes liés) ou marqueurs sont estimés simultanément tentant ainsi d'expliquer la variance génétique totale, cela grâce à une couverture dense de marqueurs sur tout le génome. L'hypothèse sous-jacente est que les caractères quantitatifs sont contrôlés par un grand nombre de QTLs à effet faible (modèle infinitésimal) et que le marquage dense permet d'estimer indirectement les effets de ces QTLs puisque ces derniers sont en déséquilibre de liaison (DL) avec au moins un marqueur. Les effets individuels de tous les loci qui contribuent à la variance génétique sont donc estimés. La somme de ces effets correspond alors à la valeur génétique estimée (GEBV) de l'individu (Meuwissen *et al.* 2001). Cette GEBV correspond à la valeur génétique de l'individu en sélection pour le caractère d'intérêt. Parce que la SG inclut tous les marqueurs, les effets majeurs et polygéniques peuvent être capturés et peuvent potentiellement expliquer plus de variance génétique (Solberg *et al.* 2008).

Plusieurs étapes sont nécessaires avant de sélectionner les individus du programme de sélection. Deux populations vont être utilisées dans ce processus de sélection. La population de sélection ou de validation, formée par les individus à sélectionner, est entièrement génotypée. La population de calibration ou d'entraînement du modèle est composée d'individus génotypés et phénotypés pour le caractère agronomique d'intérêt. Cette population de calibration va être utilisée pour estimer les paramètres (effets des marqueurs) du modèle d'apprentissage, ce même modèle étant ensuite utilisé pour prédire la GEBV des individus de la population de sélection. Enfin, le sélectionneur pourra choisir les meilleurs candidats en se basant sur les GEBV (Figure 1-10) (Heffner *et al.* 2009).

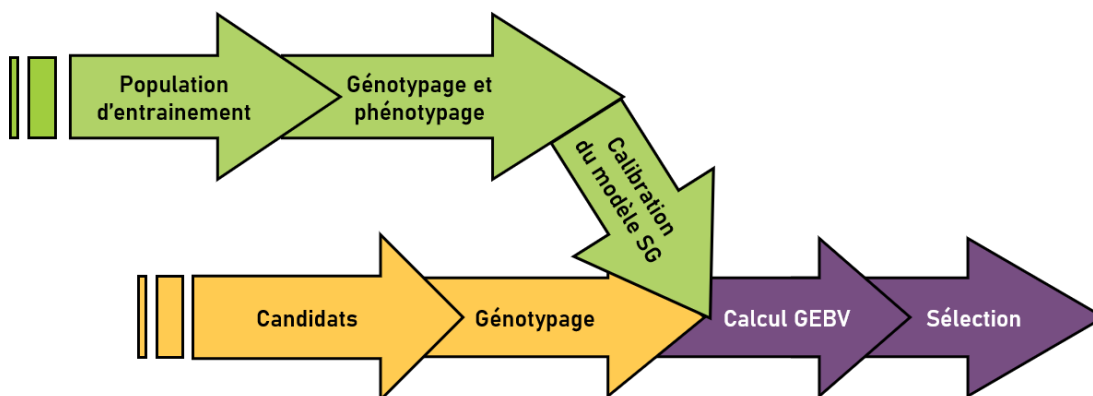


Figure 1-10. Processus de sélection génomique utilisant une population d'entraînement pour estimer la valeur génétique d'une population de candidats (adapté de Heffner *et al.* 2009).

Pour évaluer les capacités de prédiction des modèles de SG, ce n'est pas une population de candidats qui va être utilisée, mais une population dite de validation. Celle-ci est génotypée et phénotypée mais son phénotypage va être masqué pour simuler une vraie population de sélection qui seraient seulement génotypée. La capacité de prédiction est ensuite calculée comme la corrélation de Pearson entre la moyenne ajustée du caractère et la valeur prédite par le modèle. Cette capacité de prédiction va être calculée sur différentes populations d'entraînement et de validation grâce à un système de validation croisée. Cette méthode consiste à répartir aléatoirement les individus de la population de départ en k groupes de même taille. A tour de rôle, chacun des groupes servira de population de validation et les $k-1$ groupes restants constitueront la population d'entraînement. Ce processus est répété jusqu'à ce que tous les groupes soient utilisés une fois en population de validation.

3.2.2. Facteurs influençant les capacités de prédiction de la méthode

La capacité de prédiction des modèles génomiques intervient directement dans l'équation du sélectionneur, avec le paramètre r qui est la précision de sélection (cf. 2.1). Pour maximiser le gain génétique attendu, cette capacité doit donc aussi être maximisée. Plusieurs facteurs vont influencer la capacité de la prédiction génomique: l'architecture génétique du caractère (Daetwyler *et al.* 2010) et son héritabilité (Hayes *et al.* 2009b), le déséquilibre de liaison (Zhong *et al.* 2009), la densité de marquage (Solberg *et al.* 2008 ; de los Campos *et al.* 2013 ; Hickey *et al.* 2014), la taille et la composition de la population d'entraînement (Albrecht *et al.* 2011 ; Heffner *et al.* 2011 ; Pszczola *et al.* 2012 ; Rincant *et al.* 2012 ; Daetwyler *et al.* 2013) et le modèle statistique (Charmet *et al.* 2020).

La complexité dans le contrôle de ces prédictions est liée au fait que ces facteurs qui influencent les capacités de prédiction ne sont pas indépendants les uns des autres. La taille de la population d'entraînement requise pour effectuer les prédictions va dépendre elle-même de l'héritabilité du caractère et du niveau de proximité génétique avec la population de sélection mais globalement, plus grande sera cette population et meilleures seront les capacités de prédictions (Heffner *et al.* 2011). L'augmentation du nombre d'individus qui calibre le modèle, permet de réduire la variance des effets estimés des marqueurs (Liu *et al.* 2011 ; de los Campos *et al.* 2013). Pour deux populations très apparentées, le nombre requis d'individus en population d'entraînement sera plus faible que pour deux populations peu apparentées (Rutkoski *et al.* 2015).

Les capacités de prédiction vont aussi augmenter avec une plus grande densité de marqueurs puisqu'un plus grand nombre de QTLs seront en DL avec les marqueurs (Solberg *et al.* 2008 ; de los Campos *et al.* 2013 ; Hickey *et al.* 2014 ; Desta et Ortiz 2014). Cependant, à partir d'une certaine densité de marqueurs, les capacités de prédiction n'augmentent plus et atteignent un plateau (de los Campos *et al.* 2013). Ce plateau est atteint plus ou moins rapidement en fonction de l'étendue du DL dans le génome, de la taille de la population d'entraînement et de la proximité génétique entre populations d'entraînement et de sélection. Hickey *et al.* (2014) ont montré, sur des simulations avec le maïs, que lorsque la proximité génétique est forte entre les deux populations, moins de marqueurs moléculaires sont alors requis. En effet des populations proches génétiquement partagent des haplotypes communs et des blocs liaisons étendus ce qui nécessite moins de marqueurs pour estimer leurs effets.

L'apparentement entre les deux populations d'entraînement et de sélection est un paramètre primordial à maîtriser pour obtenir de bonnes capacités de prédiction (Habier *et al.* 2007; Crossa *et al.* 2014; Charmet *et al.* 2014). Plusieurs algorithmes d'optimisation permettent de choisir au mieux les individus qui constitueront la population d'entraînement du modèle. Le critère de variance d'erreur de prédiction (PEV) est largement utilisé en génétique quantitative pour calculer les précisions de prédiction des valeurs génétiques estimées. Il est donc possible de choisir une population d'entraînement qui minimise ce PEV, mais la variance génétique à l'intérieur de la population d'entraînement ne sera pas considérée (Isidro *et al.* 2015). Rincent *et al.* (2012) ont proposé d'utiliser le coefficient de détermination moyen (CDmean) qui maximise directement la moyenne des fiabilités attendues des contrastes entre chaque individu en sélection et la moyenne de population. Ce critère donne en général de meilleures capacités de prédiction que le PEV ou la

moyenne des PEVs.

L'intérêt de ces méthodes d'optimisation est de limiter le phénotypage aux individus les plus informatifs pour le modèle de prédiction et donc de réduire les coûts liés au phénotypage. La taille de la population d'entraînement est donc réduite et la capacité de prédiction du modèle est maintenue.

3.2.3. Modèles utilisés en prédiction génomique

3.2.3.1. Prise en compte de l'effet génétique additif dans les modèles de prédiction

Les modèles statistiques qui vont être utilisés pour effectuer les prédictions génomiques vont aussi impacter les capacités de prédiction de ces dernières. D'un point de vue statistique, les modèles linéaires classiques ne peuvent pas être utilisés en prédiction génomique à cause d'une part, du grand nombre de variables explicatives par rapport au nombre d'observations ($p \gg n$) et d'autre part, de la multi colinéarité entre les marqueurs moléculaires (Lorenz *et al.* 2011). Pour surmonter ce problème, de nombreux modèles ont été comparés au cours de cette dernière décennie (Heslot *et al.* 2012; Pérez-Rodríguez *et al.* 2012; Desta et Ortiz 2014; Charmet *et al.* 2020). Globalement, le critère qui différencie ces modèles, mise à part la méthode d'inférence (fréquentiste et bayésienne), est la distribution des effets des marqueurs de l'ensemble du génome. Les méthodes choisies sont des méthodes de régression pénalisée ou des méthodes semi paramétriques. Les méthodes bayésiennes diffèrent dans la distribution a priori des effets des marqueurs et du poids de ces effets dans la part de variance du caractère à prédire.

Le modèle de prédiction génomique le plus répandu est le G-BLUP ou Genomic Best Linear Unbiased Prediction (Habier *et al.* 2013). Ce modèle mixte consiste à prédire l'effet génétique des individus grâce à une matrice d'apparentement entre individus (appelée "Kinship"), calculée à partir des marqueurs moléculaires. L'utilisation de matrices génomiques permet de réduire la dimension des variables explicatives aux dimensions du nombre d'observations. Il existe un équivalent: la régression ridge (RR-BLUP), qui consiste à estimer l'effet de chaque marqueur au lieu de calculer une kinship entre individus. En RR-BLUP, les effets des marqueurs sont identiquement distribués et les effets sont équitablement pénalisés vers zéro. Pour opérer de la sélection de variable et privilégier l'effet de certains marqueurs par rapport à d'autres, il est possible d'utiliser le modèle LASSO qui pénalise plus fortement les effets des marqueurs conduisant certains effets à être réduit à zéro. Ce type de modèle peut être intéressant pour prédire

des caractères oligogéniques, contrôlés par quelques gènes à effet majeur. En inférence bayésienne, on retrouve les modèles de types régression ridge bayésienne (BRR) et le lasso bayésien (BL). Cependant BL ne fait pas de la sélection de variable puisqu'aucun effet de marqueurs n'est réduit à zéro. Il faudra alors privilégier le modèle Bayes-C qui permet de réduire les effets de marqueurs à zéro. On trouve encore d'autres modèles comme Bayes A et B qui diffèrent dans les distributions a priori des effets des marqueurs moléculaires.

La plupart des études de SG qui ont comparé différents modèles de prédiction génomique, ont montré qu'il n'y avait, en moyenne, sur de nombreux caractères, pas ou peu de différences en termes de capacité de prédiction entre les modèles (Heslot *et al.* 2012). Le choix d'un modèle plutôt qu'un autre est alors lié à l'architecture génétique du caractère à prédire et le temps de calcul du modèle, le G-BLUP étant le plus rapide comparé aux autres modèles de sélection de variables ou d'inférence bayésienne (Juliana *et al.* 2017).

3.2.3.2. *Prise en compte de l'effet d'épistasie dans les modèles de prédiction*

Plusieurs approches de SG ont été développées pour modéliser les effets génétiques principaux et les effets d'épistasie (Xu 2007; Wang *et al.* 2012). Ces approches présentent des réussites très variables d'une étude à l'autre, l'effet d'épistasie étant généralement omis dans les modèles paramétriques (Jiang et Reif 2015). Les marqueurs moléculaires décrivent bien la part des effets additifs sur la variance génétique mais très mal la part des effets de dominance ou d'épistasie (Hill *et al.* 2008). En théorie, l'épistasie peut être estimée avec un grand nombre de marqueurs moléculaires mais le nombre de degrés de libertés pour estimer les effets d'un grand nombre d'interactions épistatiques sera insuffisant et les capacités de calculs limitées (Sweeney *et al.* 2019). De nombreux modèles ont été développés pour dépasser cette limite. Jiang et Reif (2015) ont proposé d'étendre le G-BLUP en ajoutant une matrice de similarité épistatique basée sur les marqueurs moléculaires (EG-BLUP). Des modèles semi paramétriques ont aussi été proposés en prédiction génomique, pour capturer les effets non additifs des gènes. Les modèles RKHS font référence à un espace de Hilbert à noyau reproduisant, dans lequel une fonction noyau convertit les variables prédictives en distances entre les observations afin d'obtenir une matrice de dimensions $n \times n$ (n étant le nombre d'observations) utilisable dans un modèle linéaire (Lorenz *et al.* 2011). Les travaux de Gianola *et al.* (2006) ont montré qu'un modèle qui tient compte de l'effet génétique classique et qui inclut également un terme RKHS pour capturer les effets non-additifs, a donné de meilleures capacités de prédiction qu'un modèle RR-BLUP pour prédire des caractères

simulés présentant une grande part de variance non additive. C'est parce que les méthodes de noyaux sont plus flexibles et ne supposent pas la linéarité entre les variables à prédire et les variables explicatives, qu'elles sont plus efficaces à capturer tous les effets non additifs simultanément. La comparaison entre RKHS et EG-BLUP a montré que ces modèles, qui tiennent compte explicitement de l'effet épistatique, améliorent les capacités de prédiction des modèles de prédictions génomiques. Ceci est particulièrement vérifié dans le cas de prédiction de matériel issu d'autofécondation, comme les lignées de blé, et non d'hybrides, comme chez le maïs (Jiang et Reif 2015). Les méthodes de *machine learning* comme les *support vector machine* (SVM) ou la *random forest* (RF), ont aussi un intérêt en prédiction génomique puisque ces méthodes permettent de construire des modèles de prédiction non linéaire. Ils sont donc utiles lorsque la relation entre les prédicteurs et la variable à expliquer est aussi non linéaire. C'est particulièrement vrai lorsque l'effet épistatique explique une grande part de la variance du caractère d'intérêt (Jannink 2010). Malgré ces avancées dans la modélisation de l'effet épistatique, ces derniers sont toujours limités pour capturer de l'épistasie physiologique (Sackton et Hartl 2016). Un faible niveau d'épistasie statistique trouvé dans les études de populations n'implique pas que l'épistasie physiologique soit également faible. Il semble donc que les interactions sous-jacentes à l'épistasie physiologique soient peu susceptibles d'être découvertes en recherchant de l'épistasie statistique.

3.2.3.3. *Prise en compte de l'interaction génotype-environnement dans les modèles de prédiction*

Pour des caractères complexes comme le rendement, où l'effet de l'interaction génotype-environnement est non négligeable sur la variance du caractère, il est important de modéliser l'effet du GxE pour améliorer les capacités de prédiction du modèle, surtout dans le cadre de prédictions multi-environnementales. (Burgueño *et al.* 2012) ont constaté que l'intégration d'informations partagées par différents environnements pouvait améliorer la capacité de prédiction du modèle de 17,5 % par rapport aux modèles avec un unique effet génétique. Pour ce faire, une matrice de covariance environnementale a été introduite dans le modèle de prédiction. Cela implique que ces modèles sont applicables uniquement aux environnements dans lesquels des phénotypes ont été mesurés. Pour modéliser l'interaction GxE dans les prédictions génomiques, plusieurs études ont proposé de tenir compte de covariables environnementales qui décrivent des caractéristiques de l'environnement (Heslot *et al.* 2014 ; Jarquín *et al.* 2014 ; Malosetti *et al.* 2016 ; Lado *et al.* 2016 ; Ly *et al.* 2018). C'est le cas par exemple de données pédoclimatiques comme la température, les

radiations, les caractéristiques du sol. Ces données peuvent également être utilisées pour calculer des indices de stress (Ly *et al.* 2017 ; Rincent *et al.* 2019). Des environnements avec des covariables environnementales similaires sont supposés interagir de la même manière avec les génotypes. C'est ainsi que pour effectuer des prédictions dans un nouvel environnement inconnu, le modèle peut prédire l'effet de ce nouvel environnement sur les génotypes. Toutefois, les gains en capacité de prédiction obtenus avec cette modélisation du GxE sont limités. Ceci peut être dû au fait que cette modélisation ne tient pas compte de la complexité des interactions entre la plante et l'environnement. De même, le fait que ce ne soient pas nécessairement les mêmes covariables environnementales qui affectent les différentes variétés n'est, en général, pas non plus pris en compte.

3.2.3.4. *Intégration de la sélection génomique dans les programmes d'amélioration végétale*

Le développement rapide de la sélection génomique dans l'amélioration des plantes a été rendu possible avec le développement de nouvelles technologies de séquençage. C'est le cas des nouvelles générations de séquençage (NGS) qui offrent une approche robuste et rapide pour obtenir un génotypage de haute-densité (Poland et Rife 2012). La première application des NGS, est l'utilisation de marqueurs moléculaires à polymorphisme nucléotidique (*single-nucleotide polymorphism*, SNP). Ce type de polymorphisme est très répandu dans l'ensemble du génome et permet donc une bonne couverture de ce dernier en marqueurs moléculaires. Associé à ces nouvelles sources de polymorphisme, le coût du génotypage a drastiquement diminué depuis l'arrivée des NGS. C'est notamment le cas pour le séquençage humain (Figure 1-11). Cette réduction du coût a ainsi permis l'utilisation des méthodes de prédiction en amélioration des plantes (Lorenz *et al.* 2011 ; Poland et Rife 2012). Ce coût de génotypage est encore probablement amené à diminuer ce qui permettra aux sélectionneurs d'utiliser le séquençage pour toutes les études génétiques à mener pour leur programme de sélection.

D'abord utilisée dans le monde animal sur les vaches laitières, la SG a indéniablement amélioré l'efficacité de la sélection (Hayes *et al.* 2009a). En amélioration végétale (Lorenzana et Bernardo 2009; Heffner *et al.* 2010 ; Asoro *et al.* 2011 ; Burgueño *et al.* 2012), elle peut être utilisée pour prédire différents caractères d'intérêt tels que le rendement, les résistances aux maladies et la date de floraison et ce dans différentes conditions environnementales (Pérez-Rodríguez *et al.* 2017).

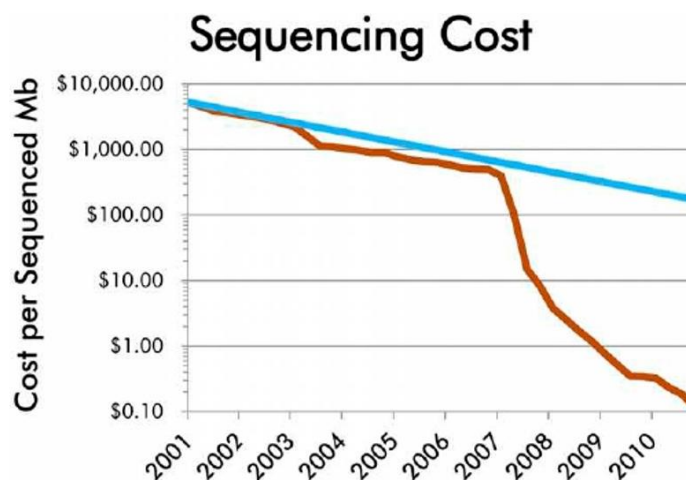


Figure 1-11. Évolution du coût de séquençage du génome humain (orange) depuis 2001 (Poland et Rife 2012). La droite bleue correspond à l'évolution du coût de séquençage si les technologies de séquençage avaient suivi la loi de Moore. La chute du prix de séquençage en 2007 correspond à l'introduction des NGS. Les données proviennent de l'institut national de recherche sur le génome humain (Wetterstrand 2012).

L'un des objectifs de la sélection génomique est d'obtenir de meilleur gain génétique à un coût plus faible et plus rapidement qu'avec la sélection phénotypique (Bassi *et al.* 2016; Crossa *et al.* 2017). Le gain génétique attendu de la SG par unité de temps est défini par l'équation du sélectionneur (cf. section 2.1). La précision de sélection est directement liée à la capacité de prédiction de la SG. Il est donc important de maximiser cette précision pour ne pas réduire le gain génétique attendu. C'est donc l'équilibre entre bénéfice de la méthode et coût de la méthode qui va guider le sélectionneur dans ses stratégies d'applications de la SG.

Bernardo et Yu (2007) ont été les premiers à montrer sur des études de simulations que la SG donne de meilleures capacités de prédiction des valeurs génétiques prédites que ce que peut donner la SAM. L'efficacité de la SG a été évaluée sur des données issues des programmes de sélection du centre international d'amélioration du maïs et du blé (CIMMYT), et a montré de meilleures capacités de prédiction qu'avec un modèle utilisant les informations de pedigree entre les lignées (de los Campos *et al.* 2009 ; Crossa *et al.* 2010). Heffner *et al.* (2010) ont aussi comparé par simulation, les gains obtenus en SAM ou en SG pour des programmes de sélection de maïs et de blé tendre, en tenant compte des coûts et de la durée des cycles de sélection. Ils ont montré que la

SG peut surpasser la SAM en termes de gain génétique annuel. Ils ont aussi proposé un schéma comparant l'utilisation de la SAM et la SG dans un programme d'amélioration du blé tendre, et montrer que la SG permet d'accélérer le cycle de sélection (Figure 1-12).

Si l'on considère, depuis les années 2010, que la SG est prometteuse pour prédire des caractères complexes comme le rendement, caractère majeur en sélection du blé tendre (Poland *et al.* 2012; Lado *et al.* 2013; Zhao *et al.* 2015; Michel *et al.* 2017; Song *et al.* 2017; Norman *et al.* 2017) très peu d'études ont validé le réel gain génétique obtenu dans de vrais programmes de sélection. Quelques études, sur des populations-biparentales, ont mesuré le gain génétique obtenu avec l'utilisation de la sélection génomique dans des cycles rapides de sélection, pour améliorer le rendement. Ils ont montré, pour le maïs, que la SG permet d'améliorer le gain génétique par unité de temps. Sur blé tendre, Rutkoski *et al.* (2015) ont obtenu des gains similaires entre SG et la sélection phénotypique. Crossa *et al.* (2017) ont aussi réuni les différents gains génétiques obtenus sur des population bi-parentales de maïs du CIMMYT, et expliquent que le gain génétique obtenu par cycle avec la SG (0.086 t/ha) est plus important que celui obtenu par la sélection assistée par marqueurs (0.051 t/ha).

La SG peut être utilisée à différentes étapes du schéma de sélection. Or des schémas de sélection intégrant la sélection génomique dans les programmes de sélection du blé tendre sont peu disponibles dans la littérature. (Bassi *et al.* 2016) ont proposé un certain nombre de schémas intégrant la SG à la génération F2, F3, F4 et F7. La SG a un intérêt tout particulier pour assister le sélectionneur dans la sélection des lignées à des générations précoces du programme de sélection. En effet en génération F3 et avant, peu de sélection est appliquée aux lignées puisque leur caractérisation est très sommaire. La sélection se fait principalement sur des caractères visuels de conformité de la plante. La SG a donc un rôle à jouer là où des milliers de lignées de génération précoce, sélectionnées au hasard, pourraient être génotypées pour prédire différents caractères d'intérêt agronomique (caractère de qualité, rendement, résistance aux maladies). Ces prédictions peuvent être utilisées pour sélectionner les meilleures lignées qui seront plus à même d'avoir de bonnes performances à des générations plus avancées (Malosetti *et al.* 2016).

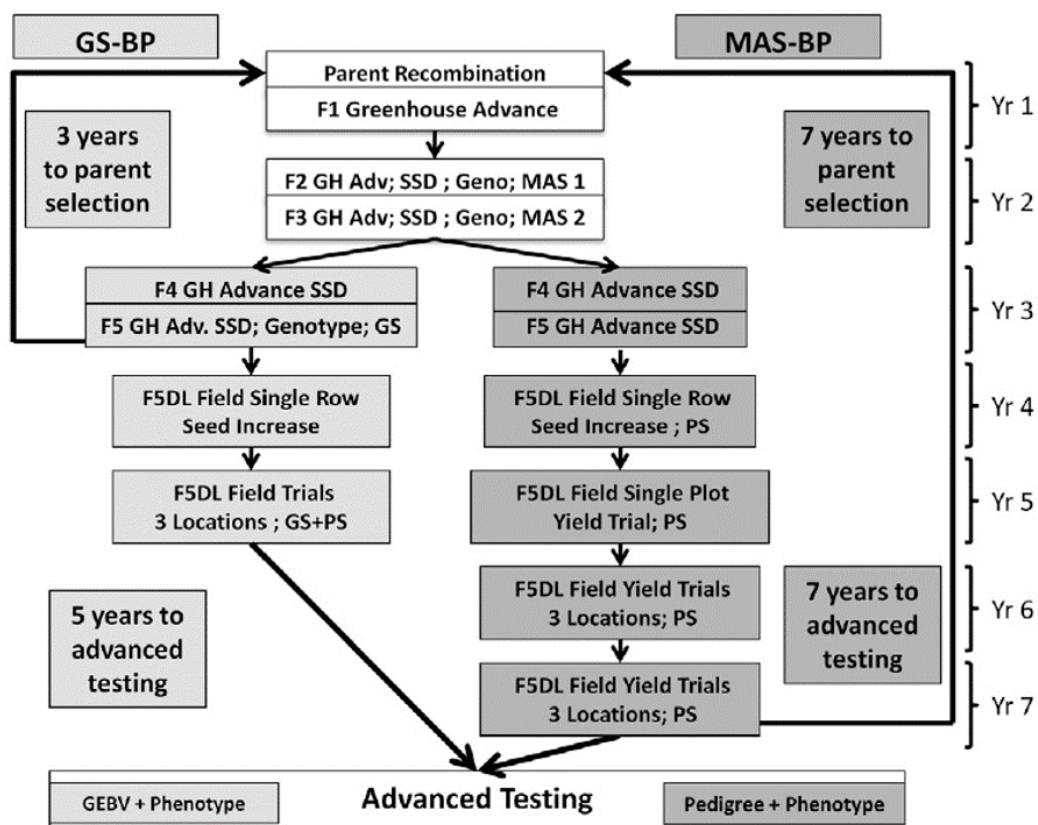


Figure 1-12. Comparaison d'un programme de sélection du blé tendre, basé sur la sélection assistée par marqueurs (MAS-BP, gris foncé) ou sur la sélection génomique (GS-BP, gris clair). Les étapes communes de sélection sont présentées en blanc. GH = serre ; Adv = avancée ; Geno = génotypage ; PS = sélection phénotypique ; F5DL = lignées dérivées de F5 (d'après Heffner *et al.* 2010).

Au stade des premiers essais du programme de sélection, les lignées doivent être évaluées sur plusieurs années et dans différents sites pour connaître leur performance potentielle de rendement dans une large gamme d'environnements. Ces réseaux d'essais sont difficiles à mettre en place du fait de la quantité de semences nécessaire pour évaluer chaque lignée dans différents environnements, et du coût du phénotypage des micro-parcelles d'essais. Les prédictions génomiques sont donc avantageuses pour différents scénarios de sparse-testing, pour lesquels toutes les lignées ne sont pas évaluées dans tous les environnements (Crossa *et al.* 2017). Il est encore possible de prédire les performances de lignées qui n'ont été évaluées dans aucun des environnements du réseau d'essais. Enfin, avec les prédictions génomiques, il est possible d'effectuer des prédictions dans un nouvel environnement ne faisant pas partie du réseau d'essais. Ces prédictions nécessitent cependant la modélisation du GxE au sein du réseau d'essais, ce que les marqueurs moléculaires font difficilement.

Parmi les utilisations plus récentes des prédictions génomiques, ces dernières peuvent être aussi utilisées en amont du programme de sélection, pour la sélection des parents utilisés dans les futurs croisements. Le choix des parents est un choix décisif que le sélectionneur doit faire pour maximiser la variance génétique de la descendance. La valeur des croisements peut être prédite grâce aux marqueurs moléculaires (Zhong et Jannink 2007) et différentes stratégies ont été développées pour sélectionner les meilleurs croisements (Lado *et al.* 2017 ; Gorjanc *et al.* 2018 ; Allier *et al.* 2019).

Depuis ces deux dernières décennies, la pertinence d'utiliser les méthodes de prédiction, comme la prédiction génomique, pour prédire des caractères complexes dans des programmes de sélection a été démontrée. Seulement, le coût du génotypage peut être un frein pour les plus petites structures de sélection qui n'ont pas le budget pour génotyper des milliers de candidats chaque année (R2D2 Consortium *et al.* 2021). Cela reste également un frein pour des espèces d'intérêt mineur pour lesquelles le séquençage est moins développé. D'autre part, les capacités de prédiction peuvent encore être améliorées notamment pour prédire des caractères, comme le rendement, qui sont soumis à des effets génétique et d'interaction GxE. L'utilité du marquage moléculaire est de pouvoir connecter un génotype à des phénotypes et d'utiliser les prédictions pour sélectionner les individus les plus transgressifs (Poland et Rife 2012). Cependant, les marqueurs moléculaires sont des variables explicatives qui ont leurs limites notamment pour expliquer des effets d'interaction, que ce soient entre gènes ou entre génotype et environnement. C'est pourquoi depuis quelques années maintenant, de nouvelles sources d'informations sont explorées pour connecter les génotypes et phénotypes.

3.3. La sélection “genomic-like omics-based” (GLOB)

3.3.1. Notion d'endophénotypes

Depuis ces dernières années, l'utilisation d'endophénotypes en amélioration des plantes a montré un intérêt grandissant. Un endophénotype est défini comme un phénotype moléculaire intermédiaire qui influence le phénotype de l'individu (Mackay *et al.* 2009). Entre le génotype et le phénotype final, il existe différentes couches d'omiques qui peuvent capturer l'activité des gènes et les interactions, entre, et à l'intérieur, des différentes strates d'omiques. Parmi ces couches, la première, la transcriptomique, reflète la part active du génome en quantifiant l'expression des gènes. La dernière couche, la métabolomique peut être capable de capturer et d'intégrer tous les

processus de régulation et d'interaction qui se sont déroulés dans les strates supérieures (Schrag *et al.* 2018). Ces endophénotypes, qui dépendent à la fois du génotype, de l'environnement et de l'interaction entre les deux, se révèlent être des variables très informatives pour prédire des caractères complexes, soumis aux effets d'épistasie et d'interaction entre le génotype et l'environnement (Fernandez *et al.* 2016).

3.3.2. Principe de la prédiction GLOB

Les endophénotypes se révèlent être de bons prédicteurs, en amélioration des plantes, pour prédire des caractères d'intérêt agronomique comme le rendement. Quelques études ont montré que les marqueurs moléculaires utilisés en sélection génomique pouvaient être remplacés par des transcrits (Frisch *et al.* 2010 ; Fu *et al.* 2012 ; Zenke-Philippi *et al.* 2017 ; Azodi *et al.* 2020), des métabolites (Riedelsheimer *et al.* 2012 ; Xu *et al.* 2016), les deux (Guo *et al.* 2016 ; Westhues *et al.* 2017 ; Schrag *et al.* 2018) ou de petits ARN (Schrag *et al.* 2018 ; Seifert *et al.* 2018) pour la prédiction de caractères complexes dans le cas de l'amélioration du maïs. Ces prédicteurs sont collectés généralement une seule fois, en conditions contrôlées, à un stade très précoce du développement de la plante et sur différents tissus comme les feuilles ou les racines. Il est intéressant de voir que ces prédicteurs sont capables de faire aussi bien en termes de capacité de prédiction que la prédiction génomique malgré des conditions de mesures très particulières, qui sont parfois très éloignées du phénotype final acquis au champ. C'est parce que ces variables ont cette capacité à capturer de la similarité génétique entre les génotypes qu'il va être possible de mener des prédictions ou de la sélection "comme la génomique mais basée sur d'autres omiques" (*genomic-like omics based*, GLOB). Dans les prédictions GLOB menées sur maïs (Riedelsheimer *et al.* 2012 ; Guo *et al.* 2016 ; Westhues *et al.* 2017 ; Schrag *et al.* 2018 ; Azodi *et al.* 2020), blé (Ward *et al.* 2015), riz (Xu *et al.* 2016) ou encore sur *Arabidopsis thaliana* (Gärtner *et al.* 2009), il a été montré que combiner les différents types d'omiques (plus particulièrement les transcrits) avec les marqueurs moléculaires conduisait parfois à de meilleures capacités de prédiction que les modèles intégrant les marqueurs moléculaires seuls.

Les endophénotypes collectés peuvent être utilisés directement dans le modèle comme variable prédictive, ou pour calculer des matrices de similarités entre individus. La plupart des modèles développés pour la sélection génomique sont donc réutilisables avec ce type de données. Toutefois, l'acquisition des endophénotypes en routine reste très coûteuse et difficile à mettre en place à l'échelle d'un programme de sélection. C'est pourquoi Rincent *et al.* (2018) ont proposé de se

tourner vers une variable très accessible pour les sélectionneurs, non destructive et de haut débit: la spectroscopie proche infrarouge (NIRS) pour faire de la sélection non pas génomique mais phénotypique. La spectroscopie proche infrarouge permet de quantifier la présence de liaisons chimiques dans un échantillon et les valeurs des absorbances mesurées aux différentes longueurs d'onde sont directement liées à l'effet du génotype, à l'effet de l'environnement et à l'interaction entre les deux. Le NIRS a des caractéristiques communes avec les endophénotypes qui peuvent aussi être exploités en amélioration des plantes.

4. Problématique et objectifs de la thèse

Les méthodes de prédiction utilisant les marqueurs moléculaires permettent théoriquement de gagner du temps entre les cycles de sélection et d'obtenir de meilleurs gains génétiques que ceux obtenus traditionnellement par les sélectionneurs. Seulement, la sélection génomique présente des limites d'application notamment lorsque le coût du génotypage reste trop élevé pour les sélectionneurs dont l'objectif est d'obtenir le meilleur équilibre entre le gain génétique et le coût de la méthode. De plus, de nouvelles sources d'informations comme les endophénotypes, montrent un certain potentiel pour améliorer les capacités de prédiction des caractères complexes, caractères dont la variance n'est pas uniquement expliquée par la part additive des gènes. C'est dans ce contexte que la sélection phénotypique représente une opportunité pour les sélectionneurs d'investir dans une méthode peu coûteuse et qui semble au moins aussi performante que la sélection génomique.

Cette nouvelle méthode prometteuse qu'est la sélection phénotypique, a été proposée par Rincet *et al.* (2018) pour la prédiction du rendement et de la date d'épiaison sur un panel élite de diversité de blé tendre et pour la prédiction de caractères du peuplier. L'objectif de ma thèse est d'évaluer cette méthode sur du matériel de sélection et donc de rendre compte de l'intérêt de la sélection phénotypique dans des programmes de sélection du blé tendre. Plusieurs questions ont été soulevées dont notamment: La sélection phénotypique peut-elle prédire correctement différents caractères agronomiques de matériel en sélection ? La sélection phénotypique peut-elle être une alternative à la sélection génomique dans les programmes de sélection du blé tendre ? A quelles étapes du schéma de sélection peut-elle intervenir pour améliorer le gain génétique du programme de sélection ? Pour répondre à mon objectif, une stratégie en trois axes de recherches a été adoptée.

Tout d'abord, la sélection phénotypique étant une nouvelle méthode, celle-ci doit d'être correctement définie. Ainsi, dans le chapitre II, nous définirons précisément la sélection phénotypique ainsi que son cadre, en lien avec les publications qui ont traité de la méthode depuis 2018.

Puis la sélection phénotypique sera comparée à la sélection génomique dans différents scénarios de prédiction, grâce aux données des sélectionneurs que j'avais à disposition. Ainsi, mes travaux de recherches utiliseront des données de panels élites de blé tendre d'hiver, issus de différents programmes de sélection. Mon travail développé dans le chapitre III, se concentrera sur l'identification de facteurs influençant les capacités de prédiction de la sélection phénotypique appliquée à du matériel de sélection. En effet, les capacités de prédiction génomique étant influencées par un grand nombre de facteurs, en est-il de même pour la sélection phénotypique ? Ces facteurs sont-ils spécifiques ? Ces prédictions seront menées sur des prédictions intra-environnementales. Dans le chapitre IV, la prédiction phénotypique sera cette fois étudiée à l'échelle multi-environnementale. Ainsi, j'évaluerai les capacités de prédiction de la phénotypique dans différents scénarios de prédiction en réseaux d'essais pour répondre aux problématiques rencontrées par les sélectionneurs. Quelles améliorations en capacités de prédiction peut-on obtenir en intégrant la modélisation du GxE avec le NIRS ? Comment gérer les données manquantes de spectres dans des dispositifs de sparse-testing ?

Enfin, après avoir développé ces axes de recherches, les résultats que j'ai obtenus seront remis en perspective dans la discussion, pour répondre aux principales questions sur l'utilisation de la sélection phénotypique en amélioration du blé tendre. Une comparaison des gains génétiques attendus des différentes méthodes sera menée et des schémas alternatifs intégrant la sélection phénotypique dans des schémas types de sélection du blé tendre seront proposés.

Chapitre 2

Définition de la sélection phénotypique

Préambule

Dans ce chapitre, nous allons définir ce qu'est précisément la sélection phénotypique à partir de la publication de Rincent *et al.* 2018 et des différentes publications qui ont suivies. Nous ferons également la synthèse des différentes applications de la spectroscopie proche infrarouge en sélection.

Ce chapitre est présenté sous la forme d'un article de synthèse qui a été accepté comme chapitre de l'ouvrage « Complex Trait Prediction » à paraître courant 2022 dans la série « Methods in Molecular Biology, de Springer.

Phenomic selection: a new and efficient alternative to genomic selection

Pauline Robert^{1,2*}, Charlotte Brault^{3,4,5,*}, Renaud Rincenc^{1,2,#}, Vincent Segura^{3,5,#}

1 Université Paris-Saclay, INRAE, CNRS, AgroParisTech, GQE - Le Moulon, 91190, Gif-sur- Yvette, France

2 INRAE - Université Clermont-Auvergne, UMR1095, GDEC, 5 chemin de Beaulieu, 63000 Clermont-Ferrand, France

3 UMR AGAP Institut, Univ Montpellier, CIRAD, INRAE, Institut Agro, F-34398 Montpellier, France

4 Institut Français de la Vigne et du Vin, Montpellier, France

5 UMT Geno-Vigne®, IFV-INRAE-Institut Agro, Montpellier, France

* equal contribution

equal contribution

Corresponding author: vincent.segura@inrae.fr

1. Abstract

Recently, Rincenc *et al.* (2018) proposed to switch molecular markers to near-infrared (NIR) spectra for inferring relationships between individuals and further performing phenomic selection (PS), analogous to genomic selection (GS). The PS concept is similar to genomic-like omics-based (GLOB) selection, in which molecular markers are replaced by endophenotypes, such as metabolites or transcript levels, except that the phenomic information obtained for instance by near-infrared spectroscopy (NIRS) has usually a much lower cost than other omics. Though NIRS has been routinely used in breeding for several decades, especially to deal with end- product quality traits, its use to predict other traits of interest and further make selections is new. Since the seminal paper on PS, several publications have advocated the use of spectral acquisition (including NIRS and hyperspectral imaging) in plant breeding towards PS, potentially providing a scope of what is possible. In the present chapter, we first come back to the concept of PS as originally proposed and provide a classification of selected papers related to the use of phenomics in breeding. We further provide a review of the selected literature concerning the type of technology used, the preprocessing of the spectra, and the statistical modelling to make predictions. We discuss the factors that likely affect the efficiency of PS and compare it to GS in terms of predictive ability. Finally, we propose several prospects for future work and application of PS in the context of plant breeding.

Key Words: Plant breeding, phenomic selection (PS), genomic-like omics-based (GLOB) selection, genomic selection (GS), near infrared spectroscopy (NIRS), hyperspectral imaging

2. Introduction: the concept of phenomic selection and its relationship with other uses of spectra in breeding

In their recent publication, Rincent *et al.* (2018) proposed to replace genomic information by phenomic information, such as near infrared (NIR) spectra, to predict quantitative traits and further perform what they coined “phenomic selection” (PS). The use of spectroscopy in agriculture and plant breeding is not novel, however, its use as an alternative to molecular markers to build relationship matrices and further predict individual performances in the context of selection is new. PS is similar to genomic-like omics-based (GLOB) selection, in which molecular markers are replaced by endophenotypes such as transcriptomics, metabolomics, or any other omics (Riedelsheimer *et al.* 2012; Feher *et al.* 2014; Ward *et al.* 2015; Fernandez *et al.* 2016; Xu *et al.* 2016; Guo *et al.* 2016; Schrag *et al.* 2018). Endophenotypes are generally measured once and for all in controlled conditions, and used to build a predictive model for field agronomic traits. This approach is thus similar to genomic selection (GS) with genotyping in the lab replaced by endophenotypic characterization in controlled conditions. In these publications, models such as G-BLUP or multi-BLUP were generally used with the kinship matrix replaced by relationship matrices estimated with the omics data, and the predictive abilities obtained were generally similar and sometimes higher than those obtained with GS. Even if endophenotypic characterization remains costly, improved efficiency to capture non-additive effects (epistasis, genotype by environment interaction - GxE) can be highly valuable.

To decrease and scale both cost and throughput, Rincent *et al.* (2018) proposed to replace genotyping or omics characterisation by NIRS. They illustrated that NIR spectra were indeed able to capture genetic similarities, and thus resulted in accurate predictions, even for traits unrelated to the tissue on which NIR spectra were measured. A GLOB selection approach based on NIR spectra acquired in one given environment was for instance able to accurately predict yield in other environments, as long as the calibration set was phenotyped in these environments. Note, that PS, and in particular GLOB selection, is radically different from the classical NIRS use. In the classical use, NIRS predicts the chemical composition of the analysed tissue. In PS and GLOB selection, NIR spectra (or other phenomics data) are used to capture the genetic similarities between the genotypes, which allows accurate predictions of any polygenic trait.

Spectroscopy techniques, such as NIRS, measure the emission or reflection of light on a sample for a given wavenumber range, e.g. for NIR from 780 nm to 2500 nm. Various chemical bonds absorb light at different wavelengths, and this can be used in a quantitative manner.

NIRS provides a non-destructive and high-throughput measurement of living samples (where water absorbance bands do not overlap) as well as dried or crushed tissues. Absorption or reflectance values at a given wavelength are proportional to molecule concentrations, as depicted in the Beer-Lambert law. Thus, spectra variations are due to the combination of molecules in the tissue and their respective absorption bands. This chemical property has been widely used in agriculture and forestry in many species to predict traits of interest, such as those related to grain composition for cereals (Osborne 2006), milk composition for dairy cattle (Holroyd 2013), wood properties for forest trees (Tsuchikawa and Kobori 2015), quality traits for fruits and vegetables (Nicolai *et al.* 2007).

Apart from this classical use of NIRS, which can be exploited in breeding to make selections on traits related to end-product composition, other use of NIRS in breeding have been reported in the literature, which can be classified into four main categories as proposed in Table 2-1: NIR spectra as proxy of the target trait, NIR spectra as a secondary trait, PS within environments and PS across environments as a particular case of GLOB selection. In the following paragraphs, we provide a definition of each of these four categories.

The first category concerns the use of NIR spectra as a proxy of complex traits such as grain yield, water stress or chlorophyll content with vegetation indices (VIs), which are based on a few wavelength bands. Several indices were successively developed across years, the most famous one being NDVI (normalized difference vegetation index). These indices are strongly correlated with photosynthetic activity and sometimes with yield (Tucker 1979) and are still widely used to reflect biomass or yield of the analysed plants or plots (Labus *et al.* 2002; Ferrio *et al.* 2004; Babar *et al.* 2006; Hernandez *et al.* 2015; Aguate *et al.* 2017). The advantage here of using NIR spectra as a proxy trait is to avoid expensive phenotyping and to enable indirect selection of target traits before harvest.

In the second category, we have gathered studies that combined NIRS measurements with molecular markers to increase the accuracy of genomic prediction. In that case, NIR spectrum

was considered as a secondary trait to be associated with the target trait in a multivariate prediction model (Rutkoski *et al.* 2016; Sun *et al.* 2017; Hayes *et al.* 2017). It is worth mentioning that such studies report two distinct examples of application: one based on canopy reflectance (Rutkoski *et al.* 2016; Sun *et al.* 2017; Hayes *et al.* 2017), which relates to the overall plant health; and the other based on grain NIRS (Hayes *et al.* 2017), which relates to the end product or cumulative energy accumulation of the plant. In any case, this approach is particularly valuable when the predicted set is phenotyped with NIRS (trait-assisted prediction). Other authors have used NIR measurements to specifically account for GxE in the genomic prediction model (Montesinos-López *et al.* 2017a; Krause *et al.* 2019), but this requires NIRS data for each environment to be measured.

The third category includes studies that have investigated using NIR spectra to build a (hyper)spectral relationship matrix between plots/individuals, referred to as H matrix, and integrated it in the prediction equation with or without the genomic relationship matrix (kinship), referred to as G matrix (Montesinos-López *et al.* 2017a; Krause *et al.* 2019; Cuevas *et al.* 2019; Galán *et al.* 2020). The NIRS-based similarity between two plots/individuals $H(i, j)$ can be estimated with the following formula:

$$H(i, j) = \frac{\sum_{k=1}^{n_w} [S(i, k) \times S(j, k)]}{n_w},$$

with $S(i, k)$ the preprocessed absorbance or reflectance (see hereafter) measured on the i^{th} plot/individual for the k^{th} wavelength. S is centered and scaled for each wavelength and has the dimension n (number of plots/individuals) times n_w (number of wavelengths). In matrix notation

$$H = \frac{SS'}{n_w}.$$

This PS approach resulted in promising accuracies, however, in these publications, predictions were still made for a specific site-year trial, and so NIR spectra have to be measured on each plot in each environment, with the strong limitation that all varieties have to be grown in each environment.

Table 2-1. Selection of work using NIRS information in the context of breeding. The papers are sorted according to the type of prediction made with NIRS information from NIRS as a proxy trait to phenomic selection (PS). For details see the text. PLS-R: Partial Least Squares Regression, VI: Vegetation Index, UAV: Unmanned Aerial Vehicle, OLS: Ordinary Least Squares, G-BLUP: Genomic Best Linear Unbiased Prediction, LASSO: Least Absolute Shrinkage and Selection Operator, GLOB: Genomic-Like Omics-Based.

Reference	Type of technology	Statistical method	Type of prediction	Prediction setting
Ferrio <i>et al.</i> (2004)	NIR spectroscopy in the laboratory on milled grains, wavelength range: 1100-2500nm.	PLS-R	NIRS as a proxy trait	Plot/individual level
Hernandez <i>et al.</i> (2015)	NIR spectroscopy in the field on plant canopy, wavelength range: 350-2500nm.	VIs, Ridge regression		
Aguate <i>et al.</i> (2017)	Hyperspectral imaging in the field (UAV), wavelength range: 392-850nm	VIs, PLS-R, OLS, BayesB		
Hayes <i>et al.</i> (2017)	NIR spectroscopy in the laboratory on grains, wavelength range: 400-2498nm.	Multi-trait G-BLUP	NIRS as a secondary trait	Plot/individual level
Rutkoski <i>et al.</i> (2016)	Hyperspectral imaging in the field (UAV)	Multi-trait G-BLUP		Entry/genotype level within each site-year
Sun <i>et al.</i> (2017)	Hyperspectral imaging in the field (UAV)	Multi-trait G-BLUP		Entry/genotype level within each site-year
Montesinos-López <i>et al.</i> (2017a)	Hyperspectral imaging in the field (UAV), wavelength range: 392.03-850.69nm	VIs, PLS-R, OLS, Bayes B, functional regressions	PS	Entry/genotype level within each site-year
Krause <i>et al.</i> (2019)	Hyperspectral imaging in the field (UAV), wavelength range: 380-850nm	G-BLUP Single & multikernel		Entry/genotype level within each site-year
Galán <i>et al.</i> (2020)	Hyperspectral imaging in the field (UAV), wavelengths range: 410-993nm	G-BLUP Single & multikernel, LASSO and EN for feature selection		Entry/genotype level within each
Cuevas <i>et al.</i> (2019)	NIR spectroscopy in the laboratory on grains, wavelength range: 400-2500nm.	G-BLUP, Gaussian Kernel, Arc-cosine kernel		Entry/genotype level within a single environment
Lane <i>et al.</i> (2020)	NIR spectroscopy in the laboratory on grains, wavelength range: 1000-2500nm.	PLSR, G-BLUP, Functional Regression		Plot/individual level & entry/genotype level within and across environments
Rincent <i>et al.</i> (2018)	NIR spectroscopy in the laboratory on milled leaves and grains for wheat (wavelength range: 400-2500nm) and milled wood for poplar (wavelength range: 1250-2500nm)	G-BLUP, Bayesian LASSO	PS & GLOB	Entry/genotype level within and across environments

Thus, we propose here a fourth category to further make a distinction between PS applied to the plots on which NIR spectra were acquired (previous category) and PS predictions across-environments with NIR spectra measured in a reference site only, a particular case of what we called GLOB selection (scenario S2 in Rincent *et al.* (2018) and Figure 2-1). In GLOB selection, we suppose that NIRS captures genetic similarities, which means that it is sufficient to acquire the spectra once and for all in one single environment. The derived H similarity matrix is then used in replacement of the G matrix in the classical GS models. To date, we have found only two publications (Rincent *et al.* 2018; Lane *et al.* 2020), that performed GLOB selection with NIRS, and they both resulted in accurate predictions, often more accurate than GS and with dramatically reduced costs. These publications and the results obtained with GLOB selection using other kinds of omics prove that variations other than DNA markers can efficiently capture genetic similarities between genotypes and result in accurate predictions.

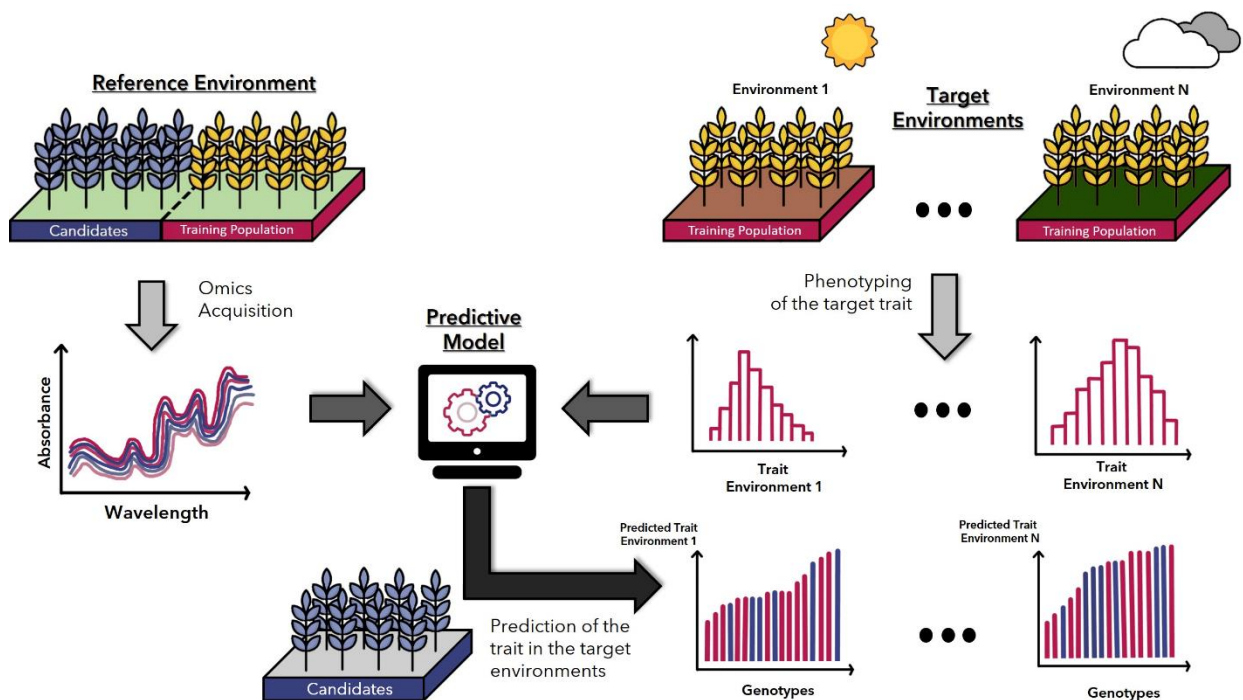


Figure 2-1. Prediction of a target trait for the selection candidates in different environments with GLOB selection. Training population is phenotyped for the target traits (e.g. productivity) in the target environments. Omics or phenomic data (e.g. NIR spectra) are collected on both the candidate and the training individuals in a same reference environment, for capturing a genetic similarity between individuals. Genotypic values of the selection candidates are predicted for the target traits in each target environment.

3. Literature review on the use of spectra in selection

Following previous definitions, we have selected a number of papers illustrating each of the four categories of use of NIRS in the context of breeding. The main features of these papers are summarized in Table 2-1, and further presented and detailed hereafter, with respect to the type of technology employed to obtain spectra, the statistical pretreatments of the spectra and the statistical model applied for phenotype prediction. We also provide in Table 2-1 a comparison of the relative performances of phenomic and genomic predictions for the very few papers which enable such a comparison. Finally, we discuss the factors that affect the predictive ability of PS.

3.1. Types of technology

Traditionally, NIRS measurements are conducted in laboratories under controlled conditions for either dried vegetative tissue (e.g. forages) or dried reproductive tissue (e.g. grain). This kind of data displays many advantages: measurements are robust, low-cost and routinely applied by breeders to predict quality traits. There are also a number of disadvantages, there is substantial extra effort needed to bring these materials from the field to the lab, and to dry them so that water absorbance (which overlaps other chemical bond absorbance) is minimized. In these laboratory conditions, spectra are constituted of many wavelengths possibly from the visible and near infrared (400 to 2500 nm approximately), constituting a dataset of hundreds of variables (Osborne 2006).

With the rise of high-throughput phenotyping, spectrum measurements have benefited from technological developments which enables the direct collection of spectra in the field possibly at several time points, like hyperspectral imaging from Unoccupied Aerial Vehicle (UAV) or direct measurements of fresh material with portable (micro-)spectrometers. Hyperspectral imaging takes images with several wavelengths for each pixel, possibly at multiple time points in the visible and in a small portion of the NIR spectrum. A reflectance measure is attributed to some groups of wavelengths (bands) or to individual wavelengths directly. The measurements at the pixel level can be integrated at the microplot level to characterize a unique variety (Aguate *et al.* 2017; Montesinos-López *et al.* 2017a; Cuevas *et al.* 2019; Galán *et al.* 2020).

Portable (micro-)spectrometers have also been developed to measure the reflectance directly in fields on undestroyed fresh material covering the visible and NIR spectrum (Hernandez *et al.* 2015).

Wavelengths can be used directly as variables in predictive models or they can be derived in several indexes, like Vis (Aguate *et al.* 2017; Montesinos-López *et al.* 2017a). VIs describe vegetation properties by summarising the information of large amounts of data to facilitate processing of camera and satellite images. However, in Aguate *et al.* (2017) the use of all the hyperspectral bands achieved better prediction than using VI individually.

Technologies used to collect NIR spectrum are numerous, each with advantages and disadvantages. On the one hand, the use of NIR spectrometers in laboratory conditions is a robust method but can be time consuming due to collection and possibly preparation of samples. On the other hand, UAV and portable (micro-)spectrometers are quick techniques to collect NIRS but the number of wavelengths available is usually reduced and measurements can be affected by environmental noise which, is harder to control in the field than in the laboratory. Depending on the application, trade-offs must be found between labor intensity, costs and spectrum quality.

To date, very few studies have tried to compare the predictive ability of different spectrum measurements, especially in the context of plant breeding. Recently, Zgouz *et al.* (2020) have reported a dataset of spectra collected on 60 sugarcane samples with 8 visible/NIR spectrometers including handheld micro-spectrometers. Such a dataset is very useful to compare different tools, although results might be context dependent, *i.e.* the most accurate model for different traits and species might be obtained with different spectrometers. Still, quantifying the gain or the loss of predictive ability for each technique will be helpful to guide in using one technique rather than another for a specific objective. Other techniques could be used to facilitate measurement, for example to combine hyperspectral images and laboratory spectra. Instead of using a spectrometer to measure samples one by one, hyperspectral images can measure several samples at the same time. This would enhance robustness of spectra collection and reduce time of measurements. Beyond technical issues, it is also important to consider practical organizational questions, such as the period at which spectra are measured, to make sure that the predictions are available before the sowing of the next season.

3.2. Preprocessing NIR spectra

In ideal conditions, NIR spectra are based on the Beer-Lambert law and the sample absorbance is directly linked to the concentration of chemical compounds of the sample. But in practice, many factors (independent from the sample composition) will influence the measured absorbance. This is the case for instance of temperature or granulometry, which will deform the final spectrum, biasing spectra comparison. To deal with external effects, a mathematical correction or preprocessing can be applied as illustrated in Figure 2-2 for spectra collected on bread wheat grains. Mainly two external effects usually need to be corrected: additive and multiplicative effects. In additive effects, noise affects spectra irrespectively of the wavelength and usually yields a baseline shift which can be corrected with a detrend (Barnes *et al.* 1989) or a derivation (Figure 2-2, C) typically carried out through a Savitzky-Golay filter which consists in a polynomial smoothing (Savitzky and Golay 1964). The baseline shift appears when the absorbance increases with the wavelength due to the increased light intensity. Multiplicative effects typically affect spectra differently depending on the wavelength and are usually linked to an increase of the distance crossed by the photons (due to different granulometry for example). They can be corrected by a normalization (Figure 2-2, B). This effect is present when for low absorbances at a wavelength, the variability is also low and for high absorbances at a wavelength, the variability is high. Other preprocessing techniques have also been proposed to specifically deal with an external parameter known to bias spectra such as temperature or hygrometry. This is the case for instance of the method called External Parameter Orthogonalization (EPO, Roger *et al.* 2003).

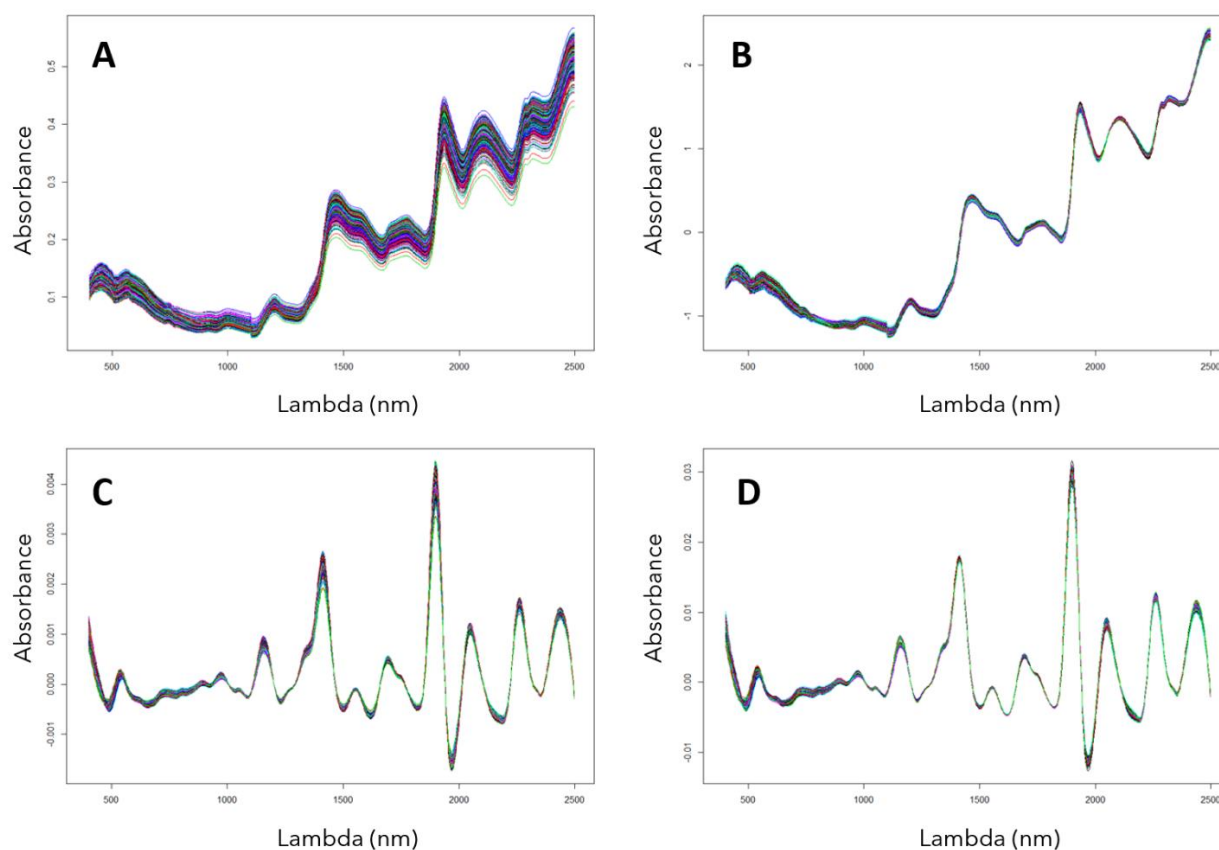


Figure 2-2. Visualization of different filters applied on a spectrum dataset. Each color represents a bread wheat variety. Spectra were collected on grains with a lab spectrometer NIRS 6500 FOSS. A. raw spectra (no preprocessing); B. normalization (standard normal variate); C. first derivative of raw spectra; D. first derivative on normalized spectra.

The preprocessing methods briefly introduced in the previous paragraph have been previously developed in the chemometrics literature. This preprocessing is routinely and widely used when applying NIRS in the classical way, *i.e.* to predict the composition of end-products. In the context of breeding and PS, further preprocessing taken from the breeding literature can be carried out to improve the ability of the spectra to predict genetic values. Such preprocessing includes building a model on the absorbance or reflectance at each wavelength considering the effects of the experimental design (e.g. blocks or spatial effects) together with genetic effects to further extract genotypic values (Lane *et al.* 2020; Galán *et al.* 2020). Genotypic values may be BLUEs or BLUPs depending on whether the genotype effect is considered as fixed or random in the model. This preprocessing typically comes from the fact that PS is carried out at the genotype level rather than at the individual or plot level, and consequently one needs to obtain a unique NIRS matrix at the genotype level for model training and prediction. It is interesting to note that if the entire

spectrum is considered rather than absorbance or reflectance at given wavelengths, such corrections are related to the orthogonalization approaches from the chemometrics literature. Indeed, recently Ryckewaert *et al.* (2020) proposed to make use of spectra replicates, typically obtained when characterizing plants under genetic trials, to reduce the repeatability error.

They developed a new preprocessing technique based on orthogonalization after an ANOVA–simultaneous component analysis (REP-ASCA).

The filters mentioned previously are not an exhaustive list but have been the most commonly used in NIRS chemometric prediction. Preprocessing can be done in numerous ways, as shown across different studies, suggesting that no one standard preprocessing approach exists. We have noticed that PS predictions were influenced by the preprocessing applied on spectra, consequently we recommend testing different filters on a subset of data to cross-validate filters efficiency, before carrying out deeper analysis.

3.3. Statistical models for phenotype prediction

NIRS reflectances or absorbances are quantitative variables, like bi-allelic markers usually coded numerically with allelic dosages, basically all models developed or used in the frame of genomic selection can also be used for PS, from the “infinitesimal” model to Bayesian models with various prior distributions or machine learning methods.

One such reference model for PS is the H-BLUP, similar to G-BLUP but with a similarity matrix (H) estimated with NIRS (Rincent *et al.* 2018; Krause *et al.* 2019; Galán *et al.* 2020). Different kernels can be used within such a framework, including Gaussian kernel or arc-cosine kernel (Cuevas *et al.* 2019). As with molecular markers, this model can be equivalent to a ridge-regression on the wavelengths, provided the H matrix is computed accordingly, as demonstrated hereafter. The predictive ability of the H-BLUP model can be measured with cross-validation, as with G-BLUP or other GS models.

The H-BLUP model is defined as: $y = \mu + u + e$, with $\text{var}(y) = H\sigma^2 + I\sigma^2$, and where y is a vector of phenotypes, H is the NIR spectra-based similarity matrix as defined above, μ is the intercept, u and e are random genetic and residual effects, respectively.

The RRN-BLUP model (Ridge Regression NIRS BLUP) is defined as: $y = \mu + Sv + e$, with $var(y) = SS'\sigma_s^2 + I\sigma_e^2$, and where S is the matrix of preprocessed, centred and scaled NIRS as defined above.

The mean of y is equal to μ in both models, thus H-BLUP and RRN-BLUP are equivalent if $H\sigma_u^2 = SS'\sigma_s^2$, which is for instance the case when $H = \frac{SS'}{n_w}$ and $\sigma_s^2 = \frac{\sigma_u^2}{n_w}$.

Functional regression models seem particularly interesting for PS, as they model the linear trend of the spectra (Montesinos-López *et al.* 2017a). Different kinds of functional regressions were proposed such as functional B-Spline, functional Fourier (Montesinos-López *et al.* 2017a,) and Bayesian functional (Montesinos-López *et al.* 2017a). H-BLUP and functional regression models have proven to yield accurate predictions while reducing computational time by diminishing the number of parameters to estimate. This could be important if several spectra from different environments are available, resulting in a high number of predictors.

Partial Least Squares (PLS) regression, classically used in chemometrics, or variable selection approaches (such as LASSO or BayesB) can also be used to tackle multicollinearity and high dimensionality. PLS regression consists of condensing the information contained across all wavelengths into a few orthogonal variables that maximize the covariation between the predictor matrix and the response variable. In LASSO and Bayes B, it is assumed that only a portion of the variables has an effect on the trait. Variable selection seems promising for PS, because the spectrum could be restricted to its most heritable parts (Ferragina *et al.* 2015; Galán *et al.* 2020). However, it should be noted that the preselection of wavelengths using vegetation indices or with knowledge on the genomic heritability of the wavelengths generally result in lower prediction accuracies than when using the full spectrum (Aguate *et al.* 2017; Galán *et al.* 2020).

In GS the choice of the prediction model can be guided by the expected genetic architecture of the predicted trait. The choice of a PS model adapted to a given trait cannot yet rely on such assumptions, and it is not clear how the optimal prediction model can be related to the trait characteristics. The various models tested in the literature sometimes resulted in contrasted prediction accuracies, but in general sophisticated models were not better than a simple H-BLUP. Models relying on a mixture of distributions such as BayesR (Erbe *et al.* 2012) are accurate for contrasted genetic architecture in GS, it would be interesting to test them in PS.

In any case, alternative prediction models should be compared using cross-validations within the calibration set.

Contrary to molecular markers in GS, in PS several spectra corresponding to different replicates of genotypes possibly across different environments can be available to build predictive models. In this case, one possibility for calibration is to test each spectrum in order to determine the one which yields the most accurate predictions. Another possibility is to make use of all the available spectra. Lane *et al.* (2020) proposed in the frame of the H-BLUP model to compute the mean of the relationship matrices calculated from each spectrum individually. It is noteworthy that this proposition is equivalent to computing the relationship matrix from a large combined spectra matrix, providing that the individual spectra matrices have the same number of wavelengths, as shown hereafter.

The similarity matrix $H_T(i, j)$ computed with the combined spectra matrix S_T (in which all spectra matrices are included one next to the other) is given by:

$$H_T(i, j) = \frac{\sum_{k=1}^{n_t} [S_T(i, k) \times S_T(j, k)]}{n_t}$$

$$H_T(i, j) = \frac{\sum_{p=1}^{n_w} [S_1(i, p) \times S_1(j, p)] + \dots + \sum_{p=1}^{n_w} [S_{n_l}(i, p) \times S_{n_l}(j, p)]}{n_l \times n_w}$$

$$H_T(i, j) = \frac{1}{n_l} \times \left[\frac{\sum_{p=1}^{n_w} [S_1(i, p) \times S_1(j, p)]}{n_w} + \dots + \frac{\sum_{p=1}^{n_w} [S_{n_l}(i, p) \times S_{n_l}(j, p)]}{n_w} \right]$$

$$H_T(i, j) = \frac{1}{n_l} \times \left[\sum_{u=1}^T H_u(i, j) \right]$$

S_T has dimension n (number of individuals) times $n_t = n_l \times n_w$, with n_l the number of spectra (e.g. number of environments in which NIRS was acquired) and n_w the number of wavelengths of each spectrum (we consider that all spectra have the same wavelengths). $S_{ud}(i, k)$ is the absorbance or reflectance measured on the i^{th} individual for the k^{th} wavelength in the u^{th} NIR preprocessed spectrum. S_{ud} has the dimension n (number of individuals) times n_w (number of wavelengths). $H_u(i, j)$ is the similarity between individuals i and j estimated with one given u spectrum.

3.4. Relative performance of PS versus GS

There are very few studies that compare PS (and in particular GLOB selection) with GS (Table 2-2). Although Lane *et al.* (2020) was one of the two studies that implemented GLOB prediction with spectra following our definition, it could not be included in this comparison because they did not apply GS. Table 2-2 illustrates that PS and GLOB selection have been mainly implemented on cereal species, probably because of the widespread and routine use of NIR measurements on grains to predict protein content. Krause *et al.* (2019) and Galán *et al.* (2020) reported similar accuracies for GS and PS, while Cuevas *et al.* (2019) showed lower accuracies for PS compared to GS (0.37 and 0.46, respectively). The highest PS accuracy compared with GS was observed in Rincent *et al.* (2018) in wheat. The lowest PS accuracy compared to GS was observed in Rincent *et al.* (2018) in poplar for which NIR spectra were collected on wood for a reduced range of wavelengths. From these data, it is apparent that PS had comparable or higher accuracies than GS in most cases. Even in cases where PS is less accurate than GS, as NIR measurements are high-throughput and low-cost compared to genotyping, PS could still provide higher genetic gains than GS, as demonstrated in Rincent *et al.* (2018). In our ongoing research we compared GS and PS at different generations of elite bread wheat selection. We found that PS could be as accurate as GS and even better when applied to early generations. Further work on other species is clearly needed to deepen this comparison and provide valuable information on the factors and conditions (e.g. tissue, environment) that determine the predictive ability of NIRS.

By considerably reducing the costs of implementation, PS is a tool of choice to improve the balance between costs and benefits in comparison with GS. PS would be particularly valuable for orphan crops for which genotyping is expensive, or for major crops for which phenomic data are already routinely collected (e.g. maize and wheat). In the latter case, phenomic prediction already opens new possibilities in existing breeding programs without any additional cost, and with predictive abilities similar to those obtained with genomic prediction (Rincent *et al.* 2018).

Table 2-2. Comparative predictive ability reported for prediction based on G and H matrices. GY: grain yield, DTHD: days to heading date, DMY: dry matter yield, HD: heading date, HT: height, CIRC: circumference, BF: bud flush, BS: bud set, RUST: resistance to rust. ORL: Orléans site (France), SAV: Savigliano site (Italy). Several predictive abilities were available in the study of Krause *et al.* (2019), one per flying date. When several methods were tested, the one with the highest predictive ability was kept for each modality. When the prediction was made in several environments, years or on different tissues for PS, the range of predictive ability is indicated in brackets.

References	Species	Traits	Mean predictive ability	
			G matrix	H matrix
Krause <i>et al.</i> (2019)	Wheat	GY (without DTHD correction)	0.41 (0.19-0.6)	0.42 (0-0.75)
Galán <i>et al.</i> (2020)	Winter rye	DMY	0.6	0.59
Cuevas <i>et al.</i> (2019)	Wheat	GY	0.46	0.37
Rincent <i>et al.</i> (2018)	Wheat S1	HD	0.57	0.84 (0.78-0.88)
		GY	0.38	0.51 (0.37-0.62)
	Wheat S2	HD	0.61	0.78 (0.67-0.83)
		GY	0.46	0.44 (0.28-0.53)
	Poplar	HT.ORL	0.56	0.64 (S1) / 0.49 (S2)
		CIRC.ORL	0.61	0.72 (S1) / 0.48 (S2)
		CIRC.SAV	0.77	0.80 (S1) / 0.46 (S2)
		BF.ORL	0.72	0.15 (S1) / 0.33 (S2)
		BF.SAV	0.73	0.32 (S1) / 0.09 (S2)
		BS.ORL	0.73	0.53 (S1) / 0.44 (S2)
		BS.SAV	0.59	0.34 (S1) / 0.45 (S2)
RUST.ORL	0.61	0.45 (S1) / 0.33 (S2)		

3.5. Factors affecting PS predictive ability

In the past, several kinds of omics were used to make genomic-like predictions with promising results (Schrag *et al.* 2018). NIRS captures an integrative signal, and is biologically more difficult to interpret than other omics, which describe each molecule individually (e.g. transcriptomics, proteomics, metabolomics). However, because prediction models do not necessarily need to be interpreted biologically, NIRS can be used to make predictions using “black-box” models.

There are two factors that contribute to the predictive ability and consequently to the success of PS: (i) the ability to capture target trait proxies, and (ii) the ability to infer genetic relatedness.

The former depends on the physiological connectedness between the target trait and the composition and features of the tissue analysed with NIRS. This is for example the case when NIR spectra collected on wood powder is used to predict wood properties, or when NIR spectra collected on fruits is used to predict fruit composition. In these cases, PS should be nearly equivalent or superior to the traditional way of using NIRS (prediction of the tissue composition), the only difference being that when doing PS we usually work at the genotype level because we aim at ranking and selecting the best genotypes while in the traditional use of NIRS we make predictions at the plot or plant level (Lane *et al.* 2020). But we could think of more indirect relationships between the target trait and NIR spectra to explain its predictive ability, for instance in wheat the good predictive ability of PS for yield could be due to the fact that NIRS is a very good predictor of grain composition, that is often negatively correlated to yield. This could also be the case for maturity: the spectra are influenced by the maturity of the plants, and this maturity is sometimes correlated to yield (Krause *et al.* 2019). However, it is important to stress that even in the absence of any direct relationship between the predicted trait and the tissue analysed with NIRS, PS can still be accurate. It was for instance shown in (Rincent *et al.* (2018) that NIR spectra collected on leaves in one environment could be used to estimate a covariance matrix resulting in accurate prediction of yield in a completely independent environment. In this particular example, the correlation between yield in the environment in which NIR spectra were obtained and yield in the predicted environment (by cross-validation within the predicted environment) was as low as 0.16, whereas PS predictive ability was above 0.5. This means that NIRS derived relationship matrices were able to capture genetic relatedness between lines valuable for predicting yield. This was further demonstrated by the fact that genomic heritability was

significant for many wavelengths. A further demonstration could be done with a simulation study, by estimating the predictive ability of PS for traits simulated with genotype data. In this case, the predictive ability of PS averaged over a large number of simulated traits would provide an evaluation of the ability of NIRS to infer genetic relationships for predicting quantitative traits unrelated to the tissue composition.

PS is a recent research topic, and further investigations are required to use it in an optimal way. One can expect that, as for genomic selection, prediction accuracy will be strongly dependent on the target trait and its heritability, as well as the size and composition of the training set. In comparison to GS, prediction accuracy obtained with PS can also be affected by the origin of the spectrum (tissue, environments, kind of sensors). This is similar to the choice of a SNP array and marker filtering in GS, but the effect of the origin of the spectrum appears to be more pronounced. First results suggest that NIR spectra collected under plant stress conditions are more efficient (Rincint *et al.* 2018; Lane *et al.* 2020), but other experiments are required before it can be understood if this is the rule or the exception. An interesting result is that, in practice, the combination of different NIR spectra (collected on different tissues or different environments) leads to predictive abilities at least as good as those obtained with the best NIR spectrum taken alone (Rincint *et al.* 2018, unpublished results). This means that in some cases it is not necessary to identify the best conditions to obtain NIR spectra, but simply to aggregate all the spectra collected (e.g. spectra obtained on the same genotypes at the different steps of the breeding program). As shown in the present review, aggregating NIRS matrices prior to computing the H matrix is equivalent to averaging H matrices estimated with individual NIRS matrices and is thus quite straightforward. But, in any case, the choice of the tissue, timing and sensors could and should also be optimized. We can think that NIR spectra collected on homogeneous, representative samples (leaf powder, seed sample or flour) are more useful than NIR spectra obtained on a tiny area of raw material. The way NIR spectra are collected should also be optimized in terms of practical feasibility. For instance, in wheat, it would be much more feasible to measure NIR spectra during the growing season, than on seeds after harvest, because the few weeks between harvest and sowing of the next generation is labour intensive and so NIR spectra acquisition would be difficult during this period.

4. Prospects

As introduced above, there are numerous ways of using PS in breeding. We particularly foresee several applications to be addressed with PS and that we detail hereafter. Some of them are quite direct applications, which can already be deployed in breeding programs (A-F), while others are prospects which deserve research investments prior to their adoption in breeding in a relatively longer term (G). Although many of the prospects presented here are also shared with GS, we have tried in what follows to underline their specificity with respect to PS.

4.1. Prebreeding: screening diversity collections at low cost

Gene banks are a reservoir of genetic diversity in which genes of tolerance to biotic or abiotic stress can be discovered. These collections will become crucial as the genetic diversity in the breeding program will not be sufficient to face upcoming changes due to the evolution of management practices or climate change. However, the identification of promising genes or individuals require the phenotyping of the collection for the target trait, which could be too expensive given the number of accessions stored in these collections. Yu *et al.* (2016) and Crossa *et al.* (2016), have proposed to run GS to screen these collections. However, this requires genotyping the full collection, which is also expensive considering the large size of the gene banks. Another option would be to measure NIR spectra on each accession, phenotype a subset, and predict the remaining accessions using PS. In most genebanks, accessions are regularly sown in nurseries to produce new seeds, as the germination rate decreases with time. NIRS could thus either be collected on seeds directly in gene banks or on other tissues in these nurseries. The same approach could also be extended to other species like perennials which are usually not conserved as seeds but as living plants in the field.

4.2. Sparse testing: experimental design optimization in breeding programs

For most species, selection candidates are evaluated in multi-environment trials to estimate their stability and productivity in contrasted environments. This is an expensive step as the number of variety/environment combinations can rapidly become very high. To increase the number of environments or varieties with the same costs, one option would be to run a sparse testing design, in which all varieties are evaluated at least in one environment, but with a given proportion of varieties/environments not tested. Sparse testing is sometimes imposed to breeders

because part of a trial is accidentally destroyed, or not harvested for some reason.

Sparse testing is a scenario for which genomic predictions are particularly accurate, in comparison to predicting a completely new variety or a new environment (Jarquín *et al.* 2014; Ly *et al.* 2018; Jarquín *et al.* 2020). Our proposition is to use PS to predict the unobserved variety/environment combinations. In wheat, this approach is already applicable as NIR spectra are usually collected in nurseries the year before the multi-environment trials. Our first unpublished results on sparse testing show that the prediction accuracy of PS under this scenario can be as high as to the one obtained with GS.

4.3. Combining reduction of generation time (speed breeding) and performance prediction (PS) to increase genetic progress

One challenge in breeding is to accelerate programs to quickly release new varieties. Breeding is often constrained to one to two generations per year for annual crops, which limits genetic progress. Several methods have been proposed to reduce generation time, including recently “speed breeding” (Watson *et al.* 2018). Speed breeding consists in controlling photoperiod and temperature to get optimal growing conditions and accelerate the time elapsed from seed to seed, allowing up to six spring wheat generations per year. Thus, a great number of segregation and recombination events can occur in a short time, allowing to rapidly produce varieties combining favourable alleles. PS could have two applications that could work particularly well if combined with speed breeding. The first application could be implemented in the speed breeding process itself. During this process, each plant is unique and is phenotypically quite different from what it would look in the field, which makes direct selection impossible except for phenology, height and some disease resistance traits. But molecular markers or NIR spectra could be measured on the plants or on the seeds which would allow predicting performance traits using GS (Watson *et al.* 2019), or at a lower cost PS. This would considerably increase genetic progress by reducing both the generation interval (speed breeding) and the phenotyping process (partially replaced by predictions). The second application would be to apply PS to reduce the field trials size after the speed breeding process. Thousands of genotypes can be produced by speed breeding, and it would be difficult to phenotype all of them in field experiments (after multiplication in the nurseries). As NIR spectra can be collected at the end of the process or in the nurseries, PS could be used to predict genotype performance in the nurseries or with sparse testing in the following field experiments as described above.

4.4. GxE prediction

Unlike molecular markers, NIR spectra are directly influenced by the response of the plants to the environmental conditions. This seems likely to result in a lack of stability of the spectra in different environments. It also may mean that the spectra are able to capture the genetic responses to a given environment (GxE), as illustrated in Rincent *et al.* (2018), which opens new perspectives of application for PS. This GxE variance could be exploited to enhance predictive ability of local adaptation. One possibility to improve the GxE prediction models is to use NIRS collected in each environment to estimate environment specific covariance matrices. Krause *et al.* (2019) compared different models (single and multi-kernel) using molecular marker, pedigree or NIR spectra, to predict wheat grain yield in multi-environment trials. They found that the best multi-kernel integrating GxE was the one with the hyperspectral matrix. Using NIRS enhanced predictive accuracy of GxE compared to models that use molecular markers or pedigree. Similar results were observed by Montesinos-López *et al.* (2017a) and Lane *et al.* (2020) where the interaction between NIR bands and environments was integrated in the models. The introduction of NIR information in the prediction model allowed increasing GxE prediction accuracy, which was not the case with molecular markers.

This multi-kernel method requires collecting the spectra of all genotypes in all the environments of the multi-environment trials, which is not possible with classical sparse testing designs. One possibility would be to grow a nursery in parallel of each trial, which would be much cheaper than a real trial. In this case, the trial will be dedicated to phenotype the training population, and the nursery to collect spectra on both training and predicted varieties. Krause *et al.* (2019) proposed to reduce the size of the microplots, to measure NIR spectra on all varieties in all environments using a UAV. The objective is to find a compromise between the number of candidate varieties predicted with PS and the number of training varieties phenotyped for the target traits.

Another potential application would be to use NIR spectra collected in the different environments to estimate similarities between them, as proposed in Jarquín *et al.* (2014) and Heslot *et al.* (2014) with environmental covariates. In comparison to classical environmental covariates (e.g. temperature, hydric balance, radiation), NIRS has the advantage of capturing signals from the plant that could be influenced by the experienced stresses. GxE models enriched with NIRS would be particularly useful to make predictions in multi-environment trials. This application of PS

could potentially allow making predictions in new environments (no other phenotype than NIR spectra are collected in these environments) by estimating the similarity between the new environment and the calibration environments. NIR spectra can also be used as a covariate in the predictive model to characterize the GxE. Lane *et al.* (2020) proposed to consider the GxE by using wavelengths as covariates in the predictive model through functional regression. They predicted yields of known hybrids in an unknown environment and found that considering GxE with covariates worked better than regular H-BLUP. The different models tested in these studies underline that NIRS can be a good predictor of GxE, and in many different ways. These results are very promising to enhance predictive ability in the context of multi-environment trials.

4.5. Making use of historical NIRS data in prediction

In GS the enrichment of the training set with historical data from multiple environments can improve predictive ability by increasing the size of the training set and by limiting the effect of atypical years that are difficult to predict with reduced and traditional datasets (Sweeney *et al.* 2019). The use of historical data might be more complicated with PS, as NIR spectra are likely to be more specific to the environment in which they are measured. To estimate the NIR similarity matrix (H) between historical varieties characterized with NIRS over successive years (the varieties changing from one year to another), the effect of environment and GxE should be accounted for and corrected from the spectra. One option would be to use check varieties for which NIR spectra would be collected each year to determine the transformation from the spectra collected in one environment to the spectra collected in another environment.

4.6. The case of perennials

Genetic resources for perennials are usually conserved as collections in nurseries. These collections typically include individuals from different species, populations, genotypes as well as clones. In this case, we foresee several potential applications of PS. First, as already mentioned in the case of genebanks, NIRS could be used to rapidly screen these resources for target traits that are typically difficult to evaluate on a large number of individuals, such as fruit quality or phenology. Second, in a more advanced breeding context, one could imagine that progenies from controlled crosses could be planted near a well characterized reference panel of widely cultivated varieties. NIRS data could be collected at the same time on both reference and candidate individuals to limit environmental heterogeneity between NIRS matrices from training and test sets, and a model could be trained by using phenotypic data potentially obtained in other sites for

the reference panel in order to predict the performance of the candidates in these particular sites. This scenario corresponds to GLOB prediction, which proved to be quite efficient for some traits in wheat and poplar in (Rincent *et al.* 2018).

4.7. Other applications

4.7.1. Genotype inference

Whalen *et al.* (2020) described a new application of NIRS for breeding beyond PS: the use of High-Throughput Phenotypes (HTP) such as NIRS to infer genotypes. They illustrated their approach with simulations, in which spectra resulted from an additive genetic model, with 100 QTLs per chromosome and heritability of wavelength ranging between 0.1 and 0.7. From that, they fit a model which links the segregation states at each locus with the HTP, for the training population. Then, this model was applied to predict the segregation states of non-genotyped varieties. They concluded that under certain conditions it was possible to infer the genotypes of individuals derived from biparental crosses. The HTP-enabled genomic prediction (with genotypes inferred from HTP data) yielded higher accuracies than PS, the best accuracies being from classic genomic prediction models with real genotypes. However, it is important to note that PS does not only capture additive genetic effects, but also epistatic and Gx \times E. This advantage of PS over GS could not be illustrated here, as only additive effects were simulated. Nevertheless, this application seems very promising for breeding, especially for species with high genotyping cost. Furthermore, as PS relies on global relatedness between genotypes derived from NIRS, it could be difficult to predict traits with mono- or oligogenic genetic architecture. Hence this transition proposed by Whalen *et al.* (2020) from NIRS to marker data allows to apply GS but also QTL detection. This shows that the use of NIRS instead of genotyping in breeding may be adapted to contrasted genetic architectures. A demonstration of its usefulness with real data remains to be conducted.

4.7.2. Hybrid prediction

As NIR spectrum is a phenotype, it captures both additive and interaction effects. For this reason, it is possible that the NIRS covariance matrix could be used to predict hybrid performances, considering both general and specific combining abilities. The idea would be to collect NIR spectra on the hybrids to estimate a covariance matrix taking interaction effects into account. One option would be to collect NIRS data on large collections of hybrids in nurseries,

and phenotype only part of the hybrids in classical multi-environment trials, possibly with the sparse testing approach described above.

4.7.3. *Progeny sorting*

In some programs, breeders are interested in quickly characterizing progenies from controlled crosses with respect to their resemblance to their parents. This is typically the case when crossing an established widely-used variety with a donor genotype with the aim to introgress particular features from the donor to the variety. An example of this would be the case of grapevine for which a current challenge is to quickly breed varieties with resistance to biotic factors while maintaining some established quality for wine making. In this case when a given variety is crossed with a disease resistant genotype, it is usually quite straightforward using marker assisted selection to select with molecular markers the progenies that carry disease resistant genes, but for wine quality it is much more complicated. The goal would be to select among the resistant progenies those that are more similar to the parental wine-making variety, which using molecular markers is a very difficult task without considering prior knowledge on the genetic architecture of the traits. In this particular case NIRS could be useful to provide a distance between the resistant progenies and their parents.

5. Conclusion

We have reviewed the different approaches that have so far been proposed in literature to predict agronomic traits with NIRS, from prediction at the plot/individual level to PS and GLOB selection. NIRS has been intensively used to make predictions at the plot or genotype level, considering NIRS as a secondary trait or as a yield proxy. The originality of PS and GLOB selection as defined in Rincent *et al.* (2018) is that NIR spectra are used in a similar way as molecular markers in GS. They indeed supposed that NIR spectra were able to capture genetic relationships between individuals. PS resulted in good predictive abilities (often similar or even higher than those obtained with GS), even when the predicted trait was completely independent from the tissue analyzed with NIRS (e.g. different environments). The high-throughput and low-cost of this approach makes it interesting to increase breeding efficiency in comparison to GS, particularly for species for which genotyping is expensive, or for crops for which NIRS data are already routinely collected in the breeding programs. We have also proposed different promising applications of PS in breeding and prebreeding among which some can readily be applicable, while others require further work in order to test and optimize this approach.

Funding

The preparation of this chapter was partly supported by the project OASIs funded by the French Ministry of Agriculture and Food (“fond CASDAR, appel à projets Semences et sélection végétale 2020, projet n° C-2020-5”) and by the BreedWheat project funded by the French National Research Agency under Investment for the Future (ANR-10-BTBR-03).

Partial funding of Charlotte Brault’s PhD was provided by the Association Nationale de la Recherche et de la Technologie (ANRT, grant number 2018/0577), IFV and Inter-Rhône. Pauline Robert’s PhD was jointly funded by Agri-Obtentions, Florimond-Desprez, and the Association Nationale de la Recherche et de la Technologie (ANRT, grant number 2019/0060).

Aknowledgements

We would like to thank the reviewer and editors for their suggestions and useful remarks on a previous version of this chapter. We also thank Jacques Le Gouis for proposing the “GLOB” denomination.

Chapitre 3

**Identification et optimisation des facteurs influençant les
capacités de prédictions phénomiques**

Préambule

Dans ce premier chapitre d'analyse, nous allons évaluer les prédictions phénotypiques sur des données issues des programmes de sélection du blé tendre de Florimond Desprez, Agri-Obtentions et INRAE. Nous les comparerons aux prédictions génomiques, et aux prédictions de modèles qui combinent à la fois les marqueurs moléculaires et les spectres NIR. Pour faciliter cette première évaluation de la sélection phénotypique, nous conduirons les prédictions dans un seul environnement à la fois.

Depuis le développement de la sélection génomique, de nombreuses publications ont étudié les différents facteurs qui peuvent influencer les capacités de prédiction de la méthode. Ce chapitre s'applique aussi à évaluer différents facteurs qui peuvent influencer les capacités de la prédiction phénotypique.

Ce chapitre est présenté sous la forme d'un article scientifique qui a été publié dans la revue *Theoretical and Applied Genetics* en janvier 2022.

Phenomic selection in wheat breeding: identification and optimisation of factors influencing prediction accuracy and comparison to genomic selection

P. Robert^{1,2,3,4}, J. Auzanneau³, E. Goudemand⁴, F.X. Oury², B. Rolland⁵, E. Heumez⁶, S. Bouchet², J. Le Gouis², R. Rincent^{1,2, #}

¹Université Paris-Saclay, INRAE, CNRS, AgroParisTech, GQE - Le Moulon, 91190, Gif-sur-Yvette, France

²INRAE - Université Clermont-Auvergne, UMR1095, GDEC, 5 chemin de Beaulieu, 63000 Clermont-Ferrand, France

³Agri-Obtentions, Ferme de Gauvilliers, 78660, Orsonville, France

⁴Florimond-Desprez Veuve & Fils SAS, 3 rue Florimond-Desprez, BP 41, 59242, Cappelle-en-Pévèle, France

⁵INRAE–Agrocampus Ouest-Université Rennes 1, UMR1349, IGEPP, Domaine de la Motte, 35653 Le Rheu, France

⁶INRAE, UE 972, Grandes Cultures Innovation Environnement, 2 Chaussée Brunehaut, 80200 Estrées-Mons, France

Corresponding authors: renaud.rincent@inrae.fr

ORCID: P. Robert: 0000-0002-6314-180X / J. Le Gouis: 0000-0001-5726-4902 / S. Bouchet: 0000-0001-5868-3359 / R. Rincent: 0000-0003-0885-0969

1. Abstract

Key message

Phenomic selection is a promising alternative or complement to genomic selection in wheat breeding. Models combining spectra from different environments maximise the predictive ability of grain yield and heading date of wheat breeding lines.

Phenomic selection (PS) is a recent breeding approach similar to genomic selection (GS) except that genotyping is replaced by near infrared (NIR) spectroscopy. PS can potentially account for non-additive effects and has the major advantage of being low cost and high throughput. Factors influencing GS predictive abilities have been intensively studied, but little is known about PS. We tested and compared the abilities of PS and GS to predict grain yield and heading date from several datasets of bread wheat lines corresponding to the first or second years of trial evaluation from two breeding companies and one research institute in France. We evaluated several factors affecting PS predictive abilities including the possibility of combining spectra collected in different environments. A simple H-BLUP model predicted both traits with prediction ability from 0.26 to 0.62, and with an efficient computation time. Our results showed that the environments in which lines are grown had a crucial impact on predictive ability based on the

spectra acquired and was specific to the trait considered. Models combining NIR spectra from different environments were the best PS models and were at least as accurate as GS in most of the datasets. Furthermore, a GH-BLUP model combining genotyping and NIR spectra was the best model of all (prediction ability from 0.31 to 0.73). We demonstrated also that as for GS, the size and the composition of the training set has a crucial impact on predictive ability. PS could therefore replace or complement GS for efficient wheat breeding programs.

Key words

Bread wheat, Genomic selection (GS), Genomic-like omics-based (GLOB) prediction, Near infrared spectroscopy (NIRS), Phenomic selection (PS), Plant breeding.

Declarations

Funding: This work was funded by Agri-Obtentions, Florimond-Desprez, and the Association Nationale de la Recherche et de la Technologie (ANRT, grant number 2019/0060)

Conflicts of interest/Competing interests: On behalf of all authors, the corresponding author states that there is no conflict of interest.

Availability of data and material: The datasets generated during and/or analysed during the current study are not publicly available due breeding programs privacy but are available from the corresponding author on reasonable request.

Code availability: Code used to lead the analysis of this study is available from the corresponding author on request.

Author contribution statement: JA, FXO, BR and EH designed the field trials and collected the phenotypic data from Agri-Obtentions and INRAE. EGD provided the phenotypic data and genotyping data from Florimond Desprez company. SB provided the genotyping data from Agri-Obtentions and INRAE and participate in discussions of this study. RR initiated the project and with JLG supervised the study and helped improving the manuscript. PR analysed the data and wrote the manuscript. All authors approved the final manuscript.

2. Introduction

In plant breeding, phenotyping is a key element for breeders to select candidates. The creation of new varieties that outperform the control varieties relies on the ability of breeders to evaluate numerous candidates to maximise the chance of success. It is particularly difficult for complex traits interacting with the environmental conditions, as in this case the best variety is not the same from one environment to another (Allard and Bradshaw 1964). Phenotyping is costly, time consuming and requires a large amount of seed, so the number of candidates phenotyped is one of the main limiting factors. Prediction approaches such as genomic selection (GS) have been proposed in the last two decades to overcome this constraint. The aim of GS is to predict the performances of un-phenotyped candidates based on molecular markers (Whittaker *et al.* 2000; Meuwissen *et al.* 2001). In GS, a genotyped and phenotyped training set of varieties is used to train a statistical model. This model can then be applied to predict the genomic estimated breeding values of a predicted set using genotyping information alone. GS can improve genetic gain by increasing the number of genotypes under selection, which increases selection intensity, or by reducing the duration of the breeding cycles. The efficiency of GS depends on the prediction accuracy of the model, which can be influenced by multiple factors: the genetic architecture of the trait (Daetwyler *et al.* 2010) and its heritability (Hayes *et al.* 2009b), linkage disequilibrium (Zhong *et al.* 2009), marker density (Solberg *et al.* 2008; de los Campos *et al.* 2013; Hickey *et al.* 2014), the size and composition of the training set (Albrecht *et al.* 2011; Heffner *et al.* 2011; Pszczola *et al.* 2012; Rincent *et al.* 2012; Daetwyler *et al.* 2013), and the statistical model (Charmet *et al.* 2020). The efficiency of GS has been evaluated in breeding (Hayes *et al.* 2009a; Jannink *et al.* 2010; Crossa *et al.* 2010, 2017) and pre-breeding programs (Gorjanc *et al.* 2016; Yu *et al.* 2016) with promising results, but implementation of the approach remains difficult for some species due to the expense of genotyping tools. Small breeding companies may not have the budget for genotyping thousands of candidates each year (R2D2 Consortium *et al.* 2021). Furthermore, in GS, the biological processes occurring between the genotype level and the phenotype level are ignored. One can then wonder if other types of information could better link genotypes to phenotypes considering non-additive effect for example.

Endophenotype refers to an intermediate trait level between the genome and the phenotype, resulting from the association of a genotype, an environment and the interaction between them (Fernandez *et al.* 2016). Some studies have shown that molecular markers can be replaced by endophenotypes such as transcripts (Frisch *et al.* 2010; Fu *et al.* 2012; Zenke-Philippi *et al.* 2017; Azodi *et al.* 2020), metabolites (Riedelsheimer *et al.* 2012; Xu *et al.* 2016), or both (Guo *et al.* 2016; Westhues *et al.* 2017; Schrag *et al.* 2018), or small RNA (Schrag *et al.* 2018; Seifert *et al.* 2018) when predicting complex traits in maize hybrids. By depicting interactions between genes and the environment and the plant regulatory system, these types of omics data are thought to capture non-additive relationships that are difficult to model with molecular markers alone, and the resulting predictions are often at least as accurate. The omics data are usually collected at an early stage of plant development from different tissues (leaves, roots) of genotypes grown in controlled conditions. These (potentially controlled) conditions of omics acquisition may be radically different from the field trial environment in which the training set is phenotyped. This genomic-like omics-based (GLOB) prediction approach is however effective in capturing genetic similarity potentially related to both additive and non-additive effects.

The routine acquisition of endophenotypic data is expensive, so it is still challenging to implement them in breeding. Rincent *et al.* (2018) proposed a low-cost and efficient alternative, called phenomic selection (PS), in which genotyping or other omics are replaced by near infrared spectra. Near infrared spectroscopy (NIRS) is a non-destructive method to measure the reflectance or absorbance of a biological sample at different wavelengths in the visible and near-infrared spectrum. According to the Beer-Lambert law, reflectance and absorbance are directly linked to the molecular composition of the sample, which for plants is influenced by genetics, environmental factors and their interaction. NIR spectra are already routinely collected in cereal breeding, for instance, to predict grain composition and quality (Osborne 2006). The accuracies of PS and GS in predicting different traits in bread wheat and poplar populations were compared in two scenarios (Rincent *et al.* 2018). In the first scenario, predictions were made within one environment, that is, the plants or plots that were predicted are those from which the NIR spectra were acquired. The second scenario was a GLOB approach in which spectra were used instead of endophenotypes, so spectra were collected in at least one environment that may have differed from the environments in which the training set was phenotyped. Spectra were used to estimate genetic similarities between varieties. PS outperformed GS in predicting wheat grain yield (GY) in both scenarios. A simple mixed model was used in which the similarity matrix was computed with molecular marker

information for GS (G-BLUP) and with NIR spectra for PS (H-BLUP). As molecular markers are considered as quantitative variables in GS (with regression according to allele counts), most GS models including RR-BLUP, G-BLUP, BayesA, B, C and $C\pi$ can be used in PS by replacing genotyping data by NIRS measurements. More specifically for PS, other types of models, such as functional regression, can be considered. NIR spectra can be modelled as a curve in a continuous domain and so considered as a function (Montesinos-López *et al.* 2017a; Lane *et al.* 2020). In this case the wavelengths are not used to compute a kernel but as covariates in the functional model. Functional regression reduces dimensionality by transforming wavelengths into functional covariates. This kind of model is thus able to handle high-dimensional data, but is more difficult to implement than H-BLUP due to parametrisation choices needed to approximate the functions.

There are only a few studies of PS so far. In some PS studies, spectra were acquired in the same environment as the predicted trait (as in scenario 1 above). For the best statistical model based on scenario 1, PS results were promising with GY prediction accuracies ranging from 0.37 to 0.51 in wheat (Rincent *et al.* 2018; Krause *et al.* 2019; Cuevas *et al.* 2019). By comparison, a GLOB approach produced wheat GY prediction accuracies ranging from 0.28 to 0.53 (Rincent *et al.* 2018). Galán *et al.* (2020) adjusted mean spectra for each genotype by correcting environment and spatial effects and predicted dry matter yield in winter rye with prediction accuracies ranging from 0.48 to 0.59. Lane *et al.* (2020) also applied a GLOB prediction approach in which a set of spectra from one year was used to predict maize GY for another year using partial least square regression. The resulting prediction accuracies ranged from 0.19 to 0.69. They also averaged the relationship matrices computed on NIR spectra from several environments to predict maize grain yield. Predictive accuracies of H-BLUP model ranged 0.07 to 0.70 for all environments and two cross-validation scenarios.

As for GS, some factors have already been identified as impacting PS accuracy. Specifically, PS predictive ability (PA) using NIRS was strongly influenced by the trait and tissue considered, including the environment in which plants were grown (Rincent *et al.* 2018). Drought environments appeared to be more useful than irrigated environments, for example. Galán *et al.* (2020) compared different training set sizes to make predictions with H-BLUP and G-BLUP and found that prediction abilities increased the larger the training set. Cuevas *et al.* (2019) pre-processed NIR spectra with different mathematical transformations and found the predictive abilities varied according to the predictive model used.

Here we investigated the ability of PS to predict the traits of GY and heading date (HD) in elite breeding lines in winter wheat produced by two breeding companies (Agri-Obtentions and Florimond-Desprez) and one research institute (INRAE) in France. The accuracy of any predictive model used in breeding will have an impact on genetic gain and selection efficiency, so we explored which factors alter the predictive abilities of PS. The objectives were to investigate how the predictive ability of statistical models was affected by the assumptions made, or by how the NIR spectra were collected (environment or generation of plant sample) and used (singly or in combination). The second objective was to compare the predictive abilities of PS, GS, and models combining variables related to both NIR spectra and molecular markers. The effects of the size and the composition of the training set on the GS and PS predictive abilities were also assessed.

3. Materials and Methods

3.1. Phenotyping, genotyping and NIRS data

3.1.1. Genetic material and phenotypic data

We used four different datasets corresponding to the first or second year of trial evaluation of winter bread wheat candidate breeding lines. These lines (selfed and double haploids) were developed in the breeding programs of Florimond Desprez (France), and Agri-Obtentions (France) together with INRAE (France). For the selfed lines, the first year of plot trial evaluation corresponds mostly to the F₅ generation and the second year of trial evaluation to the F₆ generation. Data were grouped according to breeding program and according to the year of trial evaluation. For each dataset, the year, site, trial treatments, the number of lines, the number of lines genotyped and the traits evaluated are listed (Table 3-1). For a given breeder, the different datasets are more or less connected, as the same lines could be evaluated in successive years (see Table S1).

Set1 corresponds to the first year of trial evaluation of breeding lines developed by Florimond Desprez between 2016 and 2018. Data were collected from six environments (two sites in three years). Grain yield (GY) and heading date (HD) were evaluated for each breeding line in an augmented design with a control repeated every five microplots. NIR spectra are available for all lines in all the environments between 2016-2018 and also for the preceding year when breeding lines were in nursery, grown on three rows, for the 2017-2018 years. Genotyping of selfed lines occurred at the F₆ generation.

Set2 corresponds to the first year of trial evaluation of breeding lines developed by Agri-Obtentions and INRAE in 2018 and 2019. It is composed of eight environments, involving four sites and two years. EM site is a reference site for evaluating all the lines that were evaluated in the same year at the three other sites. GY was evaluated in an augmented design in the reference site and in randomized complete blocks for the other sites. HD data were missing for some of the environments, so to simplify the analysis, HD was not predicted for Set2. NIR spectra are available for all the genotypes and all the repetitions in all the environments. Genotyping of selfed lines occurred at the F₅ generation.

Set3 corresponds to the second year of trial evaluation of breeding lines developed by Florimond Desprez between 2016 and 2018. It is composed of nine environments, involving four sites and three years. GY and HD were evaluated in lattice or randomised block designs. NIR spectra are available for all the lines in all the environments, and also in the preceding year when breeding lines were in nursery, grown on three rows, for 2017-2018 years. Genotyping of selfed lines occurred at the F₆ generation.

Set4 corresponds to the second year of trial evaluation of breeding lines developed by Agri-Obtentions and INRAE in 2018 and 2019. This dataset is composed of seven environments, involving two sites, two treatments (intensive practices versus low input) and two years. One environment was removed from the analysis due to mis-association between genotypes and phenotypes in the data that could not be corrected. GY and HD were evaluated in each environment in three complete randomised blocks. NIR spectra are available for all the lines in all the environments and also in the preceding year when breeding lines were in trial microplots, for 2019 year. Genotyping of selfed lines occurred at the F₅ generation.

GY was adjusted to a humidity rate of 15%, and HD was scored on each plot as the date when 50% of the ears were half emerged. All the trial details can be found in Supplemental data Tables S2 & S3.

3.1.2. Genotypic data

Lines in Set1 and Set3 were genotyped with the 35K breeder Bristol array (Axiom™ Wheat Breeder's Genotyping Array). Lines in Set2 and Set4 were genotyped with the 35K BreedWheat array (Axiom™ BreedWheat Genotyping Array) which is a subset of the 280K SNP array (Rimbert *et al.* 2018). This subset corresponds to a selection of markers minimizing linkage

disequilibrium between them (Ben-Sadoun *et al.* 2020). Molecular markers were eliminated if the minor allele frequency was less than 5%, or if the heterozygosity rate or missing value rate was greater than 5%. After filtering for high-resolution SNPs, we obtained 5,824 SNPs in both Set1 and Set3, 12,303 SNPs in Set4-2018, and 19,512 SNPs for both Set2 and Set4-2019. On average over the sets, a proportion of 1.1% of missing values were imputed with the average allele frequency of the corresponding marker. Minor allele frequency, missing value and mean imputation were computed with the sommer R package (Covarrubias-Pazaran 2016).

3.1.3. NIRS data

For NIRS acquisition, all flour and grain samples were stored in dry conditions at ambient temperature. NIRS of Set1 and Set3 were collected with the NIRS 6500 FOSS spectrometer (FOSS NIR Systems, Silver Spring, MD, USA) over the range 400 to 2500 nm in steps of 2 nm. All the NIRS were collected on 10 g of wheat flour except for Set1-2019 where 40 g of grain was used. Final spectra are the average of 32 repetitions measured by the spectrometer. NIRS of Set2 were collected on 150 g of grain for each genotype with the XDS NIR Analysers FOSS spectrometer (FOSS NIR Systems, Silver Spring, MD, USA) in the range 400 to 2500 nm in steps of 2 nm. Final spectra are the average of 16 repetitions measured by the spectrometer. For these three sets, NIRS were exported with ISIScan™ and WINisi™ 4.20 (Infrasoft International). NIRS of Set4 were collected on 350 g of grain for each genotype with the MPA II FT-NIR analyser (Bruker Optics, Ettlingen, Germany) ranging from 3594.92 cm⁻¹ to 12489.60 cm⁻¹ in steps of 7.7cm⁻¹. Final spectra are the average of 64 repetitions measured by the spectrometer. NIRS were exported with the OPUS LAB software provided by Bruker. To harmonise the data, NIRS data of Set4 were converted into nm in steps of 2 nm, so the final spectra ranged from 802 to 2492 nm. Spectra were visualised to filter out any spectra with abnormal absorbances resulting from technical errors. Where possible, we computed adjusted means of NIR absorbance for each wavelength when NIRS were measured on each microplot of the experimental design using the corrective model applied for GY and HD (Table 3-2). It was necessary to transform raw spectra because of inherent noise in the absorbance measurement. We first applied the standard normal variate transformation, which scales and centres each spectrum. Then we applied the Savitzky-Golay filter (Savitzky and Golay 1964) to compute the derivative with a window length of 61, using the R package signal (Signal developers 2013). All the NIRS were processed in R software (R Core Team 2019).

Table 3-1. Summary of the four datasets resulting from the trial evaluations. All lines evaluated were either double haploid or self-pollinated lines. T, treated (equivalent to intensive practices). LI, low input. NT, not treated (equivalent to intensive practices without fungicide treatment). Summary statistics were calculated on the adjusted means (models M1-M3 in Table 3-2). GY, grain yield in t/ha. HD, heading date in number of days after the 1st of January. μ , mean. σ_{ϵ} , residual standard deviation. CV, coefficient of variation (%) as $CV = \frac{\sigma_{\epsilon}}{\mu} \times 100$. See Supplementary Table S2 for the description of the sites.

Panel	Year	Site	Treatment	Number of lines	Lines genotyped	Trait						
						GY			HD			
						μ	σ_{ϵ}	CV	μ	σ_{ϵ}	CV	
Set1	2016	HV	T	234	199	7.7	0.4	5.2	143	0.7	0.5	
		LC		269	209	8.5	0.8	9.4	121	1.4	1.2	
	2017	HV	T	280	157	10.7	0.3	2.8	140	0.8	0.6	
		LC		184	90	7.9	0.5	6.3	124	1.2	1.0	
	2018	HV	T	125	-	10.4	0.4	3.8	137	0.4	0.3	
		LT		113	-	8.2	0.5	6.1	125	1.3	1.0	
Set2	2018	CF	LI	328	-	7.5	0.3	4.0	-	-	-	
		GL	T	566	-	9.5	0.3	3.2	-	-	-	
		RE	LI	131	-	6.1	0.3	4.9	-	-	-	
		EM	LI/NT	1351	-	8.3	0.5	6.0	-	-	-	
	2019	CF	LI	264	-	5.3	0.4	7.5	-	-	-	
		GL	T	433	325	10.5	0.4	3.8	-	-	-	
		RE	LI	113	-	7.2	0.4	5.6	-	-	-	
		EM	LI/NT	1418	-	8.2	0.5	6.2	-	-	-	
Set3	2016	CP	T	95	-	9.0	0.3	3.3	145	0.9	0.6	
		HV		103	-	7.7	0.4	5.2	140	0.7	0.5	
		LC		141	-	9.1	0.5	5.5	120	4	3.3	
	2017	CP	T	77	49	12.6	0.3	2.4	144	0.7	0.5	
		HV		82	49	10.9	0.2	1.8	140	0.6	0.4	
		LC		94	57	7.8	0.4	5.1	122	1.2	1.0	
		VB		104	57	8.8	0.2	2.3	133	0.8	0.6	
	2018	CP	T	101	69	11.0	0.2	1.8	145	2.2	1.5	
		HV		102	70	10.6	0.3	2.8	136	3.1	2.3	
	Set4	2018	ML	T	-	-	9.7	0.5	5.2	141	0.2	0.1
EM			LI		82	71	8.0	0.4	5.0	139	0.1	0.1
ML					-	-	5.8	0.5	8.6	141	0.1	0.1
2019		EM	T	111	100	10.4	0.3	2.9	144	0.1	0.1	
		LM		112	101	11.5	0.4	3.5	145	0.5	0.3	
		EM	LI	112	101	8.2	0.4	4.9	144	0.1	0.1	
		LM		112	101	10.4	0.4	3.8	145	0.2	0.1	

3.2. Statistical models for adjusting data and estimating heritability

Different statistical models adapted to each experimental design were used to correct traits for environmental effects and to obtain the adjusted mean and broad-sense heritability of GY and HD (Table 3-2). In environments with an augmented design, a two-dimensional P-spline mixed model was used to compute adjusted means, simultaneously accounting for global and local spatial trends across row and column effects (Model M1). In environments with a randomised complete block, global spatial effects were corrected for block and sub-block effects using the checks and replicated lines (M2). For the other environments with a randomised complete block with row and column information, a two-dimensional P-spline mixed model including block and sub-block effects was used to adjust for global and local trends (M3). Models selected to adjust traits are described in Table S4.

Table 3-2. Statistical models for computing adjusted means for GY, HD, and absorbance at each wavelength. y is a vector of the trait observations. β is a vector of the intercept and genotype fixed effect, X is the corresponding design matrix for the fixed effect. $f(\mathbf{r}, \mathbf{c})$ stands for 2D P-splines function, \mathbf{r} is the vector of row and \mathbf{c} is the vector of column. \mathbf{u} comprises the sub vectors of random row and column effects which are mutually independent with $\mathbf{Z}_u = [\mathbf{Z}_r | \mathbf{Z}_c]$ the corresponding design matrix and the covariance matrix $\mathbf{U} = \text{diag}(\mathbf{I}_r \sigma_r^2, \mathbf{I}_c \sigma_c^2)$. \mathbf{B} is the vector of random block effects which are assumed independent i.e., $\mathbf{B} \sim N(0, I \sigma_B^2)$ and \mathbf{Z}_j the corresponding design matrix. \mathbf{SB} the vector of random sub-blocks (in corresponding blocks) effects which are assumed independent i.e., $\mathbf{SB} \sim N(0, I \sigma_{sb}^2)$ and \mathbf{Z}_k the corresponding design matrix. Finally, ϵ is the vector of independent residuals with distribution $\epsilon \sim N(0, I \sigma_\epsilon^2)$.

Model	Equation
M1	$y = X\beta + f(\mathbf{r}, \mathbf{c}) + \mathbf{Z}_u \mathbf{u} + \epsilon$
M2	$y = X\beta + \mathbf{Z}_j \mathbf{B} + \mathbf{Z}_k \mathbf{SB} + \epsilon$
M3	$y = X\beta + \mathbf{Z}_j \mathbf{B} + \mathbf{Z}_k \mathbf{SB} + f(\mathbf{r}, \mathbf{c}) + \mathbf{Z}_u \mathbf{u} + \epsilon$

After correcting for spatial effects for the environments with randomised blocks with replicates, heritability within each environment was calculated for GY and HD. The following mixed model (Equation 1) was fitted to the corrected phenotypes in order to estimate the genetic (σ_G^2) and error (σ_ϵ^2) variances. The controls were excluded from this step to focus on the variance of the breeding lines.

$$\hat{Y}_{ik} = \mu + G_i + \epsilon_{ik} \quad (1)$$

where \hat{Y}_{ik} is the trait corrected for spatial effects for genotype i and rep k , G_i the random genetic effect for genotype i , and ϵ_{ik} the residual effect. We assume that G_i and ϵ_{ik} are independent and identically distributed (i.i.d) and follow a normal distribution $G_i \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma_G^2)$ and $\epsilon_{ik} \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma_\epsilon^2)$, respectively. The different random effects of the model are assumed to be independent.

Broad-sense heritability (H^2) was estimated as follows:

$$H^2 = \frac{\hat{\sigma}_G^2}{\hat{\sigma}_G^2 + \frac{\hat{\sigma}_\epsilon^2}{Nrep}} \quad (2)$$

where $\hat{\sigma}_G^2$ and $\hat{\sigma}_\epsilon^2$ are the Restricted Maximum Likelihood of the genetic and residual variances derived by fitting Equation 1, respectively, and $Nrep$ is the average number of replicates in the corresponding environment. Details of heritability and variances are given in Table S4. Models M1 and M3 were fitted with the SpATS R package (Rodríguez-Álvarez *et al.* 2018) and Model M2 and Equation 1 were fitted with the breedR package (Facundo Muñoz and Leopoldo Sanchez 2020).

From now on, we therefore consider that Y_i stands for the adjusted mean of the considered trait for the variety i .

3.3. Calculation of the genetic and hyperspectral relationship matrices

Relationship between individuals was estimated with data from molecular markers (kinship matrix K) and from NIRS (hyperspectral relationship matrix H). For markers, the relationship matrix was estimated following the Endelman and Jannink (2012) equation:

$$K = \frac{AA'}{2 \sum p_k(1 - p_k)}$$

where A is a centred matrix with dimensions $N \times M$, N is the number of genotypes and M the number of molecular markers. For the i^{th} individual and the k^{th} marker, $A_{ik} = X_{ik} + 1 - 2p_k$ with X the genotype matrix, coded in $\{-1,0,1\}$ and p_k the frequency of allele 1 at marker k . K was computed with the rrBLUP R package (Endelman 2011).

For NIRS, the relationship matrix also called hyperspectral relationship matrix in the literature, was computed as:

$$H = \frac{Sp^*Sp^{*'}}{L}$$

where Sp^* is the centred and scaled matrix of NIRS (dimension $N \times L$), and L the number of wavelengths.

K and H were both scaled to have a sample variance of 1 to avoid biased parameter estimations due to different scalings (Kang *et al.* 2010; Forni *et al.* 2011).

3.4. Reference prediction models and prediction scenario

For genomic prediction, we applied two reference models (G-BLUP, GS-LASSO), suited to contrasted genetic architectures, to predict GY and HD:

$$\text{G-BLUP: } Y_i = \mu + G_i + \epsilon_i$$

$$\text{GS-LASSO: } Y_i = \mu + \sum_{k=1}^p x_{ik} \beta_k + \epsilon_i \text{ where}$$

$$\hat{\beta}(\lambda) = \text{Argmin}_{\beta \in \mathbb{R}^p} \{ \|Y - X\beta\|^2 + \lambda \|\beta\| \} \text{ with } \|\beta\| = \sum_{k=1}^p |\beta_k| \text{ and } \lambda > 0.$$

For G-BLUP, Y_i is the phenotype for variety i , μ is the intercept, G_i is a random polygenic effect for the variety i and follows a normal distribution $G \sim N(0, K\sigma_G^2)$. K is the relationship matrix calculated with molecular markers (see above). ϵ_i is the random residual effect which follows $\epsilon_i \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma_\epsilon^2)$. The different random effects of the model are assumed to be independent. For LASSO, Y_i is the phenotype for variety i , μ is the intercept, x_{ik} is the allele of marker k for the variety i and β_k the effect of the marker k on the trait. ϵ_i is the random residual effect which follows $\epsilon_i \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma_\epsilon^2)$.

These two models were adapted for phenomic selection. H-BLUP is an adaptation of G-BLUP with H instead of K as covariance matrix, for the random genetic effect W_i of the variety i . W_i follows a normal distribution $W \sim N(0, H\sigma_w^2)$. ϵ_i is the random residual effect which follows $\epsilon_i \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma_\epsilon^2)$.

The different random effects of the model are assumed to be independent. PS-LASSO is an adaptation of GS-LASSO using each wavelength as predictive variable s_{ik} , corresponding to the adjusted, transformed, scaled and centred absorbance of wavelength k for the variety i , and β_k as the effect of the wavelength k on the trait. ϵ_i is the random residual effect which follows $\epsilon_i \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma_\epsilon^2)$.

$$\mathbf{H-BLUP}: Y_i = \mu + W_i + \epsilon_i$$

$$\mathbf{PS-LASSO}: Y_i = \mu + \sum_{k=1}^p s_{ik} \beta_k + \epsilon_i \text{ where}$$

$$\hat{\beta}(\lambda) = \text{Argmin}_{\beta \in \mathbb{R}^p} \{ \|Y - S\beta\|^2 + \lambda \|\beta\| \} \text{ with } \|\beta\| = \sum_{k=1}^p |\beta_k| \text{ and } \lambda > 0$$

G-BLUP and H-BLUP models were run with the R package rrBLUP and the GS-LASSO and PS-LASSO models were run with the glmnet R package (Friedman *et al.* 2010)

To fairly compare all the models, we computed the predictions within each environment in a 5-fold cross-validation scheme, which consisted in randomly splitting the data in five folds of the same size. The training set was constituted of four folds and the validation set of the remaining fold. For all the models, predictive ability (PA) was computed as the Pearson's correlation between the predicted values and the adjusted means of the validation set. In total, five PA were calculated when the five folds were used once each as the validation set. This procedure was repeated 25 times to get robust results. The folds and repetitions were exactly the same for all the compared models. Comparison of models was then based on the mean and quantiles of the 125 resulting PA obtained in each environment.

3.5. Evaluation of factors influencing predictive ability of PS

3.5.1. Impact of the prediction models on predictive ability

We considered two types of penalised models with assumptions corresponding to contrasted genetic architectures (G-BLUP/H-BLUP and GS-LASSO/PS-LASSO). In this analysis, the NIR spectra used in each model were collected in the same environment as the predicted traits. The four datasets were analysed in this way to gain an overview of PS in different breeding materials.

3.5.2. Impact of NIR spectrum origin of acquisition (environment and generation)

In practice, a genotype may be characterized by multiple spectra acquired from plant material grown in different trials, nurseries or years. For example, a candidate line is generally observed over several years and sites to be characterised on different traits (e.g. disease resistances, height, precocity, grain yield, protein quality). The line is first phenotyped in nursery rows for some traits (phenology, height, disease resistance). The next year, this same line can be evaluated for performance traits in a few plot trials. If the line is selected at these two steps, it is then evaluated an additional year in more environments and with more replicates. We investigated whether the site or the year of NIR spectra acquisition had an impact on PA (Figure 3-1). We compared two types of predictive models involving a relationship matrix H derived from the NIR spectrum acquired in a single environment (Single-NIRS) or from multiple NIR spectra acquired from different sites or years (Multi-NIRS). NIRS acquisition is always made on both the training and validation sets. For the models based on a single NIR spectrum, we compared PA obtained with a NIR spectrum acquired in the same environment as the phenotyping of the training set (S), or with a NIR spectrum acquired in a different site but within the same year than the phenotyping of the training set (D) or with a NIR spectrum acquired in nursery rows or trial microplot in a previous year than the phenotyping of the training set (Y). For selfed lines, the rate of homozygosity may differ from a year to another. In this last case the environmental conditions (year, site) and the level of residual segregation are different for the plants on which NIRS is acquired than on the plants that are predicted.

Models D and Y correspond to a GLOB approach, as the environment in which the training set is phenotyped is not the same as the environment in which the NIRS are acquired.

The next objective was to assess the effect of combining NIR spectra acquired from different environments or different breeding generations to predict GY and HD. To do this, we computed H using a matrix with the NIR spectra added one next to the other from NIR spectrum 1 matrix (N_1) to NIR spectrum p matrix (N_p).

$$S_{cb} = (N_1 \quad \dots \quad N_p)$$

S_{cb} had $i \times \sum_{j=1}^p \lambda_j$ dimensions where λ_j designates the number of wavelengths for the NIR spectrum j , and i the number of genotypes.

We combined NIR spectra acquired in the same site **S** and in different sites **D** than the one in which

the training set was phenotyped (**CbSD**), or combined only those from different sites (**CbD**). Also, we combined the NIR spectrum **S** recorded in the same environment as the predicted trait with the NIR spectrum recorded on the previous year **Y** (**CbSY**). All the NIR spectra available were combined in a final model (**CbSDY**). Another method of combining information from multiple NIR spectra is to average NIR spectra measured at multiple sites for each genotype. Here we averaged the spectra from the same environment as the predicted trait and at least a different one for each genotype (**AvSD**). In this model each genotype is characterised by a single spectrum instead of the multiple spectra collected in the different environments. All the models are detailed in Table 3-3. These models correspond to GLOB approach too, as at least one NIR spectrum used was from another environment than the environment in which the training set was phenotyped.

Finally, we evaluated the PA of multi-NIRS models with an increasing number of combined NIR spectra. We first compared different models **CbSDY** which combined the **S** spectrum together with zero to four extra NIR spectra (**D** or **Y**). To dissociate the effect of including the **S** spectrum from the effect of having different numbers of **D** and **Y** spectra, we then compared different multi-NIRS models (**CbDY** or **CbD**) which combined one to four NIR spectra from **D** or **Y**, excluding the **S** spectrum.

3.5.3. *Impact of the training population size and composition on predictive ability*

In GS the size and composition of the TS have an impact on PA. We characterised this effect by testing six TS sizes (10, 50, 100, 150, 200, 250 genotypes) with a random selection of the genotypes. Furthermore, the composition of a TS can be optimised in order to minimise its size while achieving similar PA. We compared to a random selection of the TS, three algorithms selecting the individuals which would constitute the TS. These algorithms were based on molecular markers or NIR spectra (CDmean, Rincent *et al.* 2012) or NIR spectra only (HG, Honigs *et al.* 1985 and KS, Kennard and Stone 1969). More details about the methods and results can be found in Appendix 1.

3.5.4. *Comparison of models based on data from molecular markers, NIRS, or both*

To compare PS and GS predictive abilities, we compared the CbSD model to the reference model G-BLUP (**M**). Then we investigated models that combine NIR spectra and molecular marker information (GH-BLUP). Two models were developed: **CbSD+M** and **CbSY+M**, which combined the hyperspectral effect of **CbSD/CbSY** with the polygenic effect of **M**. All the models are detailed in Table 3-3. GH-BLUP models were run with R package Sommer.

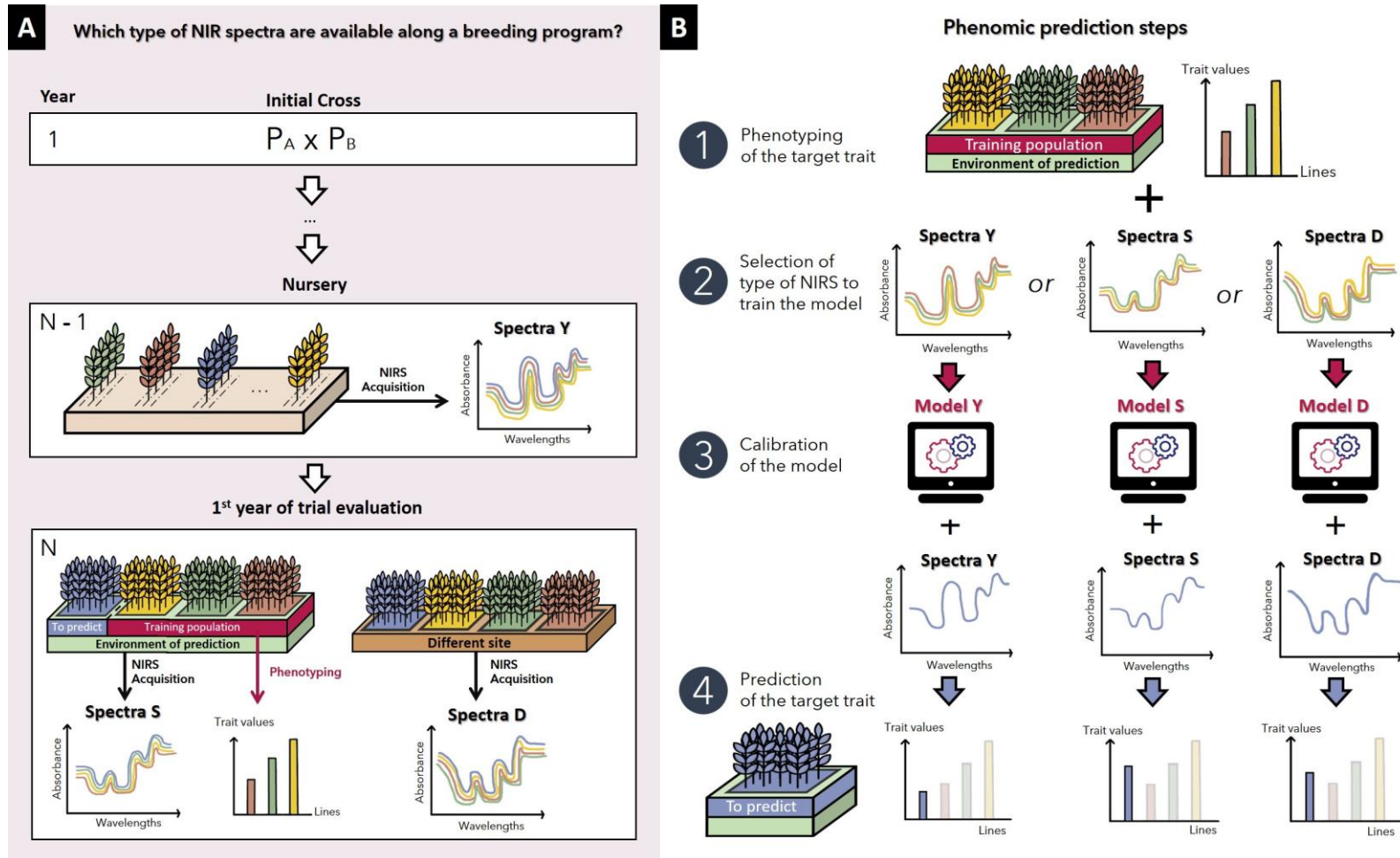


Figure 3-1. Phenomic prediction requires the use of phenotyping data and NIR spectra to predict unphenotyped lines. A wheat breeding line is evaluated for several traits along the breeding program. NIR spectra are acquired at different time in the breeding program like on grains from rows in nursery or from plots in trial to measure the protein content. In this example of a schematic breeding program, NIR spectra are collected at different stages on the same inbreds (at different generations of selfing). Three types of NIR spectra are available: spectra S from the same environment (site x year) than the phenotypic evaluation, spectra D from a different site than the phenotypic evaluation, and finally spectra Y from the previous year than the phenotypic evaluation (a). Three predictive models (S, D, Y) are calibrated with the corresponding NIR spectra (S, D, Y) as predictive variables, and the phenotyping of the training population in the environment of prediction. Finally, each model predicts the value of the target trait in the environment of prediction for the unphenotyped lines (b).

3.5.5. Construction of subsets to fairly compare the different models

Because of differences in datasets (number of environments in which lines were phenotyped, availability of NIRS collected in different sites or years, availability of genotyped lines) the models introduced above could not be applied to all the data. We therefore divided the four sets into four subsets to address each research question. Subset names referred to the type of NIRS available (S, D, Y) and/or to the genotyping availability (M).

SubSet_SD includes lines with NIR spectra acquired in the same site as the phenotyping and also in at least one other site from the same year. Each line is described by J spectra originating from J sites, allowing the evaluation of models S, D, CbSD, CbD, AvSD. **SubSet_SDY** includes lines with NIR spectra acquired in the same site as the phenotyping, and in at least one other site from the same year, as well as in the previous year, allowing the evaluation of models S, D, G, CbSD, CbD, CbSDY, AvSD. **SubSet_SDM** includes lines genotyped and with NIR spectra acquired in the same site as the phenotyping, and also in at least one other site from the same year, allowing to evaluate models M, CbSD, CbSD+M. **SubSet_SYM** includes lines genotyped and with NIR spectra acquired in the same site as the phenotyping and also in the previous year, compatible with evaluating models M, CbSY, CbSY+M. Details about the composition of the subsets are presented in Table S5.

Table 3-3. Description of the GS and PS models. Y_i is the phenotype for variety i . μ is the global mean, G_i is the random polygenic effect for variety i , W_i is the random hyperspectral effect for the variety i , ϵ_i is the residual effect, i.i.d. stands for independent and identically distributed. All the random effects are assumed independent. K is the relationship matrix derived from molecular markers and H the hyperspectral relationship matrix derived from NIRS. **M** is the model based on molecular markers. **S** is the model based on NIR spectrum acquired in the same environment as the phenotyping of the training set. **D** is the model based on NIR spectrum acquired in a different site but within the same year than the phenotyping of the training set. **Y** is the model based on NIR spectrum acquired in nursery rows or trial microplots in a previous year than the phenotyping of the training set. **CbSD** combined NIR spectra acquired in the same site **S** and in different sites **D** than the one in which the training set was phenotyped. **CbD** combined NIR spectra from only different sites. **CbSY** combined the NIR spectrum **S** recorded in the same environment as the predicted trait with the NIR spectrum recorded on the previous year **Y**. **CbDY** combined NIR spectra from different sites and a different year. **CbSDY** combined All the NIR spectra available. **AvSD** averaged NIR spectra from plant material from the same environment and from different sites for each genotype. **CbSD+M** and **CbSY+M** respectively combine models **CbSD** and **CbSY** with model **M**.

Type of model	Variable	Label	Model	Main effect	Residual effect
G-BLUP	SNP	M	$Y_i = \mu + G_i + \epsilon_i$	$G \sim N(0, K_G \sigma_g^2)$	$\epsilon_i \stackrel{i.i.d.}{\sim} N(0, \sigma_\epsilon^2)$
H-BLUP	Single-NIRS	S	$Y_i = \mu + W_i + \epsilon_i$	$W \sim N(0, H_S \sigma_w^2)$	
		D		$W \sim N(0, H_D \sigma_w^2)$	
		Y		$W \sim N(0, H_Y \sigma_w^2)$	
	Multi-NIRS	CbSD		$W \sim N(0, H_{CbSD} \sigma_w^2)$	
		CbD		$W \sim N(0, H_{CbD} \sigma_w^2)$	
		CbSY		$W \sim N(0, H_{CbSY} \sigma_w^2)$	
		CbDY		$W \sim N(0, H_{CbDY} \sigma_w^2)$	
		CbSDY		$W \sim N(0, H_{CbSDY} \sigma_w^2)$	
		AvSD		$W \sim N(0, H_{AvSD} \sigma_w^2)$	
GH-BLUP	SNP + Multi-NIRS	CbSD+M	$Y_i = \mu + G_i + W_i + \epsilon_i$	$W \sim N(0, H_{CbSD} \sigma_w^2)$ $G \sim N(0, K_G \sigma_g^2)$	
		CbSY+M		$W \sim N(0, H_{CbSY} \sigma_w^2)$ $G \sim N(0, K_G \sigma_g^2)$	

4. Results

4.1. Descriptive statistics, heritability, and correlation between traits and absorbances

Four data sets were obtained from current bread wheat breeding programs run by public and private entities in France. The data were collected from either the first or second year of trials of candidate breeding lines. We first characterized the data and NIR spectra of all datasets by evaluating the dispersion of the data, the broad-sense heritability (H^2) of traits and wavelengths and the correlation between wavelengths and phenotypic traits.

Across all datasets, the general means for GY ranged from 5.3 t/ha to 12.9 t/ha and for HD from 122 days to 145 days. We measured the dispersion of the data around the mean for each set with the coefficient of variation (CV). For GY, the CV were low to moderate ranging from 1.8% to 9.4%, and for HD the CV were low ranging from 0.1% to 2.3% (Table 3-1). When repetitions were available, H^2 were calculated for both traits and for each wavelength within each environment. For GY, H^2 values were high, ranging from 0.77 to 0.95. For HD estimated only for Set3, H^2 values were high and stable from year to year ranging from 0.97 to 0.99 (Table S4). H^2 estimated for each wavelength in each environment ranged from 0.00 to 0.99. Nearly all spectra had H^2 values between 0.88 and 0.95. Exceptions were Set4_LI_2018_EM, Set4_T_2019_LM and Set4_LI_2019_LM where the mean H^2 was poor or moderate: 0.28, 0.48, 0.66 respectively (Figure S1). Correlations estimated between adjusted wavelengths and adjusted GY across environments, fluctuated from -0.60 to 0.61 along the spectra. Correlations estimated between adjusted wavelengths and adjusted HD across environments, fluctuated from -0.56 to 0.59 (Figure S1).

4.2. Phenomic predictions of grain yield and heading date in different bread wheat breeding programs

We first compared the PA of H-BLUP and PS-LASSO for HD and GY in different steps of two breeding schemes. The breeding sets came from different years of selection, with different population sizes and genetic variances (Table S4). Overall, H-BLUP and PS-LASSO were able to accurately predict both traits in each panel, but with a wide variability in PA (Figure 3-2). Set2 gave rise to the best PA for GY with a mean of 0.57 and a standard deviation around 0.08 for both H-BLUP and PS-LASSO. Set1 and Set4 gave rise to the lowest PA for GY with both models with PA between 0.34 and 0.39. The standard deviation of PA for Set4 GY was as high as 0.22.

Conversely, Set1 gave rise to the best PA for HD with the PS-LASSO model, with an average of 0.62 and a standard deviation of 0.10. Set4 gave rise to the lowest average PA for HD with H-BLUP (0.26). Comparison between H-BLUP and PS-LASSO environment by environment are detailed in Figure S2. PS-LASSO showed the better PA for HD in the three sets, on average 13.4% better than H-BLUP (Figure 3-2). Moreover, the interquartile ranges of PS-LASSO PA were narrower than with H-BLUP. For GY, H-BLUP and PS-LASSO had similar PA with similar variances for each panel of wheat genotypes.

We calculated the frequency of selection by PS-LASSO of each wavelength in all the environments to visualize the distribution of influential wavelengths (Figure S3). For both traits, influential wavelengths were distributed across the entire visible and infrared spectra. Wavelengths of 2108 and 2496 nm were very often selected regardless of the trait or the environment considered. More specifically, for GY at 1202 nm and for HD at 400 nm, these wavelengths were very often selected regardless of the environment. Only the H-BLUP model was used for the following analyses, because it is computationally more efficient than PS-LASSO.

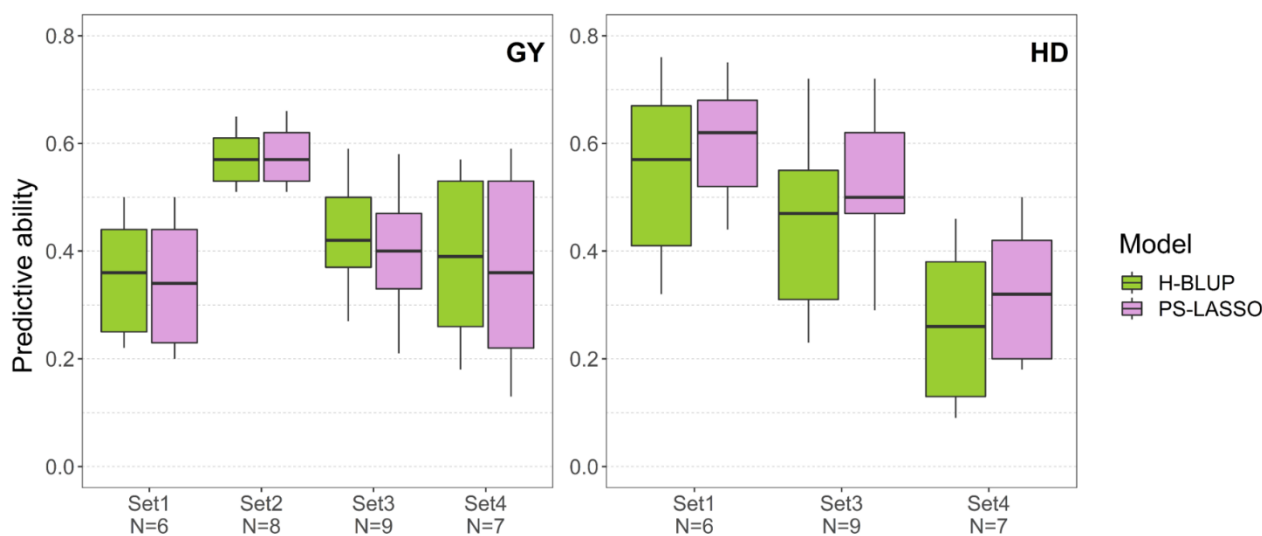


Figure 3-2. Boxplot of predictive abilities of phenomic selection for grain yield (GY) and heading date (HD) based on four different wheat breeding datasets (cf. Table 3-1) with two different models (H-BLUP, PS-LASSO). Predictions were based on a 5-fold cross-validation with 25 repetitions. Each boxplot indicates the mean (bold horizontal line), the 1st to 3rd quartiles (box), the 1st to 9th deciles (whiskers). N is the number of environments of prediction used to calculate the boxplot.

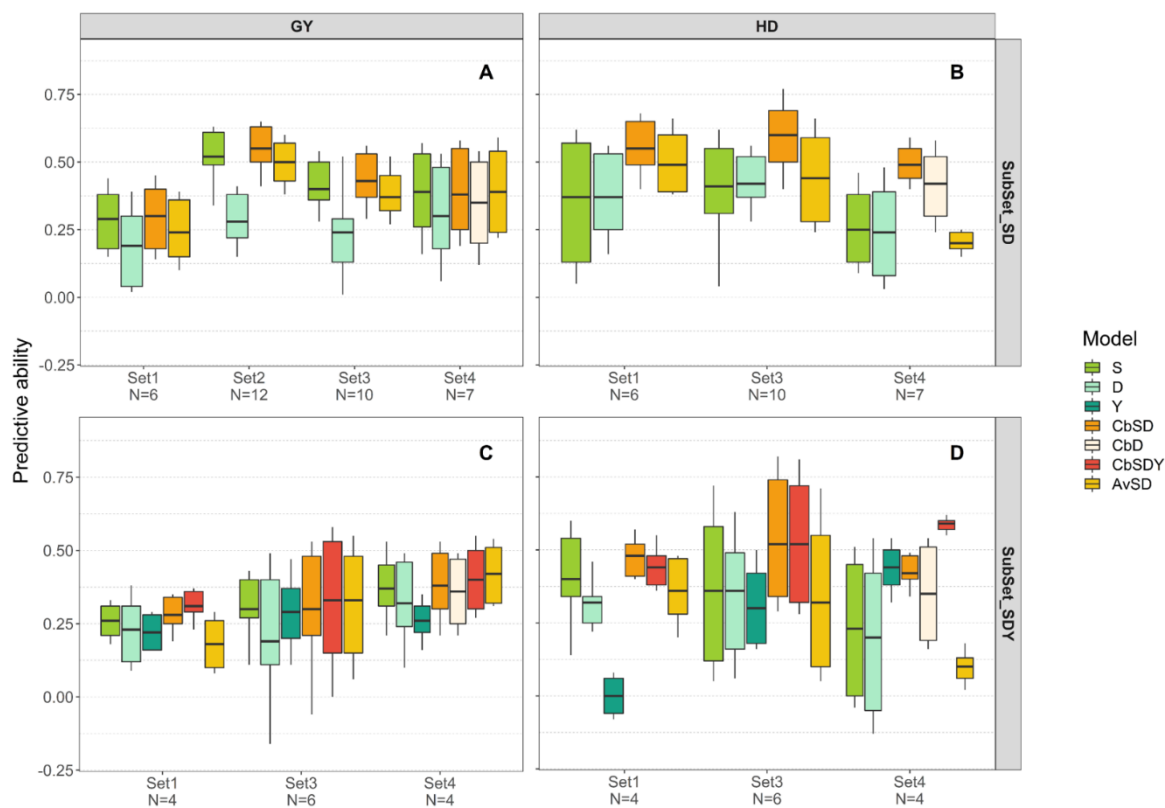


Figure 3-3. Boxplot of the predictive abilities of phenomic selection for grain yield (GY: **A,C**) and heading date (HD: **B,D**) in subgroup panels of wheat genotypes SubSet_SD (**A,B**) and SubSet_SDY (**C,D**) (cf. Table S5). Single-NIRS models S, D, and Y and multi-NIRS models CbSD, AvSD, CbD, and CbSDY were compared (cf. Table 3-3). Predictive abilities were obtained from a 5-fold cross-validation with 25 repetitions. Each boxplot indicates the mean (bold horizontal line), the 1st to 3rd quartile (box) and the 1st to 9th deciles (whiskers). N was the number of environments of prediction used to calculate the boxplot.

4.3. Impact of NIR spectrum origin of acquisition on predictive ability

We tested different models relying on a single spectrum or multiple spectra collected on the same genetic material grown in different environments to calculate the similarity between genotypes (Figure 3-3). The environment (site or year) of the plant material which produced the grain or flour that was analysed by NIRS had a consequence on the PA of the H-BLUP model. For models based on a single spectrum as predictive variable, the NIRS may have been measured on plant material grown on the same trial in which the trait was phenotyped (S), on a different site (D), or from the previous year (Y). As expected, in the respective models S was on average better than D and Y for predicting GY in each panel (Figures 3-3, A, C). For HD, S and D models gave rise to similar average PA (Figure 3-3, B).

We also compared different models combining multiple NIR spectra. For SubSet_SD, predictions based on CbSD were as good as those based on the S model for GY (Figure 3-3, A) and were even better for HD (Figure 3-3, B). For SubSet_SDY genotypes, prediction based on this same model was variable but was mostly as good as the best single NIRS model for both traits (Figures 3-3, C, D). In direct comparison, using the AvSD model resulted in lower PA than CbSD in 78% of cases and in particular for HD, on average 38% less of the CbSD PA.

The multi-NIRS models resulted in dissimilar and variable PA. CbSDY produced the best predictions in almost all panels and was at least as good as CbSD for both traits (Figures 3-3, C, D). Specifically, for Set4, the CbD model was compared to CbSD. PA were higher with CbSD than with CbD for HD. Using CbD resulted in similar PA for GY as with CbSD. This was unexpected because, for example, the PA reduction between S and D was almost 22% for GY (Figure 3-3, A), but the loss of PA between CbD and CbSD was only 11%. This was not compared on the other sets due to a lack of extra NIR spectra D.

As the PA were averaged by set, to ensure that the tendencies found for S, D and CbSD held true for all the specific panel-environment (P-E) combinations, we compared the deviation between the PA of S and D models. For GY and HD, around 10% and 50% of the combinations respectively gave better predictions using NIR spectra originating from a trial different than the one in which the predicted trait was measured (Figure S4, A). These P-E combinations were selected to compare the PA of the S, D and CbSD models (Figure S4, B). In these particular P-E combinations, the PA of the S model was on average 43% less for GY and 69% less for HD than the PA of the D model. The multi-NIRS model CbSD resulted in PA as high as the best single NIRS models. For HD, the same tendency was observed, but for some P-E combinations CbSD resulted in much higher PA than using D or S models.

4.4. Influence of the number of NIRS combined in multiple NIRS models on predictive ability

Combining NIR spectra acquired in multiple environments (CbSD or CbSDY) enhanced the PA of PS models. We investigated how the addition of NIR spectra acquired in several environments in the prediction model affected the PA (Figure 3-4).

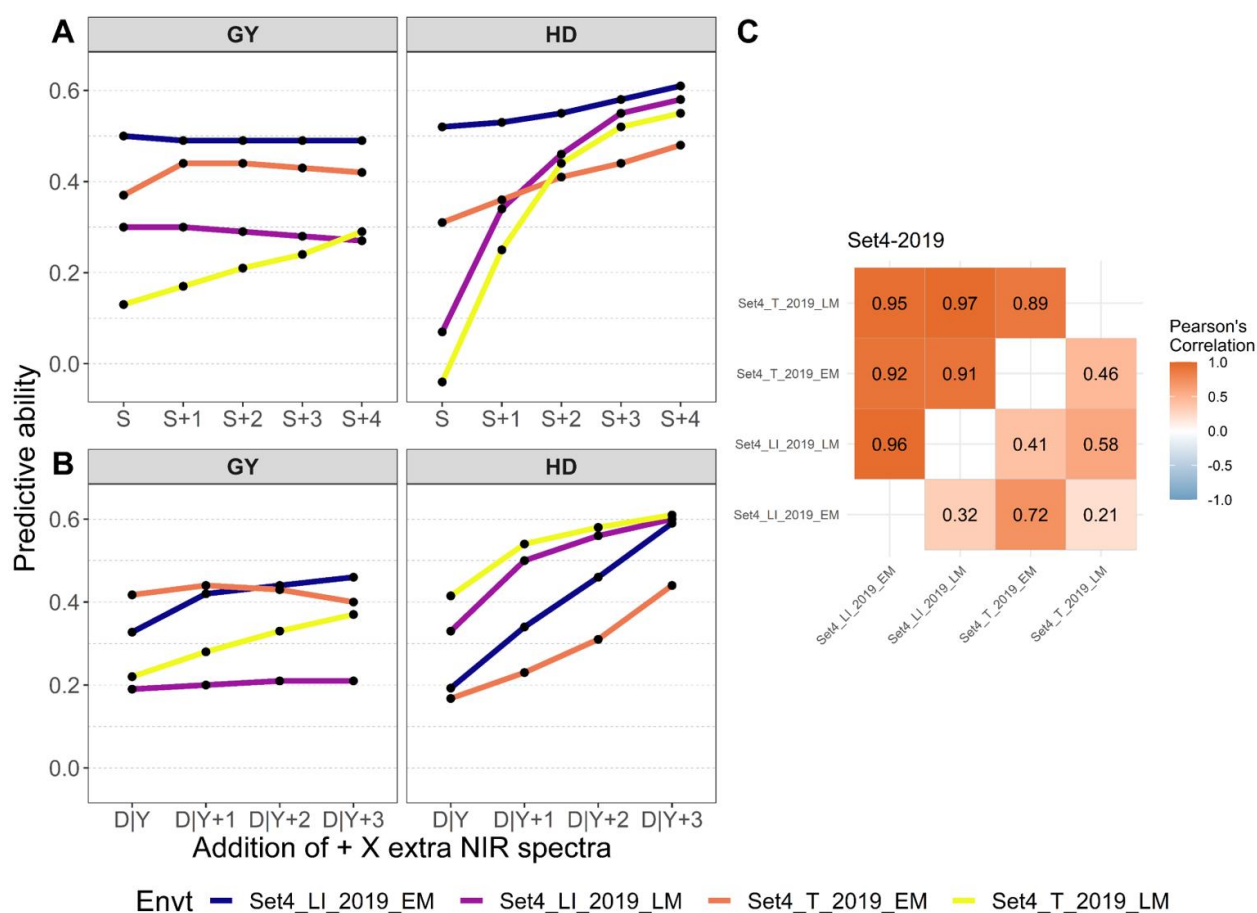


Figure 3-4. Predictive ability plotted as a function of the number of combined NIR spectra for grain yield (GY) and heading date (HD) for the panel-environment combinations of Set4 genotypes from year 2019 (A,B). When an environment is defined as environment of prediction, it has its own spectra S. The other environments correspond to spectra D. One spectra Y was also used and was drawn from the Set2_LI_2018_EM the previous year. In plot A, S corresponds to model S without adding any extra spectra in the model. Multi-NIRS models are S+1 to S+4 which correspond to the addition of spectra S plus 1 to 4 extra NIR spectra D1, D2, D3 or Y. In plot B, D|Y corresponds to model D or Y without adding any extra spectra in the model. Multi-NIRS models are D|Y +1 to D|Y +3 which correspond to the addition of spectra D or Y plus 1 to 3 extra NIR spectra excluding S. To give a concrete example, blue line corresponds to the environment of prediction Set4_LI_2019_EM. All the combinations of spectra are the following ones: S correspond to spectra from Set4_LI_2019_EM, D1 corresponds to Set4_LI_2019_LM, D2 corresponds to Set4_T_2019_EM, D3 corresponds to Set4_T_2019_LM, Y corresponds to Set2_LI_2018_EM. Each dot corresponds to the mean of all the possible NIRS combinations of the same number of extra NIR spectra. Predictive abilities were obtained from a 5-fold cross-validation with 25 repetitions. Matrices of correlation between environments for Set4-2019 (C), showing the Pearson correlation for heading date (upper matrix) and grain yield (lower matrix).

The analysis was conducted on all the datasets, but we focused on the set4-2019 in which had up to five NIR spectra could be combined. For both scenario and traits, on average, adding more NIR spectra in the multi-NIRS model enhanced PA, but the marginal gain in PA decreased as more NIRS were introduced. For GY, on average for both scenarios, successive addition of one more NIR spectrum to the original NIR spectrum in the model increased PA by 12.0%, 3.7%, 1.4% and 2.0%, respectively. The same tendency was observed with HD, but with higher gains of 57.0%, 22.0%, 15.0%, and 2.5%, respectively. The variability observed in all the panel-environment combinations are detailed in Figure S5 and S6. Looking at the variability on all the data of the scenario S and extra NIR spectra, two tendencies are observed according to the predicted trait. For GY, the results are very variable according to the data: the average gain using at least two NIR spectra in a combined model is +0.01 on the predictive ability with a maximum decrease of -0.07 and a maximum increase of +0.14. For HD, the average gain using at least two NIR spectra in a combined model is +0.17 on the predictive ability.

4.5. Comparison of GS and PS predictive abilities

We performed predictions with the multi-NIRS models (CbSD, CbSY) and the G-BLUP model based on molecular markers (M). We also tested a GH-BLUP model combining information from both markers and NIRS (CbSD+M, CbSY+M). For both traits, multi-NIRS models were on average better than M in SubSet_SDM and gave variable PA in SubSet_SYM (Figure 3-5). On average, considering all the subsets, in 48% of cases PS outperformed GS in predicting GY and in 62% of cases PS outperformed GS in predicting HD. Interestingly, GH-BLUP models combining both marker and NIRS data were the best models for both traits in all cases except one, and sometimes resulted in marked improvements, such as with SubSet_SDM Set4 that was 30% better than the M and CbSD models for HD prediction.

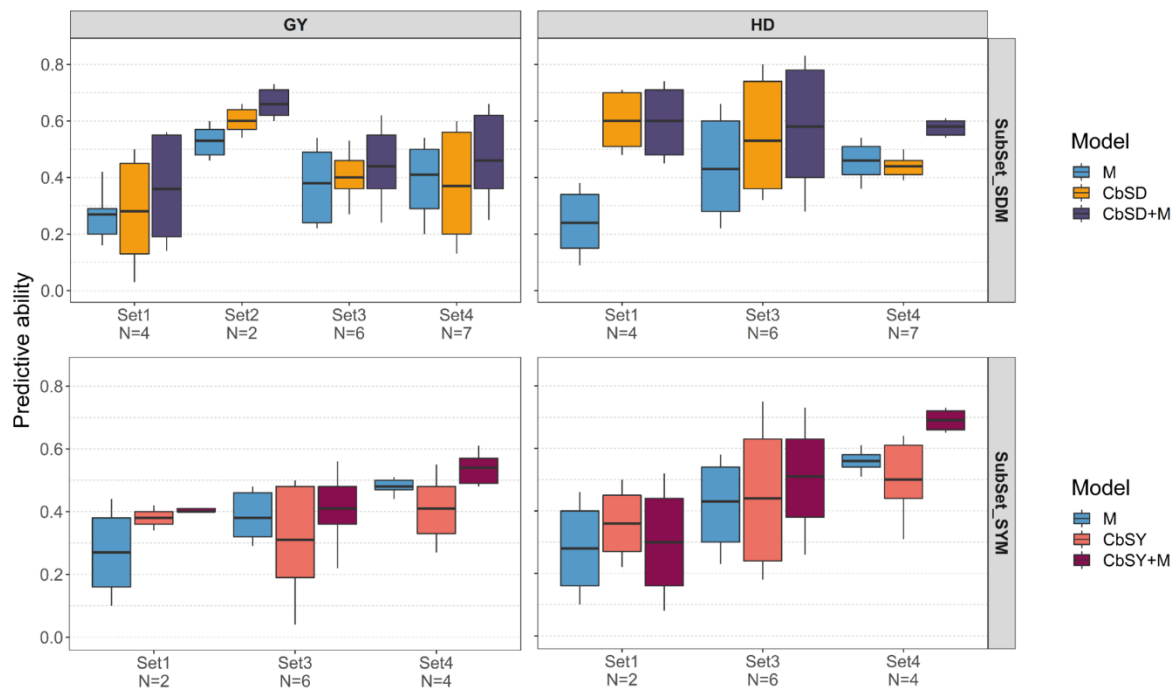


Figure 3-5. Boxplots of the predictive abilities of GS and PS for grain yield (GY) and heading date (HD) using different wheat breeding datasets (SubSet_SDM and SubSet_SYM) (cf. Table S5). Phenomic selection H-BLUP models CbSD and CbSY were compared to the G-BLUP genomic selection model M. GH-BLUP models including both marker and NIRS data were also tested (CbSD+M, CbSY+M). Predictive abilities were obtained from a 5-fold cross-validation with 25 repetitions. Each boxplot indicates the mean (bold horizontal line), the 1st to 3rd quartile (box), and the 1st to 9th decile (whiskers).

5. Discussion

5.1. Phenomic selection is an accurate complement or alternative to genomic selection in elite breeding material

The results presented here showed that phenomic selection was able to predict two traits with dissimilar genetic architectures for different sets of elite bread wheat breeding lines from different breeding companies. PS may therefore be useful for direct applications in breeding.

To evaluate the performance of this method, we compared PS predictions to those obtained with a reference GS model. We developed multi-NIRS models to make GLOB prediction for wheat breeding material. Predictions were at least as accurate as the GS model in 58% of the cases for GY and in 65% of cases for HD (Figure 3-5). These promising results were consistent with previous results using PS to predict GY and HD in a panel of registered bread wheat varieties (Rincent *et al.* 2018). The few studies to have implemented PS with H-BLUP were mainly applied to cereal species (Montesinos-López *et al.* 2017a; Rincent *et al.* 2018; Krause *et al.* 2019; Cuevas *et al.* 2019; Galán *et al.* 2020; Lane *et al.* 2020), facilitated by the routine NIRS use in these crops to predict grain quality traits such as protein content (Osborne 2006). In most of the studies just cited, the traits were phenotyped and NIRS data were acquired on the same plots. PS was mostly as good and sometimes better than GS (Rincent *et al.* 2018; Krause *et al.* 2019; Galán *et al.* 2020). Even if GS sometimes outperforms PS (e.g. Cuevas *et al.* 2019), PS can still be an attractive method due to the low cost of NIRS acquisition (Rincent *et al.* 2018).

We assessed a model that used both types of similarity matrix, that is, with genomic and NIRS data. Interestingly, this model systematically outperformed each of the basic models (G-BLUP and H-BLUP), meaning that molecular markers and NIR spectra bring complementary information (Figure 3-5). These results are consistent with studies which combined molecular marker scores and NIRS with the addition of pedigree information (Krause *et al.* 2019) or plant height phenotype (Galán *et al.* 2020). The same was not true for a model combining NIRS, molecular marker and pedigree information, as PS and GS predictive abilities were similar in Cuevas *et al.* (2019). In GLOB studies conducted on maize (Riedelsheimer *et al.* 2012; Guo *et al.* 2016; Westhues *et al.* 2017; Schrag *et al.* 2018; Azodi *et al.* 2020), wheat (Ward *et al.* 2015), rice (Xu *et al.* 2016) or *Arabidopsis thaliana* (Gärtner *et al.* 2009), combining information on transcripts and/or metabolites with DNA markers, led to more accurate predictions than classic GS models.

However, Riedelsheimer *et al.* (2012) and Xu *et al.* (2016), found that combining multi-omics data did not further improve genomic prediction in comparison to GLOB prediction with metabolomics data alone. Possibly GLOB prediction blending genomic and metabolomic information is better than GS when fewer DNA markers are available.

Our results confirmed the potential of GLOB prediction and the combination of different omics to predict complex traits. The reasoning is that omics data other than genomics reflect the expression of genes or biological pathways, which are different information than the genotype. Being able to account for genotype by environment (GxE) interactions and non-additive effects like epistasis or dominance may lead to improvement in complex trait predictions, which is more difficult with molecular markers alone.

5.2. Factors influencing PS predictive ability

Our study explored different factors thought to impact the PA of PS. We explored factors that have been identified in GS such as the type of predictive model, or factors specific to PS such as the number and origin of spectra used to compute the similarity matrix.

5.2.1. *H-BLUP* was well suited to predicting both traits with a reasonable computational demand

We compared the H-BLUP model, which considers absorbances at all the wavelengths of a spectrum to compute a relationship matrix between individuals, and the PS-LASSO model, which selects wavelengths that contribute the most to the trait variance. Furthermore, we predicted two traits GY and HD, which do not have the same genetic architecture. GY is controlled by many genes with small effects and is submitted to high levels of GxE interactions. HD is mainly controlled by major genes, is very heritable (e.g. Hanocq *et al.* 2007; Griffiths *et al.* 2009) and rarely shows cross-over GxE interactions with standard sowing dates.

We observed that the prediction for GY was similar for both models (Figure 3-2). However, for HD, PS-LASSO was slightly more accurate than H-BLUP. These results suggest that specific wavelengths could be linked to the HD trait. We compared the wavelengths selected by the LASSO model for each environment in parallel with the PA of the model for each environment (Figure S3). As the NIR bands are signatures of specific chemical bonds, investigating the organic molecules giving rise to influential bands would help us understand which secondary traits are correlated to the trait to be predicted. NIR readouts can give an overview of the organic molecules

identified (Xiaobo *et al.* 2010). The peaks identified in the present study are characteristic of several common atomic bonds (e.g. H₂O, CH, CH₂, CONH₂) making interpretation difficult.

NIR spectra are composed of multiple quantitative variables, which capture some genetic variability (Posada *et al.* 2009; Rincent *et al.* 2018). Thus, models developed for GS using molecular markers as explicative variables can be transposed to using NIRS variables. In this study we found that H-BLUP was efficient at predicting both traits with a low computational demand. We have evaluated only two models here among the numerous ones available (Charmet *et al.* 2020). H-BLUP performs well for different traits with an efficient computation time and it is user-friendly.

5.2.2. NIR spectrum origin of acquisition impacted differently the predictive abilities for HD and GY

Unlike molecular markers, NIR spectra are not independent of the environment. As for GY or HD, spectra are phenotypes that are the result of the genotype, the environment and the GxE interaction. The calculation of spectra heritability within environments revealed that the heritability varied from one environment to another and from one wavelength to another (Figure S1). Spectra with a medium or low average heritability also tended to be worse for predicting GY and HD. The quality of spectra measurement should be considered too. When comparing single NIR spectra, our first assumption was that the NIR spectrum from the same environment as the predicted trait would be the best predictor (S model). We found that, D and Y models sometimes lead to similar or higher PA than model S, especially for the prediction of HD. This is a significant result showing that according to the trait, the origin of the spectrum, that is the environment in which the specific plant material is analysed with NIRS, can have a major impact on prediction.

One factor which can explain that is the poor quality of certain NIR spectra in terms of heritability and thus in terms of predictive abilities. Krause *et al.* (2019) also found that models with predictive abilities close to zero were the ones with lower spectra heritabilities. Another possibility would be that for certain environments, NIR spectra are strongly influenced by the phenology and are thus able to predict HD accurately in any environment, HD being a major phenologic step of bread wheat. We know that NIR spectra are able to capture a significant portion of additive genetic variance (Rincent *et al.* 2018). Furthermore, we assume that the NIRS are able to indirectly capture some secondary traits correlated to the predicted trait. For example, GY is faintly or highly negatively correlated to the protein content (-0.1 to -0.64 in our data) and the NIR spectra are an

excellent predictor of protein content, so this correlation would indirectly predict GY in some situations. Both factors can explain why PS is able to predict complex traits. GY is a difficult trait to select for, because of the impact of the GxE interaction on the trait variance. For this trait, genetic and GxE interaction information are both required to reach good predictions. We suppose that spectra contain the GxE variance information by capturing the GxE effect of the phenotype via the metabolites and other molecules in response to the different stresses and growing conditions. Thus, for GY, only the spectrum from the same environment as the phenotype is expected to depict the responses of the plant through the gene expression and regulation that occurred in this particular environment. Unlike GY, HD of winter bread wheat has a very low level of GxE when the varieties are sown in autumn. We suppose that spectra collected in any environment are useful to predict this trait because they all capture genetic variance and because a relationship between a spectrum and HD obtained in one environment will be useful in any other environment.

5.2.3. Multi-NIRS models ensured accurate predictions regardless of the trait and the environment

We explored two ways of dealing with spectral data, focusing on the second: identify the best spectra *a priori* to calculate the similarity matrix or consider all the available spectra together. We proposed several models combining spectra from different environments or generations. We also compared two ways of computing the similarity matrix. The multi-NIRS models (CbSD, CbD, CbSY, CbDY, CbSDY), which correspond to a GLOB approach, are equivalent to the one used by Lane *et al.* (2020) where all the similarity matrices specific to each environment are averaged when all the spectra had the same number of wavelengths. The multi-NIRS model AvSD differs as the spectra were averaged by genotype followed by computation of the similarity matrix. We showed that combining spectra gave better predictions than averaging spectra. The difference is mainly due to the loss of information characterising genotypes in different environments when averaging the spectra by genotype.

Combining multiple NIRS from different environments or generations resulted in more robust PA in the models developed here. Our multi-NIR spectra models CbSD and CbSDY outperformed the best single NIRS model for GY in 47% of cases and the best single NIRS model for HD in 68% of cases (Figure 3-3). This suggests that adding other spectra to the predictive model, from another environment or generation, allowed a better estimate of the similarity between genotypes than a

single spectrum. More investigation is still required to incorporate spectra from different tissues or at different time points of the growing cycle to bring different useful information to the predictive model. Similarly to molecular markers in GS (Norman *et al.* 2018), adding more and more predictors eventually leads to a plateau in PA, probably as the information added becomes redundant (Figure 3-4). We noticed that depending on the predicted trait, the plateau was reached with fewer predictors. For GY, the gain was marginal when more than two NIR spectra were combined, whereas for HD, the plateau was still not reached with four extra spectra. For HD, we suppose that adding more spectra into the model allows more genetic variance to be captured reflecting the more indirect relationships between the underlying tissue composition and HD. For GY, we suppose that the GxE interaction which explains much of the trait variance, is environment specific and can be captured only with spectra from the same environment as the phenotype. Thus, adding more spectra from different environments will not bring more useful information in the predictive model. More data are however required to fully understand the benefit of combining NIR spectra for different traits.

5.2.4. Other factors such as size and composition of the training set, the scanned organ and spectral acquisition method should be considered

We also studied the effect of size and composition of the training set on PA in phenomic selection (Appendix 1). Our results revealed that like GS (Norman *et al.* 2018), the gain in accuracy reached a plateau for PS. There is a point where adding other individuals to train the model does not bring novel information. This suggests that optimisation of the training set could generate the best possible prediction while minimising its size. One interesting observation is that the multi-omics models gave the best prediction when we considered larger training sets and the plateau was reached with larger training sets than in the other models. This may be because useful genomic information and phenomic information are not carried by the same genotypes and thus increasing the training set size still increases the PA. When determining the composition of the training set, there is a trade-off to be made between maximising the predictive ability and decreasing the number of individuals to be phenotyped. Our results revealed that optimisation algorithms such as CDmean (Rincent *et al.* 2012) gave better PA than random selection, with both molecular markers and NIRS data. Because the CDmean directly targets the expected PA, this criterion is better suited to predictive models than deterministic algorithms (Honigs or Kennard and Stone algorithm) (Kennard and Stone 1969; Honigs *et al.* 1985) which select the individuals to cover the absorbance variability across the spectrum. It would be interesting to find a criterion

which could handle molecular markers and NIRS at the same time to select the individuals in the training set.

In our study we could not compare the effect of the tissue used for NIRS collection or the effect of the spectrometer and protocol used to collect NIRS. However, it is interesting that other studies used hyperspectral imaging of the canopy (UAV) and obtained PS predictions even better than GS (Krause *et al.* 2019; Galán *et al.* 2020). In this case the number of wavelengths compared to lab measurements is halved, with wavelengths from the visible part of spectrum only. The nature of the plant material in canopies is quite different from grain samples, particularly as the canopy is still growing in the environment or the trial. The tissue sample for NIRS acquisition can have a substantial impact on PA, for example, using leaf spectra was less accurate than using grain spectra for the prediction of GY in Rincent *et al.* (2018). Furthermore in PS, or in any application of NIRS, spectra need to be pre-treated with specific filters to remove irrelevant information due to additive and multiplicative external effects (Blanco and Villarroya 2002). Mathematical transformations are applied to spectra correcting the absorbance value of each wavelength. Such transformations impact the PA of the method used downstream and cannot be generalised to other data (Cuevas *et al.* 2019). In practice, different filters should be tested in a cross-validation to select the best set of pre-treatments.

6. Conclusion

We explored whether PS can be implemented in wheat breeding programs by identifying factors which impact the PA of the method. Our main results showed that the choice of the spectra used in the predictive model was crucial and was specific to the predicted trait. The H-BLUP models developed here which combined multiple spectra to compute the similarity matrix are meaningful and can be readily handled by breeders. As in GS, the size of the training set had a crucial impact on PA and optimisation of the training set with CDmean criterion giving better PA than random selection. Finally, amid all these factors, PS was mostly at least as accurate as GS for both traits. PS is a good alternative to GS, in particular when NIRS are already routinely measured (e.g. in all wheat breeding programs). For these species, PS could be implemented without any additional cost. In breeding programs in which genotypic data are already routinely collected, PS combining NIRS and molecular marker data could potentially be advantageous and more accurate as we found. Integrating different kinds of high-throughput phenotyping and multi-omics information in statistical models is a promising approach.

Acknowledgements

The authors thank the work in experimental units by INRAE (Clermont-Ferrand, Estrées-Mons, Le Moulon, Rennes), breeders from Agri-Obtentions and Florimond Desprez. The authors are grateful to Agri-Obtentions, Florimond-Desprez, and the Association Nationale de la Recherche et de la Technologie (ANRT, grant number 2019/0060) which supported this PhD work. The authors also thank Bastian Alexandre and Rachel Carol (Bioscience Editing, France) for the proofreading of this work, and Tristan Mary-Huard for the careful reading of the equations. Finally, the authors thank the two anonymous reviewers for the helpful comments on this work.

7. Appendix 1

We describe in this section the analysis of two factors which also impacted the predictive ability of the phenomic selection: the size and composition of the training set (TS).

In GS the size and composition of the TS have an impact on PA. We characterised this effect by testing six TS sizes (10, 50, 100, 150, 200, 250 genotypes) on two specific sites from the dataset Set2-2019: GL and EM. These datasets had the larger number of lines genotyped to test different TS sizes. For this, we randomly split the data in five folds of the same size. One-fold constituted the validation set and the remaining folds are the genotypes potentially included in the TS. Among the latter, we randomly sampled a definite number of lines to constitute the final TS with the corresponding size. The same procedure was followed for all the folds and all the TS sizes and was repeated 25 times to give 125 predictive abilities for each TS size. We thus compared the PA of the four models **M**, **S**, **CbSD**, and **CbSD+M** representing the G-BLUP, H-BLUP and GH-BLUP model types.

The composition of a TS can be optimised in order to minimise its size while retaining similar PA. In one scenario the size of the TS was arbitrarily defined, and either random or optimized procedures were used to select among all genotypes available, the one which will constitute the TS. The validation set was composed of the genotypes not included in the TS. We compared three optimisation algorithms to define the TS for a particular size. We tested the CDmean algorithm (**CDmean**) (Rincent *et al.* 2012), developed originally for GS, and two algorithms developed to optimize NIRS calibration equations in chemiometry, Honigs (**HG**) (Honigs *et al.* 1985) and Kennard-Stone (**KS**) (Kennard and Stone 1969). CDmean was applied with custom R code, while HG and KS were applied with the prospectr R package (Stevens and Ramirez-Lopez 2020). Algorithm performance was compared to randomly select (RD) training sets. PA for each TS size was averaged over 50 repetitions for CDmean or 125 repetitions for RD. There was no repetition for HG and KS as they are both deterministic. To compare these TS selection approaches, we used the datasets Set2-2019-EM and Set2-2019-GL, in which many varieties were genotyped.

We found that the effect of increasing TS size on the PA of PS and GS was substantial (Figure 3-6). For each model and both environments, PA increased with the TS size. However, from 50 to 250, the PA increased only slightly. For the smallest TS, the S model was slightly better than the others. Regardless of the TS size, the GS model M was less accurate than the PS models. Model CbSD+M, combining both marker and NIRS data, outperformed the other models for the larger size of TS.

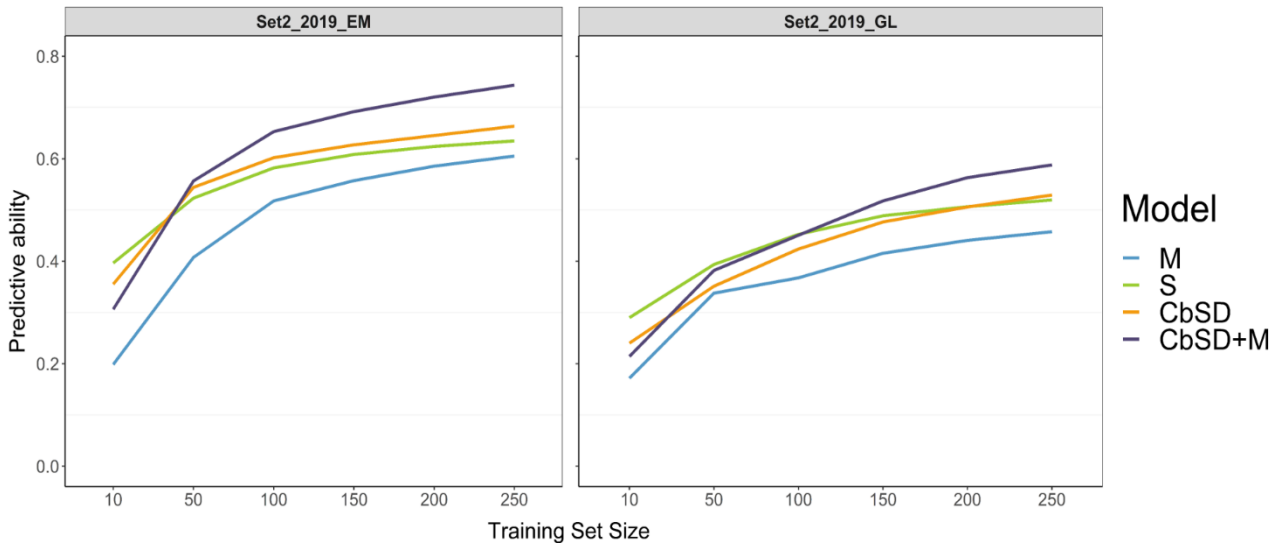


Figure 3-6. Comparison of predictive abilities of GS and PS for GY as a function of the size of the training population for four predictive models. M refers to a G-BLUP GS model, S and CbSD to H-BLUP models, and CbSD+M to a GH-BLUP model. CbSD combined NIRS from two environments. Predictions were performed with Set2 data from two environments, Set2_2019_EM and Set2_2019_GL for each TP size. Lines indicate mean predictive abilities of a 5-fold cross-validation with 25 repetitions.

To optimise TS composition for a given size, we compared three different optimisation algorithms to determine the composition of the TS for performing GS or PS (Figure 3-7). When applying GS (G-BLUP model), TS optimized with CDmean computed with the kinship matrix (CDmean_K) performed better than random TS (RD) for both environments with an average gain of +22%, a maximum of +30% and a minimum of -4.5%. When applying PS (both H-BLUP models), TS optimized with CDmean_H computed on the NIR similarity matrix also performed slightly better than randomly sampled TS (RD). Finally, for the GH-BLUP model combining molecular markers and NIRS, CDmean_K performed slightly better than CDmean_H and RD. KS and HG performed very variably as a function of the TS size in Set2_2019_GL and with lower PA than RD in Set2_2019_EM when applying H-BLUP and GH-BLUP models.

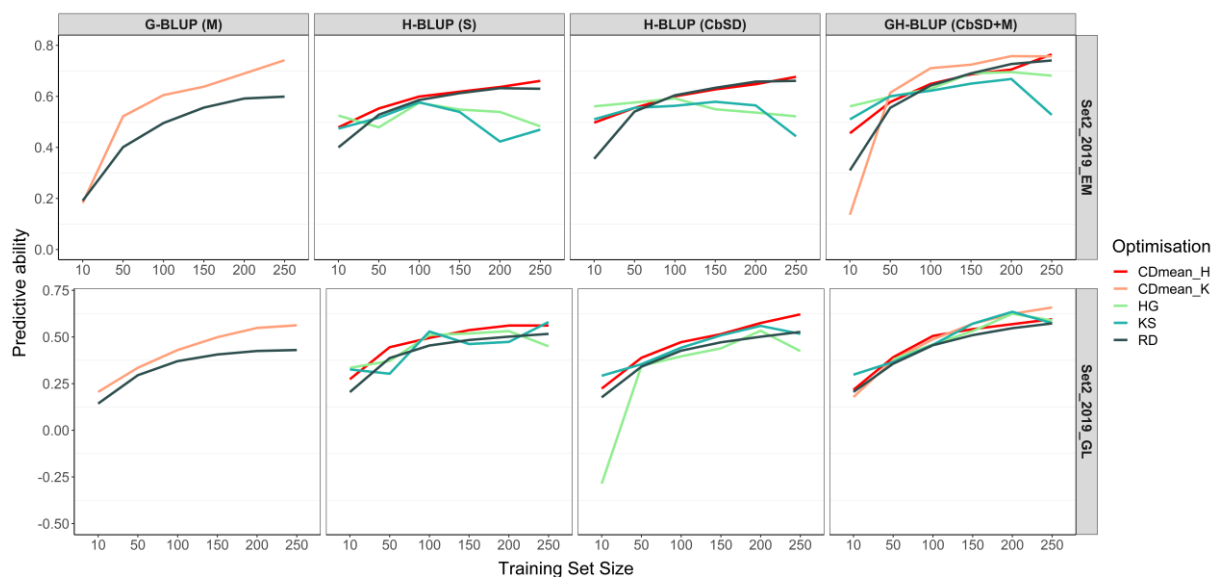


Figure 3-7. Comparison of predictive abilities for GY as a function of the size of the training population for four predictive models and three optimisation algorithms. Algorithms using NIR spectra were Honigs (HG), Kennard-Stone (KS) and CDmean_H using the hyperspectral similarity matrix (H). CDmean_K was conducted on kinship (K) based on molecular markers. Optimisations are compared to a random selection (RD). Predictions were performed using Set2 data from two environments: 2019_EM and 2019_GL. Predictive models used were M, S, CbSD, CbSD+M (cf. Table 3-3). Lines represent the predictive ability with respect to TP size. Predictive ability was averaged on 50 repetitions for CDmean_K and CDmean_H, and on 125 repetitions for RD. Predictive ability for HG and KS were obtained once because they are deterministic criteria.

Chapitre 4

**Prédiction de l'interaction génotype-environnement dans
le contexte d'essais multi-environnementaux**

Préambule

Dans ce deuxième chapitre d'analyse, nous allons étendre les prédictions phénotypiques aux scénarios de prédiction multi-environnementale (MET). L'un des enjeux dans les prédictions MET est de pouvoir correctement modéliser l'effet du GXE. En effet, le rendement est un caractère complexe dont la variance est influencée par l'effet du génotype, mais aussi l'effet du GxE. Il a été largement montré que cette modélisation est difficile à faire en prédiction génomique, car les marqueurs moléculaires sont indépendants de l'environnement. Ce chapitre s'applique donc à montrer qu'il est possible de modéliser l'effet du GxE grâce aux spectres NIR. Dans les scénarios de prédiction MET, nous allons proposer des scénarios qui modélisent ce GxE grâce à l'acquisition des spectres dans chacun des environnements du réseau d'essais. L'acquisition de ces spectres peut être une contrainte que nous lèverons en proposant des méthodes pour prédire des spectres NIR manquants au sein du réseau d'essais.

Les différents scénarios de prédiction développés permettront de comparer la sélection phénotypique à la sélection génomique, et de comparer également des modèles qui combinent à la fois les marqueurs moléculaires et les spectres NIR.

Ce chapitre est présenté sous la forme d'un article scientifique qui a été soumis à la revue *Theoretical and Applied Genetics* le 18 février 2022.

Phenomic selection in wheat breeding: prediction of the genotype-by-environment interaction in multi-environment breeding trials

Pauline Robert^{1,2,3,4}, Ellen Goudemand⁴, Jérôme Auzanneau³, François-Xavier Oury², Bernard Rolland⁵, Emmanuel Heumez⁶, Sophie Bouchet², Antoine Caillebotte¹, Tristan Mary-Huard^{1,7}, Jacques Le Gouis², Renaud Rincent^{1,2, #}

1 Université Paris-Saclay, INRAE, CNRS, AgroParisTech, GQE - Le Moulon, 91190, Gif-sur-Yvette, France

2 INRAE - Université Clermont-Auvergne, UMR1095, GDEC, 5 chemin de Beaulieu, 63000 Clermont-Ferrand, France

3 Agri-Obtentions, Ferme de Gauvilliers, 78660, Orsonville, France

4 Florimond-Desprez Veuve & Fils SAS, 3 rue Florimond-Desprez, BP 41, 59242, Cappelle-en-Pévèle, France

5 INRAE–Agrocampus Ouest-Université Rennes 1, UMR1349, IGEPP, Domaine de la Motte, 35653 Le Rheu, France

6 INRAE, UE 972, Grandes Cultures Innovation Environnement, 2 Chaussée Brunehaut, 80200 Estrées-Mons, France

7 MIA, INRAE, AgroParisTech, Université Paris-Saclay, 75005 Paris, France

Corresponding authors: renaud.rincent@inrae.fr

ORCID: P. Robert: 0000-0002-6314-180X / J. Le Gouis: 0000-0001-5726-4902 / S. Bouchet: 0000-0001-5868-3359

/ R. Rincent: 0000-0003-0885-0969

1. Abstract

Key message

Phenomic prediction of wheat grain yield and heading date in different multi-environmental trial scenarios is accurate. Modelling the genotype by environment interaction effect using phenomic data is a potentially low-cost complement to genomic prediction.

The performance of wheat cultivars in multi-environmental trials (MET) is difficult to predict because of the variability in genotype by environment interactions (GxE). Phenomic selection is better than genomic selection in modelling the GxE effect because it accounts for non-additive effects. Here, phenomic data are near infrared (NIR) spectra obtained from plant material used to compute relationship matrices describing the variance in the chemical composition. While phenomic modelling has recently been shown to accurately predict grain yield in single environments, it is not known how accurate it is for MET. We used four datasets from two winter wheat breeding programs in France to test and compare the abilities of phenomic and genomic models to predict grain yield and heading date in different MET scenarios. We also compared

different methods to model the GxE using different covariance matrices. On average, phenomic and genomic prediction abilities of grain yield and heading date are similar in all the different MET scenarios. Better predictive abilities were obtained when GxE effects were modelled with NIR spectra than without them, and it was better to use all the spectra of all genotypes in all environments for modelling the GxE. To facilitate the uptake of phenomic prediction by breeders, we tested MET designs where the NIR spectra were measured only on the genotype-environment combinations phenotyped for the target trait. Missing spectra were predicted with a nearest neighbour method and a weighted multivariate ridge regression. Intermediate predictive abilities for grain yield were obtained in a sparse testing scenario and for new genotypes, which shows that phenomic selection is an efficient and practicable prediction method for dealing with GxE.

Key words

Bread wheat, Genomic selection (GS), Multi-Environment Trial (MET), Near infrared spectroscopy (NIRS), Genotype by environment interaction (GxE), Phenomic selection (PS), Plant breeding, *Triticum aestivum*.

Declarations

Funding: This work was funded by Agri-Obtentions, Florimond-Desprez, and the Association Nationale de la Recherche et de la Technologie (ANRT, grant number 2019/0060).

Conflicts of interest/Competing interests: Co-author Ellen Goudemand was employed by Florimond-Desprez, and co-author Jérôme Auzanneau was employed by Agri-Obtentions. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Availability of data and material: The datasets generated during and/or analysed during the current study are not publicly available due to the breeding program privacy policy, but are available from the corresponding author on reasonable request.

Code availability: Code used to run the analysis is available from the corresponding author on request.

Author contribution statement: JA, FXO, BR and EH designed the field trials and collected the phenotypic data from Agri-Obtentions and INRAE. EGD provided the phenotypic data and genotyping data from the Florimond Desprez company. SB provided the genotyping data from

Agri-Obtentions and INRAE and participated in discussions of this study. AC and TMH developed the method of multivariate weighted ridge regression prediction of NIR spectra. RR initiated the project and with JLG supervised the study and helped improving the manuscript. PR analysed the data and wrote the manuscript. All authors approved the final manuscript.

2. Introduction

To select superior crop cultivars, plant breeders mainly compare the recorded phenotypes of candidates then select the best ones. However, phenotypic information only documents the relative performance of a cultivar in that environment, which depends on the genome in interaction with the environment (GxE). The relative performance of cultivars commonly differs in different environments, leading to differences in how cultivars are ranked (Comstock and Moll 1963; Allard and Bradshaw 1964). To take GxE into account, breeders evaluate candidate lines across several environments (combinations of year \times site \times management). These multi-environment trials (MET) enable breeders to identify high-performing selection candidates in each environment. METs are very expensive because of the phenotyping costs, so only a limited number of candidates and environments can be tested at a time, which slows genetic progress.

Genomic selection (GS) (Bernardo 1994; Whittaker *et al.* 2000; Meuwissen *et al.* 2001) is a method developed to predict the performances of candidates using a predictive model calibrated with a genotyped and phenotyped training population, potentially taking GxE into account. Single-environment predictions on breeding material were promising, because the predictive abilities (PA) were higher than those obtained using pedigree information only (de los Campos *et al.* 2009; de los Campos *et al.* 2010; Crossa *et al.* 2010). In the past decade, specific GS models have been proposed for predicting GxE. Burgueño *et al.* (2012) found that integrating the information that was common to different environments could improve the PA of the model by 17.5% compared to models with a main genetic effect only. To do this, an environmental covariance matrix was introduced in the prediction model. However, these models were still only applicable to environments in which some phenotypes had been measured. Other models were proposed that relied on environmental covariates (EC), either pedoclimatic data (e.g. temperature, radiation, soil characteristics) or stress indexes (Ly *et al.* 2017; Rincent *et al.* 2019). Environments with similar EC values are assumed to interact with genotypes in a similar way, so the model can be used to predict plant behaviour in new environments (Heslot *et al.* 2014; Jarquín *et al.* 2014; Malosetti *et al.* 2016; Lado *et al.* 2016; Ly *et al.* 2018). However, further improvements in predictions based

on ECs are marginal, as modelling GxE this way does not account for complex interactions between plants and environments. Some ECs may not be responsible for GxE while the ECs responsible for the GxE may differ from one genotype to another. The response of a plant to its environment can be characterised at the molecular level using endophenotypes such as transcripts, proteins or metabolites. An endophenotype is essentially a measurable trait that is intermediate between the genotype and the final phenotype. It has been shown that transcripts (Frisch *et al.* 2010; Fu *et al.* 2012; Zenke-Philippi *et al.* 2017; Azodi *et al.* 2020), and a combination of transcripts and metabolites (Guo *et al.* 2016; Westhues *et al.* 2017; Schrag *et al.* 2018; Wang *et al.* 2019) can be used in accurate genomic-like omics based (GLOB) prediction of the performances of hybrid maize or hybrid rice, especially for complex traits. Prediction of grain yield (GY) using such omics data was on average just as accurate as those obtained with GS models using molecular markers. However, when only information on metabolites was used to inform the model, PA were lower than when using molecular markers to predict GY in hybrid maize and in barley (Riedelsheimer *et al.* 2012; Xu *et al.* 2016; Gemmer *et al.* 2020). Practically, GLOB prediction is similar to genomic prediction with the difference that scoring molecular markers is replaced by omics data collection for the chosen endophenotypes (Robert *et al.* 2022a). To limit costs, the omics data are generally collected once and for all on material grown in controlled conditions. The control conditions are radically different than the environments of MET, so endophenotype data may be of limited use in predicting GxE.

Phenomic selection (PS) is a recently developed method (Rincent *et al.*, 2018) similar to GS and GLOB, that replaces molecular markers or endophenotypes with information from near infrared spectroscopy (NIRS). By measuring the reflectance or absorbance of a tissue sample at different wavelengths of light from visible to near infrared (NIR), NIRS quantifies the chemical bonds and hence the molecular composition of the tissue. Like the endophenotypes, this molecular composition is under genetic control and is a consequence of the expression of the genotype in a particular environment. NIRS, which is already routinely used by cereal breeders to predict quality traits such as grain protein content, has the advantage of being low-cost, high throughput and non-destructive (Osborne 2006). PS was shown to outperform GS in two prediction scenarios. The first used NIR spectra acquired in the same environment as the phenotyping, and the second used NIR spectra acquired in a different environment than the one in which the training population was phenotyped (Rincent *et al.*, 2018). For single environment scenarios, PS has shown promising results for the prediction of GY in bread wheat (Rincent *et al.* 2018; Krause *et al.* 2019; Cuevas *et*

al. 2019; Robert *et al.* 2022a), rye (Galán *et al.* 2020), maize (Lane *et al.* 2020), triticale (Zhu *et al.* 2021b) and soybean (Parmley *et al.* 2019; Zhu *et al.* 2021a). More recently, Robert *et al.* (2022a) found that it might be preferable to use the NIRS data from the same environment as the target trait (e.g. for GY) or to combine several spectra from various environments (e.g. for heading date (HD)) depending on the interactivity of the underlying genes with the environment.

Like endophenotypes, NIR spectra are the result of a genotype, an environment, and the interaction between them. Therefore, it is reasonable to expect that the spectra capture the genetic effect and the GxE even for complex traits such as GY. By comparing PA of models based on NIR spectra collected in the same or in different environments from the one in which the training population was phenotyped, it was shown that the information from NIRS was indeed able to capture part of the GxE (Rincent *et al.*, 2018; Robert *et al.*, 2022a). A few examples of phenomic prediction in the context of MET have now been reported. Krause *et al.* (2019) compared different best linear unbiased prediction (BLUP) models in GS (G-BLUP) and in PS (H-BLUP) and a model combining molecular markers and NIR spectra (multi-omics model) on MET data. They found that the maximal PA obtained was with the multi-omics model where the main genetic effect is modelled by markers or pedigree and the GxE effect is modelled by NIR spectra collected in each environment. Lane *et al.* (2020) integrated the wavelength-by-environment interaction effect with all spectra collected in each environment of a MET in a H-BLUP model or in functional regression. They then tested these models in a MET scenario where one environment was new in the trial network and no genotypes were phenotyped. On average, the PA reached 0.40 for H-BLUP and 0.53 for the functional regression. These two demonstrations show the promise of integrating information from NIR spectra with GxE estimates.

Our aim here was to compare different PS, GS and multi-omics models (combining molecular markers and NIR spectra) for predicting GY and HD of wheat breeding candidates in different MET scenarios. More precisely, we investigated whether NIR spectra can model the GxE effect in predictive models by comparing different covariance structures based on spectral data. For the modelling of GxE using NIR spectra to be practicable, breeders would ideally acquire spectra in all environments of the MET and for all genotypes, whether they were used to calibrate the model or not, which would require specific nurseries for growing the reference genotypes in each environment. To overcome this limitation, we also tested two approaches for predicting missing spectra and evaluated them in two original MET scenarios.

3. Materials and Methods

3.1. Plant material, genotyping and NIRS acquisition

The plant material, genotyping and NIR spectra acquisition in the present study have been previously described in Robert *et al.* (2022a). Briefly, four datasets of winter bread wheat breeding lines were used: Set1-2016, Set2-2019, Set4-2018, Set4-2019. Set1 corresponds to lines developed in the breeding program of Florimond Desprez (France) and the other sets to lines developed in the breeding program of Agri-Obtentions (France) in collaboration with INRAE (France). Set1 and Set2 correspond to the first year of trial evaluation for the candidate lines and the two Set4 sets to the second year of trial evaluation. These datasets were chosen based on the number of environments of the MET where NIR spectra were acquired and the number of candidate lines phenotyped and genotyped. Each dataset was analysed independently.

In more detail, Set1-2016 is composed of 152 candidate lines genotyped with the 35K breeder Bristol array (Axiom™ Wheat Breeder's Genotyping Array). NIR spectra of 10 g of wheat flour were acquired with the NIR 6500 FOSS spectrometer (FOSS NIR Systems, Silver Spring, MD, USA) over the range 400 to 2500 nm in steps of 2 nm. Final spectra are the average of 32 repeated measurements. Set2-2019 is composed of 325 lines genotyped with the 35K BreedWheat array (Axiom™ BreedWheat Genotyping Array) which is a subset of the TaBW410k SNPs array (Kitt *et al.* 2021) including 280k SNPs from the TaBW280k SNP array (Rimbert *et al.* 2018). NIR spectra of 150 g of grain were acquired with the XDS NIR Analysers FOSS spectrometer (FOSS NIR Systems, Silver Spring, MD, USA) in the range 400 to 2500 nm in steps of 2 nm. Final spectra are the average of 16 repeated measurements. Set4-2018 and Set4-2019 are composed of 71 and 100 lines, respectively, genotyped with the 35K BreedWheat array (Axiom™ BreedWheat Genotyping Array) which is a subset of the TaBW410k SNP array including 280k SNPs from the TaBW280k SNP array (Rimbert *et al.* 2018). NIR spectra on 350 g of grain were acquired with the MPA II FT-NIR analyser (Bruker Optics, Ettlingen, Germany) ranging from 3594.92 cm⁻¹ to 12489.60 cm⁻¹ in steps of 7.7 cm⁻¹. Final spectra are the average of 64 repeated measurements.

3.2. Analysis of the phenotypes measured in the METs, genotypic data and NIRS data

3.2.1. Description of the different METs and the phenotypic data

The trial information and designs of the different METs (two to four environments) recorded in the datasets are described in Robert *et al.* (2022a). Briefly, Set1-2016 is composed of two environments (two sites \times one year) with augmented design trials in Houville (HV) and Lectoure (LC) for the year 2016. Lines were phenotyped for GY and HD. Set2-2019 is composed of two environments (two sites \times one year) with augmented design trials in Estrée-Mons (EM) and randomised block trials in Genlis (GL) for the year 2019. Intensive management practices were used at both of the latter sites, except that lines were not treated with fungicide at EM. Lines were phenotyped for GY. Set4-2018 and Set4-2019 are composed of four environments (two sites \times two treatments) with three complete randomised block trials in EM and Le Moulon (LM) and intensive practices (T) or low input treatment (LI) at each, in the years 2018 and 2019. Set4-2018 and Set4-2019 are thus independent datasets. For Set4-2018, one environment was removed from the analysis due to a mis-association between genotypes and phenotypes, so only three environments were analysed here. Lines in Set4-2018 and Set4-2019 were phenotyped for GY and HD.

3.2.2. Adjustment of the phenotypic and NIRS data and estimation of the variance components

When possible, GY, HD and the NIR spectra at each wavelength were adjusted with specific spatial models described in Robert *et al.* (2022a). Briefly, statistical models were used to compute adjusted means corrected for spatial effects, or block and replicate effects, or all three depending on the experimental design.

NIR spectra quality control was required to harmonise the NIRS data of Set4 with a conversion into nm in steps of 2 nm and to define a common window from 802 to 2492 nm for all the NIR spectra. Spectra were visualised to filter out any spectra with abnormal absorbances resulting from technical errors. For trials which included repetitions, adjusted means of NIR absorbance were computed as for GY and HD. Finally, all the spectra were pre-treated (first derivative of the normalized spectra) to eliminate noise inherent in the absorbance measurement. The different corrections are described in Robert *et al.* (2022a).

For the datasets with replicates (Set4-2018 and Set4-2019) we estimated the different variance components for GY, HD and NIRS wavelength readouts based on the following equation:

$$\hat{Y}_{ijk} = \mu + E_j + G_i + GE_{ij} + \epsilon_{ijk} \quad (1)$$

\hat{Y}_{ijk} is the trait value corrected for spatial effects for the replicate k of the genotype i in the environment j ; μ is the intercept; E_j , G_i and GE_{ij} are random environmental, genetic and interaction effects, respectively; and ϵ_{ijk} is the random residual effect. We assumed that the random effects are independent and identically distributed (i.i.d) and normally distributed such that $E_j \stackrel{i.i.d.}{\sim} N(0, \sigma_E^2)$, $G_i \stackrel{i.i.d.}{\sim} N(0, \sigma_G^2)$, $GE_{ij} \stackrel{i.i.d.}{\sim} N(0, \sigma_{GE}^2)$ and $\epsilon_{ijk} \stackrel{i.i.d.}{\sim} N(0, \sigma_\epsilon^2)$. Broad-sense heritabilities within each environment for traits and wavelengths were described in Robert *et al.* (2022a).

For all the datasets without replicates we followed the Rincent *et al.* (2018) approach where the genomic variance components were estimated from the following bivariate mixed model across two environments:

$$\hat{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = X\beta + Zu + e \quad (2)$$

where y_1 and y_2 are the absorbances for a given wavelength in each environment, corrected for spatial effects. X and Z are the design matrices of the corresponding effect, β is the vector of fixed environment effect, u is the random polygenic effect with $\text{var}(u) = \begin{bmatrix} \sigma_{u1}^2 & \sigma_{u12} \\ \sigma_{u21} & \sigma_{u2}^2 \end{bmatrix} \otimes K$, and e is the vector of the residuals with

$\text{var}(e) = \begin{bmatrix} \sigma_{e1}^2 & 0 \\ 0 & \sigma_{e2}^2 \end{bmatrix} \otimes I$, K being the genomic relationship matrix described below, and I the identity matrix.

The variance components were then estimated with the approach of Yamada *et al.* (1998):

$$\sigma_G^2 = \hat{\sigma}_{u12}, \sigma_{G \times E}^2 = \frac{1}{2}(\hat{\sigma}_{u1} + \hat{\sigma}_{u2}) - \hat{\sigma}_{u12} \text{ and } \sigma_\epsilon^2 = \frac{1}{2}(\hat{\sigma}_{e1}^2 + \hat{\sigma}_{e2}^2)$$

This decomposition was conducted on paired environments from two different sites. For Set1-2016 and Set2-2019, the two environments of each dataset were used to compute the variance components. For Set4-2018, only the data from a low input treatment were used. Finally, for Set4-

2019, data were paired by environments from the same treatment (intensive practices or low input) but from different sites (EM and LM).

From now on, we consider that Y_{ij} stands for the adjusted mean of the considered trait for the genotype i in environment j .

3.2.3. *Quality analysis of the genomic data*

Molecular marker results were checked for quality. Markers were eliminated if the minor allele frequency was less than 5%, or the heterozygosity rate or missing value rate were greater than 5%. A total of 5 824 SNPs for Set1-2016, 12 303 SNPs for Set4-2018, and 19 512 SNPs for both Set2-2019 and Set4-2019 remained after filtering. On average, 1.1% of the SNPs were missing and were imputed with the average allele frequency of the corresponding marker. All the data quality analysis was conducted with the sommer R package (Covarrubias-Pazarán 2016).

3.3. Estimation of the covariance matrices describing the similarities between genotypes or environments

The predictive models developed further in this study rely on covariance matrices to describe the covariance between the genotypes and between the GxE interactions. Relationship matrices between varieties were estimated based on molecular markers (kinship matrix K) or NIR spectra (hyperspectral relationship matrix H).

3.3.1. *Genomic and hyperspectral relationship matrices*

The kinship matrix K was computed following the Endelman and Jannink (2012) equation:

$$K = \frac{AA'}{2 \sum p_k(1 - p_k)}$$

where A is a centred genotypic matrix with dimensions $N_G \times M$, N_G is the number of genotypes, and M the number of molecular markers. For the i^{th} individual and the k^{th} marker, $A_{ik} = X_{ik} + 1 - 2p_k$ with X the genotype matrix, coded in $\{-1,0,1\}$ and p_k the frequency of allele 1 at marker k . K was computed with the rrBLUP R package (Endelman 2011).

The hyperspectral relationship matrix specific to environment j was computed as:

$$H_j = \frac{Sp_j^* Sp_j^{*'}}{L}$$

where Sp_j^* is the centred and scaled matrix of NIR spectra (dimension $N_G \times L$) from environment j , L the number of wavelengths, and j the environmental index with $j \in \{1, \dots, N_E\}$, N_E being the number of environments of the MET.

The combined hyperspectral relationship matrix was computed as

$$H_{cb} = \frac{S_{cb}^* S_{cb}^{*'}}{L}$$

where S_{cb}^* is the centred and scaled matrix of NIR spectra added one next to the other from NIR spectrum 1 matrix (S_1) to NIR spectrum N_E matrix (S_{N_E})

$$S_{cb} = (S_1 \quad \dots \quad S_{N_E})$$

S_{cb} had $N_G \times \sum_{j=1}^{N_E} \lambda_j$ dimensions where λ_j designates the number of wavelengths for the NIR spectrum j , and N_G the number of genotypes.

The K , H_j and H_{cb} matrices have the same dimensions ($N_G \times N_G$).

3.3.2. Genotype-by-environment covariance matrix

We then decided to characterise the covariance between all the genotype-environment combinations based on NIR spectra. It was shown in literature that environmental covariates can be used to estimate such covariances. We assumed that wavelengths constituting the NIR spectrum are similar to EC and relevant to describe GxE because they capture the expression of the genotype in a specific environment.

Each wavelength of a spectrum corresponds to an EC in which the absorbance is dependent on an environment j and a genotype i . The matrix of the ECs is then all the NIR spectra of the MET, called S_{all} with dimension $N_{GE} \times L$, N_{GE} corresponding to the number of genotype-environment combinations.

$$S_{all} = \begin{pmatrix} S_1 \\ \dots \\ S_{N_E} \end{pmatrix}$$

The GxE interaction similarity matrix P was computed based on the full matrix of NIR spectra S_{all} . First, we calculated the Euclidian distance ($D_{S_{\text{all}}}$) between the different combinations of genotype \times environment terms with S_{all} . Then, we estimated the genotype \times environment covariance matrix following the equation:

$$P = 1_{\text{GE}} - \frac{D_{S_{\text{all}}}}{\max(D_{S_{\text{all}}})}$$

where 1_{GE} is a matrix of 1 of size $(N_G \times N_E, N_G \times N_E)$. Here we assume that two genotype-environment combinations will covary if the two corresponding spectra are similar. The underlying hypotheses is that spectra capture the similarity between genotypes as well as the genetic similarity between environments.

K , H_j , H_{cb} and P were scaled to have a sample variance of 1 to avoid biased parameter estimations due to different scales (Kang *et al.* 2010; Forni *et al.* 2011).

3.4. Scenarios of prediction and corresponding cross-validation schemes

Six prediction scenarios were considered to simulate the different prediction objectives of genomic and phenomic selection (Figure 4-1). In four scenarios, all the NIR spectra are available for all varieties in each environment of the MET, whether phenotyped or not for the target trait. In the other two scenarios, one environment of the MET is the environment of reference where all the varieties have their NIR spectrum acquired, while in the other environments the NIRS are collected only on the varieties phenotyped for the target trait. The phenotype and NIR spectra of other varieties are hence missing in the non-reference environments. For each prediction scenario a cross-validation scheme is described below.

For the first category of scenarios in which NIR spectra are available for all varieties in each environment, **oGoE** corresponds to a sparse testing scenario. This scenario was tested in a CVRandom scheme with 5-fold cross-validation, which consisted in randomly splitting the MET data in five folds of the same size. **nGoE** corresponds to new varieties in observed environments, which mimics the evaluation of new varieties in the MET without any phenotyping of the target trait. Concretely, missing data were attributed in all the environments for the varieties designated as new. This scenario was tested in a CVNewG scheme with 5-fold cross-validations, in which

varieties are randomly split in five groups. **oGnE** corresponds to observed varieties in a new environment as it mimics the prediction in a new environment. The predicted environment is new in the sense that the target trait is not measured, but the NIR spectra are known. This scenario was tested in a CVNewE scheme with a leave-one-environment-out scheme, in which one environment of the MET is predicted with the others. Finally, **nGnE** corresponds to new varieties in a new environment. This scenario was tested in a CVNewGE scheme with a leave-one-environment-out scheme and a 5-fold cross-validation to determine the varieties considered as new.

For the second category of scenarios, in which NIR spectra are only available for the plots phenotyped for the target trait, **oGoEref** and **nGoEref** correspond to the same scenarios as **oGoE** and **nGoE** respectively, with the exception that one environment was randomly designated the environment of reference in which the phenotyping and NIR spectra are available for all varieties. In the other environments, the new varieties were not observed at all (neither for the target trait nor for NIRS). This mimics the situation where traditionally a breeder evaluates all the varieties in a single reference environment with sparse testing in the other environments. The scenario **oGoEref** was tested in a CVRandom_Ref scheme with one environment considered as reference and a 5-fold cross-validation, by randomly splitting the rest of the MET data in five folds of the same size. The scenario **nGoEref** was tested in a CVNewG_Ref scheme with one reference environment and a 5-fold cross-validation, by randomly splitting varieties in five folds of the same size.

To compare the performance of the different predictive models in the different cross-validation schemes, we calculated the predictive ability (PA) of each model as the Pearson's correlation coefficient between the predicted values and the adjusted means of the target trait (GY or HD) on the validation set, for each environment of prediction in the MET. For CVRandom, CVNewG and CVNewGE $5 \times N_E$ PA were calculated once the five folds were designated as the validation sets. For CVNewE, N_E PA were calculated because we applied a leave-one-environment-out method. In scenarios **oGoEref** and **nGoEref**, the PA was not calculated for the reference environment. For CVRandom_Ref and CVNewG_Ref, $5 \times (N_E - 1) \times N_E$ PA were then calculated on the 5-fold cross-validation multiplied by the number of environments, which are in turn considered as the new environment or reference environment. All the different cross-validation schemes were repeated 50 times to get stable results, except for the CVNewE, which is a simple leave-one-environment-out validation.

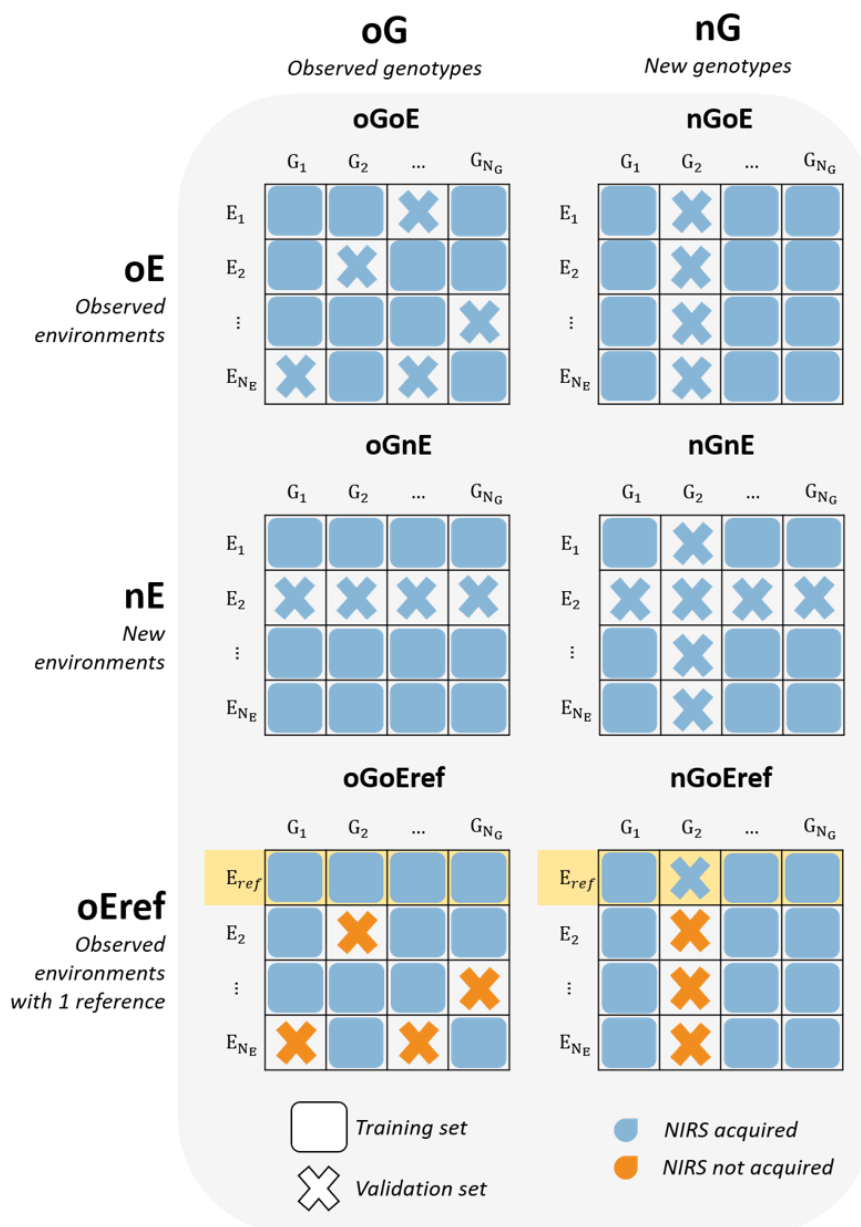


Figure 4-1. Schematic overview of the six prediction scenarios representing the associated missing data to be predicted in the MET. G_i corresponds to the genotype i and E_j corresponds to the environment j of the MET. Filled squares represent data of the training set and crosses represent data of the validation set. The training sets were always phenotyped for the target trait whereas the validation sets were not phenotyped for the target trait. Blue shading indicates that a NIR spectrum was acquired from genotype i in environment j whereas orange shading indicates that no NIR spectrum was acquired. A yellow background indicates the reference environment. oGoE corresponds to prediction in a sparse testing scenario where some combinations of genotypes and environments are unobserved. oGnE corresponds to a new environment in which no phenotype is available. nGoE corresponds to the prediction of new genotypes in observed environments. nGnE corresponds to new genotypes to be predicted in a new environment. For these four scenarios, spectra are acquired on both the training and validation sets. oGoEref and nGoEref correspond to two scenarios where one environment is considered as a reference where NIR is acquired on all genotypes. oGoEref therefore is similar to scenario oGoE except that NIR spectra are not acquired on the validation set, while nGoEref is similar to scenario nGoE except that NIR spectra are not acquired on the validation set apart from in the reference environment.

3.5. Models developed for the different multi-environment prediction scenarios

3.5.1. Genomic prediction models

In the different prediction scenarios, we compared models with different levels of complexity, using molecular markers, NIR spectra or both.

For genomic prediction, we used two reference models based on the genomic relationship matrix (kinship) allowing to share information between varieties but not between environments (Jarquín *et al.* 2014).

$$\mathbf{EG}: Y_{ij} = \mu + E_j + G_i + \epsilon_{ij} \quad (3)$$

$$\mathbf{EG_GxE}: Y_{ij} = \mu + E_j + G_i + GE_{ij} + \epsilon_{ij} \quad (4)$$

Y_{ij} is the phenotype for variety i in environment j ; μ is the intercept; G_i is a random polygenic effect of the variety i following a normal distribution $G \sim N(0, K\sigma_G^2)$; and GE_{ij} is a random effect corresponding to the interaction between variety i and environment j following a normal distribution $GE \sim N(0, [ZKZ'] \odot [Z_E Z_E'] \sigma_{GE}^2)$. So, in this model no information is shared between environments in the GE term. K is the relationship matrix calculated with molecular markers (see above), Z the incidence matrix for the genetic effect and Z_E the incidence matrix for the effects of the environments. \odot corresponds to the Hadamard product. Finally, ϵ_i is the random residual effect that follows the normal distribution $\epsilon_{ij} \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma_\epsilon^2)$. We assumed a same residual variance for all environments. The different random effects of the models are assumed to be independent.

3.5.2. Single-NIRS phenomic models

For phenomic prediction we used two kinds of predictive models, involving a NIR spectrum acquired in only one environment (Single-NIRS) or from multiple NIR spectra acquired in different environments of the MET (Multi-NIRS). Like the genomic prediction Models (3) and (4), we compared two reference models based on the hyperspectral relationship matrix computed from a single environment.

$$\mathbf{EW}: Y_{ij} = \mu + E_j + W_i + \epsilon_{ij} \quad (5)$$

$$\mathbf{EW_WxE}: Y_{ij} = \mu + E_j + W_i + WE_{ij} + \epsilon_{ij} \quad (6)$$

Y_{ij} is the phenotype for variety i in environment j ; μ is the intercept; W_i is a random genetic effect of the variety i following a normal distribution $W \sim N(0, H_S \sigma_W^2)$; and WE_{ij} is a random effect corresponding to the interaction between variety i and environment j following a normal distribution $WE \sim N(0, [ZH_S Z'] \odot [Z_E Z_E'] \sigma_{GE}^2)$. H_S is the relationship matrix calculated with NIR spectra (see above) from the environment S with $S \in \{1, \dots, N_E\}$. Finally, ϵ_i is the random residual effect following a normal distribution $\epsilon_{ij} \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma_\epsilon^2)$. The different random effects of the models are assumed to be independent. In Models (5) and (6) the H_S matrix of each environment was tested successively, resulting in N_E Models (5) and (6).

We tested several other models to evaluate how well NIR spectra capture and predict GxE interactions. First, we specified in the WE_{ij} covariance matrix a block-diagonal structure (Malosetti *et al.* 2016), in which each block corresponds to the H_j matrix of the corresponding environment. In this model we consider that within one environment, varieties covary according to their hyperspectral similarity in the same environment. The assumption is that NIR spectra capture information on local adaptation. The corresponding model is:

$$\mathbf{EW_W_{diag}xE}: Y_{ij} = \mu + E_j + W_i + WE_{d_{ij}} + \epsilon_{ij} \quad (7)$$

The description of the different effects is the same as in Model (6), and $WE_{d_{ij}}$ follows a normal

distribution $WE_{d_{ij}} \sim N(0, H_{diag} \sigma_{WE_{d_{ij}}}^2)$, with $H_{diag} = \begin{pmatrix} H_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & H_{N_E} \end{pmatrix}$ and H_1 to H_{N_E} correspond to

the N_E hyperspectral relationship matrices available in the N_E environments of the MET.

Secondly, we described GxE by using the P similarity matrix based on all the NIR spectra:

$$\mathbf{EW_P}: Y_{ij} = \mu + E_j + W_i + WE_{p_{ij}} + \epsilon_{ij} \quad (8)$$

The description of the different effects is the same as in Model (6), and $WE_{p_{ij}}$ follows a normal distribution $WE_{p_{ij}} \sim N(0, P \sigma_{WE_{p_{ij}}}^2)$. P is the genotype-by-environment similarity matrix computed with the NIR spectra (see above).

3.5.3. Multi-NIRS phenomic models

The next objective was to assess the effect of using a hyperspectral relationship matrix combining NIR spectra acquired from different environments to predict GY and HD. To do this, we used the H_{cb} matrix (the hyperspectral relationship matrix computed with S_{cb}) described above. Multi-NIRS models were adapted from the single-NIRS Models (5), (6), (7) and (8) with models called \mathbf{EW}_{cb} , $\mathbf{EW}_{cb_W_{cb}xE}$, $\mathbf{EW}_{cb_W_{diag}xE}$, and \mathbf{EW}_{cb_P} respectively.

3.5.4. Multi-omics models

We developed models combining information from molecular markers and NIR spectra. We tested different models including genetic effects involving covariance matrices K or H_{cb} and GxE involving covariance matrices K , H_{cb} , H_{diag} or P .

$$\mathbf{EG_P}: Y_{ij} = \mu + E_j + G_i + P_{ij} + \epsilon_{ij} \quad (9)$$

$$\mathbf{EGW}_{cb}: Y_{ij} = \mu + E_j + G_i + W_{cb_i} + \epsilon_{ij} \quad (10)$$

$$\mathbf{EGW}_{cb_GxE}: Y_{ij} = \mu + E_j + G_i + W_{cb_i} + GE_{ij} + \epsilon_{ij} \quad (11)$$

$$\mathbf{EGW}_{cb_W_{cb}xE}: Y_{ij} = \mu + E_j + G_i + W_{cb_i} + W_{cb}E_{ij} + \epsilon_{ij} \quad (12)$$

$$\mathbf{EGW}_{cb_GxE_W_{cb}xE}: Y_{ij} = \mu + E_j + G_i + W_{cb_i} + GE_{ij} + W_{cb}E_{ij} + \epsilon_{ij} \quad (13)$$

The description of the different effects is the same as in the previous Models, and W_{cb_i} is a random genetic effect of the variety i following a normal distribution $W_{cb} \sim N(0, H_{cb} \sigma_W^2)$, $W_{cb}E_{ij}$ is a random effect corresponding to the interaction between variety i and environment j following a normal distribution $W_{cb}E \sim N(0, [ZH_{cb}Z'] \odot [Z_E Z_E'] \sigma_{GE}^2)$. H_{cb} is the relationship matrix calculated with S_{cb} NIR spectra. Finally, ϵ_i is the random residual effect which follows the normal distribution $\epsilon_{ij} \stackrel{i.i.d.}{\sim} N(0, \sigma_\epsilon^2)$. The different random effects of the models are assumed to be independent.

Alternatives to models (11) and (12) were defined by replacing $W_{cb}E_{ij}$ with $\mathbf{W}_{diag}xE$ (models $\mathbf{EGW}_{cb_W_{diag}xE}$ and $\mathbf{EGW}_{cb_GxE_W_{diag}xE}$), or with $WE_p \sim N(0, P \sigma_{WEp}^2)$ (models \mathbf{EGW}_{cb_P} and $\mathbf{EGW}_{cb_GxE_P}$).

All the models presented here were fitted with the R package BGLR (Pérez and de los Campos 2014) with a burn-in of 2 000 iterations on a running total of 10 000 iterations to reach convergence. Starting parameters for df_0 and S_0 were chosen as 5 and 0.5 by default. All the different models are described in Table 4-1.

3.6. Imputation of missing spectra for unphenotyped varieties

All the models developed above assume that NIR spectra have been obtained from all the varieties, whether phenotyped or not, in all the environments of the MET. However, in more ambitious scenarios such as **oGoEref** and **nGoEref**, the genotype-environment combinations to be predicted are not characterised by NIRS, except in the reference environment. It is necessary to predict the missing spectra for the models involving local NIR spectra (Multi-NIRS and Multi-omics).

We used two approaches to predict the missing NIR spectra of the genotype-environment combinations to be predicted using the spectra of all genotypes in the reference environment, and the spectra of the training genotypes in the environment to be predicted. These predicted spectra were then used to compute the environment-specific or the combined hyperspectral relationship matrices compatible with all the phenomic models presented above. For the single-NIRS models, the NIR spectra were acquired on all genotypes only in the reference environment. So, Models (5) and (6) were adapted using H_{ref} instead of H_S , leading to models \mathbf{EW}_{ref} and $\mathbf{EW}_{ref} - \mathbf{W}_{ref} \mathbf{x} \mathbf{E}$, respectively. For multi-NIRS models and the modelling of the GxE interaction with local or combined spectra, missing spectra were added by imputation using one of two methods.

The first method is a nearest neighbour (NN) method. For any given environment each missing genotype is attributed the same spectrum as the most similar genotype. This similarity is based on the lowest Euclidian distance between the NIR spectra in the environment of reference (Figure S1).

The second method is a weighted multivariate ridge regression (WMRR) following the equation:

$$Y = X\beta + \epsilon \quad (13)$$

Y is the matrix of spectra to be predicted and X the matrix of spectra from the reference environment, and both matrices have dimensions $N_G \times L$. β is the matrix of estimated wavelength effects with dimensions $L \times L$, while ϵ is the matrix of random error effect such as $\text{vec}(\epsilon) \sim N(0, \sigma_\epsilon^2 I_{N_G \times L})$.

We weighted the wavelengths so that some are more explicative than others. For this, we applied a ridge regression to the training set to predict the target trait in the environment to be predicted. This was possible because NIR spectra were acquired on the training set in the environment to be predicted. We then used the estimated β (absolute value) to define a weight matrix Q :

$$Q = \begin{pmatrix} q_1^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & q_L^2 \end{pmatrix}$$

To minimise the norm: $\|Y - XB\|^2 = \|YQ - XBQ\|^2$, the estimator B was calculated with the following equation:

$$\hat{B} = [(Q \otimes X)^T (Q \otimes X) + \lambda I_{L \times L}]^{-1} (Q \otimes X)^T (Q \otimes I_{N_G}) Y_{vec}$$

\otimes stands for the Kronecker product.

The shrinkage parameter λ was selected based on a 10-fold cross-validation. Multivariate ridge regression and estimation of the weights were run with the *glmnet* package (Friedman *et al.* 2010).

Contrary to the NN method, the WMRR predicts a new spectrum for all the genotypes (training and validation sets) in the specific environment. The predicted spectra of both calibration and validation sets were used to compute the hyperspectral matrix.

Table 4-1. Summary of all genomic, phenomic and multi-omics models. For each model the different effects and associated covariances matrices are indicated, as well as the prediction scenarios in which they were tested. Effects were either included (×) or excluded (-). For random effects, the covariance matrices are indicated. Models were tested on of three scenarios (O, NN, WMRR) or not (-). For **oGoE**, **nGoE**, **oGnE** and **nGnE**, models used a H matrix computed on observed NIRS (O), whereas for **oGoEref** and **nGoEref**, the multi-NIRS models and multi-omics models, used an H matrix computed on all spectra imputed by the nearest neighbour (NN) method or by the weighted multivariate ridge regression (WMRR).

Models	Model effects					Prediction scenarios					
	E	G	W	G×E	W×E	oG oE	nG oE	oG nE	nG nE	oG oEref	nG oEref
Genomic models											
EG	×	K	-	-	-	O	O	O	O	O	O
EG_GxE	×	K	-	$[ZKZ'] \odot [Z_E Z_E']$	-	O	O	-	-	O	O
Single NIRS models											
EW	×	-	H_S	-	-	O	O	O	O	-	-
EW_WxE	×	-	H_S	-	$[ZH_S Z'] \odot [Z_E Z_E']$	O	O	-	-	-	-
EW_WdiagxE	×	-	H_S	-	H_{diag}	O	O	-	-	-	-
EW_P	×	-	H_S	-	P	O	O	O	O	-	-
EW_ref	×	-	H_{ref}	-	-	-	-	-	-	O	O
EW_ref_WrefxE	×	-	H_{ref}	-	$[ZH_{ref} Z'] \odot [Z_E Z_E']$	-	-	-	-	O	O
Multi-NIRS models											
EW _{cb}	×	-	H_{cb}	-	-	O	O	O	O	NN / WMRR	NN / WMRR
EW _{cb} -W _{cb} xE	×	-	H_{cb}	-	$[ZH_{cb} Z'] \odot [Z_E Z_E']$	O	O	-	-	NN / WMRR	NN / WMRR
EW _{cb} -W _{diag} xE	×	-	H_{cb}	-	H_{diag}	O	O	-	-	NN / WMRR	NN / WMRR
EW _{cb} -P	×	-	H_{cb}	-	P	O	O	O	O	NN / WMRR	NN / WMRR
Multi-omics models											
EG_P	×	K	-	-	P	O	O	O	O	NN / WMRR	NN / WMRR
EGW _{cb}	×	K	H_{cb}	-	-	O	O	O	O	NN / WMRR	NN / WMRR
EGW _{cb} -GxE	×	K	H_{cb}	$[ZKZ'] \odot [Z_E Z_E']$	-	O	O	-	-	NN / WMRR	NN / WMRR
EGW _{cb} -W _{cb} xE	×	K	H_{cb}	-	$[ZH_{cb} Z'] \odot [Z_E Z_E']$	O	O	-	-	NN / WMRR	NN / WMRR
EGW _{cb} -W _{diag} xE	×	K	H_{cb}	-	H_{diag}	O	O	-	-	NN / WMRR	NN / WMRR
EGW _{cb} -P	×	K	H_{cb}	-	P	O	O	O	O	NN / WMRR	NN / WMRR
EGW _{cb} -GxE-W _{cb} xE	×	K	H_{cb}	$[ZKZ'] \odot [Z_E Z_E']$	$[ZH_{cb} Z'] \odot [Z_E Z_E']$	O	O	-	-	NN / WMRR	NN / WMRR
EGW _{cb} -GxE-W _{diag} xE	×	K	H_{cb}	$[ZKZ'] \odot [Z_E Z_E']$	H_{diag}	O	O	-	-	NN / WMRR	NN / WMRR
EGW _{cb} -GxE-P	×	K	H_{cb}	$[ZKZ'] \odot [Z_E Z_E']$	P	O	O	-	-	NN / WMRR	NN / WMRR

4. Results

4.1. Characterisation of the METs based on phenotypes

We characterised and compared four previously established sets of data from METS of two winter wheat breeding programs in France. Within each dataset, correlations were calculated between the adjusted means for GY for each pair of environments in the MET, that is, two environments in each of Set1-2016 and Set2-2019, three for Set4-2018, and four for Set4-2019 (Figure 4-2). GY correlations between environments were very variable, from as low as 0.01 in Set1 to a maximum of 0.72 in Set4-2019. Globally in Set4-2019, the MET with the most environments, the correlations ranged from 0.21 to 0.72. Where possible, correlations were also calculated for the adjusted means for HD between each pair of environments within a MET (Figure 4-2). Unlike GY, the HD correlations were strong and did not vary widely. For example, in Set4-2019 the correlation ranged from 0.89 to 0.97.

For Set4-2019, the variance in GY and HD was decomposed using Equation (1). For Set4-2018, only the variance for GY was decomposed (Table S1, B). The other two sets did not have any replicates for GY and HD so the decomposition of variance could not be done. For GY, the variance in the GxE effect in the two Set4 sets was low for Set4-2019 (12.81) and high for Set4-2018 (47.58) compared to the genetic effect (16.77 and 12.76, respectively). For GY, the environment explained most of the variance. For HD, the variance was mainly explained by the genetic effect.

4.2. Decomposition of the variance along NIR spectrum

We decomposed the genetic variance for all spectra based on the genomic kinship (Figure 4-3). Spectra from Set1-2016 had the lowest proportion of genomic variance (18% on average) and Set2-2019 had the highest proportion (67% on average). In all sets, the total variance in absorbances included a non-negligible proportion of GxE variance. Two datasets, Set1-2016 and Set4-2018-LI had on average, more GxE variance (30% and 28%, respectively) than genetic variance (18% and 23%, respectively).

We also decomposed the variance of the absorbances based on Equation (1) and using NIRS acquisition replicates for Set4-2018 and Set4-2019 (Figure S2). The E effect explains most of the absorbance variances. On average across all wavelengths, the E variance represented 56.8%

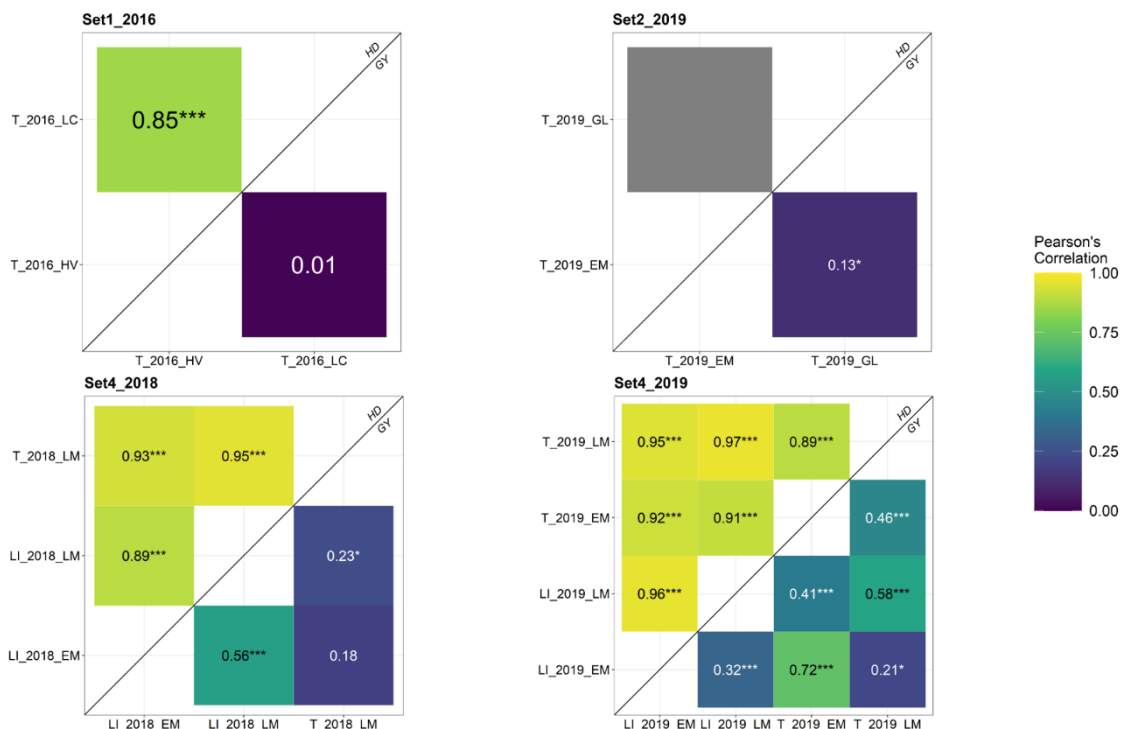


Figure 4-2. Correlation coefficient matrices between trial environments for heading date adjusted means (upper left) and grain yield adjusted means (lower right). A trial environment is defined as a combination of treatment × year × site. Asterisks indicate the significance level: *P-value <0.05, **P-value <0.01 and ***P-value <0.001.

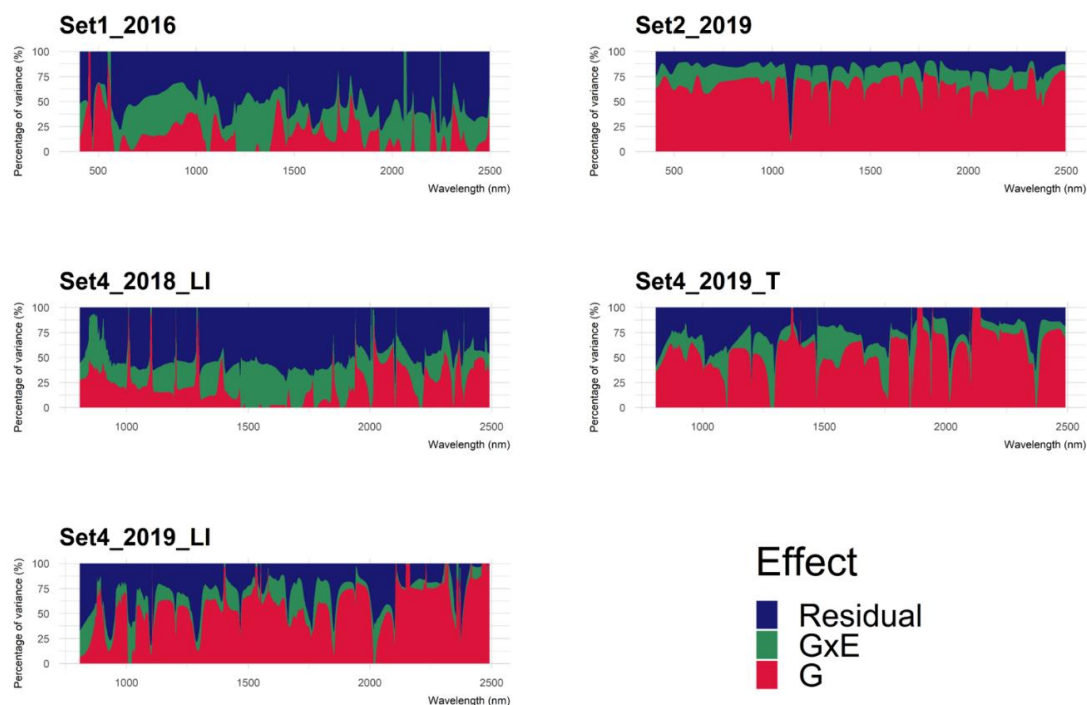


Figure 4-3. Proportion of the genomic (G), the genomic x environment (GxE) and residual variances across the NIR spectra of winter wheat grains from 5 different MET sets.

4.2.1. Comparison of the scenarios oGoE and nGoE

For the oGoE and nGoE scenarios NIRS data were available for all genotype-environment combinations, and predictions were sought for certain unobserved combinations or new genotypes in all environments, respectively. The PA of models for GY were highly variable between each other, between the two scenarios and between datasets. Taking the average PA for the four datasets and comparing CVRandom (Table 4-2) and CVNewG (Table 4-3) validation scenarios, the highest PA were reached by the models $EGW_{cb_GxE_W_{cb}xE}$ and $EGW_{cb_GxE_W_{diag}xE}$ in both the CVRandom (range 0.4-0.68) and in the CVNewG (range 0.42-0.68) validations. The PA of the models for HD were higher than the PA of the models for GY in oGoE and nGoE scenarios and for the three datasets. On average, several models reached very high PA in validation scenario CVRandom (range 0.81-0.94) as well as high PA in CVNewG (0.57-0.74).

We considered first the models without a GxE effect. On average, for the prediction of GY, single NIRS model EW gave similar PA as multi-NIRS model EW_{cb} in both scenarios. For the prediction of HD, EW_{cb} gave higher PA than EW with an average gain of 16%. The best phenomic models in the CVRandom scenario gave similar PA to genomic models for the prediction of GY and HD. In scenario CVNewG, the best phenomic model was better than the best genomic model for the prediction of GY (0.44 versus 0.38) and for the prediction of HD (0.51 versus 0.43). Combining information from molecular markers and NIR spectra gave the best PA for GY in both scenarios (0.54 for CVRandom and 0.50 for CVNewG) and was equivalent to genomic and phenomic models for HD in scenario CVRandom. For CVNewG, multi-omics models performed better than genomic and phenomic models for the prediction of HD (0.64 versus 0.43 and 0.51).

For GY, the predictive models that included the GxE effect always gave better predictions than those without. For the prediction of HD, the PA were similar for both kinds of models. We therefore focused on GY prediction only. In genomic models, a single covariance matrix was used to integrate the GxE effect, which increased the gain in PA by 24% on average. For phenomic models integrating the GxE effect we compared the results of using different covariance matrix structures. In single NIRS models, the model with a diagonal structure $EW_W_{diag}xE$ gave the best PA for both scenarios (0.44 and 0.43, respectively), compared to other covariance structures (EW_WxE and EW_P). In multi-NIRS models, all the models with a GxE effect performed similarly. Finally, the multi-omics models with higher PA were the ones combining all main genetic effects (G and W_{cb}) and all the GxE interaction effects (GxE and $W_{cb}xE$ or $W_{diag}xE$).

4.2.2. Comparison in the scenarios oGnE and nGnE

In the scenarios oGnE and nGnE, there are NIRS data for every genotype-environment combination but some environments do not have any observations, and for nGnE some genotypes are not observed. The PA for GY were much lower in oGnE (Tables S2) and nGnE (Tables S3) than in oGoE and nGoE. Considering the average of the different sets and comparing CVNewE and CVNewGE validation scenarios, the highest PA were reached with the multi-omics model EG_P (0.32 and 0.27, respectively), higher than with the EG model (0.24 and 0.16, respectively). Models integrating the GxE effect (modelled by P) had better PA than those without it, with an average gain of 36% and 57% for the corresponding scenarios. EG_GxE was not compared to the other models because the GxE could not be estimated. This was also the case for models with $W_{cb} \times E$ and $W_{diag} \times E$. The PA for HD were similar in scenarios oGnE and nGnE as to scenarios oGoE and nGoE. The best models were EG_ W_{cb} and EG_ W_{cb} _P with PA reaching 0.91 for CVNewE and around 0.62 for CVNewGE. Adding the GxE effect in the models did not increase the PA for HD.

4.3. Comparison of genomic, phenomic and multi-omics models in two original MET scenarios for the prediction of GY and HD

We compared the previous developed models in two original scenarios oGoEref (Table 4-4) and nGoEref (Table 4-5) where the NIR spectra were missing for the validation set in the environments for which GY was to be predicted. We imputed the missing spectra by the nearest neighbour (NN) method or by weighted multivariate ridge regressions (WMRR). The WMRR gave, on average, better PA for multi-NIRS and multi-omics models in both scenarios.

For the best multi-NIRS models, the average gain using WMRR instead of the NN spectra was around 38% and 82% in CVRandom_Ref and CVNewG_Ref, respectively. For the best multi-omics models, the gain was around 12% and 19%. The results of both these imputing methods were lower than the optimum for both CV schemes when all the spectra were known. For the best multi-NIRS model with WMRR, the loss in PA was around 28% for both scenarios and for the best multi-omics model, the loss in PA was around 11% for CVRandom_Ref and 14% for CVNewG_Ref. For the prediction of HD, the two methods of imputation performed similarly for CVRandom_Ref while WMRR performed better than NN for CVNewG_Ref (Tables S4-S5).

Using multi-NIRS models with imputed spectra gave slightly better PA than using the single NIRS

model based on spectra of reference, while using multi-omics models gave slightly better PA than genomic models. The best multi-omics models integrated all the G, W, GxE and WxE effects (Tables 4-4, Table 4-5) in both scenarios. On average, $EGW_{cb_GxE_W_{cb_xE}}$ with predicted spectra gave the best PA for both scenarios, 10% better than genomic models. Models including the GxE effect gave higher PA than those without. The different types of covariance structures to model the GxE effect performed similarly except for the P covariance matrix which gave slightly lower PA than the diagonal or combined structures. Finally, the PA of the best model in CVRandom_Ref and in CVNewG_Ref were lower than those of the best model in CVRandom and CVNewG with a decrease of 11% and 14% respectively.

Table 4-2. Predictive abilities of the different genomic, single NIRS, multi-NIRS and multi-omics models in scenario oGoE for GY and HD. CVRandom is a five-fold cross-validation scheme in which random data were missing from the dataset. This CV corresponds to a sparse-testing design. Averaged predictive abilities based on 50 repetitions are presented for each set, and averaged over all sets for GY or three sets for HD. The best predictive abilities by set and by model category (based on Average) are shown in bold.

<i>oGoE scenario (CVRandom)</i>	<i>GY prediction</i>					<i>HD prediction</i>			
	Set1 2016	Set2 2019	Set4 2018	Set4 2019	Average	Set2 2019	Set4 2018	Set4 2019	Average
Genomic models									
EG	0.13	0.35	0.31	0.55	0.34	0.76	0.91	0.94	0.87
EG_GxE	0.17	0.55	0.44	0.6	0.44	0.72	0.92	0.94	0.86
Single NIRS models									
EW	0.19	0.38	0.28	0.38	0.31	0.69	0.79	0.93	0.80
EW_WxE	0.29	0.46	0.28	0.41	0.36	0.67	0.73	0.93	0.78
EW_W _{diag} xE	0.36	0.59	0.36	0.45	0.44	0.69	0.68	0.92	0.76
EW_P	0.35	0.57	0.32	0.43	0.42	0.70	0.75	0.93	0.79
Multi-NIRS models									
EW _{cb}	0.21	0.43	0.33	0.45	0.36	0.78	0.87	0.93	0.86
EW _{cb} _W _{cb} xE	0.38	0.6	0.33	0.49	0.45	0.76	0.88	0.93	0.86
EW _{cb} _W _{diag} xE	0.38	0.6	0.37	0.49	0.46	0.76	0.87	0.93	0.85
EW _{cb} _P	0.35	0.58	0.33	0.46	0.43	0.78	0.88	0.93	0.86
Multi-omics models									
EG_P	0.32	0.57	0.35	0.59	0.46	0.76	0.92	0.94	0.87
EGW _{cb}	0.18	0.43	0.32	0.57	0.38	0.83	0.91	0.94	0.89
EGW _{cb} _W _{cb} xE	0.37	0.61	0.35	0.64	0.49	0.82	0.93	0.94	0.9
EGW _{cb} _W _{diag} xE	0.36	0.61	0.4	0.64	0.5	0.82	0.92	0.94	0.89
EGW _{cb} _P	0.33	0.58	0.34	0.59	0.46	0.82	0.92	0.94	0.89
EGW _{cb} _GxE	0.22	0.6	0.43	0.62	0.47	0.82	0.92	0.94	0.89
EGW _{cb} _GxE_W _{cb} xE	0.4	0.67	0.43	0.66	0.54	0.81	0.93	0.94	0.89
EGW _{cb} _GxE_W _{diag} xE	0.38	0.68	0.45	0.66	0.54	0.81	0.92	0.94	0.89
EGW _{cb} _GxE_P	0.34	0.66	0.43	0.63	0.52	0.81	0.92	0.94	0.89

Table 4-3. Predictive abilities of genomic, single NIRS, multi-NIRS and multi-omics models in scenario **nGoE** for GY and HD. CVNewG is a five-fold cross-validation scheme in which random genotypes are unphenotyped in all the environments of the trial network. Predictive abilities based on 50 repetitions are presented by set and are averaged over all sets for GY or three sets for HD. The best predictive abilities by set and model category (based on Average) is shown in bold.

<i>nGoE scenario (CVNewG)</i>	<i>GY prediction</i>					<i>HD prediction</i>			
	Set1 2016	Set2 2019	Set4 2018	Set4 2019	Average	Set2 2019	Set4 2018	Set4 2019	Average
Genomic models									
EG	0.25	0.40	0.24	0.37	0.32	0.36	0.42	0.52	0.43
EG_GxE	0.23	0.54	0.34	0.43	0.38	0.36	0.42	0.52	0.43
Single NIRS models									
EW	0.22	0.38	0.22	0.31	0.28	0.63	0.29	0.32	0.41
EW_WxE	0.30	0.46	0.25	0.37	0.34	0.62	0.30	0.32	0.41
EW_W _{diag} xE	0.37	0.59	0.34	0.41	0.43	0.66	0.31	0.32	0.43
EW_P	0.36	0.57	0.28	0.36	0.39	0.64	0.30	0.32	0.42
Multi-NIRS models									
EW _{cb}	0.25	0.44	0.27	0.36	0.33	0.73	0.36	0.43	0.51
EW _{cb} -W _{cb} xE	0.39	0.59	0.30	0.43	0.43	0.72	0.36	0.42	0.50
EW _{cb} -W _{diag} xE	0.38	0.60	0.34	0.42	0.44	0.73	0.36	0.43	0.51
EW _{cb} -P	0.38	0.58	0.29	0.37	0.40	0.72	0.37	0.42	0.50
Multi-omics models									
EG_P	0.38	0.59	0.31	0.46	0.44	0.43	0.4	0.54	0.46
EGW _{cb}	0.29	0.48	0.26	0.47	0.38	0.74	0.58	0.60	0.64
EGW _{cb} -W _{cb} xE	0.41	0.63	0.29	0.53	0.46	0.74	0.58	0.60	0.64
EGW _{cb} -W _{diag} xE	0.40	0.63	0.33	0.53	0.47	0.75	0.56	0.59	0.63
EGW _{cb} -P	0.39	0.60	0.29	0.48	0.44	0.74	0.57	0.59	0.63
EGW _{cb} -GxE	0.29	0.61	0.34	0.52	0.44	0.74	0.57	0.59	0.63
EGW _{cb} -GxE-W _{cb} xE	0.42	0.67	0.36	0.56	0.50	0.74	0.57	0.59	0.63
EGW _{cb} -GxE-W _{diag} xE	0.41	0.68	0.38	0.55	0.50	0.74	0.56	0.59	0.63
EGW _{cb} -GxE-P	0.38	0.66	0.36	0.52	0.48	0.74	0.56	0.59	0.63

5. Discussion

5.1. Phenomic prediction and genomic prediction performed similarly and were outperformed by multi-omics prediction in MET designs

We studied the PA of genomic, phenomic and multi-omics models in classic MET designs for the prediction of GY and HD. We adapted the models proposed by Robert *et al.* (2022a) to the MET context, by modelling the genetic and environmental effects, and in some models the G×E effect too. For the prediction of GY, we found that single-NIRS models gave slightly lower PA than multi-NIRS models. For the prediction of HD, the single-NIRS models gave much lower PA than multi-NIRS models. These results are consistent with the results found by Robert *et al.* (2022a), in which combining several spectra from different environments increased the PA for HD, but not necessarily for GY. We suppose that this difference is due to different effects on the variance of the trait, being mainly additive for HD and additive and interactive (G×E) for GY. Also, the relationship between spectra and HD is more likely to be similar in the different environments of the MET knowing that the phenotypic correlations were very strong between environments for this trait (Figure 4-2). GY is strongly influenced by the response of the plants to the specific environmental conditions in which they are grown. NIR spectra are apparently able to capture part of this response, so may be useful in predicting G×E.

We compared the phenomic to the genomic prediction models in each MET scenario. In scenarios oGoE and oGnE, there was little difference in PA of phenomic and genomic prediction models for either trait. These results are consistent with those of Krause *et al.* (2019) who found that predictions of GY with NIR spectra in a multi-environment context were similar to the predictions with molecular markers. In scenarios nGoE and nGnE, when the behaviour of a new genotype (nG) has to be predicted, no phenotyping is available in a MET. Phenomic prediction gave slightly better PA than genomic prediction for both traits. The genomic models appeared to be more sensitive to missing information than the phenomic models. In single environment predictions, several studies also found that PA of models using NIR spectra were similar or higher than for models using molecular markers (Rincent *et al.* 2018; Krause *et al.* 2019; Galán *et al.* 2020; Zhu *et al.* 2021b; Robert *et al.* 2022a). In the present study, the multi-omics models combining information from molecular markers and NIR spectra were the best models whatever the MET scenario. Krause *et al.* (2019) also reported that the multi-kernel model performed better than single kernel models. In addition, Galán *et al.* (2021) found that a bivariate model across

environments, incorporating molecular markers, height and NIR spectra gave better PA than the G-BLUP or H-BLUP alone.

5.2. Modelling the G×E effect with NIR spectra improves the ability to predict grain yield

Modelling the G×E effect is challenging for complex traits like GY in multiple environments, because it is influenced by both additive genetic and G×E effects (Burgueño *et al.* 2012). We proposed different ways to model G×E based on NIR spectra. Our results showed that modelling G×E with molecular markers or with NIR spectra always improved the PA for GY, consistent with several reports from the past decade (Heslot *et al.* 2013; Jarquín *et al.* 2014, 2017; Lopez-Cruz *et al.* 2015; Lado *et al.* 2016; Cuevas *et al.* 2016; Pérez-Rodríguez *et al.* 2017; Ly *et al.* 2018; Rincent *et al.* 2019; Robert *et al.* 2020). However, for HD there was no improvement in PA, probably because the interaction effect only has a slight influence on the trait variance. This was not unexpected considering the very high correlation for HD between environments (Figure 4-2) and the genetic decomposition of variance (Table S1b). Indeed, all experiments in the MET were autumn sown and we do not expect a high G×E for winter wheat genotypes in these conditions.

Different covariance structures computed with NIR spectra ($W_{cb \times E}$, $W_{diag \times E}$ and P) were used to model the G×E effect. In scenarios oGoE and nGoE, $W_{cb \times E}$ and $W_{diag \times E}$ performed similarly, and models with P had a slightly lower PA. These results showed that considering all the NIR spectra in the same matrix (W_{cb}) or specifying each NIR spectra for each environment (W_{diag}) contribute the same information towards describing the W×E effect, while no information is shared between environments. By contrast, the P structure considers the covariance between each genotype-environment combination of the MET, allowing the sharing of information between environments using NIR spectra. Our results showed the P matrix to be less efficient than the other structures in modelling the G×E interaction. However, in scenarios in which a new environment is predicted (oGnE and nGnE), only the P matrix allows sharing information between environments. The P matrix can be used in these scenarios to improve the predictive ability of the model. We found that models with this G×E effect gave better PA than models without it, which confirms that NIR spectra capture part of the genetic similarity between environments.

Table 4-4. Predictive abilities of the genomic, single NIRS, multi-NIRS and multi-omics models in scenario **oGoEref** for GY. CVRandom_Ref is a leave-one-environment-out scheme followed by a five-fold cross-validation in which there is sparse testing in all the environments of the MET except the reference environment. Predictive abilities based on 50 repetitions are presented for each set and as the average over all sets. The optimum models are shown in red type, and the best predictive ability by set and by model category (based on Average) in bold type.

<i>oGoEref scenario</i> (CVRandom_Ref)	<i>GY prediction</i>				
	Set1 2016	Set2 2019	Set4 2018	Set4 2019	Average
Models					
Genomic models					
EG	0.12	0.31	0.32	0.55	0.32
EG_GxE	0.17	0.54	0.46	0.6	0.44
Single NIRS models					
EW _{ref}	0.11	0.27	0.27	0.38	0.26
EW _{ref} -W _{ref} xE	0.24	0.35	0.24	0.40	0.31
Multi-NIRS models					
<i>Optimum</i>					
EW _{cb}	0.17	0.40	0.36	0.45	0.35
EW _{cb} -W _{cb} xE	0.38	0.60	0.33	0.49	0.45
EW _{cb} -W _{diag} xE	0.37	0.60	0.37	0.50	0.46
EW _{cb} -P	0.34	0.55	0.34	0.48	0.43
<i>Nearest Neighbour</i>					
EW _{cb}	0.07	0.19	0.28	0.40	0.24
EW _{cb} -W _{cb} xE	0.18	0.20	0.21	0.38	0.24
EW _{cb} -W _{diag} xE	0.12	0.16	0.14	0.35	0.19
EW _{cb} -P	0.06	0.16	0.21	0.33	0.19
<i>Weighted Multivariate Ridge Regression</i>					
EW _{cb}	0.13	0.28	0.27	0.38	0.26
EW _{cb} -W _{cb} xE	0.28	0.40	0.24	0.41	0.33
EW _{cb} -W _{diag} xE	0.27	0.41	0.24	0.41	0.33
EW _{cb} -P	0.22	0.33	0.25	0.40	0.30
Multi-omics models					
<i>Optimum</i>					
EGW _{cb} -GxE	0.19	0.59	0.46	0.62	0.46
EGW _{cb} -GxE-W _{cb} xE	0.39	0.67	0.44	0.66	0.54
EGW _{cb} -GxE-W _{diag} xE	0.37	0.68	0.45	0.66	0.54
EGW _{cb} -GxE-P	0.33	0.65	0.44	0.64	0.52
<i>Nearest Neighbour</i>					
EGW _{cb} -GxE	0.13	0.52	0.45	0.61	0.43
EGW _{cb} -GxE-W _{cb} xE	0.21	0.41	0.40	0.61	0.41
EGW _{cb} -GxE-W _{diag} xE	0.18	0.38	0.35	0.57	0.37
EGW _{cb} -GxE-P	0.12	0.42	0.42	0.57	0.38
<i>Weighted Multivariate Ridge Regression</i>					
EGW _{cb} -GxE	0.15	0.55	0.45	0.61	0.44
EGW _{cb} -GxE-W _{cb} xE	0.29	0.57	0.44	0.64	0.48
EGW _{cb} -GxE-W _{diag} xE	0.28	0.57	0.42	0.64	0.48
EGW _{cb} -GxE-P	0.21	0.56	0.43	0.62	0.46

Table 4-5. Predictive abilities of genomic, single NIRS, multi-omics and multi-NIRS models in scenario **nGoEref** for GY. CVRandom_Ref and CVNewG_Ref are leave-one-environment-out schemes followed by five-fold cross-validation in which new genotypes are unphenotyped in all the environments of the MET except the reference environment. Predictive abilities calculated based on 50 repetitions are presented by set and averaged over all sets. The optimum models are shown in red, and the best predictive abilities by set and by model category (based on Average) in bold.

<i>nGoEref scenario</i> (CVNewG_Ref)	<i>GY prediction</i>				
	Set1 2016	Set2 2019	Set4 2018	Set4 2019	Average
Genomic models					
EG	0.25	0.39	0.24	0.38	0.32
EG_GxE	0.23	0.54	0.34	0.44	0.39
Single NIRS models					
EW _{ref}	0.16	0.29	0.19	0.30	0.23
EW _{ref} -W _{ref} xE	0.25	0.35	0.20	0.37	0.29
Multi-NIRS models					
<i>Optimum</i>					
EW _{cb}	0.25	0.44	0.27	0.36	0.33
EW _{cb} -W _{cb} xE	0.40	0.60	0.29	0.43	0.43
EW _{cb} -W _{diag} xE	0.39	0.60	0.32	0.42	0.43
EW _{cb} -P	0.36	0.55	0.30	0.38	0.40
<i>Nearest Neighbour</i>					
EW _{cb}	0.11	0.18	0.11	0.17	0.14
EW _{cb} -W _{cb} xE	0.19	0.20	0.09	0.19	0.17
EW _{cb} -W _{diag} xE	0.15	0.16	0.06	0.19	0.14
EW _{cb} -P	0.08	0.15	0.08	0.16	0.12
<i>Weighted Multivariate Ridge Regression</i>					
EW _{cb}	0.19	0.31	0.17	0.30	0.24
EW _{cb} -W _{cb} xE	0.29	0.40	0.19	0.37	0.31
EW _{cb} -W _{diag} xE	0.29	0.41	0.18	0.36	0.31
EW _{cb} -P	0.25	0.34	0.14	0.32	0.26
Multi-omics models					
<i>Optimum</i>					
EGW _{cb} -GxE	0.29	0.61	0.33	0.52	0.44
EGW _{cb} -GxE-W _{cb} xE	0.43	0.67	0.33	0.56	0.50
EGW _{cb} -GxE-W _{diag} xE	0.41	0.68	0.35	0.55	0.50
EGW _{cb} -GxE-P	0.38	0.65	0.36	0.53	0.48
<i>Nearest Neighbour</i>					
EGW _{cb} -GxE	0.21	0.51	0.30	0.42	0.36
EGW _{cb} -GxE-W _{cb} xE	0.26	0.40	0.27	0.41	0.34
EGW _{cb} -GxE-W _{diag} xE	0.22	0.37	0.23	0.41	0.31
EGW _{cb} -GxE-P	0.17	0.42	0.27	0.40	0.32
<i>Weighted Multivariate Ridge Regression</i>					
EGW _{cb} -GxE	0.24	0.56	0.33	0.48	0.40
EGW _{cb} -GxE-W _{cb} xE	0.33	0.57	0.31	0.51	0.43
EGW _{cb} -GxE-W _{diag} xE	0.32	0.58	0.31	0.51	0.43
EGW _{cb} -GxE-P	0.28	0.57	0.31	0.49	0.41

The PA were not too low (0.32 and 0.27, Tables S2-S3) considering the difficulty of predicting GY in new environments. For some sets, the PA markedly increased when the P matrix was used to model GxE. For example, the PA of model EG was 0.03 for Set2-2019 while the PA of model EG_P was 0.27 (Table S3). We found that using a single NIRS model was less effective than using all spectra available to model GxE. The main genetic effect can be modelled by an H matrix computed on a single environment, but GxE is better modelled by combining all the NIR spectra available. The optimal situation for predicting GxE is therefore to have NIR data for each predicted environment.

Is it better to model GxE with information from NIR spectra or from molecular markers? Unlike molecular markers, the spectra characterise the effects of both the genotype and the environment on the plant material tested, so they may provide more information. After trying different combinations to model the main genetic effect with molecular markers (G) or NIRS (W) and the GxE effect with molecular markers (GxE) or NIRS (P), we found that the best combination was G for the main genetic effect and P for the GxE effect. However, the increase in PA is moderate. Similar conclusions were reported by Krause *et al.* (2019) for predicting GY in wheat. Montesinos-López *et al.* (2017b) also reported that modelling GxE by the interaction between spectral band and the environment (BxE) resulted in higher PA than models without this term. To assess the benefit of using NIR spectra to model the GxE effect, larger datasets with numerous and contrasted environments would need to be studied.

Our assumption here was that NIR spectra can be considered as simple EC for estimating the similarity between environments. They are indeed the results of the complex regulation and expression of genes in a specific environment, as confirmed by the large amount of GxE along the spectra (Figure 4-3). However, the environmental factors responsible for the absorbance variance are not necessarily the same as the ones responsible for the GY variance. It might be informative to select wavelengths which capture most of the GxE variance for the GY trait. For example, to weight the contribution of the different wavelengths, Additive Main effects and Multiplicative Interaction (AMMI) decomposition can determine the markers and ECs important for predicting the interactions (Rincent *et al.*, 2019).

5.3. Original MET designs enable breeders to use PS at minimal cost

To extend the application of PS for breeders, we compared different MET scenarios with different assumptions. oGoE, nGoE, oGnE and nGnE are classic MET scenarios in which GS has been applied. We first tested our models in a sparse testing scenario (oGoE) and in a scenario where a new line would not have been phenotyped in the MET (nGoE). It is acknowledged that the datasets are very small, both in terms of the number of environments and the number of breeding lines. Despite this, we demonstrate the advantage of MET prediction by PS in the high PA for GY and HD achieved by models developed in oGoE and nGoE. We then tested our models in more challenging scenarios where no genotype was phenotyped in the environment of prediction (oGnE) and new genotypes were not phenotyped in the MET (nGnE). As expected, PA were lower than the previous scenarios. Similarly, Lane *et al.* (2020) found that for a CV0 scenario equivalent to oGnE, the PA of a multi-NIRS model including the GxE interaction for maize GY was around 0.54. The maize trial covered 4 environments (two years x two treatments) and around 300 hybrids. For wheat, we found similar PA for GY with the largest dataset, Set4-2019, but this level of accuracy was not attained with the sets with fewer genotypes (Set4-2018) or with just two environments (Set1-2016 and Set2-2019). This may be partly because the environmental conditions of the unobserved environment did not fall within the range of those of the training environments (Jarquín *et al.* 2017). For Set1-2016 and Set2-2019, the correlation of GY between environments is close to null (Figure 4-2), and the prediction of GY cannot be accurate when the training and validating environment differ so much.

In designing these scenarios, we suppose that breeders acquire NIR spectra on all lines (training and validation) in each environment of the MET. Practically this could be done by dedicating nursery rows in each environment to observation (i.e. scoring diseases and lodging) and NIR spectra acquisition of all lines, in parallel to trial plots dedicated to the measurement of GY of the training lines. This is feasible if the grains or tissue samples are in a good state with no deterioration due to biotic or abiotic factors. Even in the event of disease, we assume that biotic stress will influence the phenotype of the plant and the NIR spectra, but it would still be possible to compute the covariance between genotypes. In our study of Set2_2019, the two environments are different sites with intensive management practices, with the difference that GL was treated with fungicide but EM was not. The PA reached by our models for this dataset were the highest for GY in most scenarios (Tables 4-2 to 4-5). For an easier and cheaper application of PS, we also proposed two

new experimental designs (oGoEref and nGoEref) in which NIR spectra are acquired only on the plots evaluated for GY, which means there is no need for a nursery in parallel to the trials. The principle is to observe and acquire NIRs from all lines in one environment. This environment serves as the reference which is usually chosen by the breeder to observe all the candidate lines for selection. Scenario oGoEref resembles a sparse testing scenario and scenario nGoEref resembles a new genotype scenario, except that they use one environment as reference to acquire NIR spectra and the phenotypes on all genotypes.

To compare multi-NIRS models and multi-omics models in these scenarios, we had to impute the missing spectra. Our results showed that WMRR gave better PA for GY than the NN imputation method (Table 4-4, Table 4-5). This may be because NN assumes that if two lines have a similar spectrum in one environment, then they will have similar spectra in the other environments. This is clearly not in accordance with the observation of strong GxE across the spectra (Figure 4-3). Two genotypes can have similar spectra in one environment and dissimilar spectra in another due to the GxE effect, which is the reason why they are able to capture GxE of the predicted trait more efficiently than markers. Inversely WMRR was expected to consider GxE in the spectra by predicting the specific spectrum of a specific genotype in a specific environment, and thus better impute the missing spectra. The expected maximum PA for GY were however still not reached when the WMRR method was used. The correlations between predicted and acquired spectra were very variable from one set to another, but for Set4-2019 the correlation across spectra was high (Figures S3, FigureS4). Set4-2019 was the only set for which the best model using predicted spectra almost reached the PA of the optimum (0.64 instead of 0.66 in Table 4-4 for CVRandom_Ref and 0.51 instead of 0.56 in Table 4-5 for CVNewG_Ref). The objective with the WMRR method was not to accurately predict the entire spectra, but to accurately predict the wavelengths indicative of major effects on the target trait. However, all the wavelengths are somewhat explicative, so it was not possible to isolate distinct wavelengths or peaks. PA for GY were slightly higher compared to an unweighted multivariate ridge regression method (not shown here). For the prediction of HD, we found similar PA for the two methods, lower than for the optimum.

The WMRR method is promising because it circumvents the potential obstacle of acquiring NIR spectra on all genotypes in each environment. It would be more convenient for breeders to directly apply PS in already defined MET designs and without any additional cost. There is scope to improve the method as the optimum has not been reached yet. Two factors which could be adjusted

is the size and the composition of the training set for predicting missing spectra. As we know, these factors influence the PA of different traits (Albrecht *et al.* 2011; Heffner *et al.* 2011; Pszczola *et al.* 2012; Rincenc *et al.* 2012; Daetwyler *et al.* 2013) in GS as well as in PS (Parmley *et al.* 2019; Galán *et al.* 2020; Zhu *et al.* 2021a; Robert *et al.* 2022a). We assume that the prediction of the absorbances will also be influenced by these factors. Other methods of imputation could be considered. We tried the single-step method of Legarra *et al.* (2009) to overcome the issue of incomplete NIRS acquisition. We used the NIR spectra acquired in the environment of reference combined with the incomplete NIR spectra in another environment to have a hybrid covariance matrix between all genotypes (Michel *et al.* 2021). The hybrid matrix has the advantage of being quick to compute. However, PA obtained with the hybrid matrix were slightly lower compared to PA of our WMRR method.

5.4. Phenomic selection is a promising tool to improve breeding programs

More and more studies have tried to apply PS to annual crops like bread wheat (Montesinos-López *et al.* 2017b; Rincenc *et al.* 2018; Krause *et al.* 2019; Cuevas *et al.* 2019; Robert *et al.* 2022a), maize (Lane *et al.* 2020), rye (Galán *et al.* 2020, 2021), triticale (Zhu *et al.* 2021b) or soybean (Parmley *et al.* 2019; Zhu *et al.* 2021a), or perennial crops like poplar (Rincenc *et al.* 2018) and recently grapevine (Brault *et al.* 2021). PS has the advantage, compared to other types of predictors, of being more flexible as to which data acquired. NIR spectra can be collected on different tissues (leaves, wood, grain, fruits, roots) at multiple stages of the growing cycle, which is similar to other omics methods but less costly. Conventionally, NIR spectra are collected in controlled conditions in a laboratory, with coverage of wavelengths from the visible and near infrared (Osborne 2006). With the rapid development of high throughput phenotyping and new sensors, NIR spectra can be collected at moderate to low cost directly in the fields with an uninhabited aerial vehicle, portable (micro)spectrometer, or directly in the harvester. The method is also very straightforward for breeders who are used to the prediction models already developed for GS. Multiple applications are possible in breeding like sparse testing in single environments (Rincenc *et al.* 2012; Cuevas *et al.* 2019; Robert *et al.* 2022a) or in MET designs as we showed here. PS is also useful to predict breeding material genetically distant from the training set (Galán *et al.* 2021). Other avenues do not have been explored yet but have great potential to interest breeders (Opinion and review in Robert *et al.* 2022b) like the screening of diversity collections at lower cost.

6. Conclusion

We explored whether PS can be used by breeders in different multi-environmental contexts to predict GY and HD. Our main results showed that the PS models performed as well as GS models in the prediction of GY for classic MET designs. For the prediction of HD, GS performed slightly better than PS because HD variance is mainly additive. We compared different ways to model GxE and found that models integrating the GxE effect modelled with NIR spectra were more accurate than the models without this term. We incorporated a P matrix characterising the covariance between the different genotype-environment combinations of the MET, which effectively shares information between environments. For scenarios with new environments, this gave better predictions than models without the P matrix. Finally, we designed two new MET scenarios specific to PS in which NIR spectra were missing for the genotypes to be predicted. Among the different methods available for spectra imputation, the WMRR gave the best PA but not the optimum. The models developed in this paper as well as the original MET designs would enable breeders to use PS for diverse objectives and at minimal cost.

Acknowledgements

The authors thank INRAE experimental units (UE PHACC Clermont-Ferrand, UE GCIE Estrées-Mons, IE UMR GQE Le Moulon, UE Domaine de la Motte Rennes), breeders from Agri-Obtentions and Florimond Desprez. The authors are grateful to Agri-Obtentions, Florimond-Desprez, and the Association Nationale de la Recherche et de la Technologie (ANRT, grant number 2019/0060) which supported this PhD work. The authors also thank Rachel Carol (Bioscience Editing, France) for proofreading.

Chapitre 5

Discussion générale

Préambule

Cette thèse effectuée en partenariat avec Agri-Obtentions, Florimond Desprez et INRAE, a pour objet d'étudier un nouveau type de prédiction pour améliorer les programmes de sélection du blé tendre. L'enjeu est de proposer aux sélectionneurs une nouvelle méthode performante, c'est-à-dire facilement utilisable, précise et peu chère. Après avoir défini les objectifs de ce travail et les axes de recherche dans les précédents chapitres, je vais maintenant discuter de ses apports, de ses limites ainsi que des perspectives de ce travail.

1. La sélection phénotypique: une nouvelle approche de prédiction pour l'amélioration des plantes

Dans cette première partie je vais synthétiser les principaux résultats obtenus dans les différents chapitres II-III et IV, montrer les éventuelles limites rencontrées et proposer des perspectives à étudier.

1.1. L'utilisation des spectres NIR et de la sélection phénotypique pour la prédiction de caractères agronomiques

La sélection phénotypique est une méthode qui a été développée par Rincent *et al.* (2018) et qui a pour principe de remplacer les marqueurs moléculaires, utilisés dans les modèles de prédiction, par l'utilisation d'informations phénotypiques, notamment celle issue de la spectroscopie proche infrarouge. La terminologie de prédiction phénotypique n'a pas été tout de suite adoptée dans la littérature. Depuis 2018, les publications qui font référence à cette nouvelle méthode de prédiction utilisent soit la terminologie "phénotypique" soit la terminologie "hyperspectrale". Cela s'explique notamment par le fait que le terme "phénotypique" des plantes est déjà utilisé pour désigner l'acquisition de données phénotypiques de haute dimension à l'échelle de l'organisme (Houle *et al.* 2010). Dans le domaine végétal, cela se traduit par le fait de réussir à acquérir différents types de données sur l'organisme, pour différents caractères, organes, et stades de développement. Cette acquisition est rendue possible depuis une dizaine d'années, par le développement de l'imagerie et de nouveaux capteurs (Tardieu *et al.* 2017). La phénotypique regroupe donc un grand nombre de données de natures différentes caractérisant des phénotypes secondaires, dont la spectroscopie proche infrarouge. J'ai choisi de conserver cette terminologie de sélection et prédiction phénotypique plutôt que de prédiction hyperspectrale, afin de conserver un terme général et ainsi pouvoir faire le parallèle avec la terminologie de sélection génomique (qui utilise uniquement les marqueurs moléculaires). Les deux types de prédictions suivent la même démarche et sont donc directement comparables.

Dans la terminologie définie par Rincent *et al.* (2018), qui a été précisée dans le Chapitre II, cette méthode peut être utilisée avec l'entièreté des mesures possibles du phénomène. Cependant, dans mes travaux, je me suis focalisée exclusivement sur les données de type NIRS et j'ai comparé mes

résultats uniquement aux autres publications utilisant également le NIRS. Ce n'est donc qu'un petit aspect de la sélection phénotypique que je développe ici. Ces travaux pourraient tout à fait être menés sur d'autres types de données si celles-ci sont reproductibles et qu'elles présentent de la variabilité. Il a été montré que le phénotypage haut débit permet d'obtenir des données phénotypiques denses et précises, utilisables dans des modèles de prédiction (Crain *et al.* 2018). Parmley *et al.* (2019) ont ainsi proposé d'utiliser des mesures de la température et de la surface de canopée par caméra infrarouge pour prédire le rendement du soja. Ces mesures sont particulièrement intéressantes car dans certaines conditions, la température de canopée est corrélée au rendement (Rutkoski *et al.* 2016).

1.2. Des facteurs multiples qui influencent les qualités de prédiction et qui peuvent être optimisés

Les prédictions phénotypiques ou génomiques consistent à prédire la performance de candidats à la sélection pour un caractère d'intérêt. Mon premier objectif était de tester la sélection phénotypique sur du matériel en sélection et en déterminer les capacités de prédiction. Cette capacité de prédiction peut être influencée par de multiples facteurs à l'instar des facteurs identifiés comme influençant les qualités de prédictions génomiques. La sélection phénotypique est une nouvelle méthode qui a peu été étudiée jusqu'alors et qui nécessite donc d'être mieux caractérisée pour être ensuite appliquée à l'échelle de schémas de sélection. Mon deuxième objectif était donc de déterminer les facteurs influençant les qualités de prédiction phénotypique.

Pour répondre à ces objectifs, j'ai tout d'abord sélectionné un ensemble de jeux de données pour tester la sélection phénotypique. J'avais à disposition des données provenant des différents programmes de sélection de Florimond Desprez, Agri-Obtentions et INRAE. Ces jeux de données se caractérisent par des effectifs très variables et présentent peu de diversité entre individus puisque ces derniers sont issus de plusieurs générations de sélection effectuée sur la descendance de croisements entre variétés élités. Pour avoir un ordre de grandeur, le nombre de croisements dont sont issues les lignées de chacun des jeux de données varie de 50 à 350. Pour la génération F5, les effectifs varient entre 113 et 1400 individus selon les sites considérés et pour la génération F6, les effectifs sont divisés par 10 (de 82 à 141). Tous les jeux de données présentaient une caractéristique commune: le phénotypage était disponible pour chaque lignée, dans au moins un

environnement, avec les spectres NIR associés. Par contre, d'un jeu de données à l'autre, le nombre d'environnements, le nombre de spectres collectés, le nombre de lignées génotypées, le nombre de longueurs d'onde par spectre ainsi que le nombre de marqueurs moléculaires étaient variables. J'avais donc à disposition beaucoup de variabilité sur le format des jeux de données. Certains d'entre-eux pouvaient être limitants pour tester différents facteurs, notamment la taille du jeu de calibration qui n'a pu être testée que sur deux environnements du Set2. Mais avec la diversité des jeux de données disponibles j'ai pu tout de même définir des sous jeux de données adaptés aux différents facteurs à tester. Et donc, même si ces jeux de données présentent des limites d'applications, ils ont permis de montrer l'intérêt de cette méthode pour du matériel élite en sélection.

A l'issue de la construction de ces jeux de données, j'ai pu montrer dans les chapitres III et IV que la sélection phénotypique était un bon prédicteur du rendement et de la date d'épiaison pour différents panels, en première ou deuxième année de sélection. Concernant le rendement, les prédictions phénotypiques donnent en moyenne des capacités de prédiction comprises entre 0.22 et 0.44 sur l'ensemble des scénarios et des jeux de données. Pour la date d'épiaison, les moyennes de capacités de prédiction sont comprises entre 0.57 et 0.68 (Figure 5-1). Cuevas *et al.* (2019) ont obtenu des capacités de prédiction de 0.34 pour des prédictions de rendement dans un scénario "single environment" (SE). Les prédictions étaient effectuées sur des lignées avancées élites de blé tendre dans le programme de sélection du CIMMYT. Je m'attendais à ce que les qualités de prédiction soient dégradées sur les scénarios MET avec un environnement de référence (ME-oEref) et dans les scénarios MET avec des prédictions dans un nouvel environnement (MET-nE) par rapport aux autres scénarios car ces derniers sont plus difficiles à prédire (les spectres doivent être prédits dans le cas du MET-oEref et il y a peu d'information partagée entre environnements dans le cas du MET-nE). Cela se vérifie bien dans mes résultats (Figure 5-1).

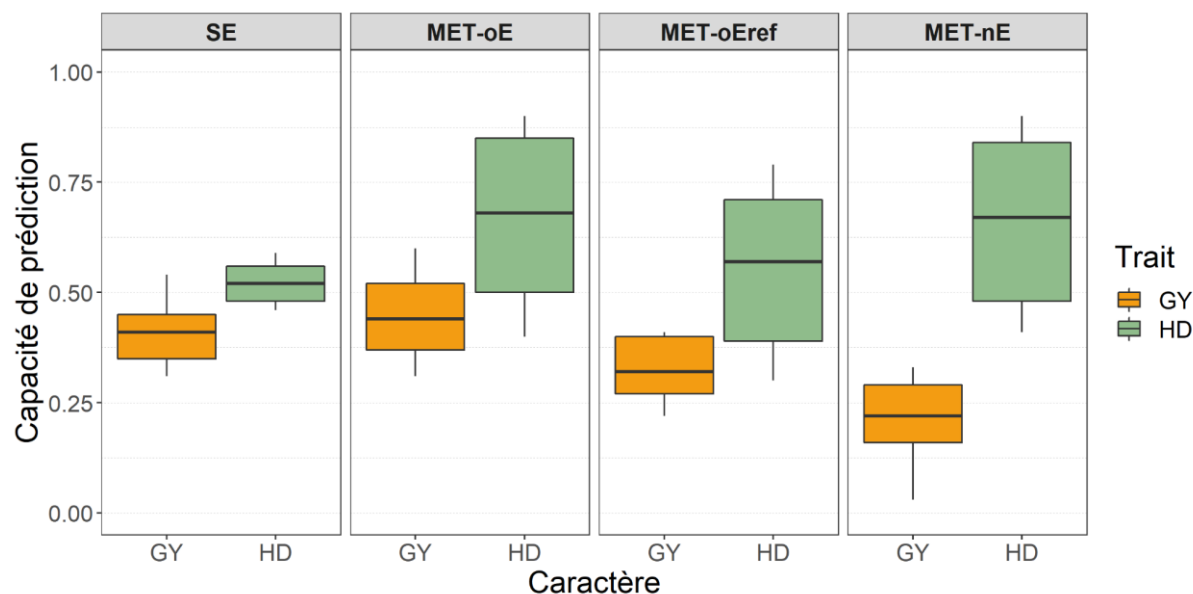


Figure 5-1. Boite à moustache des prédictions phénologiques pour le rendement (GY) et la date d'épiaison (HD sur l'ensemble des jeux de données pour les différents scénarios (SE, MET oGoE, nGoE, oGoEref, nGoEref, oGnE, nGnE). Les scénarios avec oG ou nG ont été regroupés. Pour les scénarios SE, le modèle de prédiction utilisé était le modèle CbSD (cf. Chapitre III). Pour les scénarios MET-oE, le modèle utilisé était le modèle EWcb_WcbxE. Pour les scénarios MET-oEref le modèle utilisé était le même que dans les scénarios oEref mais les spectres ont été prédits par la méthode WMRR. Enfin pour les scénarios MET-nE le modèle utilisé était le EWcb_P. Chaque boîte à moustache indique la moyenne (barre épaisse horizontale), le 1er et 3ème quartile (boîtes) et le 1er et 9ème décile (moustaches).

Par la suite, j'ai pu tester des facteurs extrinsèques aux panels, qui pouvaient potentiellement affecter les capacités de prédiction de la méthode. Dans le chapitre III, j'ai donc testé l'effet de l'origine du spectre et du nombre de spectres intégrés dans le modèle, l'effet du modèle utilisé pour faire les prédictions et l'effet de la taille et de la composition du jeu d'entraînement et, sur les capacités de prédiction phénologique.

1.2.1. L'effet de l'origine des spectres et de leur combinaison sur deux caractères présentant des architectures génétiques contrastées

J'ai étudié la prédiction de deux caractères avec une structure génétique contrastée: le rendement et la date d'épiaison. Le rendement est le caractère phare pour les prédictions car c'est un caractère complexe, difficile à sélectionner. A l'inverse, la date d'épiaison est un caractère facilement sélectionnable et peu soumis à l'effet des interactions GxE. Mes résultats ont montré que le rendement est mieux prédit par un spectre provenant du même environnement que celui dans lequel ce rendement a été mesuré et qu'ajouter des spectres d'autres environnements améliore légèrement les qualités de prédiction. Pour la date d'épiaison, le meilleur prédicteur peut aussi bien être un spectre provenant du même environnement que celui dans lequel elle a été mesurée ou celui d'un autre environnement. J'ai montré que pour ce caractère, ajouter dans le modèle plusieurs spectres issus de différents environnements améliore nettement les qualités de prédiction.

Mes travaux ont été les premiers à montrer que le nombre de spectres à intégrer dans le modèle prédictif peut être ajusté en fonction du caractère que l'on souhaite prédire. De mon point de vue, il est préférable de combiner au moins deux spectres dans le modèle prédictif, particulièrement pour la prédiction du rendement. Je justifie ce choix par le fait que certains spectres présentent une répétabilité moyenne à faible. Selon moi, cette répétabilité peut être vue comme un critère de qualité des spectres pouvant affecter les capacités de prédiction de la méthode. Je n'ai pas pu mettre en évidence de lien direct entre répétabilité et capacité du modèle faute de données suffisantes. Il serait alors intéressant d'étudier plus en détail un critère qui permettrait de juger à priori de la qualité des spectres. Pour l'instant, combiner un spectre à un autre spectre de moyenne à faible répétabilité, ne dégrade pas la qualité de prédiction du caractère final.

Pour estimer les limites d'utilisation des modèles combinant différents spectres, je pense qu'il serait intéressant d'avoir des données permettant de combiner un grand nombre de spectres de différentes natures et d'évaluer les prédictions avec ces différentes combinaisons. Avec les données disponibles, j'ai pu tester la combinaison de 2 à 5 spectres pris dans des environnements différents mais toujours sur grain. Mais qu'en est-il des spectres sur d'autres tissus ou à différents stades de croissance ? Les travaux de Rincent *et al.* (2018) ont montré que prendre les spectres sur feuilles donnait de moins bonnes capacités de prédiction du rendement que sur grain. Combiner les deux types de spectres pour prédire le rendement n'améliore pas non plus la capacité de prédiction du modèle obtenue uniquement sur grain. Krause *et al.* (2019) ont travaillé sur des

bandes acquises à différents stades de croissance, par caméra hyperspectrale embarquée sur un drone. Ils ont ainsi montré que les bandes spectrales acquises sur des parcelles au stade d'épiaison et au stade remplissage du grain donnent légèrement de meilleures capacités de prédiction du rendement qu'au stade végétatif. Il n'y a plus de différence entre stades d'acquisition lorsque les bandes sont combinées aux marqueurs moléculaires pour prédire le rendement.

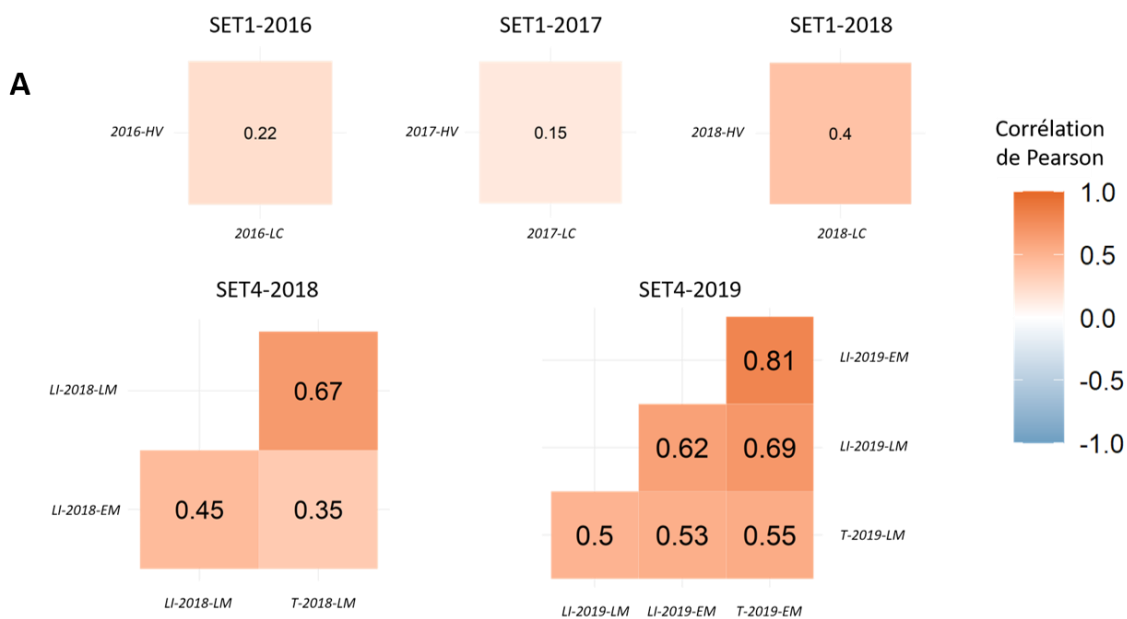
La question de prendre des spectres à différents stades et sur différents tissus est une question importante car elle permettrait d'offrir plus de flexibilité pour le sélectionneur dans l'application de la SP. Il s'agit ici de gagner du temps en faisant l'acquisition des spectres pré récolte. Par exemple: au stade de maturation pour sélectionner ensuite les plantes à récolter en pépinière, ou en pré floraison pour sélectionner des plantes à croiser. Il faut toutefois tenir compte du lien entre le stade d'acquisition du spectre et le caractère à prédire. En effet, le rendement est déterminé sur le stade final de la plante, une fois les grains remplis. Un spectre pris post récolte sera donc plus à même de capter l'expression du génome résultant des adaptations de la plante aux conditions environnementales survenues tout au long du cycle de croissance, qu'un spectre pris au stade végétatif.

Dans des travaux préliminaires, j'ai également testé les prédictions phénomiques sur d'autres caractères tels que la qualité du blé et la résistance aux maladies. Le taux de protéine du grain est très bien prédit par les prédictions phénomiques lorsque le spectre utilisé est le même que celui issu de l'environnement dans lequel est fait la prédiction du taux de protéine (Figure 5-2, B). Il est normal qu'on obtienne une capacité de prédiction de 1 ou proche de 1 car ce sont ces spectres qui ont servi à la prédiction du taux de protéine. Par contre lorsqu'on prend un spectre issu d'un autre environnement (GLOB), les capacités de prédiction sont fortement dégradées. En m'intéressant aux corrélations entre environnements pour la teneur en protéine (Figure 5-2, A), j'ai pu montrer que celle-ci était aussi très variable d'un jeu de données à l'autre, allant de 0.15 à 0.81. Ces faibles corrélations impactent directement les capacités de prédictions de type GLOB, comme dans le cas du Set1-2017 où la corrélation entre environnements pour la teneur en protéine est proche de 0.15 et la capacité moyenne de prédiction GLOB de la teneur en protéine est proche de 0.

Pour les résistances à la septoriose et à la fusariose, les capacités de prédiction sont faibles avec les spectres NIR. Ces résultats sont cohérents avec ceux de Zhu *et al.* (2021) qui ont montré sur triticale que les prédictions phénotypiques pour la résistance à l'oïdium et à la rouille jaune sont très faibles (comprises entre 0.00 et 0.20) alors que bien prédites par la prédiction génomique (entre 0.59 et 0.85). De mon point de vue, cela est lié en partie, au tissu sur lequel les spectres sont collectés. En effet, alors que le grain est le meilleur tissu pour collecter des spectres en vue de faire des prédictions de rendement, pour prédire des caractères de maladie qui touchent les feuilles de la plante, il serait plus judicieux de prendre ce tissu qui est plus à même de contenir l'expression de la résistance que le grain. De plus, d'un point de vue pratique, les caractères de résistance sont principalement monogéniques ou oligogéniques et donc contrôlés par quelques gènes à effet majeur. Cet effet de l'architecture génétique a été illustré dans les travaux de Zhu *et al.* (2021). L'introgession d'allèles favorables pour la résistance aux maladies et son suivi en sélection par SAM sont dans ce cas mieux adaptés que la sélection phénotypique.

Les prédictions phénotypiques sont particulièrement bien adaptées pour prédire les caractères complexes. L'hypothèse sous-jacente est que le NIRS peut capturer de la variance génomique et peut capturer indirectement des caractères secondaires corrélés aux caractères complexes à prédire. Pour revenir à l'exemple du rendement, on supposait que les très bonnes qualités de prédiction phénotypique sur le rendement étaient dues à la forte corrélation qu'il peut y avoir entre le rendement et la teneur en protéine, cette dernière étant très bien prédite par les spectres. Il s'avère que dans mes jeux de données, les corrélations entre rendement et teneur en protéine sont très variables (entre -0.1 et -0.64). Cette relation participe à la bonne prédiction du rendement puisque la teneur en protéine est bien prédite par les spectres, mais elle n'est pas suffisante pour expliquer entièrement les capacités de prédiction obtenues. En effet, en corrigeant les moyennes ajustées du rendement par l'effet de la teneur en protéine, j'ai pu me rendre compte que la capacité de prédiction du rendement corrigé pouvait être faiblement à fortement dégradée. Cela montre que la teneur en protéine a bien un effet sur les capacités de prédiction mais que ce n'est pas le seul et qu'il est toujours possible de prédire ce rendement corrigé malgré l'absence d'information apportée par la teneur en protéine. C'est pourquoi l'hypothèse que le spectre NIR capture indirectement des relations entre le rendement et d'autres caractères secondaires, me paraît être une hypothèse valide. Pour en comprendre un peu plus sur la biologie qui se cache derrière les absorbances mesurées aux différentes longueurs d'onde, il faudrait essayer d'identifier différentes relations entre le spectre NIR et les strates d'omiques. En collectant des données de métabolites

ou de transcrits, de spectres et de rendement sur un ensemble des génotypes mis en culture dans un environnement donné, on serait en capacité de faire le lien entre ces différents types de variables. Par exemple, on pourrait s'intéresser aux corrélations entre longueurs d'onde et les différentes données omiques. A l'instar de la protéine, il existe très certainement des longueurs d'onde explicatives pour certains types de métabolites. Vergara-Diaz *et al.* (2020), ont montré qu'il était possible de prédire des métabolites de la feuille drapeau, de l'épi et de la canopée du blé dur grâce aux spectres du proche infrarouge. Cette démarche permettrait de mieux comprendre ce qui est concrètement capté par les spectres et peut-être même d'identifier les caractères secondaires qui sont directement liés au rendement.



B

Prédiction	Set1-2016	Set1-2017	Set1-2018	Set4-2018	Set4-2019
Same Site	1	1	1	0.94	0.96
GLOB	0.35	0.01	0.4	0.25	0.6

Figure 5-2. (A) Représentation des coefficients de corrélation de Pearson entre environnements pour la teneur en protéine, sur les Set1-2016, Set1-2017, Set1-2018, Set4-2018 et Set4-2019. (B) Moyenne de capacité de prédiction pour la teneur en protéine, du modèle S (Same Site) et du modèle D (GLOB) présentés dans le chapitre III. Les moyennes ont été calculées sur la base de 25 répétitions de 5 validations croisées par environnement constituant le SE.

1.2.2. L'effet du modèle statistique sur la qualité des prédictions

J'ai par la suite étudié l'effet de différents modèles sur les capacités de prédiction. Dans le chapitre III ne sont présentés que le H-BLUP et le PS-LASSO mais j'ai aussi pu tester des modèles de types RR-BLUP (équivalent au G-BLUP), Elastic Net (sélection de variables) et RKHS (intégration des effets non additifs). J'ai observé peu de différences entre tous ces différents modèles en termes de capacité de prédiction. C'est ce qui est également observé en prédiction génomique où finalement les modèles les plus simples sont aussi souvent les plus efficaces (Charmet *et al.* 2020). J'ai opté, dans le chapitre III, pour le H-BLUP qui donne des prédictions équivalentes aux autres modèles et qui a un très bon temps de calcul. Son utilisation est très accessible pour les sélectionneurs. Il aurait été intéressant de se pencher sur d'autres types de modèles pour prendre en compte le fait que les spectres peuvent aussi être considérés comme des variables continues. C'est le cas de la régression fonctionnelle: les spectres NIR peuvent être modélisés comme une courbe dans un domaine continu et donc considérés comme une fonction (Montesinos-López *et al.* 2017a, b ; Lane *et al.* 2020). Dans ce cas, les longueurs d'onde ne sont pas utilisées pour calculer une matrice de variance covariance mais comme covariables dans le modèle fonctionnel. La régression fonctionnelle permet de réduire la dimensionnalité en transformant les longueurs d'onde en variables fonctionnelles.

1.2.3. L'effet de la taille et de la composition du jeu de calibration sur la qualité de prédiction

Enfin, je me suis intéressée à l'effet de la taille et de la composition du jeu de calibration. A travers les deux jeux de données sur lesquels j'ai pu tester ces effets, j'ai trouvé que plus la taille du jeu de calibration augmente et plus les capacités de prédictions augmentent jusqu'à atteindre un plateau (cf. Chapitre III). Cela montre que rajouter des individus au-delà de 200 n'améliore pas les prédictions car l'information amenée par les nouveaux individus est redondante avec celle déjà présente. Donc choisir les individus qui composent ce jeu de calibration est primordial car en sélectionnant les individus les plus informatifs, il est possible de réduire le nombre d'individus en calibration et donc le phénotypage à réaliser au champ. J'ai donc testé différents algorithmes pour sélectionner ces individus informatifs plutôt que de les choisir au hasard. Les algorithmes déjà développés que j'ai pu tester sont peu performants avec les spectres NIR, le CDmean étant le moins inefficace. Je pense qu'il serait intéressant de tester ces algorithmes sur les longueurs d'onde

les plus explicatives car pour le moment, j'ai toujours considéré le spectre en entier. Au niveau de l'application concrète pour un sélectionneur, il faudrait pouvoir optimiser la composition du jeu de calibration sur l'année précédente N-1 pour choisir les individus qui constitueront le jeu de calibration et qui seront donc phénotypés au champ pour l'année N. Les données mises à disposition ici n'ont pas permis de tester ce dispositif d'optimisation.

1.3. Intérêt de l'utilisation d'une variable telle que la spectroscopie proche infrarouge pour prédire les effets d'interactions génotype-environnement

L'interaction génotype-environnement est un effet qui est pris en compte par les sélectionneurs pour choisir au mieux les lignées candidates et déterminer celles qui sont les plus performantes dans différents environnements contrastés. Pour cela, ils mettent en place des réseaux d'essais pour évaluer l'effet de cette interaction notamment sur le rendement. Cette interaction est généralement difficile à prédire et limite donc les capacités de prédiction des modèles dans des essais multi-environnementaux (Burgueño *et al.* 2012). J'ai donc supposé que les spectres NIR pourraient être de bons prédicteurs de cet effet d'interaction puisque que le NIRS est la résultante de l'expression d'un génotype dans son environnement. De plus, Rincet *et al.* (2018) avaient déjà montré que les spectres NIR sont soumis à une partie de l'interaction GxE.

Dans le chapitre IV, j'ai évalué différents scénarios de prédiction pour des réseaux d'essais incomplets. J'ai sélectionné quatre jeux de données qui présentaient suffisamment de lignées (au moins 80) et d'environnements (au moins 2) pour mener des prédictions multi-environnementales. Ces données ont montré que le NIRS capture bien l'interaction GxE même sur des panels élites en sélection. Pour exploiter cette information, j'ai prédit l'effet GxE avec différentes matrices de variances-covariances. Finalement, j'ai montré qu'ajouter l'effet d'interaction GxE prédit par des spectres NIRS conduit toujours à de meilleures prédictions du rendement qu'un modèle qui n'en tient pas compte. J'ai donc pu mettre en évidence que la prise de spectres est importante dans l'ensemble des environnements du MET ainsi que sur l'ensemble des lignées pour prédire cet effet d'interaction. Cependant je m'attendais à ce que l'ajout de cet effet prédit par les spectres NIR améliore nettement plus les qualités de prédiction. Dans nos travaux, l'amélioration est légère comparée à un modèle qui ne tient pas compte de l'effet GxE (un gain maximum de 0.15 point,

tous scénarios confondus). Je pense que cela est dû aux jeux de données qui sont trop petits pour réellement voir le potentiel des spectres NIR pour prédire les effets GxE. En effet, le nombre maximum d'environnements dans nos MET est de quatre. Krause *et al.* (2019) et Lane *et al.* (2020), qui ont aussi essayé de prédire cet effet d'interaction dans leurs modèles, avaient aussi de très petits jeux de données en termes de nombre d'environnements disponibles et ont montré tout de même que les capacités de prédiction des modèles s'améliorent en tenant compte du GxE. Avec de plus gros jeux de données, il serait aussi possible d'essayer d'identifier les longueurs d'onde les plus explicatives de l'effet d'interaction GxE auquel est soumis le rendement, par une décomposition AMMI par exemple. Rincent *et al.* (2019) ont utilisé une décomposition AMMI pour déterminer quels marqueurs moléculaires et quelles covariables environnementales sont importants pour prédire les interactions. Malheureusement, je n'ai pas pu mettre en place ce type de décomposition dans mes jeux de données qui sont trop limités. Même s'il est possible de la réaliser sur deux environnements, je pense que celle-ci ne sera pas efficace surtout si les deux environnements ne sont pas corrélés.

Une autre manière de tenir compte de l'effet GxE est de caractériser les environnements du réseau d'essais. C'est ce qu'ont montré les travaux de Heslot *et al.* (2014), Jarquín *et al.* (2014), Malosetti *et al.* (2016) et Lado *et al.* (2016) pour lesquels l'utilisation de covariables environnementales était utile pour caractériser les environnements du MET et pour prédire l'interaction GxE. Avec des données comportant plus d'environnements comme ceux acquis lors du programme BreedWheat (Touzy *et al.* 2019 ; Rincent *et al.* 2019), il aurait été intéressant d'essayer de caractériser les environnements grâce aux spectres. Une approche commune dans les réseaux d'essais multi-environnementaux est de combiner des essais similaires en cluster, en utilisant la localisation géographique ou des similarités basées sur des covariables environnementales. L'idée ensuite est d'effectuer les prédictions génétiques séparément dans chacun des clusters (Piepho et Möhring 2005). L'avantage est que l'effet GxE est réduit dans des groupes d'environnements similaires. La question soulevée ensuite est de savoir s'il faut prendre en compte les spectres pris sur l'ensemble des lignées mises en culture dans les différents environnements ou si un groupe réduit de témoins serait suffisant pour faire la caractérisation environnementale. C'est le principe des génotypes révélateurs qui sont choisis a priori sur des critères d'adaptabilité, pour révéler l'intensité des facteurs limitants du rendement dans chaque environnement (Brancourt-Hulmel *et al.* 1999).

Mes travaux ont donc permis de montrer que la sélection phénotypique pouvait être utilisée sur du matériel en sélection pour prédire correctement un caractère complexe comme le rendement. J'ai pu identifier un certain nombre de facteurs qui affectent ses capacités de prédiction, que j'ai synthétisés dans la Figure 5-3. Je vais, dans les prochaines parties, me concentrer sur la comparaison avec la sélection génomique et sur la proposition de schémas de sélection intégrant sélection phénotypique et/ou génomique.

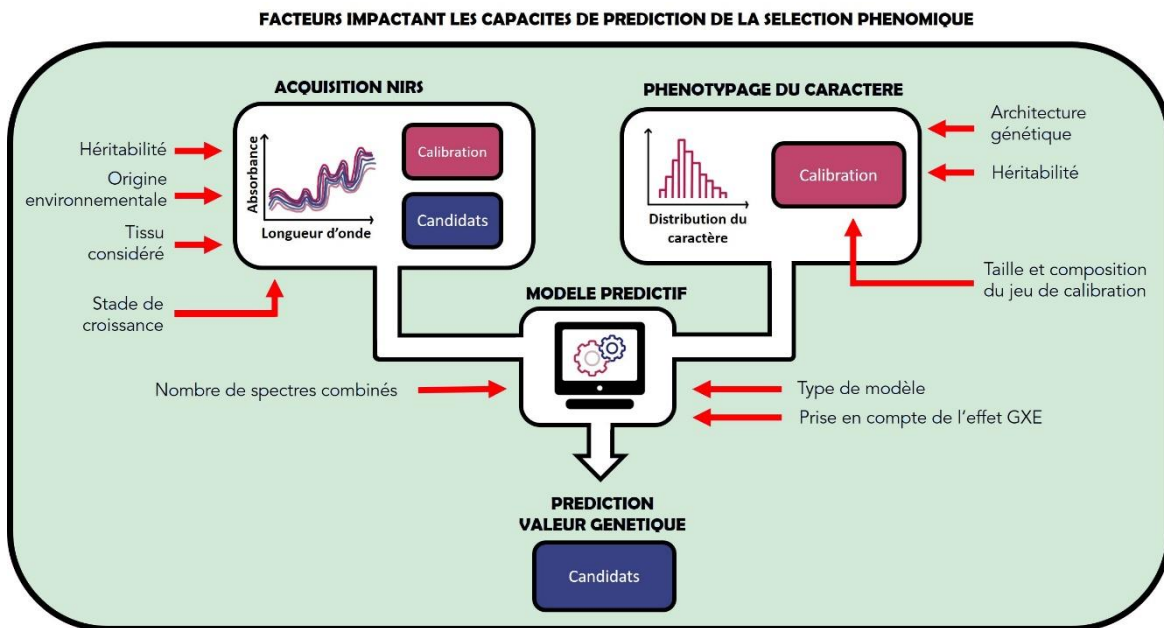


Figure 5-3. Schéma des différents facteurs, influençant les capacités de prédiction de la sélection phénotypique, identifiés dans mes travaux de recherche (Chapitre III et Chapitre IV) mais aussi dans la littérature.

2. La sélection phénotypique peut-elle remplacer la sélection génomique ?

2.1. Avantages et limites de l'utilisation des spectres NIR par rapport aux marqueurs moléculaires

Dans cette première partie, je vais d'abord me concentrer sur l'aspect économique et pratique de chacune de ces méthodes. J'étudierai par la suite leur capacité de prédiction respective.

La sélection génomique a pu se développer rapidement ces vingt dernières années grâce au développement de techniques de séquençage de nouvelle génération (NGS) permettant ainsi l'essor du génotypage haut débit et de haute densité (Lorenz *et al.* 2011 ; Poland et Rife 2012). En effet, la SG repose sur l'estimation simultanée des effets de marqueurs répartis sur l'ensemble du génome. Elle nécessite un génotypage de haute densité pour capter les potentiels QTLs en déséquilibre de liaison avec les marqueurs. Le développement de puces haute densité s'est concentré sur le polymorphisme nucléotidique (single-nucleotide polymorphism, SNP) qui est très courant sur l'ensemble du génome. L'intérêt des NGS était aussi de rendre le génotypage plus abordable. Aujourd'hui, on en sait plus sur la densité de marqueurs à atteindre pour conduire la SG en amélioration des plantes (Heffner *et al.* 2011 ; Hickey *et al.* 2014 ; Isidro *et al.* 2015 ; Rabier *et al.* 2016). Les travaux de Norman *et al.* (2018) ont permis de montrer que sur 10 375 lignées de blé tendre, utiliser 5 000 marqueurs moléculaires parmi 15 000 suffisait à atteindre le maximum de capacité de prédiction du rendement. Cela dépend toutefois de la proximité entre la population d'entraînement et la population en sélection, un plus grand nombre de marqueurs étant nécessaires lorsque les deux populations sont éloignées.

Le choix d'utiliser la sélection génomique repose sur un équilibre entre l'investissement qu'implique la méthode et le gain génétique que les sélectionneurs pourront en tirer. Malgré une diminution du nombre de SNPs nécessaires aux prédictions ainsi qu'une réduction constante des coûts du génotypage, la mise en place du génotypage en routine dans les programmes de sélection, sur des milliers d'individus chaque année, représente encore un coût que les petites structures de sélection ne peuvent pas assumer (R2D2 Consortium *et al.* 2021), pour que la SG soit réellement attractive.

A la suite de discussions avec les sélectionneurs du secteur privé, il ressort que les grosses structures ont pu investir dans des plateformes de génotypage haut débit et rentabiliser leur investissement en déployant le génotypage sur des milliers de lignées de différents programmes de sélection, parfois grâce à l'utilisation de puces multi-espèces. Pour les plus petites structures, le génotypage de haute densité est souvent sous-traité par des entreprises extérieures. Actuellement, il est difficile d'estimer le coût réel du génotypage pour un programme de sélection, notamment parce que ces coûts ne sont jamais dévoilés. Pour avoir un ordre de grandeur, Ben-Sadoun *et al.* (2021) ont estimé le coût du génotypage d'une lignée de blé tendre entre 10 € et 37 € lors de leurs simulations de schémas de sélection.

Contrairement aux marqueurs moléculaires, la spectroscopie proche infrarouge a l'avantage d'être très peu coûteuse. Celle-ci s'est largement popularisée depuis les années 90 en sélection des céréales pour prédire le taux de protéines des grains (Osborne 2006). Plus globalement, la spectroscopie proche infrarouge a été initialement utilisée pour la prédiction des teneurs en protéines, en humidité et en huile des céréales et autres graines, depuis les années 70 (Gill *et al.* 1979). A la différence des NGS, aucun coût de consommables n'est associé à la mesure. Utilisée en routine par les sélectionneurs, on estime son coût à quelques euros, si les volumes sont importants pour rentabiliser l'achat de l'appareil. Le coût est alors lié au coût de la main d'œuvre nécessaire pour l'acquisition des spectres. Les discussions avec les différents acteurs de la sélection ont permis d'estimer un coût compris entre 1€ et 10€, la fourchette haute étant liée au sous-traitement de l'acquisition par une entreprise extérieure. Il est à noter que, si l'acquisition des spectres est déjà effectuée pour la prédiction du taux de protéines des céréales, alors l'utilisation de ces spectres n'engendrera pas de coûts supplémentaires pour les sélectionneurs. Le NIRS est aussi utilisé en routine sur d'autres espèces comme le colza pour déterminer la teneur en huile des graines (Velasco *et al.* 1999).

Les spectres NIR apparaissent indéniablement comme une donnée plus accessible pour les sélectionneurs que les marqueurs moléculaires. J'ai montré dans le chapitre IV que ces spectres ont un intérêt supplémentaire par rapport aux marqueurs moléculaires car ils sont capables à la fois de capter de la variance génétique additive mais aussi de la variance d'interaction GxE. Mes travaux ont montré que cela se révèle utile dans le cadre de la prédiction de caractères complexes où l'effet GxE doit être pris en compte. Pourtant, le fait que le spectre NIR dépende à la fois du génotype et de l'environnement dans lequel ce dernier est placé, est aussi une contrainte. La

covariance, calculée entre individus à partir des spectres NIR, permet de déterminer la proximité entre ces individus. Cette proximité n'est pas uniquement liée à l'apparentement génétique mais aussi à l'effet de l'interaction entre le génotype et l'environnement. Il est donc nécessaire d'acquérir les spectres dans un même référentiel, c'est-à-dire dans un même environnement pour comparer les individus les uns aux autres.

Cette contrainte est problématique, notamment pour l'utilisation de données historiques issues des différentes années de sélection. Ces données ont la particularité d'avoir généralement très peu d'individus en commun pour faire le pont entre les années (une dizaine de témoins tout au plus) ce qui limite les possibilités de correction de spectres. Une première possibilité serait de mettre en place une acquisition standardisée de spectres sur les lignées en conditions contrôlées, identiques chaque année, en serre ou chambre de culture par exemple, pour avoir une même référence environnementale. Il faudrait ensuite déterminer les capacités de prédiction de la SP avec ce type de spectres pris en conditions contrôlées sur des années différentes. Une deuxième possibilité, qui me semble plus intéressante, est d'essayer de prédire les spectres qui n'ont pas été acquis pour les différentes combinaisons génotype-environnement manquantes, comme ce que j'ai pu proposer dans le chapitre IV, avec les régressions multivariées pondérées. Il est cependant nécessaire d'avoir des individus communs aux différentes données historiques pour faire ces prédictions. Il serait envisageable de mettre en place en routine, une prise de spectres annuelle d'un même groupe de génotypes pour faire le pont entre les différentes données. Cette deuxième possibilité me paraît plus pertinente que la première, puisqu'il n'y a alors pas besoin de mettre en place un dispositif de culture en condition contrôlée, condition qui peuvent être en plus, très éloignées des conditions de croissance au champ. Or, c'est aussi en partie ce qu'on recherche grâce à l'acquisition des spectres dans différents milieux.

Les marqueurs moléculaires utilisés en sélection génomique ont aussi d'autres applications que les prédictions génomiques. En effet, pour les sélectionneurs, il est important de conserver de la diversité génétique pour les futurs cycles de sélection. Les marqueurs moléculaires leur ont ainsi permis de caractériser la diversité génétique et la structure de populations. Cette caractérisation permet, par exemple, d'éliminer du programme du matériel génétique présentant des redondances. A plus grande échelle que le programme de sélection, les marqueurs moléculaires permettent de déterminer des core collections. L'intérêt de ces collections est de sélectionner les accessions qui représentent la plus grande diversité génétique d'une espèce en minimisant la répétitivité des

profils génétiques (Frankel 1984). Pour le blé tendre, Balfourier *et al.* (2007) se sont servis de 38 marqueurs SSR pour caractériser génétiquement 3 947 accessions mondiales de blé tendre et sélectionner finalement 372 accessions représentant toutes les régions géographiques du monde et 98% de diversité génétique aux loci étudiés.

Les marqueurs moléculaires permettent aussi de suivre l'introgession d'allèles favorables d'une lignée donneuse vers une lignée receveuse à fort potentiel agronomique (Visscher *et al.* 1996 ; Frisch *et al.* 1999). Cette sélection assistée par marqueurs permet aux sélectionneurs d'opter pour différentes stratégies d'introgession, dont notamment le pyramidage d'allèles favorables (Hospital et Charcosset 1997; Charmet *et al.* 1999). Ceci est particulièrement efficace pour l'introgession d'allèles de résistances aux maladies, qui offrent une résistance durable dans du matériel élite (Kloppers et Pretorius 1997). Enfin le choix des parents pour les croisements à réaliser est un choix décisif pour les sélectionneurs. Les marqueurs moléculaires permettent aussi de prédire la variance génétique de croisements. Zhong et Jannink (2007) ont suggéré de prédire la variance génétique de la descendance d'un croisement ou d'une population donnée, en se basant sur des simulations d'effets de marqueurs sur l'ensemble du génome. Les travaux de Lehermeier *et al.* (2017), ont montré également qu'il était possible de prédire la variance génétique d'un caractère, pour une descendance HD ou de RIL issus d'un croisement entre deux parents. Pour cela, ils ont utilisé les fréquences de recombinaison et le DL des lignées parentales.

A travers ces exemples d'utilisation des marqueurs moléculaires, il apparaît que la variance génétique du matériel en sélection est un paramètre qu'il est nécessaire de contrôler et de maximiser. J'ai vu qu'avec l'utilisation des spectres NIR, ces derniers ne sont pas uniquement soumis à l'effet génétique. Pour les différentes applications citées ci-dessus, l'utilisation des spectres NIR est plus compliquée voire impossible. Pour se rapprocher au maximum du génotype, Whalen *et al.* (2020) ont proposé d'inférer le génotype à l'aide de données phénomiques (phénotypage haut débit). Ils ont testé leur proposition à partir de données simulées. J'ai décrit leur méthode dans le chapitre II, mais, succinctement, leur approche consiste en un modèle calibré par une population d'entraînement où les états de ségrégation à chaque locus sont reliés au phénotypage haut débit. Il faut donc disposer du génotypage et du phénotypage haut-débit sur la population d'entraînement. Ensuite, ce modèle a été appliqué pour prédire les états de ségrégation des variétés non génotypées, pour lesquelles on dispose également du phénotypage haut débit. Il serait alors possible d'inférer les allèles aux marqueurs moléculaires à partir des spectres NIR et

ainsi d'appliquer la détection de QTL sur des individus non génotypés. L'utilisation du NIRS en sélection pourrait alors être adaptée à des architectures génétiques monogéniques ou oligogéniques. Toutefois, il est nécessaire que dans le spectre, des longueurs d'onde captent l'effet du gène pour que l'inférence du génotype soit correcte et puisse servir ensuite à la détection de QTL. Il est donc indispensable de valider ces simulations sur des données réelles.

2.2. Sélection génomique et sélection phénotypique: des méthodes en concurrence...

Après avoir décrit les différences économiques et pratiques de l'utilisation des spectres NIR par rapport aux marqueurs moléculaires, je vais à présent m'intéresser à leur capacité prédictive respective. Dans ce travail, j'ai porté une grande attention à pouvoir comparer les deux types de méthodes dans différents scénarios et configurations de prédiction. Dans l'ensemble des résultats obtenus et présentés dans le chapitre III et le chapitre IV, j'ai pu montrer que la sélection phénotypique faisait au moins aussi bien que la sélection génomique pour une grande partie des prédictions. Pour comparer les méthodes, je me suis principalement intéressé à la corrélation de Pearson comme critère de comparaison des modèles, moyennée sur l'ensemble des répétitions de validations croisées. Pour le sélectionneur, il est important d'avoir une méthode avec de bonnes capacités de prédiction pour identifier avec fiabilité les individus qui continueront le processus de sélection ou, au contraire, éliminer les moins bons candidats. En effet, la sélection peut se faire dans les deux sens, sachant qu'il est moins risqué d'éliminer les moins bons candidats. Il est par exemple très risqué d'utiliser une méthode de prédiction avec un $R^2=0.44$ pour identifier les quelques meilleurs candidats. Comme la méthode n'est pas fiable à 100%, il est possible de passer à côté de très bons candidats qui ne seraient pas correctement prédits par le modèle. Par contre, il est moins risqué d'éliminer les plus mauvais, car la fiabilité de la méthode est suffisante pour qu'un très bon candidat ne se retrouve pas parmi les candidats les moins bien classés. D'un point de vue de l'efficacité de la sélection, celle-ci sera moins efficace en éliminant les moins bons candidats qu'en sélectionnant les meilleurs. Dans l'exemple présenté en Figure 5-4, si le sélectionneur décide de sélectionner 10% des meilleures lignées en se basant sur les prédictions phénotypiques, il ne sélectionnera en réalité que 2 des 7 réelles meilleures lignées. A l'inverse, s'il choisit d'éliminer 10% des plus mauvaises lignées, il éliminera bien 3 des 7 réelles mauvaises lignées et il éliminera des lignées dans le groupe des lignées moyennes. Le risque pour le sélectionneur pas d'éliminer les réelles meilleures lignées avec les prédictions phénotypiques est très faible mais pas nul.

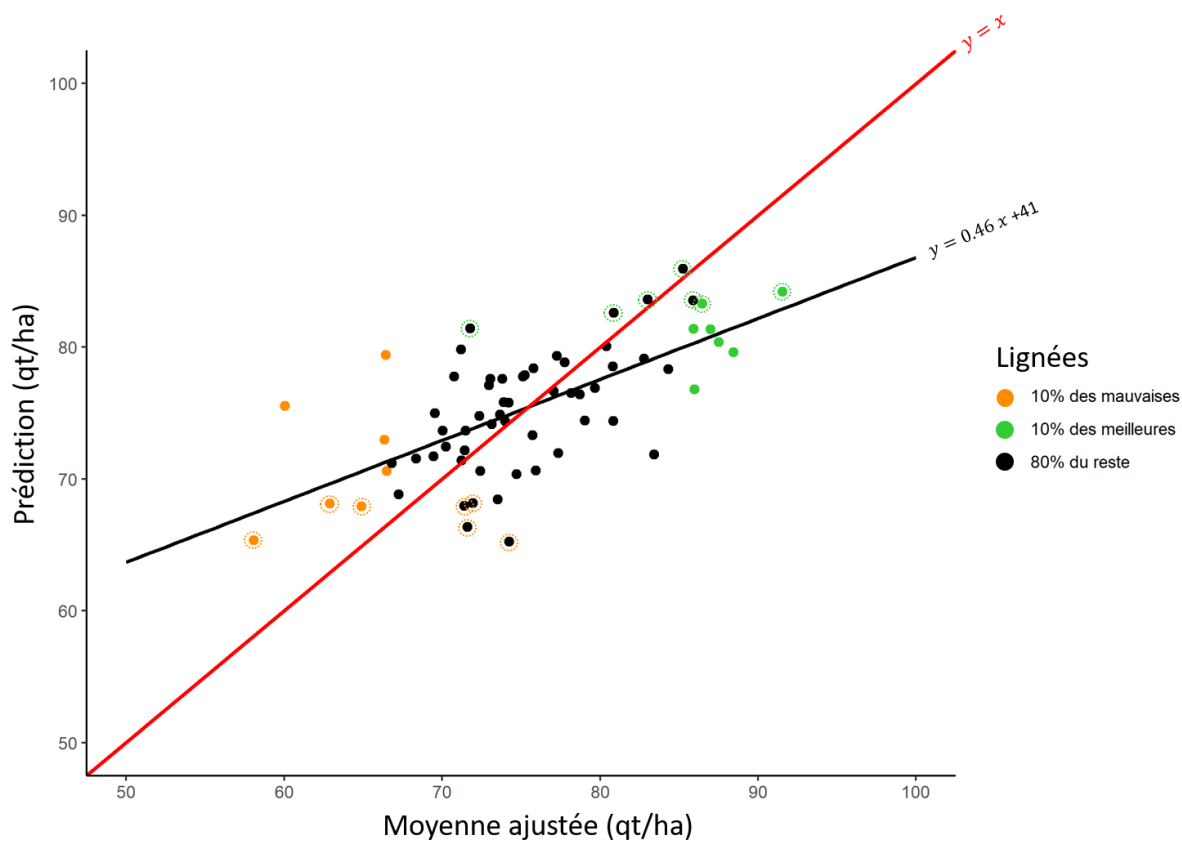


Figure 5-4. Représentation des prédictions de rendements en fonction des moyennes ajustées pour 65 lignées du jeu de données Set2_2018_CF. La droite rouge est la fiabilité optimale à atteindre pour le modèle. Basées sur la moyenne ajustée du rendement, les points verts sont les 10% des meilleures lignées et les points orange les 10% des plus mauvaises. Les cercles en pointillés verts et orange sont respectivement les 10% des meilleures lignées et les 10% les plus mauvaises basées sur les prédictions phénomiques.

Il est aussi important que la méthode donne des prédictions stables dans un même environnement de prédiction, sur les différentes validations croisées menées. Je vais donc comparer en plus des moyennes de capacités de prédiction, la stabilité des prédictions par la dispersion autour de la moyenne. J’ai sélectionné un certain nombre de modèles à prendre en compte pour calculer les moyennes de capacités de prédiction sur l’ensemble des scénarios et des données. Les modèles choisis sont explicités dans le Tableau 5-1. J’ai choisi, pour chaque méthode, le modèle qui combine le plus d’effets dans le modèle de prédiction pour la SG, la SP et le Multiomique. Pour les modèles de SP et de Multiomique, l’effet W est caractérisé par la matrice de spectres combinés.

Tableau 5-1. Choix des modèles dont la capacité de prédiction a été prise en compte pour calculer la moyenne des capacités de prédiction de la SG, la SP et du Multiomique sur l'ensemble des données. Les modèles du scénario "single-environment" (SE) sont décrits dans le Chapitre III et les modèles des scénarios "multi-environment trials" (MET) sont décrits dans le Chapitre IV. WMRR fait référence à la méthode de la régression Ridge multivariée pour prédire les spectres manquants (cf Chapitre IV).

Scénarios	SG	SP	Multiomique	
SE	M	CbSD	CbSD+M	
	oGoE nGoE	EG_GxE	EWcb_WcbxE	EGWcb_GxE_WcbxE
MET	oGoEref nGoEref	EG_GxE (WMRR)	EWcb_WcbxE (WMRR)	EGWcb_GxE_WcbxE (WMRR)
	oGnE nGnE	EG	EWcb_P	EGWcb_P

Dans le Tableau 5-2, j'ai résumé les différentes capacités de prédiction obtenues par différents modèles de SP et de SG de l'ensemble des scénarios de prédiction. Les moyennes des corrélations et leur écart-type pour chaque jeu de données sont moyennés par set. Nos résultats, montrent que sur l'ensemble des scénarios de prédiction et des sets de caractères prédits, la sélection phénomique atteint des moyennes de capacités de prédiction comprises entre 0 et 0.93, tandis que la sélection génomique atteint des moyennes de capacités de prédiction comprises entre 0.01 et 0.95. Dans l'ensemble des scénarios et sets, la SP fait aussi bien que la SG dans 50% des cas pour la prédiction du rendement et dans 33% des cas pour la date d'épiaison. Les dispersions autour de la moyenne sont très similaires entre les prédictions génomique et phénomique. Cette dispersion n'est donc pas dépendante de la méthode utilisée.

J'ai aussi résumé l'ensemble des capacités de prédiction de la SG et de la SP dans la Figure 5-5, où sont représentées des boîtes à moustaches des capacités de prédictions obtenues sur l'ensemble des jeux de données et l'ensemble des scénarios de prédictions testés. Ce graphique montre que pour le rendement, les moyennes des capacités de prédiction sont sensiblement les mêmes entre SP et SG et que pour la date d'épiaison, celles de la SP sont légèrement plus hautes que pour la SG. Il est intéressant de noter que les capacités de prédiction comprises entre le 1^{er} et le 3^{ème} quartiles (donc 50% des capacités de prédiction) sont plus étendues pour la SG que la SP. Cela

peut s'expliquer par le fait que dans les scénarios de types oGnE ou nGnE, la SG a des capacités de prédiction du rendement proches de zéro dans la moitié des jeux de données contrairement à la SP qui a de faibles capacités de prédiction uniquement pour un quart des jeux de données. En regardant non plus sur un environnement de prédiction mais sur l'ensemble des données que j'avais à disposition et sur les différents scénarios que j'ai testés, la SP semble plus stable que la SG sur l'ensemble des données. Cette stabilité est recherchée par le sélectionneur pour que la méthode de prédiction soit stable d'une année à l'autre.

En mettant en regard le nombre de lignées, le nombre de longueurs d'onde et le nombre de marqueurs moléculaires utilisés, il est difficile de déterminer clairement pour quelle raison la SP fonctionne mieux ou moins bien que la SG dans un cas ou dans un autre. La diversité des jeux de données que j'avais à disposition ne permet pas de répondre clairement à cette question. De mon point de vue, connaître a priori des critères discriminant l'une ou l'autre méthode serait un atout majeur pour le sélectionneur afin d'adapter au mieux sa stratégie de sélection en fonction des moyens qu'il a à disposition. D'autres données seraient nécessaires pour décider quelle situation est plus favorable à la sélection phénotypique plutôt qu'à la sélection génomique. Par exemple, avoir un jeu de données tel que le jeu de données BreedWheat avec des centaines de lignées toutes génotypées et phénotypées sur un grand jeu d'environnements plus ou moins contrastés. Il faudrait cependant disposer des spectres pris dans chacun des environnements pour déterminer si certains environnements sont plus favorables à l'utilisation de la SP ou de la SG. Je pense notamment aux environnements stressants qui ont montré de meilleurs résultats en termes de capacités de prédiction phénotypique pour le rendement (Rincent *et al.* 2018 ; Galán *et al.* 2020).

Pour le sélectionneur, choisir une méthode de prédiction plutôt qu'une autre ne se base pas que sur la capacité de prédiction du modèle. L'objectif est d'utiliser les méthodes de prédiction en vue d'augmenter le gain génétique des programmes de sélection. La capacité de prédiction intervient directement dans l'équation du sélectionneur puisqu'elle intervient dans la précision de sélection (r) (cf. Chapitre I). Seulement, cette précision va aussi être liée aux moyens humains et financiers nécessaires pour mettre en œuvre la méthode. Le sélectionneur va adapter sa stratégie en fonction du coût de celle-ci: il doit donc trouver un compromis entre précision et coût.

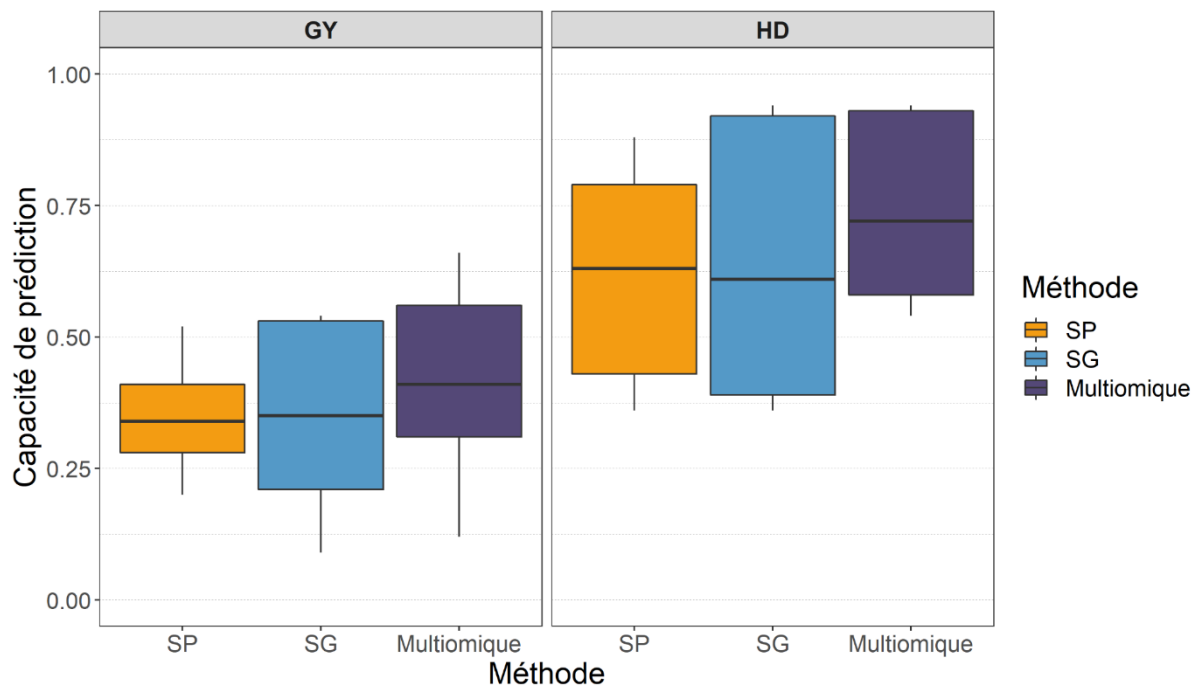


Figure 5-5. Boîtes à moustaches de la capacité de prédiction phénotypique (SP), de la capacité de prédiction génomique (SG) et de la capacité de prédiction de la méthode combinant marqueurs moléculaires et spectres NIR (Multiomique) pour le rendement (GY) et la date d'épiaison (HD) basées sur l'ensemble des jeux de données et des scénarios de prédictions. Les modèles utilisés pour calculer les moyennes de capacités de prédiction sont présentés dans la Table1. Chaque boîte à moustache indique la moyenne (barre épaisse horizontale), le 1er et 3ème quartiles (boîtes) et le 1er et 9ème déciles (moustaches).

Je vais donc comparer des simulations de coûts et fiabilités de prédiction pour la SP et la SG. La méthode employée est celle décrite par Rincenc *et al.* (2018), et pour laquelle j'ai fait varier 4 paramètres: le coût du génotypage, le coût de l'acquisition des spectres NIR, le nombre d'individus en sélection (qui sont à prédire) et la fiabilité des prédictions. La fiabilité génomique ou phénotypique a été définie comme le carré de la capacité de prédiction divisé par l'héritabilité du caractère. Cette simulation consiste à représenter le gain génétique attendu avec la SG ou la SP en un cycle de sélection selon ces différents paramètres. La vraie valeur génétique et la valeur prédite ont été simulées à partir d'une distribution normale multivariée avec les moyennes égales à 0, les variances égales à 1 et la covariance égale à la racine carrée de la fiabilité.

Le gain génétique est la différence entre la moyenne des vraies valeurs génétiques des individus sélectionnés sur la base des prédictions et la vraie valeur génétique moyenne de la population (égale à 0).

Tableau 5-2. Comparaison des capacités de prédiction entre trois méthodes de prédictions différentes (SP, SG et Multiomique). La meilleure capacité de prédiction entre la SP et la SG est notée en gras. La meilleure capacité de prédiction entre toutes les méthodes est indiquée en rouge. Les capacités de prédiction ont été déterminées sur la base des capacités de prédiction obtenues par les modèles présentés dans le Tableau 5-1.

Scénario	Set	Nombre de lignées	Nombre de longueurs d'onde	Nombre de SNPs	Capacité de prédiction					
					Rendement			Date d'épiaison		
					SP	SG	Multiomique	SP	SG	Multiomique
SE	Set1	68-152	2 100	5824	0.28 (0.19)	0.27 (0.18)	0.36 (0.17)	0.60 (0.14)	0.24 (0.21)	0.60 (0.15)
	Set2	325	2 100	19 512	0.60 (0.08)	0.53 (0.08)	0.66 (0.07)	x	x	x
	Set3	57-69	2 100	5824	0.40 (0.24)	0.38 (0.23)	0.44 (0.12)	0.53 (0.18)	0.43 (0.25)	0.58 (0.18)
	Set4	71-100	2535 - 3380	12 303 - 19 512	0.37 (0.19)	0.41 (0.2)	0.46 (0.18)	0.44 (0.17)	0.46 (0.2)	0.58 (0.17)
MET oGoE	Set1	152	2 100	5824	0.39 (0.16)	0.17 (0.16)	0.40 (0.15)	0.76 (0.07)	0.72 (0.1)	0.81 (0.06)
	Set2	325	2 100	19 512	0.60 (0.08)	0.55 (0.08)	0.68 (0.07)	x	x	x
	Set4 2018	71	2535	12 303	0.39 (0.24)	0.35 (0.24)	0.43 (0.22)	0.88 (0.08)	0.92 (0.07)	0.93 (0.06)
	Set4 2019	100	3380	19 512	0.50 (0.15)	0.6 (0.14)	0.66 (0.12)	0.93 (0.08)	0.94 (0.08)	0.94 (0.08)
MET nGoE	Set1	152	2 100	5824	0.39 (0.16)	0.24 (0.15)	0.42 (0.15)	0.72 (0.08)	0.36 (0.14)	0.74 (0.06)
	Set2	325	2 100	19 512	0.60 (0.08)	0.54 (0.1)	0.67 (0.07)	x	x	x
	Set4 2018	71	2535	12 303	0.34 (0.22)	0.34 (0.23)	0.36 (0.23)	0.37 (0.25)	0.43 (0.21)	0.57 (0.18)
	Set4 2019	100	3380	19 512	0.42 (0.16)	0.43 (0.17)	0.56 (0.14)	0.43 (0.17)	0.52 (0.2)	0.59 (0.16)

Scénario	Set	Nombre de lignées	Nombre de longueurs d'onde	Nombre de SNPs	Capacité de prédiction					
					Rendement			Date d'épiaison		
					SP	SG	Multiomique	SP	SG	Multiomique
MET oGoEref	Set1	152	2 100	5824	0.28 (0.15)	0.17 (0.16)	0.29 (0.16)	0.66 (0.09)	0.8 (0.06)	0.81 (0.06)
	Set2	325	2 100	19 512	0.41 (0.1)	0.54 (0.09)	0.57 (0.09)	x	x	x
	Set4 2018	71	2535	12 303	0.24 (0.24)	0.46 (0.25)	0.44 (0.24)	0.72 (0.2)	0.94 (0.04)	0.94 (0.04)
	Set4 2019	100	3380	19 512	0.42 (0.18)	0.6 (0.13)	0.64 (0.12)	0.87 (0.13)	0.94 (0.06)	0.94 (0.06)
MET nGoEref	Set1	152	2 100	5824	0.29 (0.14)	0.22 (0.15)	0.33 (0.14)	0.6 (0.11)	0.36 (0.15)	0.66 (0.1)
	Set2	325	2 100	19 512	0.4 (0.1)	0.54 (0.08)	0.58 (0.08)	x	x	x
	Set4 2018	71	2535	12 303	0.19 (0.25)	0.34 (0.21)	0.31 (0.24)	0.32 (0.25)	0.39 (0.2)	0.49 (0.19)
	Set4 2019	100	3380	19 512	0.37 (0.19)	0.44 (0.17)	0.51 (0.16)	0.27 (0.22)	0.42 (0.23)	0.48 (0.21)
Met oGnE	Set1	152	2 100	5824	0	0.01	-0.02	0.79	0.83	0.85
	Set2	325	2 100	19 512	0.24	0.08	0.12	x	x	x
	Set4 2018	71	2535	12 303	0.30	0.31	0.3	0.85	0.93	0.94
	Set4 2019	100	3380	19 512	0.41	0.54	0.54	0.94	0.95	0.95
Met nGnE	Set1	152	2 100	5824	0.04 (0.2)	0.17 (0.16)	0.06 (0.2)	0.64 (0.1)	0.34 (0.16)	0.69 (0.09)
	Set2	325	2 100	19 512	0.26 (0.16)	0.03 (0.21)	0.12 (0.11)	x	x	x
	Set4 2018	71	2535	12 303	0.2 (0.23)	0.09 (0.26)	0.14 (0.26)	0.42 (0.22)	0.39 (0.22)	0.54 (0.19)
	Set4 2019	100	3380	19 512	0.28 (0.19)	0.32 (0.18)	0.38 (0.17)	0.40 (0.16)	0.51 (0.19)	0.58 (0.16)

J'ai considéré un budget fixe de 50 000 € alloué à la prédiction de la performance des candidats. Ce budget correspond au génotypage de 2 500 lignées si le coût de génotypage est de 20 € ou de 5 000 lignées si le coût de génotypage est de 10 €. Je me place ici dans un ordre de grandeur d'effectifs 10 fois plus grand que ceux que j'ai eus à disposition. L'idée est d'observer les gains obtenus aussi sur des données simulées qui miment les effectifs et la sélection qui se fait plus en amont dans le réel programme de sélection. J'ai actualisé les paramètres initialement choisis par Rincent *et al.* (2018) aux nouveaux coûts de génotypage et de phénotypage et avec un budget alloué plus représentatif de ce qui peut être fait en sélection du blé tendre. Ce budget de 50 000 € est utilisé pour sélectionner les 250 meilleurs candidats à l'issue des prédictions. Les 250 lignées correspondent globalement aux effectifs sélectionnés pour les débuts d'essais en parcelle d'un programme classique. Pour le coût de l'acquisition des spectres NIR, il est difficile d'estimer le coût réel des spectres, sachant qu'une partie d'entre eux sont déjà collectés pour d'autres utilisations que la SP (généralement la teneur en protéines). Comme j'ai effectué mes prédictions avec des modèles SP qui combinent 2 à 4 spectres selon le jeu de données, j'ai considéré des coûts globaux de 2 €, 6 € ou 10 € pour l'acquisition de l'ensemble des spectres requis par le modèle. Enfin, il est à noter que j'ai toujours considéré la capacité de prédiction des modèles et non la fiabilité des modèles. Les valeurs obtenues dans le Tableau 5-2 doivent donc être prises au carré et divisées par l'héritabilité du caractère pour obtenir une estimation de la fiabilité des prédictions. L'héritabilité a été calculée lors du calcul des moyennes ajustées pour les différents caractères. Lorsque l'héritabilité était manquante pour un set, j'ai pris la moyenne des héritabilités de l'ensemble des sets.

Dans un premier temps, je vais analyser les gains génétiques simulés sur un scénario de prédiction intra environnement (Scénario SE Tableau 5-2). Dans un second temps, je m'intéresserai aux gains génétiques simulés sur un scénario de prédiction en réseau d'essais en sparse testing (Scénario MET oGoE Tableau 5-2).

Pour le scénario SE, la fiabilité de la SP est comprise entre 0.06 et 0.40 pour la prédiction du rendement et entre 0.20 et 0.37 pour la prédiction de la date d'épiaison. La fiabilité de la SG est comprise entre 0.08 et 0.32 pour la prédiction du rendement et comprise entre 0.06 et 0.22 pour la prédiction de la date d'épiaison. J'ai représenté sur les simulations de gain génétique, les fiabilités de capacités de prédiction obtenues avec les sets 1-4 pour le rendement et la date d'épiaison (Figure 5-6).

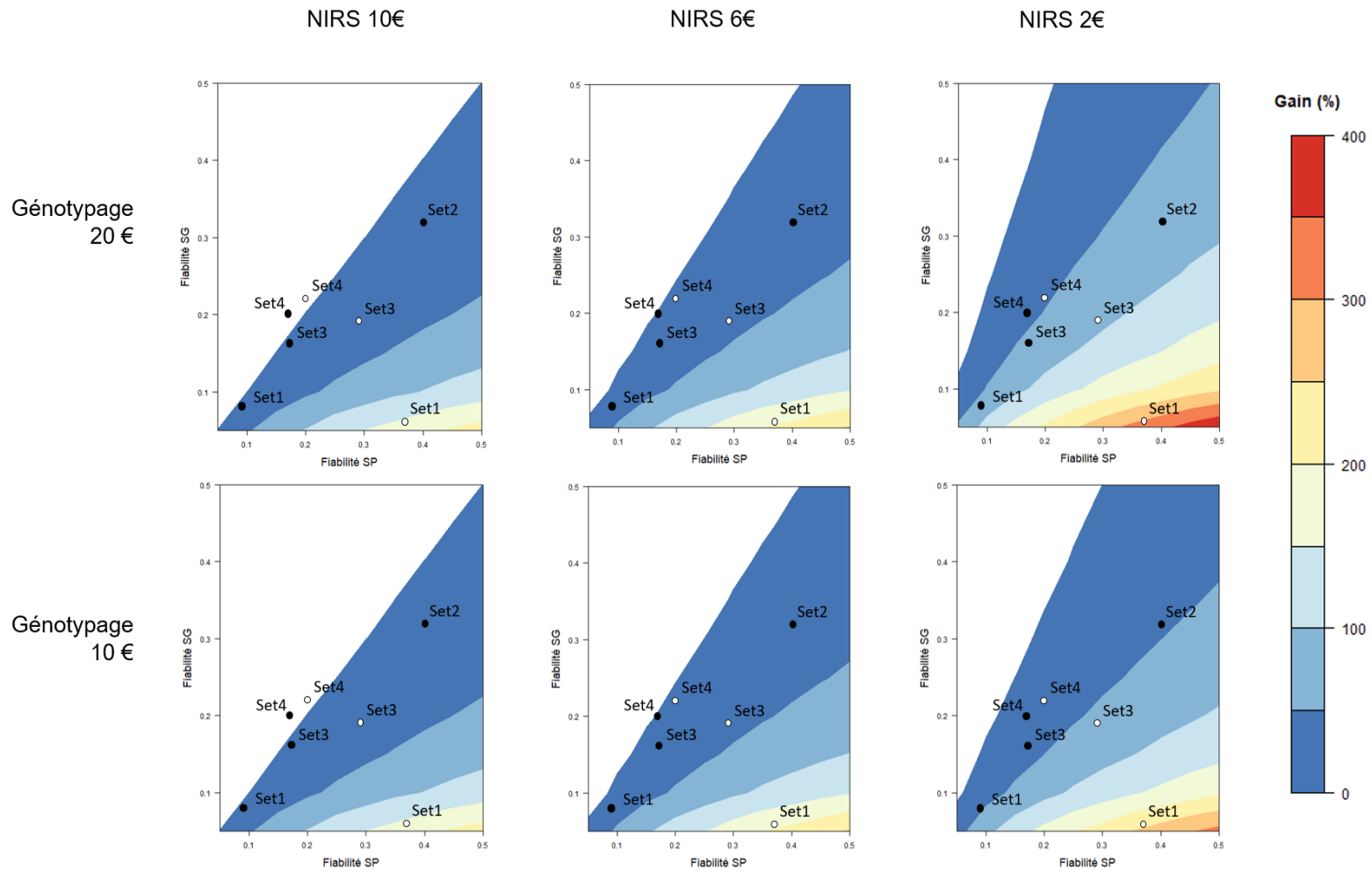


Figure 5-6. Augmentation théorique du gain génétique attendu (%) en utilisant la SP plutôt que la SG dans le scénario SE, pour différentes valeurs de fiabilité de chacune des méthodes, pour différents coûts de génotypage (20 € ou 10 €) et d'acquisition des spectres NIR (10 €, 6 €, 2 €). Le même budget de 50 000€ est consacré aux différentes simulations. Pour chaque simulation, les vraies valeurs génétiques et les valeurs prédites ont été simulées grâce à une distribution normale bivariée avec une covariance adaptée aux différentes fiabilités paramétrées. La zone blanche est la zone dans laquelle il n'y a pas de gain à utiliser la SP plutôt que la SG. Les points représentent les fiabilités obtenues pour le rendement (noir) et la date d'épiaison (blanc) dans les différents sets.

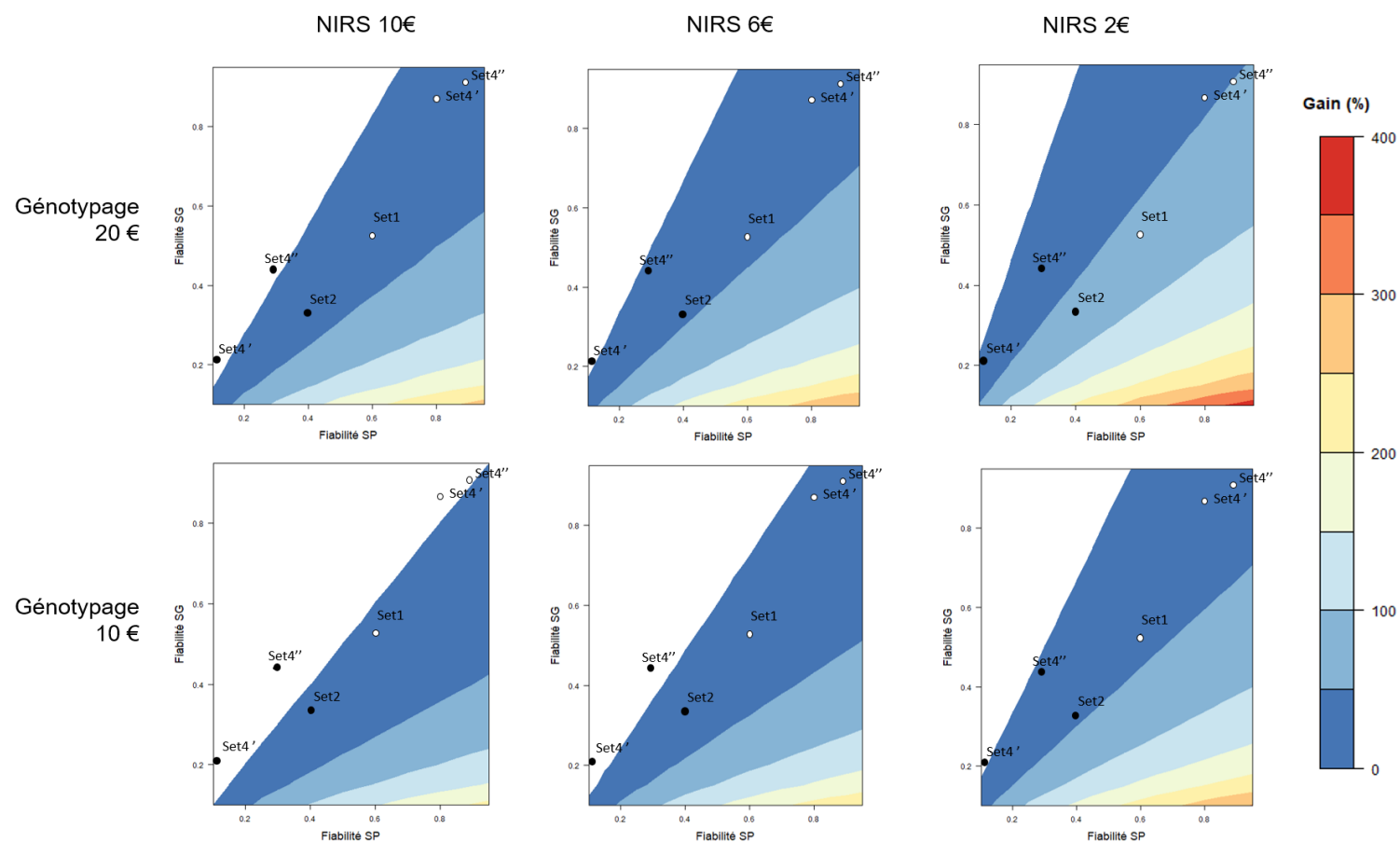


Figure 5-7. Augmentation théorique du gain génétique attendu (%) en utilisant la SP plutôt que la SG dans le scénario MET oGoE, pour différentes valeurs de fiabilité de chacune des méthodes, pour différents coûts de génotypage (20 € ou 10 €) et de coûts d'acquisition de spectres NIR (10 €, 6 €, 2 €). Le même budget de 50 000€ est consacré aux différentes simulations. Pour chaque simulation, les vraies valeurs génétiques et les valeurs prédites ont été simulées grâce à une distribution normale bivariée avec une covariance adaptée aux différentes fiabilités paramétrées. La zone blanche est la zone dans laquelle il n'y a pas de gain à utiliser la SP plutôt que la SG. Les points représentent les fiabilités obtenues pour le rendement (noir) et la date d'épiaison (blanc) dans les différents sets (Set4' correspond à Set4_2018 et Set4'' à Set4_2019). Les fiabilités de prédiction du rendement de Set1 sont manquantes, celles-ci sont trop faibles pour être représentées sur le graphique.

Les résultats obtenus ici montrent que toutes les fiabilités de SP et de SG que j'ai obtenues dans les différents sets permettent de remplacer la SG par la SP à l'exception du set 4, lorsque le génotypage et l'acquisition des spectres ont le même coût de 10 €. Globalement la SP par rapport à la SG offre un gain génétique attendu pour la prédiction du rendement, entre 0% et 100% et pour la prédiction de la date d'épiaison, un gain attendu entre 0% et 300%. Sur l'ensemble des sets, les gains sont les plus importants lorsque le NIRS a le coût le plus faible (2 €) et le génotypage a le coût le plus élevé (20 €). Ceci est attendu puisque qu'il y a un facteur 10 de différence de coût entre les deux méthodes. Dans le cadre de prédiction SE, la SP se trouve en général être une très bonne alternative à la SG pour les deux caractères à prédire.

Pour le scénario MET oGoEref, la fiabilité de la SP est comprise entre 0.09 et 0.22 pour la prédiction du rendement et entre 0.45 et 0.78 pour la prédiction de la date d'épiaison. La fiabilité de la SG est comprise entre 0.03 et 0.46 pour la prédiction du rendement et comprise entre 0.66 et 0.91 pour la prédiction de la date d'épiaison. J'ai représenté sur les simulations de gain génétique, les fiabilités de capacités de prédiction obtenues avec les sets 1, 2 et 4 pour le rendement et la date d'épiaison (Figure 5-7). Les résultats obtenus montrent que contrairement aux gains obtenus avec le scénario SE, les gains ici sont compris entre 0% et 100% pour le rendement et la date d'épiaison. Plus important encore, si le coût d'acquisition du NIRS dépasse 2 € alors certaines fiabilités ne sont plus assez importantes en SP pour que le remplacement de la SG par la SP soit rentable et ce, quel que soit le prix du coût de génotypage. Dans le cadre de prédictions dans un réseau d'essais, la SP reste une alternative à la SG mais pas pour n'importe quel coût et cela dépend aussi des jeux de données.

Dans le scénario de "*sparse testing*" oGoE, je suis partie du postulat que les spectres peuvent être pris dans tous les environnements du réseau d'essais pour toutes les lignées et cela sans surcoût supplémentaire. Ces gains sont possibles si les sélectionneurs utilisent en parallèle de leurs essais, des pépinières pour d'autres caractérisations phénotypiques (maladies ou verse par exemple). Dans ce cas-là, il n'y a pas de surcoût lié à la mise en place d'une nouvelle pépinière pour y acquérir les spectres. J'ai aussi essayé la simulation du gain génétique attendu sur le scénario oGoEref pour lequel les spectres sont prédits pour les lignées candidates. Pour ce scénario, les fiabilités de prédiction de la SP sont trop faibles par rapport à la SG pour qu'on puisse avoir un intérêt à utiliser la SP à la place de la SG. La méthode de prédiction des spectres doit être encore améliorée pour avoir de meilleures fiabilités de prédiction et concurrencer la SG dans ce type de scénario.

Ainsi, j'ai donc montré à travers ces simulations que la SP se présente comme une bonne alternative à la SG que ce soit en termes de variation de coûts ou de fiabilité de prédiction. De plus, son intérêt est encore plus grand lorsque, dans certains programmes, les spectres NIR sont déjà acquis pour d'autres applications (prédiction de la teneur en protéine par exemple). Hormis dans le cas où l'investissement a été fait pour que le génotypage soit très peu cher, alors la SP concurrence réellement la SG si ses fiabilités de prédiction sont nettement meilleures que celles de la SG. De mon point de vue, il faudrait continuer ce travail de comparaison de fiabilités de méthodes sur un plus grand nombre de données pour couvrir une large gamme de fiabilités et ainsi déterminer les gains génétiques associés.

2.3. ... ou qui peuvent être complémentaires

Mon premier objectif était de montrer que la sélection phénotypique pouvait être une alternative à la sélection génomique notamment pour les entreprises de sélection ayant des programmes d'amélioration pour lesquels l'équilibre entre coût et gain génétique de la méthode n'est pas intéressant. Je me suis intéressée dans un second temps à la combinaison de ces deux variables aux natures différentes. J'ai développé dans le chapitre III et le chapitre IV, des modèles qui intègrent à la fois l'information portée par le marquage moléculaire et par les spectres NIR. Ces modèles sont meilleurs dans la plupart des scénarios et sets que la SP ou la SG. En se référant au Tableau 5-2, dans lequel on a choisi des modèles pour comparer SP, SG et Multiomique, la méthode Multiomique donne les meilleures capacités de prédiction dans 64% des cas pour le rendement et dans 76% des cas pour la date d'épiaison. Dans les scénarios oGnE et nGnE, j'ai choisi de prendre les mêmes types de modèles pour la SP (avec l'utilisation de l'effet EWcb) pour comparer les capacités de prédiction des trois méthodes. J'ai choisi de considérer les modèles qui tiennent compte d'un maximum d'effet pour calculer les moyennes de capacité de prédiction. Or pour ces scénarios, le meilleur modèle est EG_P, qui ne considère pas l'effet hyperspectral Wcb. Comme le meilleur modèle n'est pas celui attendu dans ce cas, la méthode Multiomique n'apparaît pas comme la meilleure des méthodes.

Sur tous les scénarios et sets confondus, la méthode de la Multiomique fait en moyenne mieux de 18% que la SG pour la capacité de prédiction du rendement et de 15% pour la capacité de prédiction de la date d'épiaison (Figure 5-5). En termes de dispersion, la Multiomique réduit l'étendue entre le 1^{er} et le 3^{ème} quartile par rapport à celle de la SG, et cela pour les deux caractères. La combinaison des marqueurs moléculaires avec des spectres NIR tendrait donc à améliorer et stabiliser les capacités de prédiction génomiques.

Dans la littérature, Krause *et al.* (2019) et Galán *et al.* (2021) ont montré que combiner marqueurs moléculaires et NIRS améliore les capacités de prédiction des modèles. Les résultats présentés dans le chapitre IV ont également montré que, lors de prédictions multi-environnementales, la meilleure combinaison des variables est obtenue lorsque l'effet génétique est caractérisé par les marqueurs moléculaires et que l'effet GxE est caractérisé par les spectres NIR. Avec l'avènement du phénotypage haut-débit, il est de plus en plus intéressant de combiner différents types de variables pour améliorer les modèles de SG. D'après les études GLOB menées sur le maïs (Riedelsheimer *et al.* 2012 ; Guo *et al.* 2016 ; Westhues *et al.* 2017 ; Schrag *et al.* 2018 ; Azodi *et al.* 2020), le blé (Ward *et al.* 2015) et le riz (Xu *et al.* 2016), la combinaison d'informations sur les transcrits et/ou les métabolites avec des marqueurs d'ADN a conduit à des prédictions plus précises que les modèles de SG classiques. Ceci s'explique en partie par le fait que la variance génétique des caractères complexes peut être fractionnée en différents effets comme les effets additifs des gènes, de dominance entre allèles et d'interaction entre gènes (épistasie). Les marqueurs moléculaires décrivent bien la part des effets additifs sur la variance génétique mais très mal la part des effets de dominance ou d'épistasie (Hill *et al.* 2008). Théoriquement, l'effet d'épistasie peut être estimé avec un grand nombre de marqueurs moléculaires mais cela est rendu presque impossible à cause du grand nombre d'interactions épistatiques à estimer Sweeney *et al.* (2019). De nombreux modèles ont été développés pour dépasser cette limite, comme les modèles RKHS et de réseaux de neurones qui ont la caractéristique de traiter les effets des marqueurs de manière non linéaire (Pérez-Rodríguez *et al.* 2012), et ainsi tenir compte des interactions gènes x gènes. Malgré ces avancées, ces derniers sont toujours limités pour capturer de l'épistasie physiologique (Sackton et Hartl 2016). Avec les différents omiques et la phénomique, les effets de dominance et d'épistasie seraient indirectement capturés. C'est l'hypothèse soulevée par Schrag *et al.* (2018), pour lesquels les transcrits en tant que seconde strate biologique, sont censés refléter la partie active du génome et donc capter l'expression du de ce dernier, c'est-à-dire l'épistasie physiologique. De mon point de vue, c'est aussi le cas de la spectroscopie proche infrarouge, mais il est beaucoup plus difficile d'interpréter pour quel effet (additif, dominance, épistasie), les spectres sont explicatifs puisqu'en termes de strates d'omiques, ils quantifient indirectement les molécules des différentes strates d'omiques, plus particulièrement de la métabolomique et du protéomique.

Si en termes de capacités de prédiction, la méthode Multiomique apparaît plus intéressante que la SG seule, je vais également comparer les gains génétiques attendus si la SG est remplacée par la méthode Multiomique. J'ai simulé ces gains de la même manière que dans la partie précédente (2.2), à l'exception que le coût de la Multiomique est la somme du coût global de l'acquisition des spectres NIR et du coût global du génotypage. L'objectif est de vérifier que le gain en fiabilité de prédiction obtenu par la Multiomique se répercute en termes de gain génétique additionnel si on utilise la Multiomique au lieu de la SG. Les résultats obtenus ici dans les scénarios SE et MET, montrent que pour tous les sets et caractères, la SG peut être complétée par la SP mais dans des cas particuliers (Figure 5-8 et Figure 5-9). En effet, pour la fiabilité de prédiction du rendement et de la date d'épiaison, soit il y a un gain compris entre 0 et 50% en utilisant la Multiomique plutôt que la SG seule, soit la Multiomique est moins intéressante que la SG en termes de gain génétique attendu. Cela dépend des combinaisons de coûts et des écarts de fiabilité entre les deux méthodes.

D'après les résultats que j'ai obtenus avec la méthode Multiomique, les spectres NIR peuvent stabiliser et améliorer les capacités des prédictions génomiques, ce qui en fait un avantage considérable pour pouvoir sélectionner avec précision des candidats, comme je l'ai montré sur la Figure 5-5. En termes de gains génétiques attendus, ces derniers sont meilleurs que ceux de la SG dans le cadre du scénario SE, montrant ainsi que la méthode est très attractive pour les sélectionneurs.

La SP est donc une bonne alternative ou un complément à la sélection génomique et permet aux sélectionneurs d'adapter leur stratégie de sélection en fonction de leurs moyens. L'aspect économique est souvent peu pris en compte dans la comparaison des deux méthodes, et le travail que j'ai mené ici confirme tout l'intérêt de cette nouvelle méthode de prédiction. Il serait intéressant de mener par la suite, une méta analyse avec les différents articles qui ont comparé la SG et la SP sur différentes espèces, pour rendre compte réellement des gains génétiques qu'il est possible d'avoir en remplaçant ou en complétant la SG par la SP.

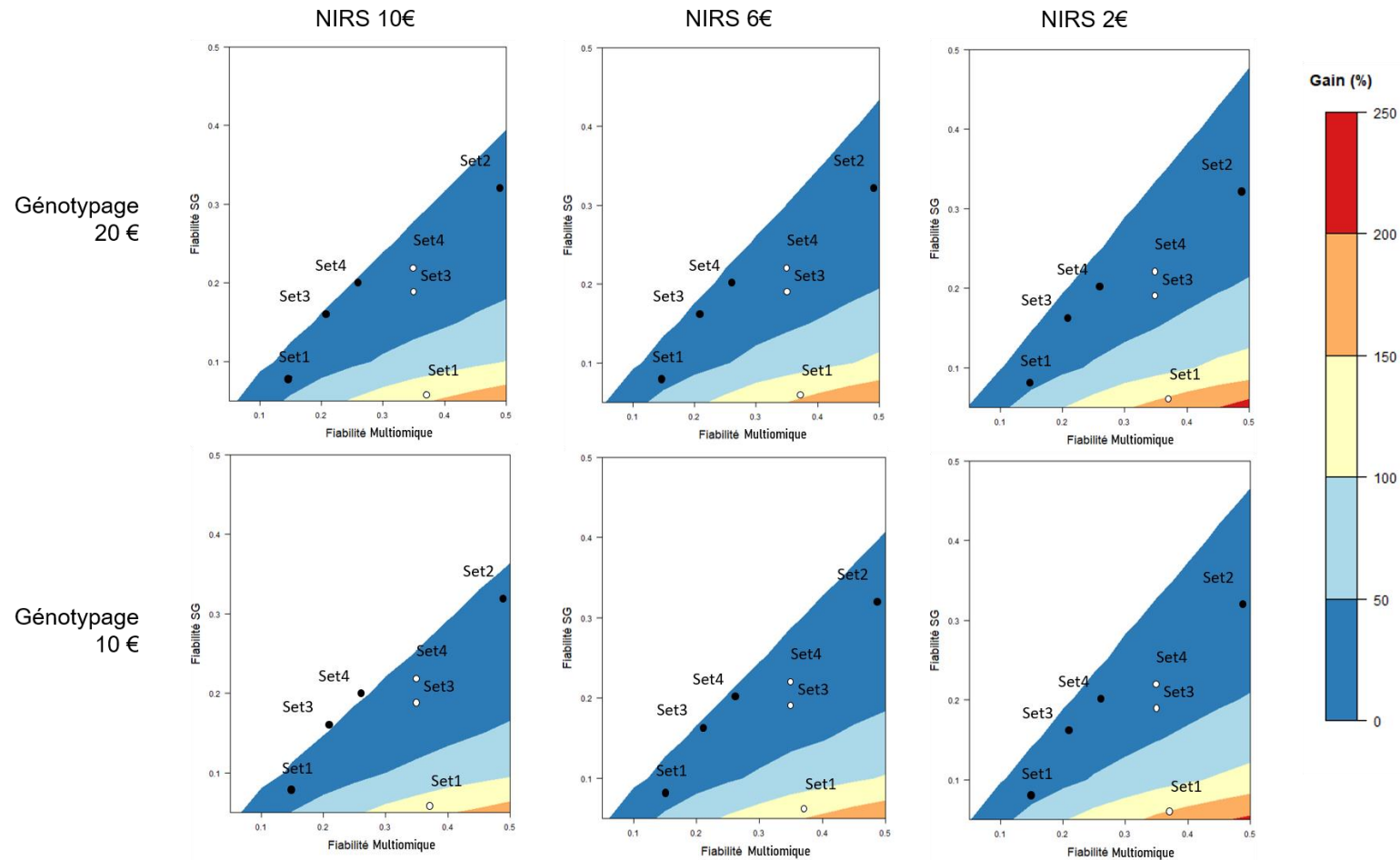


Figure 5-8. Augmentation théorique du gain génétique attendu (%) en utilisant la méthode Multiomique plutôt que la SG dans le scénario SE, pour différentes valeurs de fiabilité de chacune des méthodes, pour différents coûts de génotypage (20 € ou 10 €) et de coûts d’acquisition de spectres NIR (10 €, 6 €, 2 €). Le coût de la méthode Multiomique est donc la somme des coûts du NIRS et du génotypage (12 €, 16 €, 20 €, 22 €, 26 €, 30 €). Le même budget de 50 000€ est consacré aux différentes simulations. Pour chaque simulation, la vraie valeur génétique et la valeur estimée ont été simulées grâce à une distribution normale bivariée avec une covariance adaptée aux différentes fiabilités paramétrées. La zone blanche est la zone dans laquelle il n’y pas de gain à utiliser la méthode Multiomique plutôt que la SG. Les points représentent les fiabilités obtenues pour le rendement (noir) et la date d’épiaison (blanc) dans les différents sets.

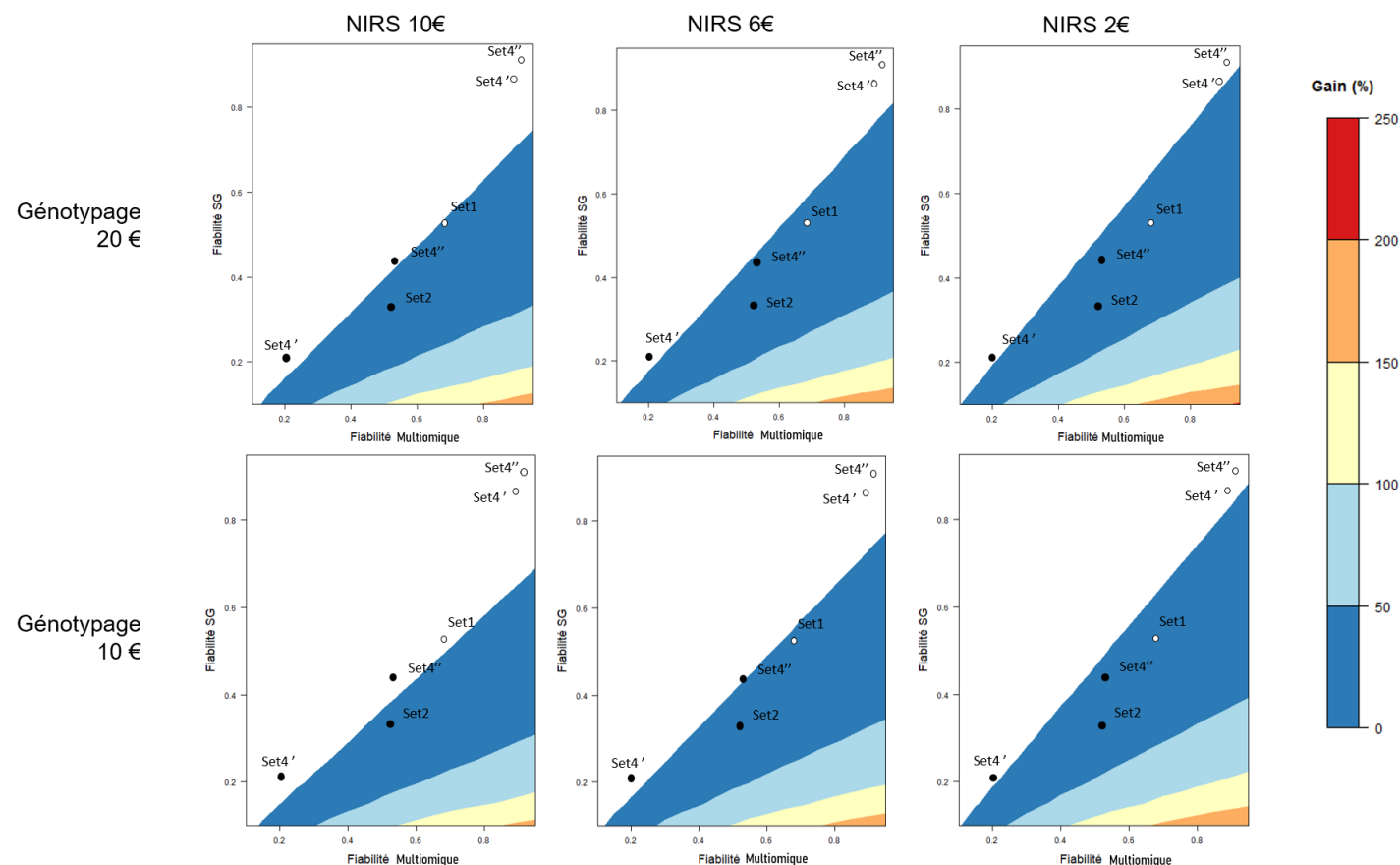


Figure 5-9. Augmentation théorique du gain génétique attendu (%) en utilisant la méthode Multiomique plutôt que la SG dans le scénario MET oGoE, pour différentes valeurs de fiabilité de chacune des méthodes et pour différents coûts de génotypage (20 € ou 10 €) et de coûts d'acquisition de spectres NIR (10 €, 6 €, 2 €). Le coût de la méthode SG+SP est donc la somme des coûts du NIR et du génotypage (12 €, 16 €, 20 €, 22 €, 26 €, 30 €). Le même budget de 50 000€ est consacré aux différentes simulations. Pour chaque simulation, la vraie valeur génétique et la valeur estimée ont été simulées grâce à une distribution normale bivariée avec une covariance adaptée aux différentes fiabilités paramétrées. La zone blanche est la zone dans laquelle il n'y pas de gain à utiliser la méthode Multiomique plutôt que la SG. Les points représentent les fiabilités obtenues pour le rendement (noir) et la date d'épiaison (blanc) dans les différents sets (Set4' correspond à Set4_2018 et Set4'' à Set4_2019). Les fiabilités de prédiction du rendement de Set1 sont hors-champs, celles-ci sont trop faibles pour être représentées sur le graphique.

3. Comment intégrer la sélection phénotypique dans les schémas de sélection du blé tendre ?

3.1. Les éléments clés identifiés pour introduire la méthode dans des programmes de sélection

Le fait qu'une méthode de prédiction telle que la SP ou la SG soit intéressante à intégrer dans un programme de sélection, repose sur sa capacité à augmenter le gain génétique par unité de temps, tout en conservant des coûts abordables. J'ai montré dans la section précédente que le gain génétique attendu par la SP était en général supérieur à celui de la SG pour un même budget alloué. Il faut ensuite que ce gain soit supérieur à celui obtenu par la sélection phénotypique. Pour les sélectionneurs, la meilleure façon de rentabiliser ces méthodes de prédiction est de les mettre en place à des générations précoces du programme de sélection (Bassi *et al.* 2016). Ces générations précoces (F2, F3 et F4) ont la particularité de présenter des effectifs de plusieurs milliers d'individus et de la variabilité génétique. A l'inverse, les générations plus avancées sont réduites en termes d'effectif et de variabilité génétique, du fait de l'élimination d'un grand nombre d'individus au cours des générations précédentes.

Aux générations précoces, la caractérisation des individus est limitée puisque la sélection se fait principalement sur des caractères visuels (aspect général de la plante, résistance aux maladies, hauteur et forme d'épis) et potentiellement sur des critères de qualité (poids spécifique, taux de protéines, etc.). Il est impossible, sur ces générations, d'envisager une évaluation du rendement car la quantité de semence est trop limitée pour mettre en place des micro-parcelles d'essais sur un ou plusieurs sites. Le coût de ce type de phénotypage, effectué sur des dizaines de milliers de lignées à ce stade, est aussi trop élevé (par micro-parcelle) et donc pas rentable. Un autre élément à prendre en compte est la fixité du matériel. En effet, pour fixer les caractères, il est nécessaire de multiplier le matériel pendant plusieurs générations pour atteindre un taux d'homozygotie acceptable. Pour gagner du temps en termes de fixité du matériel, il est possible d'avoir recours à l'haplodiploïdisation dès la génération F1 pour obtenir du matériel totalement fixé. Le sélectionneur s'affranchit donc des générations de fixation. Seulement, encore une fois, la quantité de semences issues de ces HD sera aussi limitée, ne permettant pas son implantation directe en micro parcelle d'essais.

Les méthodes prédictives sont donc très intéressantes pour aider le sélectionneur à faire ces choix sur des générations pour lesquelles la caractérisation des individus est limitée. Pour des générations plus avancées, les méthodes prédictives se trouvent être très intéressantes pour des caractères très coûteux et difficiles à acquérir. Le travail de Ben-Sadoun *et al.* (2020) a montré qu'utiliser des modèles multi caractères de sélection génomique pour prédire un caractère coûteux (note de panification), à l'aide d'un caractère peu coûteux (force boulangère), améliore les capacités de prédiction de celui-ci. Dans des travaux de prédictions génomiques des caractères de qualité boulangère, Michel *et al.* (2018) ont montré qu'il existe une forte corrélation génétique entre la teneur en protéines et différents traits de qualité boulangère du blé tendre, comme l'extensibilité, le développement et l'absorption d'eau de la pâte. Ces caractères sont directement influencés par le rapport en gliadines et gluténines, protéines responsables du réseau de gluten au sein de la pâte (Branlard 2012). La détermination du rapport gliadine/gluténine est déterminée en laboratoire, par chromatographie en phase gazeuse, et est donc difficile à obtenir. Michel *et al.* (2018), proposent d'utiliser la quantification des teneurs en protéines pour faciliter la prédiction génomique des caractères de panification. De mon point de vue, les prédictions phénotypiques pourraient être utilisées pour prédire le rapport gliadine/gluténines, qui serait ensuite utilisé dans un modèle de prédiction génomique comme caractère secondaire de caractères de qualité boulangère.

Dans les précédentes sections, j'ai listé l'ensemble des facteurs extrinsèques aux jeux d'entraînements et de prédiction, qui peuvent influencer les capacités de prédiction des modèles prédictifs en sélection phénotypique. Cependant, il faut également prendre en compte le facteur intrinsèque suivant, à savoir la proximité génétique entre les lignées à prédire et les lignées qui calibrent le modèle. En effet la capacité de prédiction des modèles génomiques dépend de l'apparentement entre les deux jeux d'individus (Habier *et al.* 2007). Pour les prédictions phénotypiques, Galán *et al.* (2021) ont montré sur le seigle, que celles-ci sont moins influencées par une grande distance génétique entre les deux panels, que les prédictions génomiques. C'est particulièrement vrai pour des caractères moyennement héritables. Ils avaient à disposition 9 familles bi-parentales et ont constitués différents scénarios de prédiction, entre panels de validation et d'entraînement. Ainsi ils ont testé des prédictions où une famille de lignées a été prédite par une population d'entraînement constituée de lignées demi-sœur et de lignées non apparentées.

Avec les données que j'avais à disposition, j'ai étudié les capacités de prédiction des modèles phénotypiques (H-BLUP) lorsque le jeu d'entraînement et le jeu de validation sont issus de différents cycles de sélection. J'ai donc prédit le rendement de lignées F6 issus des croisements de 2010 avec des modèles entraînés par des F5 issus des croisements de 2011 et inversement (Tableau 5-3). Les deux panels sont phénotypés dans le même environnement et les spectres également acquis dans cet environnement. Ce qu'il est intéressant de noter est que les capacités de prédiction sont plutôt élevées compte tenu du fait que les croisements parentaux initiaux sont différents d'un cycle à l'autre. Les représentations alléliques sont donc différentes dans chaque panel (d'entraînement ou de validation). Cette prédiction n'était possible que dans les environnements dans lesquels les deux générations ont été passées au NIRS. J'ai exploré la capacité des prédictions phénotypiques à prédire le rendement sur du matériel issu de différents cycles de sélection. En pratique, il n'y a aucun intérêt à prédire des lignées avancées en F6 qui seraient calibrées sur des lignées plus jeunes (F5) issues d'un cycle de sélection ultérieur.

Il est donc tout à fait possible de prédire des panels issus de différents cycles avec des prédictions phénotypiques. J'avais à disposition le pedigree des lignées du Set2-4. A titre indicatif, il y avait dans les deux populations, 20% parents communs. Il serait important d'étudier plus profondément ces aspects de proximité génétique entre individus comme ce qu'ont pu faire Galán *et al.* (2021) mais cette fois sur le blé tendre. Globalement, j'ai pu montrer qu'un grand nombre de facteurs influencent les capacités de prédiction, et doivent être pris en compte. Toutefois, ces différents éléments nécessitent d'être mis en pratique concrètement dans les programmes de sélection pour valider ce qui a été trouvé dans la littérature, sur ces validations croisées. L'ensemble des résultats obtenus constituent une base de travail sur laquelle les sélectionneurs peuvent s'appuyer pour commencer à mettre en place la sélection phénotypique dans les programmes de sélection.

Set	Calibration	Validation	Effectif Calibration	Effectif Validation	Environnement de prédiction	Capacite de prédiction
Set1-3	F6	F5	103	234	T_2016_HV	0.5
	F5	F6	234	103		0.54
	F6	F5	141	269	T_2016_LC	0*
	F5	F6	269	141		-0.02*
	F6	F5	82	280	T_2017_HV	0.21
	F5	F6	280	82		0.41
	F6	F5	94	184	T_2017_LC	0.15
	F5	F6	184	94		0.25
	F6	F5	102	125	T_2018_HV	0.51
	F5	F6	125	102		0.54
	F6	F5	119	113	T_2018_LC	0.31
	F5	F6	113	119		0.33
Set2-4	F6	F5	111	1418	T_2019_EM	0.33
	F5	F6	1418	111		0.41
Moyenne	F6	F5	107	374	X	0.33*
	F5	F6	374	107		0.41*

Tableau 5-3. Capacité de prédiction (PA) du rendement lorsqu'un modèle de prédiction phénotypique (H-BLUP) est entraîné avec la génération F5 pour prédire la génération F6 et inversement. Les deux générations sont expérimentées la même année dans le même lieu et représentent des cycles de sélection différents. La moyenne * de PA a été calculée sans tenir compte de l'environnement T_2016_LC où les PA sont nulles.

3.2. Proposition de schéma alternatifs intégrant la sélection phénotypique à différentes étapes du programme de sélection en généalogie différée.

Dans cette partie, je vais présenter des exemples de programmes de sélection phénotypique et leur alternative intégrant la sélection phénotypique. Je vais aussi intégrer l'utilisation des marqueurs moléculaires sous la forme de SAM ou de SG. Les exemples que je vais donner sont des propositions qui ont émergé des différentes discussions que j'ai pu avoir avec les sélectionneurs, au cours de ma thèse. A ma connaissance, aucune proposition de schémas alternatifs intégrant la sélection phénotypique n'a encore été proposée dans la littérature.

Dans deux premiers exemples (Figures 5-10 et 5-11), je propose une alternative aux schémas de sélection phénotypique, directement applicable pour les sélectionneurs.

Dans le schéma de sélection phénotypique, à l'issue des deux générations de bulk ou de SSD, les plantes choisies vont être implantées en pépinière sur une à trois lignes. C'est à cette étape du programme que l'acquisition des spectres est généralement faite pour déterminer la teneur en protéine des lignées. Il est à noter que seules les lignées F4 sélectionnées seront passées au NIRS. La sélection phénotypique peut intervenir à ce moment-là pour prédire un rendement potentiel des lignées déjà choisies par le sélectionneur et permettre d'affiner son choix. Avec les modèles développés dans le chapitre III, les prédictions pourraient être immédiatement disponibles pour le sélectionneur.

Un investissement peut être envisagé à ce stade pour passer au NIRS toute la pépinière et donc aider le sélectionneur à éliminer les lignées pour lesquelles le départage sur les critères visuels est difficile. Les lignées F4 constituent le jeu à prédire. Pour le jeu d'entraînement du modèle, je privilégierai les lignées F5 issues d'un précédent cycle de croisement et de sélection, qui seront en essais en parallèle des pépinières. Pour ce jeu d'entraînement, il est préférable d'avoir les lignées F5 implantées en essais pour réaliser le phénotypage et en pépinière pour faire la prise de spectre. Le but étant que le dispositif d'implantation des lignées n'interfère pas sur le spectre acquis et donc que lignées F5 pour la calibration et lignées F4 pour la prédiction soient conduites dans les mêmes conditions. Autrement, il faudra prévoir une correction des spectres qui tiennent compte éventuellement de l'effet dispositif. Concernant la proximité génétique des

deux jeux de données, des dégradations en capacité de prédiction sont attendues puisque ces derniers ne sont pas constitués d'individus issus des mêmes croisement parentaux. Leur proximité génétique est donc plus faible que pour des jeux constitués d'individus uniquement issus de croisements avec des parents communs. Cependant, comme montré dans la partie précédente, cette dégradation reste raisonnable et permet d'éliminer les moins bonnes lignées sans prendre trop de risques. Pour pallier au manque de proximité génétique entre les deux types de populations, il serait envisageable d'intégrer dans la population d'entraînement, les lignées parentales qui sont à l'origine de la descendance retrouvée à la génération F4. Il faudrait alors mettre en essais, plusieurs dizaines voire centaines de parents en plus des lignées en sélection. Pour minimiser le nombre de parents à phénotyper, on pourra sélectionner, sur la base d'analyses en composante principale par exemple, les parents qui capturent la plupart de la diversité génétique de l'ensemble des lignées parentales.

La sélection phénotypique peut aussi être directement appliquée à la génération F5, étape où les premiers essais ont lieu (Figures 5-10 et 5-11). Dans ce cas, les prédictions phénotypiques peuvent être menées dans différents lieux pour aider le sélectionneur à faire son choix parmi les candidats à la sélection et ne pas se baser uniquement sur le rendement dans un seul site. Les modèles que j'ai développés dans le chapitre IV, qui tiennent compte de l'effet GxE, sont des modèles directement utilisables par les sélectionneurs. Pour estimer correctement l'effet du GxE, j'ai montré qu'il était nécessaire d'acquérir les spectres dans tous les environnements de prédiction du réseau d'essais. Pour que cela soit applicable dans un programme de sélection, il faut que toutes les lignées soient en essais dans un lieu et en sparse testing dans les autres lieux (cf. scénario oGoEref chapitre IV). Très souvent, la quantité de grains récupérée sur les pépinières F4 va être limitante pour mettre en place des parcelles sur plusieurs sites. La sélection phénotypique peut donc lever cette limite. Le passage des spectres NIR peut alors se faire sur une pépinière implantée en parallèle des différents essais, ou bien, les spectres NIR sont prédits avec la méthode de la régression multivariée pondérée (cf. Chapitre IV). Il est possible alors de réduire le nombre d'années en essais multi-locaux et gagner du temps sur l'obtention de la future variété.

Une approche plus prospective serait de proposer au sélectionneur d'une part, de choisir plus d'individus à l'issue de la sélection en pépinière des F4 et d'autre part de choisir les futurs parents du nouveau cycle de sélection à cette nouvelle génération. Une pépinière de lignées F5 accueillerait 25% des individus sélectionnés en F4, pour lesquels le sélectionneur à des

doutes dans sa sélection et les 75% autres lignées sélectionnées seront placés en essais. L'idée est alors de faire des prédictions phénotypiques pour les lignées en pépinière grâce aux lignées phénotypées en essais. Avec des capacités de prédictions phénotypiques assez fiables pour sélectionner les meilleurs individus, il serait alors envisageable de rattraper des lignées de la pépinière F5, toujours basé sur les prédictions phénotypiques. Seulement la quantité de grains des lignées sélectionnées parmi les 25%, n'est pas suffisante pour ensuite faire des essais multi-locaux. Il sera nécessaire d'avoir ensuite une année de multiplication pour retrouver la quantité de grains requise. L'intérêt d'un éventuel rattrapage est donc très limité. Par contre il est toujours envisageable de sélectionner les parents des futurs croisements sur la base de prédictions phénotypiques du rendement, ce choix intervenant plus tôt dans le cycle de sélection grâce aux prédictions.

Grâce à l'utilisation de la sélection phénotypique, il est possible de gagner une année jusqu'à l'obtention des variétés, en prédisant directement le rendement des lignées dans différents sites. Il est aussi possible de gagner un an dans le choix des parents retenus pour les futurs croisements, en prédisant dès la F5, le rendement potentiel des différentes lignées dans différents lieux. Bien sûr, sous réserve que les capacités de prédiction phénotypique soient assez fiables pour sélectionner les meilleures lignées.

Dans le cadre d'un programme de sélection intégrant deux générations de SSD (F2 et F3), il est possible de faire, en complément, une sélection assistée par marqueurs pour sélectionner les plantes qui présentent les bonnes combinaisons alléliques pour un caractère très héritable défini. C'est le cas par exemple de caractères de résistance aux maladies ou de caractères de qualité boulangère (Heffner *et al.* 2010).

Dans un troisième exemple, je propose un programme de sélection prospectif qui combine à la fois les spectres NIR et les marqueurs moléculaires utilisés en sélection génomique, pour améliorer les capacités des méthodes prédictives (Figure 5-12). Avec ce type de programme, il est envisageable d'obtenir une meilleure fiabilité de prédiction pour le rendement, ce qui se traduit, pour le sélectionneur, par la possibilité d'éliminer les moins bonnes lignées mais également sélectionner les meilleures. Il serait donc nécessaire de passer au NIRS et de génotyper l'ensemble de la pépinière. Cet investissement sera rentabilisé pour plusieurs raisons. D'une part les lignées seront plus efficacement sélectionnées et ce plus tôt dans le cycle en comparaison de la sélection phénotypique classique. D'autre part, il sera possible de sélectionner les parents des futurs croisements également plus tôt dans le cycle et de même,

faire les prédictions de croisements pour déterminer les meilleures combinaisons alléliques et engendrer de la variance génétique pour les prochains cycles de sélection (Zhong et Jannink 2007). On gagne ainsi deux ans par rapport à un programme dans lequel la sélection des parents ne se ferait qu'en F6. Ce programme utilise les prédictions génomiques pour effectuer les prédictions de croisement (Bernardo 2014). Il est en effet possible de prédire la variance génétique de la descendance d'un croisement en utilisant une population appropriée de lignées génotypées et phénotypées pour calibrer le modèle (Mohammadi *et al.* 2015). Les prédictions multiomiques (utilisant spectre NIR et marqueurs moléculaires) sont quant à elles utilisées pour discriminer les individus sur le rendement.

Dans les différents programmes que j'ai présentés, j'ai toujours considéré l'application de la sélection phénotypique lorsque les spectres sont déjà disponibles pour les sélectionneurs. Actuellement, il n'est pas encore possible d'appliquer la sélection phénotypique plus tôt qu'à la génération F4, alors que la méthode aurait tout son intérêt plus en amont du programme. Pour la sortie des bulks en F3 par exemple, la récolte des plantes d'un bulk se fait en mélange, et parmi l'ensemble des grains, une quantité fixe est ressemée l'année suivante. Les grains étant choisis au hasard, une méthode de prédiction à ce moment-là aurait alors un fort potentiel. La contrainte majeure à ces générations est de pouvoir passer au NIRS des milliers de plantes avec peu de grains voir un seul grain. Cela s'appliquerait aussi à l'issue des deux générations de SSD, d'autant plus que pour la SSD menée en serre, les plantes grandissent dans des conditions contrôlées bien particulière, loin des conditions environnementales du champ.

A ma connaissance, aucune étude n'a porté sur la possibilité de faire des prédictions phénotypiques à partir de spectres obtenus sur des plantes cultivées en serre. Les limites d'applications sont donc liées à la capacité du spectromètre à acquérir les spectres sur de très petites quantités d'échantillons et à la proximité génétique du matériel F3 avec du matériel plus avancé (F5-F6) qui constituent le jeu d'entraînement du modèle. Un développement de nouveaux outils d'acquisition des spectres sera nécessaire pour obtenir les spectres sur un unique grain. L'unité d'amélioration génétique et adaptation des plantes méditerranéennes et tropicales (AGAP) de Montpellier a mis au point un spectromètre capable d'acquérir des spectres grain à grain. La réflectance est mesurée sur un grain, pour 261 longueurs d'onde de 860 nm à 1900 nm. Avec ce dispositif, ils ont montré qu'il était possible de discriminer deux variétés de blés grâce aux spectres. Il est alors possible d'accéder à l'information génétique. Encore à l'état d'expérimentation, ce type de spectromètres pourrait se développer plus largement pour d'autres applications, comme pour la prédiction phénotypique.

Les programmes de sélection que j'ai proposés ici présentent une autre limite, à savoir la constitution des jeux d'entraînements qui se fait avec du matériel en sélection. Il serait aussi intéressant de définir une population d'entraînement qui s'enrichit au fur et à mesure des cycles de sélection pour constituer un jeu d'entraînement avec de la variabilité génétique. En effet, la capacité de prédiction du modèle dépend fortement de la proximité génétique des individus qui constituent les deux jeux d'entraînement et de prédiction, puisque l'estimation des marqueurs se fait grâce à la composition allélique représentée dans le jeu d'entraînement (Habier *et al.* 2007). Pour utiliser des spectres provenant de différentes années, lieux et générations et constituer le jeu d'entraînement, j'ai montré qu'il était possible de prédire des spectres manquants (cf. Chapitre IV) dans les différents environnements ciblés. C'est une piste prometteuse pour surmonter la difficulté à utiliser les données historiques de NIRS, qu'il faudra cependant affiner par la suite puisque les capacités de prédiction avec ces spectres prédits sont moindres qu'avec les vrais spectres.

A ce jour, Aucun programme de sélection n'a encore appliqué la sélection phénotypique dans ces cycles de sélection. Les différents schémas de sélection que j'ai proposés ici, permettent donc d'offrir aux sélectionneurs, des exemples concrets d'application immédiate ou prospective de la sélection phénotypique dans les programmes de sélection du blé tendre. C'est une base sur laquelle ils peuvent s'appuyer pour ensuite adapter la méthode à leur différentes stratégies et problématiques de sélection.

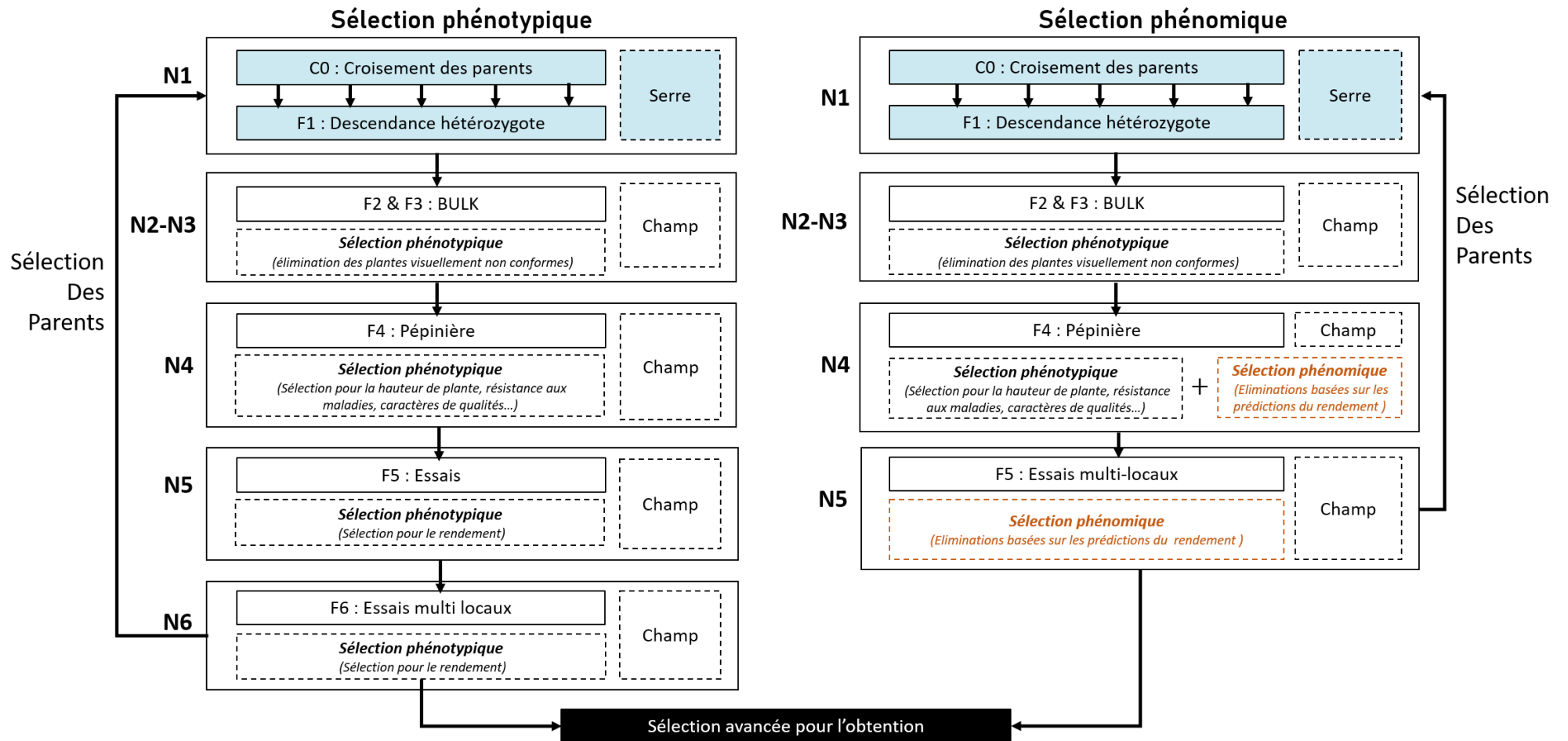


Figure 5-10. Schéma de sélection alternatif intégrant la sélection phénotypique dans le cycle de sélection. Les schémas sont basés sur des programmes “types” utilisés pour créer des variétés de lignées pures de blé tendre. “C0” correspond au croisement parental initial et “F1-F6” correspondent aux générations successives des lignées. N1-N6 correspondent aux années écoulées au cours du programme. En orange est représentée l’utilisation de la sélection phénotypique. Les rectangles en bleu représentent le passage en serre.

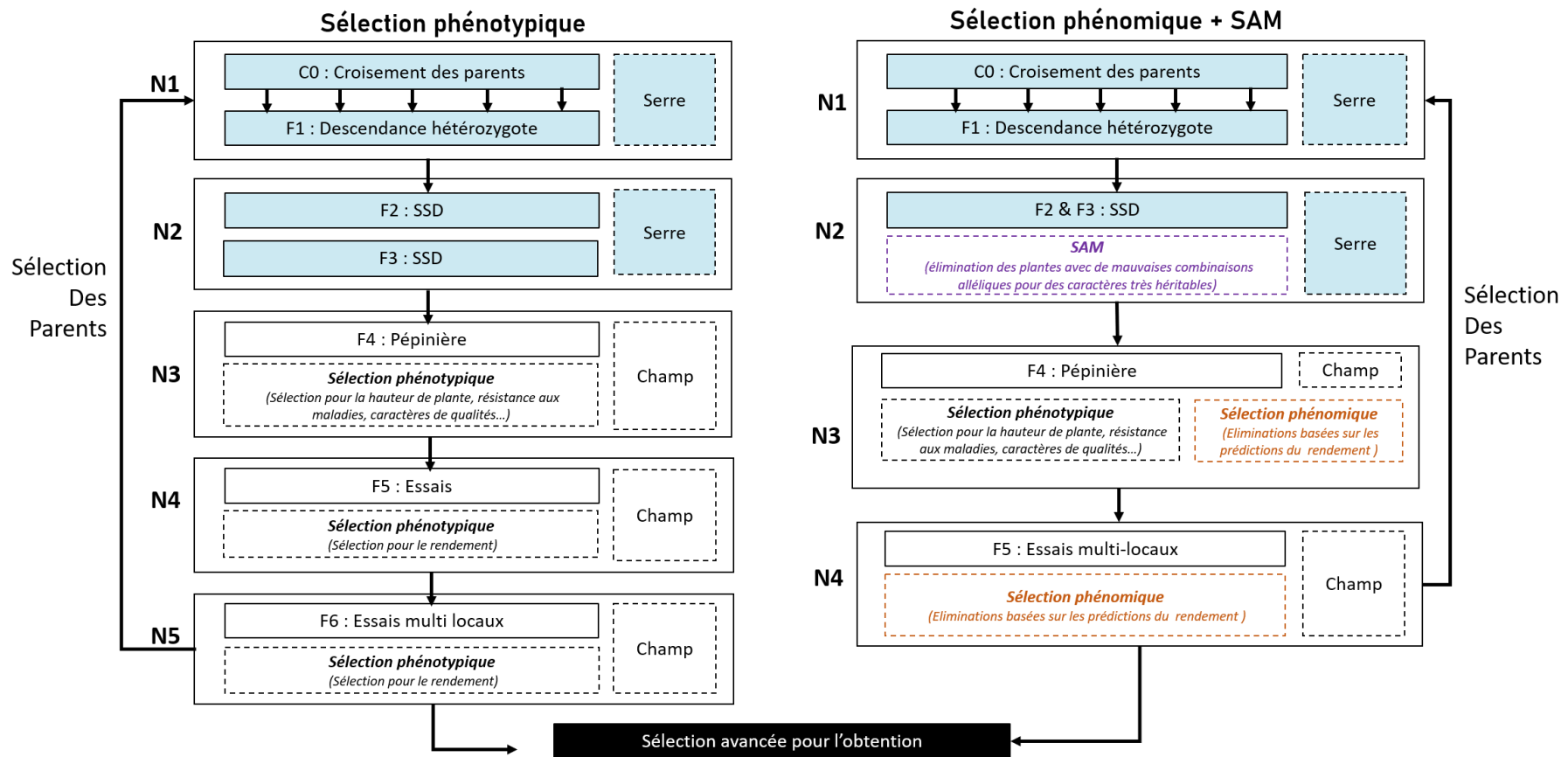


Figure 5-11. Schéma de sélection alternatif intégrant la sélection phénotypique, et la sélection assistée par marqueurs dans le cycle de sélection. Les schémas sont basés sur des programmes “types” utilisés pour créer des variétés de lignées pures de blé tendre. “Co” correspond au croisement parental initial et “F1-F6” correspondent aux générations successives des lignées. N1-N6 correspondent aux années écoulées à chaque étape du programme. En orange est représentée l’utilisation de la sélection phénotypique, en violet la sélection assistée par marqueurs (SAM). Les rectangles en bleu représentent le passage en serre.

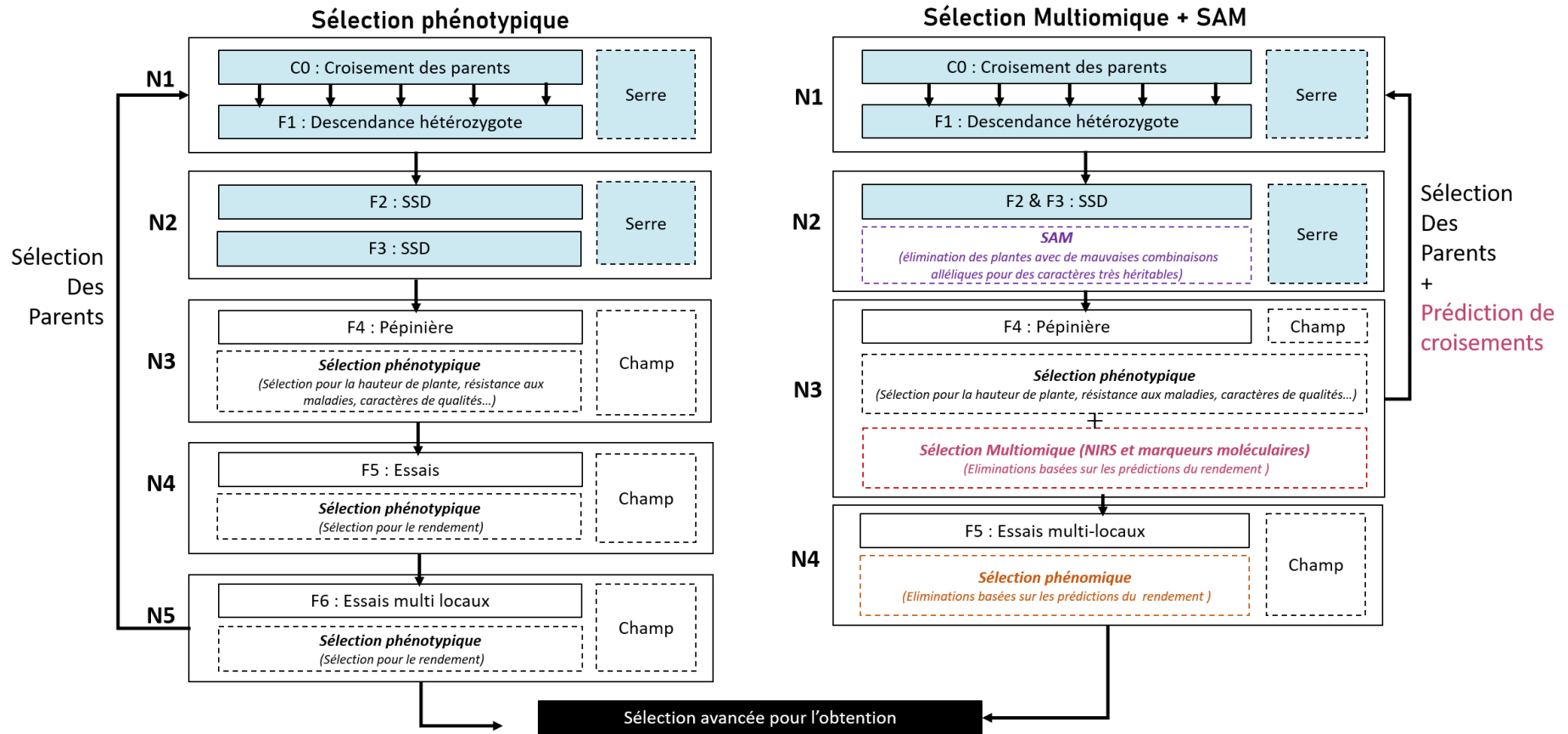


Figure 5-12. Schéma de sélection alternatif intégrant les prédictions phénotypiques, génomiques et multiomiques, ainsi que la sélection assistée par marqueurs dans le cycle de sélection. Les schémas sont basés sur des programmes “types” utilisés pour créer des variétés de lignées pures de blé tendre. “Co” correspond au croisement parental initial et “F1-F6” correspondent aux générations successives des lignées. N1-N6 correspondent aux années écoulées à chaque étape du programme. En orange est représentée l’utilisation de la sélection phénotypique, en violet la sélection assistée par marqueurs (SAM) et en rouge, l’utilisation du génotypage pour des prédictions génomiques. Les rectangles en bleu représentent le passage en serre.

4. Quelles autres perspectives d'applications de la sélection phénotypique en amélioration des plantes ?

Dans le chapitre II, différentes perspectives d'application de la sélection phénotypique sont brièvement décrites. Dans cette partie, je vais détailler plus en profondeur, une perspective d'application qui me paraît importante pour la sélection phénotypique: le pré-breeding.

J'ai montré dans les parties précédentes, que la sélection phénotypique peut être mise en pratique dans les programmes de sélection et représente une alternative ou complément à la sélection génomique. Je me suis toujours concentrée sur les différentes étapes de sélection qui succèdent aux croisements parentaux initiaux du programme de sélection. Cependant, un axe de recherche n'a pas été exploré lors de cette thèse: il s'agit de l'utilisation de la sélection phénotypique en pré-breeding. Le pré-breeding peut être défini par toutes les activités associées à l'identification de caractères ou gènes d'intérêt d'un matériel donneur, qui ne peut pas être directement utilisé dans le matériel élite, et qui doit donc être transféré vers un matériel receveur avec un meilleur fond génétique. Le pré-breeding permet donc d'obtenir des populations intermédiaires, sélectionnées sur plusieurs années, qui peuvent être utilisées par les sélectionneurs dans les croisements parentaux de leur schéma de sélection (Sharma *et al.* 2013).

La sélection de matériel élite n'exploite qu'une petite partie de la diversité génétique disponible pour l'amélioration du blé tendre. En conséquence, cette base génétique étroite qui est retrouvée dans les cultivars élites, va compromettre à long terme le gain génétique du programme de sélection. Il est donc nécessaire d'incorporer de la diversité grâce à des pools de diversité génétique, via le choix des parents à l'origine des futurs descendants sélectionnés. Ces pools peuvent être constitués d'accessions exotiques, de variétés de pays ou d'apparentées sauvages. Cette diversité est disponible dans des collections construites dans différents pays et stockées dans des banques de gènes. Environ 10% de cette diversité est utilisée par les sélectionneurs (Tadesse *et al.* 2019). Si ces derniers sont réticents à utiliser ce type de matériel, c'est à cause notamment des possibles difficultés de germination des semences, la difficulté à réaliser les croisements et le manque de caractérisation du matériel. De plus, les sélectionneurs ne veulent pas croiser du matériel élite avec du matériel peu amélioré pour les caractères agronomiques sélectionnés aujourd'hui.

Par exemple, la plupart du matériel génétique disponible dans des banques de gènes est antérieure à la révolution verte et ne possèdent donc pas les gènes de nanisme, qui est un caractère essentiel aux systèmes de production intensifs (Longin et Reif 2014).

De mon point de vue, les prédictions phénotypiques offrent un moyen peu coûteux pour caractériser ce matériel de diversité. L'idée est de phénotyper une partie du matériel de diversité et de prédire les performances du reste du matériel avec les prédictions phénotypiques. L'acquisition des spectres NIR étant facile et peu coûteuse, la caractérisation de tout ce matériel est envisageable. Cette caractérisation est d'autant plus importante qu'en pré-breeding, il arrive qu'il y ait un grand écart de performance entre les lignées donneuses issues du pré-breeding et les lignées élites. Une population intermédiaire, qui fait le pont entre les lignées issues du pré-breeding et les lignées élites issues de la sélection, est donc requise pour pallier cet écart de performance (Allier *et al.* 2020). Une meilleure caractérisation du matériel de diversité pourrait permettre de réduire plus rapidement l'écart de performance entre le matériel élite et de diversité.

Dans la littérature, très peu de publications ont traité du potentiel de la sélection génomique pour prédire des caractères d'intérêt agronomique de matériel génétique issu de banques de gènes. Crossa *et al.* (2016) ont montré sur le blé tendre, que les capacités de prédictions génomiques étaient prometteuses pour la prédiction de différents caractères (date d'épiaison, poids de milles grains, etc.) de milliers d'accessions mexicaines et iraniennes de type variété de pays. Ces prédictions peuvent permettre d'utiliser au mieux le matériel de diversité avec le matériel élite en sélection. Yu *et al.* (2016) ont également démontré sur sorgho, qu'il est possible d'intégrer les prédictions génomiques dans l'évaluation de matériels génétiques issus de banque de gènes. Par comparaison avec la sélection génomique, je pense que la sélection phénotypique serait encore plus efficace pour effectuer cette caractérisation sachant que les prédictions phénotypiques semblent être moins affectées par le faible apparentement entre les différents jeux de calibration et de prédiction (Galán *et al.* 2021) que les prédictions génomiques. C'est un atout qu'il faudrait exploiter pour ce type de prédiction.

Pour réduire l'écart de performance entre les populations de pré-breeding et les populations élites, Allier *et al.* (2020) propose de passer par une population intermédiaire de transition, appelée population de pont ou de "bridging" en anglais. Celle-ci est constituée de parents issus des populations élites et des meilleurs descendants de pré-breeding. La meilleure descendance issue de ces croisements parentaux sera ensuite utilisée dans les croisements du schéma de

sélection. Allier *et al.* (2019) ont proposé un critère de sélection pour choisir de manière optimale les croisements parentaux à effectuer, et cela basé sur les effets des marqueurs moléculaires. L'objectif de ce critère est de trouver un équilibre entre diversité génétique et gain génétique attendu de la descendance. Le génotypage reste donc nécessaire pour mettre en place ce type d'approche. Je pense cependant qu'il est possible de faire, en amont du bridging, une première filtration phénotypique des lignées de pré-breeding qui ont de très mauvaises caractéristiques agronomiques prédites.

Pour gagner du temps dans l'élaboration des populations de pré-breeding, qui nécessitent plusieurs années d'amélioration avant d'être croisées avec du matériel génétique élite, il est possible de mettre en place un système de générations accélérées encore appelé "speed breeding". Cette méthode permet d'augmenter le nombre de recombinaisons durant le développement des lignées et permet une sélection phénotypique rapide (Watson *et al.* 2018). Grâce aux chambres de culture, la durée de la photopériode peut être étendue et les températures contrôlées pour accélérer le cycle de croissance de la plante. C'est d'ailleurs ce qui est déjà mis en place en serre pour l'obtention des générations accélérées de SSD (Voss-Fels *et al.* 2019). Suite à ma proposition de schéma de sélection alternatif (cf. 3.2), il serait envisageable de mettre en place des prédictions phénotypiques avec des spectres acquis en condition contrôlée. Mais comme déjà mentionné, aucune étude n'a encore évalué les capacités de prédictions phénotypiques dans ces conditions.

Enfin sur des ressources génétiques, le phénotypage au champ qui permettrait d'obtenir les principaux caractères agronomiques d'intérêt (rendement, date d'épiaison, hauteur de plante, résistances aux maladies) est souvent difficile du fait de l'inadaptation du matériel. L'utilisation de plateformes de phénotypage en condition contrôlée peut faciliter la caractérisation de matériel de diversité en mesurant de manière précise des variables inaccessibles sans outils de mesure spécifique. La plateforme de phénotypage PhenoArc située à INRAE de Montpellier, permet la mesure de caractères associés à l'adaptation de la plante au changement climatique (Bosquet *et al.* 2015). Il est possible d'avoir accès à un grand nombre de variables comme celles liées aux rythmes de développement de la plante, au taux de croissance et à l'architecture de la plante, ou encore la capacité d'interception de la lumière et l'efficacité d'utilisation de l'eau. Ces variables sont obtenues soit par l'utilisation de l'imagerie qui, grâce aux différents clichés pris permet de reconstruire un avatar digital de chaque plante, ou par des modèles écophysologiques utilisant les paramètres de croissance des plantes (température de l'air,

incidence de la lumière etc.). Ce genre de plateforme permettrait donc de caractériser du matériel de diversité dans des scénarios futurs avec différents effets du changement climatique, et ainsi sélectionner des lignées avec des caractères source d'adaptabilité ou de résistance à introduire dans du matériel élite.

Ainsi, il me semblerait intéressant de mener des études plus approfondies sur l'utilisation de la sélection phénotypique en pré-breeding pour d'une part, en connaître un peu plus sur les capacités de la phénotypique à tirer bénéfice du matériel de diversité et d'autre part, explorer ces aspects de pré-breeding qui ont encore été peu étudiés jusqu'à maintenant. Par exemple, lors du programme Breedwheat, 4600 accessions de blé tendre présentes dans le center de ressources biologiques de Clermont-Ferrand ont été échantillonnées pour représenter la diversité mondiale présente dans le centre de ressources biologiques de Clermont-Ferrand. De ces 4600 accessions, le panel BWP3 a été constitué pour représenter la diversité mondiale de blé d'hiver, adaptée aux conditions de cultures européennes (Paux *et al.* 2022). Ce panel constitué de 450 accessions, a été phénotypé pour des caractères de résistance aux maladies, de tolérance à la sécheresse et au déficit en fertilisation azotée. Avec ce panel, il serait possible d'entraîner un modèle prédictif pour prédire les performances potentielles du reste des accessions qui n'ont pas été phénotypées. Une étude conjointe entre prédiction phénotypique et prédiction génomique serait intéressante pour déterminer si effectivement les prédictions phénotypiques sont moins affectées par des grandes distances génétiques entre accessions, que les prédictions génomiques.

Le pré-breeding est une étape à part entière du schéma de sélection, qui est encore peu étudiée aujourd'hui et qui peut être optimisée à l'instar des autres étapes de sélection. Comme j'ai pu le démontrer précédemment, les prédictions phénotypiques auraient un rôle à jouer comme alternative ou complément aux prédictions génomiques en pré-breeding.

L'ensemble de ce travail de thèse a permis de mettre en lumière une nouvelle méthode de sélection pour les programmes d'amélioration du blé tendre (Figure 5-13). Il reste toutefois plusieurs pistes à explorer qui nécessitent des recherches plus approfondies. La mise en place de la sélection phénotypique plus en amont du schéma de sélection et en pré-breeding, pourra faire l'objet de nouveaux travaux de recherche.

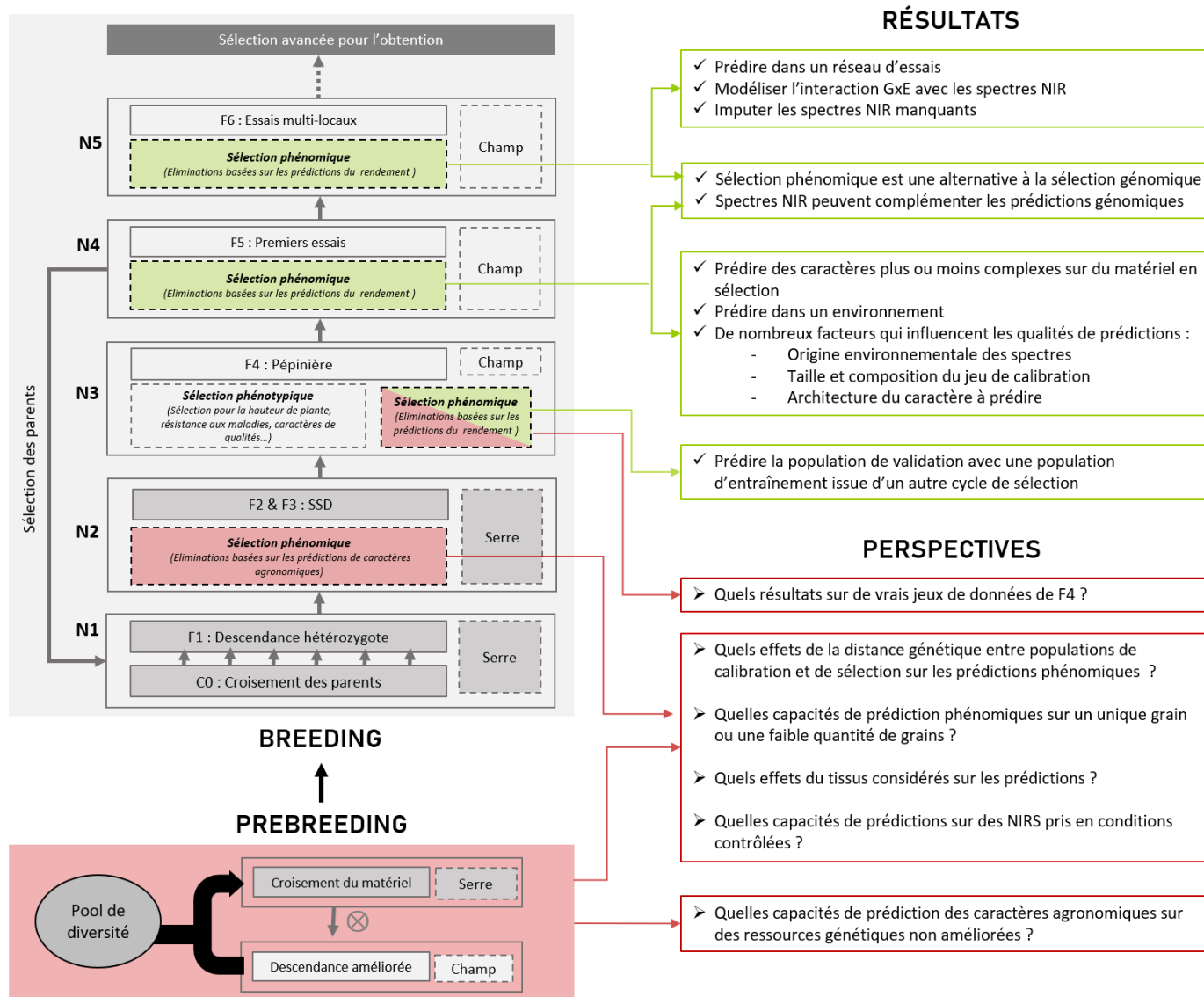


Figure 5-13. La sélection phénomique en sélection du blé tendre: les résultats majeurs de mes travaux de recherches (en vert) et perspectives de pistes de recherches (en rouge)

Références bibliographiques

Références bibliographiques

- Acevedo E, Silva P, Silva H (2002) Wheat growth and physiology. Bread wheat, improvement and production:39–70
- Agreste (2021) GraPha2021
- Aguate FM, Trachsel S, Pérez LG, et al (2017) Use of Hyperspectral Image Data Outperforms Vegetation Indices in Prediction of Maize Yield. *Crop Sci* 57:2517–2524. <https://doi.org/10.2135/cropsci2017.01.0007>
- Albrecht T, Wimmer V, Auinger H-J, et al (2011) Genome-based prediction of testcross values in maize. *Theor Appl Genet* 123:339–350. <https://doi.org/10.1007/s00122-011-1587-7>
- Allard RW, Bradshaw AD (1964) Implications of Genotype-Environmental Interactions in Applied Plant Breeding ¹. *Crop Sci* 4:503–508. <https://doi.org/10.2135/cropsci1964.0011183X000400050021x>
- Allier A, Moreau L, Charcosset A, et al (2019) Usefulness Criterion and Post-selection Parental Contributions in Multi-parental Crosses: Application to Polygenic Trait Introgression. *G3 GenesGenomesGenetics* 9:1469–1479. <https://doi.org/10.1534/g3.119.400129>
- Allier A, Teyssède S, Lehermeier C, et al (2020) Optimized breeding strategies to harness genetic resources with different performance levels. *BMC Genomics* 21:349. <https://doi.org/10.1186/s12864-020-6756-0>
- Asoro FG, Newell MA, Beavis WD, et al (2011) Accuracy and Training Population Design for Genomic Selection on Quantitative Traits in Elite North American Oats. *Plant Genome* 4:132–144. <https://doi.org/10.3835/plantgenome2011.02.0007>
- Azodi CB, Pardo J, VanBuren R, et al (2020) Transcriptome-Based Prediction of Complex Traits in Maize. *Plant Cell* 32:139–151. <https://doi.org/10.1105/tpc.19.00332>
- Babar MA, Reynolds MP, Ginkel M van, et al (2006) Spectral Reflectance to Estimate Genetic Variation for In-Season Biomass, Leaf Chlorophyll, and Canopy Temperature in Wheat. *Crop Sci* 46:1046–1057. <https://doi.org/10.2135/cropsci2005.0211>
- Balfourier F, Roussel V, Strelchenko P, et al (2007) A worldwide bread wheat core collection arrayed in a 384-well plate. *Theor Appl Genet* 114:1265–1275. <https://doi.org/10.1007/s00122-007-0517-1>
- Barnes RJ, Dhanoa MS, Lister SJ (1989) Standard Normal Variate Transformation and De-Trending of Near-Infrared Diffuse Reflectance Spectra. *Appl Spectrosc* 43:772–777. <https://doi.org/10.1366/0003702894202201>
- Bassi FM, Bentley AR, Charmet G, et al (2016) Breeding schemes for the implementation of genomic selection in wheat (*Triticum* spp .). *Plant Sci* 242:23–36. <https://doi.org/10.1016/j.plantsci.2015.08.021>
- Belderok B, Mesdag J, Mesdag H, Donner DA (2000) Bread-making quality of wheat: a century of breeding in Europe, Springer Science&Business Media.
- Ben-Sadoun S, Fugerey-Scarbel A, Auzanneau J, et al (2021) Integration of Genomic Selection into Winter-Type Bread Wheat Breeding Schemes: A Simulation Pipeline including Economic Constraints. *Crop Breed Genet Genomics*. <https://doi.org/10.20900/cbagg20210008>
- Ben-Sadoun S, Rincet R, Auzanneau J, et al (2020) Economical optimization of a breeding scheme by selective phenotyping of the calibration set in a multi-trait context: application to bread making quality. *Theor Appl Genet* 133:2197–2212. <https://doi.org/10.1007/s00122-020-03590-4>
- Bernardo R (1994) Prediction of Maize Single-Cross Performance Using RFLPs and Information from Related Hybrids. *Crop Sci* 34:20–25. <https://doi.org/10.2135/cropsci1994.0011183X003400010003x>
- Bernardo R (2008) Molecular Markers and Selection for Complex Traits in Plants: Learning from the Last 20 Years. *Crop Sci* 48:1649–1664. <https://doi.org/10.2135/cropsci2008.03.0131>
- Bernardo R (2014) Genomewide Selection of Parental Inbreds: Classes of Loci and Virtual Biparental Populations. *Crop Sci* 54:2586–2595. <https://doi.org/10.2135/cropsci2014.01.0088>
- Bernardo R, Charcosset A (2006) Usefulness of Gene Information in Marker-Assisted Recurrent Selection: A Simulation Appraisal. *Crop Sci* 46:614–621. <https://doi.org/10.2135/cropsci2005.05-0088>
- Bernardo R, Yu J (2007) Prospects for Genomewide Selection for Quantitative Traits in Maize. *Crop Sci* 47:1082–1090. <https://doi.org/10.2135/cropsci2006.11.0690>
- Blanco M, Villarroja I (2002) NIR spectroscopy: a rapid-response analytical tool. *TrAC Trends Anal Chem* 21:240–250. [https://doi.org/10.1016/S0165-9936\(02\)00404-1](https://doi.org/10.1016/S0165-9936(02)00404-1)
- Bosquet LC, Bricet N, Fournier C, et al (2015) PHENOARCH, a multiscale phenotyping platform for plant architecture, growth rate, water use efficiency and radiation use efficiency. *Recent Prog Drought Toler Genet Model* 2
- Brancourt-Hulmel M, Lecomte C, Meynard J-M (1999) A Diagnosis of Yield-Limiting Factors on Probe Genotypes for Characterizing Environments in Winter Wheat Trials. *Crop Sci* 39:1798–1808. <https://doi.org/10.2135/cropsci1999.3961798x>
- Branlard G (2012) La qualité du gluten : variations de sa composition et de ses propriétés. *Médecine Nutr* 48:19–23. <https://doi.org/10.1051/mnut/201248401>
- Brault C, Lazerges J, Doligez A, et al (2021) Interest of phenomic prediction as an alternative to genomic prediction in grapevine. *Genetics*

- Brisson N, Gate P, Gouache D, et al (2010) Why are wheat yields stagnating in Europe? A comprehensive data analysis for France. *Field Crops Res* 119:201–212. <https://doi.org/10.1016/j.fcr.2010.07.012>
- Burgueño J, de los Campos G, Weigel K, Crossa J (2012) Genomic Prediction of Breeding Values when Modeling Genotype × Environment Interaction using Pedigree and Dense Molecular Markers. *Crop Sci* 52:707–719. <https://doi.org/10.2135/cropsci2011.06.0299>
- Charmet G, Robert N, Perretant MR, et al (1999) Marker-assisted recurrent selection for cumulating additive and interactive QTLs in recombinant inbred lines: *Theor Appl Genet* 99:1143–1148. <https://doi.org/10.1007/s001220051318>
- Charmet G, Storlie E, Oury FX, et al (2014) Genome-wide prediction of three important traits in bread wheat. *Mol Breed* 34:1843–1852. <https://doi.org/10.1007/s11032-014-0143-y>
- Charmet G, Tran L-G, Auzanneau J, et al (2020) BWGS: A R package for genomic selection and its application to a wheat breeding programme. *PLOS ONE* 15:e0222733. <https://doi.org/10.1371/journal.pone.0222733>
- Cheverud JM, Routman EJ (1995) Epistasis and its contribution to genetic variance components. *Genetics* 139:1455–1461. <https://doi.org/10.1093/genetics/139.3.1455>
- Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica* 142:169–196. <https://doi.org/10.1007/s10681-005-1681-5>
- Comstock RE, Moll RH (1963) Genotype x Environment Interactions. *Stat Genet Plant Breed*
- Covarrubias-Pazarán G (2016) Genome-Assisted Prediction of Quantitative Traits Using the R Package sommer. *PLOS ONE* 11:e0156744. <https://doi.org/10.1371/journal.pone.0156744>
- Crain J, Mondal S, Rutkoski J, et al (2018) Combining High-Throughput Phenotyping and Genomic Information to Increase Prediction and Selection Accuracy in Wheat Breeding. *Plant Genome* 11:170043. <https://doi.org/10.3835/plantgenome2017.05.0043>
- Crossa J, Campos G de los, Pérez P, et al (2010) Prediction of Genetic Values of Quantitative Traits in Plant Breeding Using Pedigree and Molecular Markers. *Genetics* 186:713–724. <https://doi.org/10.1534/genetics.110.118521>
- Crossa J, Jarquín D, Franco J, et al (2016) Genomic Prediction of Gene Bank Wheat Landraces. *G3 GenesGenomesGenetics* 6:1819–1834. <https://doi.org/10.1534/g3.116.029637>
- Crossa J, Pérez P, Hickey J, et al (2014) Genomic prediction in CIMMYT maize and wheat breeding programs. *Heredity* 112:48–60. <https://doi.org/10.1038/hdy.2013.16>
- Crossa J, Pérez-Rodríguez P, Cuevas J, et al (2017) Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. *Trends Plant Sci* 22:961–975. <https://doi.org/10.1016/j.tplants.2017.08.011>
- Cuevas J, Crossa J, Soberanis V, et al (2016) Genomic Prediction of Genotype × Environment Interaction Kernel Regression Models. *Plant Genome* 9:plantgenome2016.03.0024. <https://doi.org/10.3835/plantgenome2016.03.0024>
- Cuevas J, Montesinos-López O, Juliana P, et al (2019) Deep Kernel for Genomic and Near Infrared Predictions in Multi-environment Breeding Trials. *G3amp58 GenesGenomesGenetics* 9:2913–2924. <https://doi.org/10.1534/g3.119.400493>
- Daetwyler HD, Calus MPL, Pong-Wong R, et al (2013) Genomic Prediction in Animals and Plants: Simulation of Data, Validation, Reporting, and Benchmarking. *Genetics* 193:347–365. <https://doi.org/10.1534/genetics.112.147983>
- Daetwyler HD, Pong-Wong R, Villanueva B, Woolliams JA (2010) The Impact of Genetic Architecture on Genome-Wide Evaluation Methods. *Genetics* 185:1021–1031. <https://doi.org/10.1534/genetics.110.116855>
- De Los Campos G, Gianola D, Rosa GJM, et al (2010) Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert spaces methods. *Genet Res* 92:295–308. <https://doi.org/10.1017/S0016672310000285>
- de los Campos G, Hickey JM, Pong-Wong R, et al (2013) Whole-Genome Regression and Prediction Methods Applied to Plant and Animal Breeding. *Genetics* 193:327–345. <https://doi.org/10.1534/genetics.112.143313>
- de los Campos G, Naya H, Gianola D, et al (2009) Predicting Quantitative Traits With Regression Models for Dense Molecular Markers and Pedigree. *Genetics* 182:375–385. <https://doi.org/10.1534/genetics.109.101501>
- Desta ZA, Ortiz R (2014) Genomic selection: genome-wide prediction in plant improvement. *Trends Plant Sci* 19:592–601. <https://doi.org/10.1016/j.tplants.2014.05.006>
- Doré C, Varoquaux F (2006) Histoire et amélioration de cinquante plantes cultivées, Quae
- Dubcovsky J, Dvorak J (2007) Genome Plasticity a Key Factor in the Success of Polyploid Wheat Under Domestication. *Science* 316:1862–1866. <https://doi.org/10.1126/science.1143986>
- Endelman JB (2011) Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. *Plant*

Références bibliographiques

- Genome 4:250–255. <https://doi.org/10.3835/plantgenome2011.08.0024>
- Endelman JB, Jannink J-L (2012) Shrinkage Estimation of the Realized Relationship Matrix. *G3 GenesGenomesGenetics* 2:1405–1413. <https://doi.org/10.1534/g3.112.004259>
- Erbe M, Hayes BJ, Matukumalli LK, et al (2012) Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. *J Dairy Sci* 95:4114–4129. <https://doi.org/10.3168/jds.2011-5019>
- Falconer DS (1996) Introduction to quantitative genetics, Pearson Education India
- FAO (2022) FAOSTAT
- FAO (2016) La situation mondiale de l'alimentation et de l'agriculture : changement climatique, agriculture et sécurité alimentaire
- Feher K, Lisek J, Römisch-Margl L, et al (2014) Deducing Hybrid Performance from Parental Metabolic Profiles of Young Primary Roots of Maize by Using a Multivariate Diallel Approach. *PLOS ONE* 9:e85435. <https://doi.org/10.1371/journal.pone.0085435>
- Fernandez O, Urrutia M, Bernillon S, et al (2016) Fortune telling: metabolic markers of plant performance. *Metabolomics* 12:158. <https://doi.org/10.1007/s11306-016-1099-1>
- Ferragina A, de los Campos G, Vazquez AI, et al (2015) Bayesian regression models outperform partial least squares methods for predicting milk components and technological properties using infrared spectral data. *J Dairy Sci* 98:8133–8151. <https://doi.org/10.3168/jds.2014-9143>
- Ferrio JP, Bertran E, Nachit MM, et al (2004) Estimation of grain yield by near-infrared reflectance spectroscopy in durum wheat. *Euphytica* 137:373–380. <https://doi.org/10.1023/B:EUPH.0000040523.52707.1e>
- Forni S, Aguilar I, Misztal I (2011) Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. *Genet Sel Evol* 43:1. <https://doi.org/10.1186/1297-9686-43-1>
- FranceAgriMer (2021) Fiche filière - blé tendre
- Frankel OH (1984) Genetic perspectives of germplasm conservation. *Genet Manip Impact Man Soc* 161:170
- Friedman J, Hastie T, Tibshirani R (2010) Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw* 33:. <https://doi.org/10.18637/jss.v033.i01>
- Frisch M, Bohn M, Melchinger AA (1999) Minimum Sample Size and Optimal Positioning of Flanking Markers in Marker-Assisted Backcrossing for Transfer of a Target Gene. *Crop Sci* 39:967–975. <https://doi.org/10.2135/cropsci1999.0011183X003900040003x>
- Frisch M, Thiemann A, Fu J, et al (2010) Transcriptome-based distance measures for grouping of germplasm and prediction of hybrid performance in maize. *Theor Appl Genet* 120:441–450. <https://doi.org/10.1007/s00122-009-1204-1>
- Fu J, Falke KC, Thiemann A, et al (2012) Partial least squares regression, support vector machine regression, and transcriptome-based distances for prediction of maize hybrid performance with gene expression data. *Theor Appl Genet* 124:825–833. <https://doi.org/10.1007/s00122-011-1747-9>
- Galán RJ, Bernal-Vasquez A-M, Jebsen C, et al (2020) Integration of genotypic, hyperspectral, and phenotypic data to improve biomass yield prediction in hybrid rye. *Theor Appl Genet* 133:3001–3015. <https://doi.org/10.1007/s00122-020-03651-8>
- Galán RJ, Bernal-Vasquez A-M, Jebsen C, et al (2021) Early prediction of biomass in hybrid rye based on hyperspectral data surpasses genomic predictability in less-related breeding material. *Theor Appl Genet* 134:1409–1422. <https://doi.org/10.1007/s00122-021-03779-1>
- Gallais A (2015) Comprendre l'amélioration des plantes: Enjeux, méthodes, objectifs et critères de sélection, Quae
- Gallais A (2011) Méthodes de création de variétés en amélioration des plantes, Quae
- Gärtner T, Steinfath M, Andorf S, et al (2009) Improved Heterosis Prediction by Combining Information on DNA- and Metabolic Markers. *PLoS ONE* 4:e5220. <https://doi.org/10.1371/journal.pone.0005220>
- Gemmer MR, Richter C, Jiang Y, et al (2020) Can metabolic prediction be an alternative to genomic prediction in barley? *PLOS ONE* 15:e0234052. <https://doi.org/10.1371/journal.pone.0234052>
- GEVES (2019) Evaluation de la Valeur Agronomique Technologique et Environnementale (VATE) des nouvelles variétés à l'inscription au Catalogue Français - blé tendre
- GEVES (2022) Le CTPS
- Gianola D, Fernando RL, Stella A (2006) Genomic-Assisted Prediction of Genetic Value With Semiparametric Procedures. *Genetics* 173:1761–1776. <https://doi.org/10.1534/genetics.105.049510>
- Gill AA, Starr C, Smith DB (1979) Lysine and nitrogen measurement by infra-red reflectance analysis as an aid to barley breeding. *J Agric Sci* 93:727–733. <https://doi.org/10.1017/S0021859600039149>
- Gorjanc G, Gaynor RC, Hickey JM (2018) Optimal cross selection for long-term genetic gain in two-part programs with rapid recurrent genomic selection. *Theor Appl Genet* 131:1953–1966. <https://doi.org/10.1007/s00122-018-3125-3>

- Gorjanc G, Jenko J, Hearne SJ, Hickey JM (2016) Initiating maize pre-breeding programs using genomic selection to harness polygenic variation from landrace populations. *BMC Genomics* 17:30. <https://doi.org/10.1186/s12864-015-2345-z>
- Griffiths S, Sharp R, Foote TN, et al (2006) Molecular characterization of Ph1 as a major chromosome pairing locus in polyploid wheat. *Nature* 439:749–752. <https://doi.org/10.1038/nature04434>
- Griffiths S, Simmonds J, Leverington M, et al (2009) Meta-QTL analysis of the genetic control of ear emergence in elite European winter wheat germplasm. *Theor Appl Genet* 119:383–395. <https://doi.org/10.1007/s00122-009-1046-x>
- Guo Z, Magwire MM, Basten CJ, et al (2016) Evaluation of the utility of gene expression and metabolic information for genomic prediction in maize. *Theor Appl Genet* 129:2413–2427. <https://doi.org/10.1007/s00122-016-2780-5>
- Habier D, Fernando RL, Dekkers JCM (2007) The Impact of Genetic Relationship Information on Genome-Assisted Breeding Values. *Genetics* 177:2389–2397. <https://doi.org/10.1534/genetics.107.081190>
- Habier D, Fernando RL, Garrick DJ (2013) Genomic BLUP Decoded: A Look into the Black Box of Genomic Prediction. *Genetics* 194:597–607. <https://doi.org/10.1534/genetics.113.152207>
- Hanocq E, Laperche A, Jaminon O, et al (2007) Most significant genome regions involved in the control of earliness traits in bread wheat, as revealed by QTL meta-analysis. *Theor Appl Genet* 114:569–584. <https://doi.org/10.1007/s00122-006-0459-z>
- Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME (2009a) Invited review: Genomic selection in dairy cattle: Progress and challenges. *J Dairy Sci* 92:433–443. <https://doi.org/10.3168/jds.2008-1646>
- Hayes BJ, Visscher PM, Goddard ME (2009b) Increased accuracy of artificial selection by using the realized relationship matrix. *Genet Res* 91:47–60. <https://doi.org/10.1017/S0016672308009981>
- Hayes BJ, Panozzo J, Walker CK, et al (2017) Accelerating wheat breeding for end-use quality with multi-trait genomic predictions incorporating near infrared and nuclear magnetic resonance-derived phenotypes. *Theor Appl Genet* 130:2505–2519. <https://doi.org/10.1007/s00122-017-2972-7>
- Heffner EL, Jannink J-L, Iwata H, et al (2011) Genomic Selection Accuracy for Grain Quality Traits in Biparental Wheat Populations. *Crop Sci* 51:2597–2606. <https://doi.org/10.2135/cropsci2011.05.0253>
- Heffner EL, Lorenz AJ, Jannink J, Sorrells ME (2010) Plant Breeding with Genomic Selection: Gain per Unit Time and Cost. *Crop Sci* 50:1681–1690. <https://doi.org/10.2135/cropsci2009.11.0662>
- Heffner EL, Sorrells ME, Jannink J-L (2009) Genomic Selection for Crop Improvement. *Crop Sci* 49:1. <https://doi.org/10.2135/cropsci2008.08.0512>
- Hernandez J, Lobos G, Matus I, et al (2015) Using Ridge Regression Models to Estimate Grain Yield from Field Spectral Data in Bread Wheat (*Triticum Aestivum* L.) Grown under Three Water Regimes. *Remote Sens* 7:2109–2126. <https://doi.org/10.3390/rs70202109>
- Heslot N, Akdemir D, Sorrells ME, Jannink J-L (2014) Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. *Theor Appl Genet* 127:463–480. <https://doi.org/10.1007/s00122-013-2231-5>
- Heslot N, Jannink J-L, Sorrells ME (2013) Using Genomic Prediction to Characterize Environments and Optimize Prediction Accuracy in Applied Breeding Data. *Crop Sci* 53:921–933. <https://doi.org/10.2135/cropsci2012.07.0420>
- Heslot N, Yang H-P, Sorrells ME, Jannink J-L (2012) Genomic Selection in Plant Breeding: A Comparison of Models. *Crop Sci* 52:146. <https://doi.org/10.2135/cropsci2011.06.0297>
- Hickey JM, Dreisigacker S, Crossa J, et al (2014) Evaluation of Genomic Selection Training Population Designs and Genotyping Strategies in Plant Breeding Programs Using Simulation. *Crop Sci* 54:1476–1488. <https://doi.org/10.2135/cropsci2013.03.0195>
- Hill WG, Goddard ME, Visscher PM (2008) Data and Theory Point to Mainly Additive Genetic Variance for Complex Traits. *PLoS Genet* 4:e1000008. <https://doi.org/10.1371/journal.pgen.1000008>
- Holroyd SE (2013) The Use of near Infrared Spectroscopy on Milk and Milk Products. *J Infrared Spectrosc* 21:311–322. <https://doi.org/10.1255/jnirs.1055>
- Honigs DE, Hieftje GM, Mark HL, Hirschfeld TB (1985) Unique-sample selection via near-infrared spectral subtraction. *Anal Chem* 57:2299–2303. <https://doi.org/10.1021/ac00289a029>
- Hospital F, Charcosset A (1997) Marker-Assisted Introgression of Quantitative Trait Loci. *Genetics* 147:1469–1485. <https://doi.org/10.1093/genetics/147.3.1469>
- Houle D, Govindaraju DR, Omholt S (2010) Phenomics: the next challenge. *Nat Rev Genet* 11:855–866. <https://doi.org/10.1038/nrg2897>
- Isidro J, Jannink J-L, Akdemir D, et al (2015) Training set optimization under population structure in genomic selection. *Theor Appl Genet* 128:145–158. <https://doi.org/10.1007/s00122-014-2418-4>
- Jannink J-L (2010) Dynamics of long-term genomic selection. *Genet Sel Evol* 42:35. <https://doi.org/10.1186/1297-9686-42-35>
- Jannink J-L, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. *Brief*

Références bibliographiques

- Funct Genomics 9:166–177. <https://doi.org/10.1093/bfpg/elq001>
- Jarquín D, Crossa J, Lacaze X, et al (2014) A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theor Appl Genet* 127:595–607. <https://doi.org/10.1007/s00122-013-2243-1>
- Jarquín D, Howard R, Crossa J, et al (2020) Genomic Prediction Enhanced Sparse Testing for Multi-environment Trials. *G3amp58 GenesGenomesGenetics* 10:2725–2739. <https://doi.org/10.1534/g3.120.401349>
- Jarquín D, Lemes da Silva C, Gaynor RC, et al (2017) Increasing Genomic-Enabled Prediction Accuracy by Modeling Genotype × Environment Interactions in Kansas Wheat. *Plant Genome* 10:plantgenome2016.12.0130. <https://doi.org/10.3835/plantgenome2016.12.0130>
- Jiang Y, Reif JC (2015) Modeling Epistasis in Genomic Selection. *Genetics* 201:759–768. <https://doi.org/10.1534/genetics.115.177907>
- Juliana P, Singh RP, Singh PK, et al (2017) Comparison of Models and Whole-Genome Profiling Approaches for Genomic-Enabled Prediction of Septoria Tritici Blotch, Stagonospora Nodorum Blotch, and Tan Spot Resistance in Wheat. *Plant Genome* 10:. <https://doi.org/10.3835/plantgenome2016.08.0082>
- Kang HM, Sul JH, Service SK, et al (2010) Variance component model to account for sample structure in genome-wide association studies. *Nat Genet* 42:348–354. <https://doi.org/10.1038/ng.548>
- Kennard RW, Stone LA (1969) Computer Aided Design of Experiments. *Technometrics* 11:137. <https://doi.org/10.2307/1266770>
- Kloppers FJ, Pretorius ZA (1997) Effects of combinations amongst genes *Lr13*, *Lr34* and *Lr37* on components of resistance in wheat to leaf rust. *Plant Pathol* 46:737–750. <https://doi.org/10.1046/j.1365-3059.1997.d01-58.x>
- Krause MR, González-Pérez L, Crossa J, et al (2019) Hyperspectral Reflectance-Derived Relationship Matrices for Genomic Prediction of Grain Yield in Wheat. *G3amp58 GenesGenomesGenetics* g3.200856.2018. <https://doi.org/10.1534/g3.118.200856>
- Labus MP, Nielsen GA, Lawrence RL, et al (2002) Wheat yield estimates using multi-temporal NDVI satellite imagery. *Int J Remote Sens* 23:4169–4180. <https://doi.org/10.1080/01431160110107653>
- Lado B, Barrios PG, Quincke M, et al (2016) Modeling Genotype × Environment Interaction for Genomic Selection with Unbalanced Data from a Wheat Breeding Program. *Crop Sci* 56:2165–2179. <https://doi.org/10.2135/cropsci2015.04.0207>
- Lado B, Battenfield S, Guzmán C, et al (2017) Strategies for Selecting Crosses Using Genomic Prediction in Two Wheat Breeding Programs. *Plant Genome* 10:. <https://doi.org/10.3835/plantgenome2016.12.0128>
- Lado B, Matus I, Rodríguez A, et al (2013) Increased Genomic Prediction Accuracy in Wheat Breeding Through Spatial Adjustment of Field Trial Data. *G3amp58 GenesGenomesGenetics* 3:2105–2114. <https://doi.org/10.1534/g3.113.007807>
- Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124:743–756. <https://doi.org/10.1093/genetics/124.3.743>
- Lane HM, Murray SC, Montesinos-López OA, et al (2020) Phenomic selection and prediction of maize grain yield from near-infrared reflectance spectroscopy of kernels. *Plant Phenome J* 3:e20002. <https://doi.org/10.1002/ppj2.20002>
- Legarra A, Aguilar I, Misztal I (2009) A relationship matrix including full pedigree and genomic information. *J Dairy Sci* 92:4656–4663. <https://doi.org/10.3168/jds.2009-2061>
- Lehermeier C, Teyssède S, Schön C-C (2017) Genetic Gain Increases by Applying the Usefulness Criterion with Improved Variance Prediction in Selection of Crosses. *Genetics* 207:1651–1661. <https://doi.org/10.1534/genetics.117.300403>
- Liu Z, Seefried FR, Reinhardt F, et al (2011) Impacts of both reference population size and inclusion of a residual polygenic effect on the accuracy of genomic prediction. *Genet Sel Evol* 43:19. <https://doi.org/10.1186/1297-9686-43-19>
- Longin CFH, Reif JC (2014) Redesigning the exploitation of wheat genetic resources. *Trends Plant Sci* 19:631–636. <https://doi.org/10.1016/j.tplants.2014.06.012>
- Lopez-Cruz M, Crossa J, Bonnett D, et al (2015) Increased Prediction Accuracy in Wheat Breeding Trials Using a Marker × Environment Interaction Genomic Selection Model. *G3 GenesGenomesGenetics* 5:569–582. <https://doi.org/10.1534/g3.114.016097>
- Lorenz AJ, Chao S, Asoro FG, et al (2011) Genomic Selection in Plant Breeding. In: *Advances in Agronomy*. Elsevier, pp 77–123
- Lorenzana RE, Bernardo R (2009) Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. *Theor Appl Genet* 120:151–161. <https://doi.org/10.1007/s00122-009-1166-3>
- Ly D, Chenu K, Gauffreteau A, et al (2017) Nitrogen nutrition index predicted by a crop model improves the genomic prediction of grain number for a bread wheat core collection. *Field Crops Res* 214:331–340.

- <https://doi.org/10.1016/j.fcr.2017.09.024>
- Ly D, Huet S, Gauffreteau A, et al (2018) Whole-genome prediction of reaction norms to environmental stress in bread wheat (*Triticum aestivum* L.) by genomic random regression. *Field Crops Res* 216:32–41. <https://doi.org/10.1016/j.fcr.2017.08.020>
- Mackay TFC, Stone EA, Ayroles JF (2009) The genetics of quantitative traits: challenges and prospects. *Nat Rev Genet* 10:565–577. <https://doi.org/10.1038/nrg2612>
- Malosetti M, Bustos-Korts D, Boer MP, van Eeuwijk FA (2016) Predicting Responses in Multiple Environments: Issues in Relation to Genotype × Environment Interactions. *Crop Sci* 56:2210. <https://doi.org/10.2135/cropsci2015.05.0311>
- Marcussen T, Sandve SR, Heier L, et al (2014) Ancient hybridizations among the ancestral genomes of bread wheat. *Science* 345:1250092. <https://doi.org/10.1126/science.1250092>
- Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics* 157:1819–1829. <https://doi.org/10.1093/genetics/157.4.1819>
- Michel S, Ametz C, Gungor H, et al (2017) Genomic assisted selection for enhancing line breeding: merging genomic and phenotypic selection in winter wheat breeding programs with preliminary yield trials. *Theor Appl Genet* 130:363–376. <https://doi.org/10.1007/s00122-016-2818-8>
- Michel S, Kummer C, Gallee M, et al (2018) Improving the baking quality of bread wheat by genomic selection in early generations. *Theor Appl Genet* 131:477–493. <https://doi.org/10.1007/s00122-017-2998-x>
- Michel S, Wagner C, Nosenko T, et al (2021) Merging Genomics and Transcriptomics for Predicting Fusarium Head Blight Resistance in Wheat. *Genes* 12:114. <https://doi.org/10.3390/genes12010114>
- Mohammadi M, Tiede T, Smith KP (2015) PopVar: A Genome-Wide Procedure for Predicting Genetic Variance and Correlated Response in Biparental Breeding Populations. *Crop Sci* 55:2068–2077. <https://doi.org/10.2135/cropsci2015.01.0030>
- Montesinos-López OA, Montesinos-López A, Crossa J, et al (2017a) Predicting grain yield using canopy hyperspectral reflectance in wheat breeding data. *Plant Methods* 13:4. <https://doi.org/10.1186/s13007-016-0154-2>
- Montesinos-López A, Montesinos-López OA, Cuevas J, et al (2017b) Genomic Bayesian functional regression models with interactions for predicting wheat grain yield using hyper-spectral image data. *Plant Methods* 13:62. <https://doi.org/10.1186/s13007-017-0212-4>
- Moreau L, Charcosset A, Gallais A (2004) Experimental evaluation of several cycles of marker-assisted selection in maize. *Euphytica* 137:111–118. <https://doi.org/10.1023/B:EUPH.0000040508.01402.21>
- Muñoz PR, Resende MFR, Gezan SA, et al (2014) Unraveling Additive from Nonadditive Effects Using Genomic Relationship Matrices. *Genetics* 198:1759–1768. <https://doi.org/10.1534/genetics.114.171322>
- Nicolai BM, Beullens K, Bobelyn E, et al (2007) Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. *Postharvest Biol Technol* 46:99–118. <https://doi.org/10.1016/j.postharvbio.2007.06.024>
- Norman A, Taylor J, Edwards J, Kuchel H (2018) Optimising Genomic Selection in Wheat: Effect of Marker Density, Population Size and Population Structure on Prediction Accuracy. *G3* 8:2889–2899. <https://doi.org/10.1534/g3.118.200311>
- Norman A, Taylor J, Tanaka E, et al (2017) Increased genomic prediction accuracy in wheat breeding using a large Australian panel. *Theor Appl Genet* 130:2543–2555. <https://doi.org/10.1007/s00122-017-2975-4>
- Osborne BG (2006) Applications of near Infrared Spectroscopy in Quality Screening of Early-Generation Material in Cereal Breeding Programmes. *J Infrared Spectrosc* 14:93–101. <https://doi.org/10.1255/jnirs.595>
- Oury F-X, Godin C, Mailliard A, et al (2012) A study of genetic progress due to selection reveals a negative effect of climate change on bread wheat yield in France. *Eur J Agron* 40:28–38. <https://doi.org/10.1016/j.eja.2012.02.007>
- Parmley K, Nagasubramanian K, Sarkar S, et al (2019) Development of Optimized Phenomic Predictors for Efficient Plant Breeding Decisions Using Phenomic-Assisted Selection in Soybean. *Plant Phenomics* 2019:1–15. <https://doi.org/10.34133/2019/5809404>
- Paux E, Lafarge S, Balfourier F, et al (2022) Breeding for Economically and Environmentally Sustainable Wheat Varieties: An Integrated Approach from Genomics to Selection. *Biology* 11:149. <https://doi.org/10.3390/biology11010149>
- Pérez P, de los Campos G (2014) Genome-Wide Regression and Prediction with the BGLR Statistical Package. *Genetics* 198:483–495. <https://doi.org/10.1534/genetics.114.164442>
- Pérez-Rodríguez P, Crossa J, Rutkoski J, et al (2017) Single-Step Genomic and Pedigree Genotype × Environment Interaction Models for Predicting Wheat Lines in International Environments. *Plant Genome* 10:plantgenome2016.09.0089. <https://doi.org/10.3835/plantgenome2016.09.0089>
- Pérez-Rodríguez P, Gianola D, González-Camacho JM, et al (2012) Comparison Between Linear and Non-parametric Regression Models for Genome-Enabled Prediction in Wheat. *G3* 2:2889–2899. <https://doi.org/10.1534/g3.112.200311>

Références bibliographiques

- GenesGenomesGenetics 2:1595–1605. <https://doi.org/10.1534/g3.112.003665>
- Piepho HP, Möhring J (2005) Best Linear Unbiased Prediction of Cultivar Effects for Subdivided Target Regions. *Crop Sci* 45:1151–1159. <https://doi.org/10.2135/cropsci2004.0398>
- Poland J, Endelman J, Dawson J, et al (2012) Genomic Selection in Wheat Breeding using Genotyping-by-Sequencing. *Plant Genome* 5:plantgenome2012.06.0006. <https://doi.org/10.3835/plantgenome2012.06.0006>
- Poland JA, Rife TW (2012) Genotyping-by-Sequencing for Plant Breeding and Genetics. *Plant Genome* 5:plantgenome2012.05.0005. <https://doi.org/10.3835/plantgenome2012.05.0005>
- Posada H, Ferrand M, Davrieux F, et al (2009) Stability across environments of the coffee variety near infrared spectral signature. *Heredity* 102:113–119. <https://doi.org/10.1038/hdy.2008.88>
- Pretty J (2008) Agricultural sustainability: concepts, principles and evidence. *Philos Trans R Soc B Biol Sci* 363:447–465. <https://doi.org/10.1098/rstb.2007.2163>
- Pszczola M, Strabel T, Mulder HA, Calus MPL (2012) Reliability of direct genomic values for animals with different relationships within and to the reference population. *J Dairy Sci* 95:389–400. <https://doi.org/10.3168/jds.2011-4338>
- R2D2 Consortium, Fugeray-Scarbel A, Bastien C, et al (2021) Why and How to Switch to Genomic Selection: Lessons From Plant and Animal Breeding Experience. *Front Genet* 0: <https://doi.org/10.3389/fgene.2021.629737>
- Rabier C-E, Barre P, Asp T, et al (2016) On the Accuracy of Genomic Selection. *PLOS ONE* 11:e0156086. <https://doi.org/10.1371/journal.pone.0156086>
- Ray DK, Mueller ND, West PC, Foley JA (2013) Yield Trends Are Insufficient to Double Global Crop Production by 2050. *PLoS ONE* 8:e66428. <https://doi.org/10.1371/journal.pone.0066428>
- Reynolds M, Bonnett D, Chapman SC, et al (2011) Raising yield potential of wheat. I. Overview of a consortium approach and breeding strategies. *J Exp Bot* 62:439–452. <https://doi.org/10.1093/jxb/erq311>
- Riedelshemer C, Czedik-Eysenberg A, Grieder C, et al (2012) Genomic and metabolic prediction of complex heterotic traits in hybrid maize. *Nat Genet* 44:217–220. <https://doi.org/10.1038/ng.1033>
- Rimbert H, Darrier B, Navarro J, et al (2018) High throughput SNP discovery and genotyping in hexaploid wheat. *PLOS ONE* 13:e0186329. <https://doi.org/10.1371/journal.pone.0186329>
- Rincent R, Charpentier J-P, Faivre-Rampant P, et al (2018) Phenomic Selection Is a Low-Cost and High-Throughput Method Based on Indirect Predictions: Proof of Concept on Wheat and Poplar. *G3* 8:200760. <https://doi.org/10.1534/g3.118.200760>
- Rincent R, Laloë D, Nicolas S, et al (2012) Maximizing the Reliability of Genomic Selection by Optimizing the Calibration Set of Reference Individuals: Comparison of Methods in Two Diverse Groups of Maize Inbreds (*Zea mays* L.). *Genetics* 192:715–728. <https://doi.org/10.1534/genetics.112.141473>
- Rincent R, Malosetti M, Ababaei B, et al (2019) Using crop growth model stress covariates and AMMI decomposition to better predict genotype-by-environment interactions. *Theor Appl Genet* 132:3399–3411. <https://doi.org/10.1007/s00122-019-03432-y>
- Robert P, Auzanneau J, Goudemand E, et al (2022a) Phenomic selection in wheat breeding: identification and optimisation of factors influencing prediction accuracy and comparison to genomic selection. *Theor Appl Genet*. <https://doi.org/10.1007/s00122-021-04005-8>
- Robert P, Brault C, Rincent R, Segura V (2022b) Phenomic selection: a new and efficient alternative to genomic selection. (Accepted in *Methods in Molecular Biology* edited by Springer, Soon Published)
- Robert P, Le Gouis J, The BreedWheat Consortium, Rincent R (2020) Combining Crop Growth Modeling With Trait-Assisted Prediction Improved the Prediction of Genotype by Environment Interactions. *Front Plant Sci* 11:827. <https://doi.org/10.3389/fpls.2020.00827>
- Rodríguez-Álvarez MX, Boer MP, van Eeuwijk FA, Eilers PHC (2018) Correcting for spatial heterogeneity in plant breeding experiments with P-splines. *Spat Stat* 23:52–71. <https://doi.org/10.1016/j.spasta.2017.10.003>
- Roger J-M, Chauchard F, Bellon-Maurel V (2003) EPO–PLS external parameter orthogonalisation of PLS application to temperature-independent measurement of sugar content of intact fruits. *Chemom Intell Lab Syst* 66:191–204. [https://doi.org/10.1016/S0169-7439\(03\)00051-0](https://doi.org/10.1016/S0169-7439(03)00051-0)
- Rutkoski J, Poland J, Mondal S, et al (2016) Canopy Temperature and Vegetation Indices from High-Throughput Phenotyping Improve Accuracy of Pedigree and Genomic Selection for Grain Yield in Wheat. *G3* 6:2799–2808. <https://doi.org/10.1534/g3.116.032888>
- Rutkoski J, Singh RP, Huerta-Espino J, et al (2015) Efficient Use of Historical Data for Genomic Selection: A Case Study of Stem Rust Resistance in Wheat. *Plant Genome* 8: <https://doi.org/10.3835/plantgenome2014.09.0046>
- Ryckewaert M, Gorretta N, Henriot F, et al (2020) Reduction of repeatability error for analysis of variance-Simultaneous Component Analysis (REP-ASCA): Application to NIR spectroscopy on coffee sample.

- Anal Chim Acta 1101:23–31. <https://doi.org/10.1016/j.aca.2019.12.024>
- Sackton TB, Hartl DL (2016) Genotypic Context and Epistasis in Individuals and Populations. *Cell* 166:279–287. <https://doi.org/10.1016/j.cell.2016.06.047>
- Salamini F, Özkan H, Brandolini A, et al (2002) Genetics and geography of wild cereal domestication in the near east. *Nat Rev Genet* 3:429–441. <https://doi.org/10.1038/nrg817>
- Savitzky Abraham, Golay MJE (1964) Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal Chem* 36:1627–1639. <https://doi.org/10.1021/ac60214a047>
- Schrag TA, Westhues M, Schipprack W, et al (2018) Beyond Genomic Prediction: Combining Different Types of *omics* Data Can Improve Prediction of Hybrid Performance in Maize. *Genetics* 208:1373–1385. <https://doi.org/10.1534/genetics.117.300374>
- Seifert F, Thiemann A, Schrag TA, et al (2018) Small RNA-based prediction of hybrid performance in maize. *BMC Genomics* 19:371. <https://doi.org/10.1186/s12864-018-4708-8>
- Semenov MA, Shewry PR (2011) Modelling predicts that heat stress, not drought, will increase vulnerability of wheat in Europe. *Sci Rep* 1:66. <https://doi.org/10.1038/srep00066>
- Sharma S, Upadhyaya HD, Varshney RK, Gowda CLL (2013) Pre-breeding for diversification of primary gene pool and genetic enhancement of grain legumes. *Front Plant Sci* 4:. <https://doi.org/10.3389/fpls.2013.00309>
- Shewry PR (2009) Wheat. *J Exp Bot* 60:1537–1553. <https://doi.org/10.1093/jxb/erp058>
- Shewry PR, Halford NG, Lafiandra D (2003) Genetics of Wheat Gluten Proteins. In: *Advances in Genetics*. Elsevier, pp 111–184
- Shewry PR, Hey SJ (2015) The contribution of wheat to human diet and health. *Food Energy Secur* 4:178–202. <https://doi.org/10.1002/fes3.64>
- Solberg TR, Sonesson AK, Woolliams JA, Meuwissen THE (2008) Genomic selection using different marker types and densities. *J Anim Sci* 86:2447–2454. <https://doi.org/10.2527/jas.2007-0010>
- Song J, Carver BF, Powers C, et al (2017) Practical application of genomic selection in a doubled-haploid winter wheat breeding program. *Mol Breed* 37:117. <https://doi.org/10.1007/s11032-017-0715-8>
- Sun J, Rutkoski JE, Poland JA, et al (2017) Multitrait, Random Regression, or Simple Repeatability Model in High-Throughput Phenotyping Data Improve Genomic Prediction for Wheat Grain Yield. *Plant Genome* 10:0. <https://doi.org/10.3835/plantgenome2016.11.0111>
- Sweeney DW, Sun J, Taagen E, Sorrells ME (2019) Genomic Selection in Wheat. In: *Applications of Genetic and Genomic Research in Cereals*. Elsevier, pp 273–302
- Tadesse W, Sanchez-Garcia M, Assefa SG, et al (2019) Genetic Gains in Wheat Breeding and Its Role in Feeding the World. *Crop Breed Genet Genomics*. <https://doi.org/10.20900/cbgg20190005>
- Tardieu F, Cabrera-Bosquet L, Pridmore T, Bennett M (2017) Plant Phenomics, From Sensors to Knowledge. *Curr Biol* 27:R770–R783. <https://doi.org/10.1016/j.cub.2017.05.055>
- The International Wheat Genome Sequencing Consortium (IWGSC), Appels R, Eversole K, et al (2018) Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361:eaar7191. <https://doi.org/10.1126/science.aar7191>
- Touzy G, Rincant R, Bogard M, et al (2019) Using environmental clustering to identify specific drought tolerance QTLs in bread wheat (*T. aestivum* L.). *Theor Appl Genet* 132:2859–2880. <https://doi.org/10.1007/s00122-019-03393-2>
- Tsuchikawa S, Kobori H (2015) A review of recent application of near infrared spectroscopy to wood science and technology. *J Wood Sci* 61:213–220. <https://doi.org/10.1007/s10086-015-1467-x>
- Tucker CJ (1979) Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens Environ* 8:127–150. [https://doi.org/10.1016/0034-4257\(79\)90013-0](https://doi.org/10.1016/0034-4257(79)90013-0)
- Velasco L, Möllers C, Becker HC (1999) Estimation of seed weight, oil content and fatty acid composition in intact single seeds of rapeseed (*Brassica napus* L.) by near-infrared reflectance spectroscopy. *Euphytica* 106:79–85. <https://doi.org/10.1023/A:1003592115110>
- Vergara-Diaz O, Vatter T, Kefauver SC, et al (2020) Assessing durum wheat ear and leaf metabolomes in the field through hyperspectral data. *Plant J* 102:615–630. <https://doi.org/10.1111/tpj.14636>
- Visscher PM, Haley CS, Thompson R (1996) Marker-Assisted Introgression in Backcross Breeding Programs. *Genetics* 144:1923–1932. <https://doi.org/10.1093/genetics/144.4.1923>
- Wang D, Salah El-Basyoni I, Stephen Baenziger P, et al (2012) Prediction of genetic values of quantitative traits with epistatic effects in plant breeding populations. *Heredity* 109:313–319. <https://doi.org/10.1038/hdy.2012.44>
- Wang S, Wei J, Li R, et al (2019) Identification of optimal prediction models using multi-omic data for selecting hybrid rice. *Heredity* 123:395–406. <https://doi.org/10.1038/s41437-019-0210-6>
- Ward J, Rakszegi M, Bedő Z, et al (2015) Differentially penalized regression to predict agronomic traits from metabolites and markers in wheat. *BMC Genet* 16:19. <https://doi.org/10.1186/s12863-015-0169-0>
- Watson A, Ghosh S, Williams MJ, et al (2018) Speed breeding is a powerful tool to accelerate crop research and

Références bibliographiques

- breeding. *Nat Plants* 4:23–29. <https://doi.org/10.1038/s41477-017-0083-8>
- Watson A, Hickey LT, Christopher J, et al (2019) Multivariate Genomic Selection and Potential of Rapid Indirect Selection with Speed Breeding in Spring Wheat. *Crop Sci* 59:1945–1959. <https://doi.org/10.2135/cropsci2018.12.0757>
- Westhues M, Schrag TA, Heuer C, et al (2017) Omics-based hybrid prediction in maize. *Theor Appl Genet* 130:1927–1939. <https://doi.org/10.1007/s00122-017-2934-0>
- Wetterstrand KA (2012) DNA sequencing costs: data from the NHGRI large-scale genome sequencing program.
- Whalen A, Gaynor C, Hickey JM (2020) Using high-throughput phenotypes to enable genomic selection by inferring genotypes. *Genetics*
- Whittaker JC, Thompson R, Denham MC (2000) Marker-assisted selection using ridge regression. *Genet Res* 75:249–252. <https://doi.org/10.1017/S0016672399004462>
- Xiaobo Z, Jiewen Z, Povey MJW, et al (2010) Variables selection methods in near-infrared spectroscopy. *Anal Chim Acta* 667:14–32. <https://doi.org/10.1016/j.aca.2010.03.048>
- Xu S (2007) An Empirical Bayes Method for Estimating Epistatic Effects of Quantitative Trait Loci. *Biometrics* 63:513–521. <https://doi.org/10.1111/j.1541-0420.2006.00711.x>
- Xu S, Xu Y, Gong L, Zhang Q (2016) Metabolomic prediction of yield in hybrid rice. *Plant J* 88:219–227. <https://doi.org/10.1111/tbj.13242>
- Xu Y, Crouch JH (2008) Marker-Assisted Selection in Plant Breeding: From Publications to Practice. *Crop Sci* 48:391–407. <https://doi.org/10.2135/cropsci2007.04.0191>
- Yu X, Li X, Guo T, et al (2016) Genomic prediction contributing to a promising global strategy to turbocharge gene banks. *Nat Plants* 2:. <https://doi.org/10.1038/nplants.2016.150>
- Zadoks JC, Chang TT, Konzak CF (1974) A decimal code for the growth stages of cereals. *Weed Res* 14:415–421. <https://doi.org/10.1111/j.1365-3180.1974.tb01084.x>
- Zenke-Philippi C, Frisch M, Thiemann A, et al (2017) Transcriptome-based prediction of hybrid performance with unbalanced data from a maize breeding programme. *Plant Breed* 136:331–337. <https://doi.org/10.1111/pbr.12482>
- Zgouz A, Hérán D, Barthès B, et al (2020) Dataset of visible-near infrared handheld and micro-spectrometers – comparison of the prediction accuracy of sugarcane properties. *Data Brief* 31:106013. <https://doi.org/10.1016/j.dib.2020.106013>
- Zhao Y, Li Z, Liu G, et al (2015) Genome-based establishment of a high-yielding heterotic pattern for hybrid wheat breeding. *Proc Natl Acad Sci* 112:15624–15629. <https://doi.org/10.1073/pnas.1514547112>
- Zhong S, Dekkers JCM, Fernando RL, Jannink J-L (2009) Factors Affecting Accuracy From Genomic Selection in Populations Derived From Multiple Inbred Lines: A Barley Case Study. *Genetics* 182:355–364. <https://doi.org/10.1534/genetics.108.098277>
- Zhong S, Jannink J-L (2007) Using Quantitative Trait Loci Results to Discriminate Among Crosses on the Basis of Their Progeny Mean and Variance. *Genetics* 177:567–576. <https://doi.org/10.1534/genetics.107.075358>
- Zhu X, Leiser WL, Hahn V, Würschum T (2021a) Phenomic selection is competitive with genomic selection for breeding of complex traits. *Plant Phenome J* 4:. <https://doi.org/10.1002/ppj2.20027>
- Zhu X, Maurer HP, Jenz M, et al (2021b) The performance of phenomic selection depends on the genetic architecture of the target trait. *Theor Appl Genet*. <https://doi.org/10.1007/s00122-021-03997-7>

Annexes

Annexe 1: données supplémentaires du chapitre 3

Liste des tableaux

Table S1. Summary of the datasets structure in term of common lines between environments

Table S2. Description of trial design and characteristic of the datasets

Table S3. Description of trial management for each dataset

Table S4. Description of the genetic, residual variance and broad-sense Heritability within each dataset

Table S5. Description of the four subsets constructed from the four initial Sets

Liste des figures

Figure S1. Representation of the broad-sense Heritability, correlation with phenotypic trait (Grain Yield and Heading Date) for each wavelength within each environment

Figure S2. Boxplot of predictive abilities of phenomic selection for grain yield (GY) and heading date (HD) for each panel-environment combination from four different wheat breeding datasets

Figure S3. Representation of the selection frequency of each wavelength in each Set-Environment by the PS-LASSO model for Grain Yield and Heading Date

Figure S4. Barplot of grain yield (GY) and heading date (HD) predictive ability on all Set-Environment combinations

Figure S5. Boxplot of predictive abilities as a function of the number of extra NIR spectra combined to spectra S for grain yield (GY) and heading date (HD) for all the panel-environment combinations

Table S1. Summary of the datasets structure in term of common lines between environments. For each breeder, the diagonal in bold is the number of lines evaluated in the corresponding trial, the upper triangle is the number of common lines in a pair of environments and the lower triangle is the number of common lines genotyped.

From FD company		2016			2017				2018		
		CP	HV	LC	CP	HV	LC	VB	CP	HV	LC
2016	CP	95	87	66	22	23	18	25	22	22	21
	HV	2	335	246	66	78	62	65	41	48	48
	LC	1	153	405	55	68	91	79	39	45	70
2017	CP	2	46	41	77	76	41	50	42	42	30
	HV	2	49	46	51	360	161	52	85	97	48
	LC	1	41	63	26	94	275	79	51	56	90
	VB	2	40	58	27	27	58	104	31	31	55
2018	CP	1	20	22	24	66	35	11	101	100	35
	HV	1	20	22	24	67	36	11	70	224	92
	LC	1	22	38	17	28	56	35	15	38	227

AO-INRAE program			2018						2019						
			LI				T		LI				T		
			CF	RE	EM	LM	GL	LM	CF	RE	EM	LM	GL	EM	LM
2018	LI	CF	328	3	305	0	5	0	0	1	30	26	2	26	26
		RE	0	131	129	0	0	0	0	2	28	25	1	24	25
		EM	24	29	1433	82	1	2	0	4	128	82	4	81	82
		LM	0	0	2	82	0	2	0	0	6	5	0	5	5
	T	GL	0	3	452	0	566	0	0	2	6	0	4	0	0
		LM	0	0	82	5	0	5	0	0	112	112	0	111	112
2019	LI	CF	0	0	0	0	0	0	264	0	259	0	0	0	0
		RE	0	0	0	0	0	0	0	113	112	0	0	0	0
		EM	26	25	82	0	0	0	23	19	1518	112	325	100	101
		LM	26	25	77	0	0	0	0	0	101	112	0	100	101
	T	GL	0	0	0	0	0	0	0	1	367	0	433	0	0
		EM	26	24	76	0	0	0	0	0	111	111	0	111	111
		LM	26	25	77	0	0	0	0	0	112	112	0	100	112

Table S2. Description of trial design and characteristic of the datasets defined in 2.1. **T**: Intensive practices. **NT**: Not treated **LI**: low input. **T** corresponds to an intensive management in order to evaluate the yield potential without any constraint except the climatic conditions. **NT** corresponds to a crop management without fungicide. **LI** corresponds to a crop management with lower input levels than the T management (lower sowing density, no fungicide and no growth regulator). **Lat.** and **Long.** stand for latitude and longitude in decimal degrees. **Sowing density** and **Microplot size** are expressed in plant per m² and m² respectively.

Set	Year	Site	Experimental design	Location	Code treatment	Lat.	Long.	Sowing Date	Harvest Date	Sowing density	Microplotsize
Set 1	2016	HV	Augmented design	Houville	T	48.44	1.64	19/10/2015	22/07/2016	250	6.7
		LC		Lectoure		43.9	0.62	21/10/2015	15/07/2016	330	7.5
	2017	HV		Houville		48.44	1.64	11/10/2016	16/07/2017	250	6.7
		LC		Lectoure		43.9	0.62	27/10/2016	28/06/2017	330	7.5
	2018	HV		Houville		48.44	1.64	13/10/2017	12/07/2018	250	6.7
		LC		Lectoure		43.9	0.62	24/10/2017	03/07/2018	330	7.5
Set 2	2018	CF	Randomized complete block	Clermont-Ferrand	LI	45.78	3.08	25/10/2017	11/07/2018	250	7.5
		GL		Genlis	T	47.27	5.15	19/10/2017	18/07/2018	300	5
		RE		Le Rheu	LI	48.1	-1.79	25/10/2017	30/07 – 02/08/2018	265	6.48
		EM	Augmented design	Estrées-Mons	LI/NT	49.86	3.04	11/10/2017	13-16/07/2018	250	6.85
	2019	CF	Randomized complete block	Clermont-Ferrand	LI	45.78	3.08	5/11/2018	11/07/2019	250	7.5
		GL		Genlis	T	47.27	5.15	17/10/2018	18/07/2019	300	5
		RE		Le Rheu	LI	48.1	-1.79	24 & 25/10/2018	24 – 29/07/2019	265	6.48
		EM	Augmented design	Estrées-Mons	LI/NT	49.86	3.04	10/10/2018	19-24/07/2019	250	7.27
Set 3	2016	CP	Randomized complete block	Cappelle	T	50.51	3.17	13/10/2015	01/08/2016	300	7.5
		HV		Houville		48.44	1.64	19/10/2015	22/07/2016	250	6.5
		LC		Lectoure		43.9	0.62	21/10/2015	15/07/2016	330	7.5
	2017	CP		Cappelle		50.51	3.17	13/10/2016	25/07/2017	300	7.5
		HV		Houville		48.44	1.64	11/10/2016	16/07/2017	250	6.5
		LC		Lectoure		43.9	0.62	27/10/2016	28/06/2017	330	7.5
	2018	VB		Villamblain		48.01	1.64	14/10/2016	07/07/2017	250	7.5
		CP		Cappelle		50.51	3.17	16/10/2017	24/07/2018	300	7.5
	HV	Houville	48.44	1.64	13/10/2017	12/07/2018	250	6.7			
Set 4	2018	LM	Randomized complete block	Le Moulon	T	48.7	2.14	27 & 31/10/2017	19 & 20/07/2018	220-240	7.5
		EM		Estrées-Mons	LI	49.86	3.04	11/10/2017	13-16/07/2018	250	7
		LM		Le Moulon		48.7	2.14	27 & 31/10/2017	12 & 17/07/2018	160	7.5
	2019	EM		Estrées-Mons	T	49.86	3.04	17/10/2018	24-26/07/2019	150	7.5
		LM		Le Moulon		48.7	2.14	24/10/2018	30/07/2019	220-240	7.5
		EM		Estrées-Mons	LI	49.86	3.04	10/10/2018	19-24/07/2019	250	7.5
		LM		Le Moulon		48.7	2.14	24/10/2018	26/07/2019	160	7.5

Table S3. Description of trial management for each dataset. Number of applications for each intervention are detailed as well as the total dose (kg/ha) of nitrogen fertilisation (N) applied in the trial.

Set	Year	Site	Location	Code treatment	Herbicide	Fungicide	Growth regulator	Insecticide	Fertilisation	
Set 1	2016	HV	Houville	T	x	2-3	1	x	2-3 N= 180-200 kg/ha	
		LC	Lectoure							
	2017	HV	Houville							
		LC	Lectoure							
	2018	HV	Houville							
		LC	Lectoure							
Set 2	2018	CF	Clermont-Ferrand	LI	1-2	x	x	x	2 N= 60 kg/ha	
		GL	Genlis	T	2	4	x	x	4 N= 240 kg/ha	
		RE	Le Rheu	LI	x	x	x	x	2 N= 100 kg/ha	
		EM	Estrées-Mons	LI/NT	2/2	0/0	0/1	0/0	1/2 N= 70/130 kg/ha	
	2019	CF	Clermont-Ferrand	LI	1-2	x	x	x	2-3 N= 110 kg/ha	
		GL	Genlis	T	2	3	x	1	3 N= 190 kg/ha	
		RE	Le Rheu	LI	x	x	x	x	2 N= 100 kg/ha	
		EM	Estrées-Mons	LI/NT	1/1	0/0	0/1	0/0	2/4 N= 100/160 kg/ha	
	Set 3	2016	CP	Cappelle	T	x	2-3	1	x	2-3 N= 180-200 kg/ha
			HV	Houville						
			LC	Lectoure						
		2017	CP	Cappelle						
HV			Houville							
LC			Lectoure							
2018		VB	Villamblain							
		CP	Cappelle							
2018	HV	Houville								
Set 4	2018	LM	Le Moulon	T	2	3	x	x	3 N= 190 kg/ha	
		EM	Estrées-Mons	Li	2	0	0	0	1 N= 70 kg/ha	
		LM	Le Moulon		2	x	x	x	2 N= 130 kg/ha	
	2019	EM	Estrées-Mons	T	1	2	1	0	4 N=160 kg/ha	
		LM	Le Moulon	LI	1	3	x	x	3 N= 160 kg/ha	
		EM	Estrées-Mons		1	0	0	0	2 N= 100 kg/ha	
		LM	Le Moulon		1	x	x	x	2 N= 110 kg/ha	

Table S4. Description of the genetic (σ_g^2) and residual (σ_e^2) variances obtained with Equation 1 (e.g. Statistical models for adjusting data and estimating heritability) for each environment of the four datasets, with the adjusted model associated. Broad-sense heritability (H^2) described in Equation 2 is also reported. (-) denotes missing value due to a lack of repetition. **T**: treated. **LI**: low input. **NT**: Not treated. Genotype number referred to the number of lines in each environment, used for the analysis. See Table S2 for the description of the sites.

Panel	Year	Site	Treatment	Number of lines	Trait									
					GY				HD					
					Model	σ_g^2	σ_e^2	H^2	Model	σ_g^2	σ_e^2	H^2		
Set1	2016	HV	T	234	M1				M1					
		LC		269										
	2017	HV	T	280										
		LC		184										
	2018	HV	T	125										
		LT		113										
Set2	2018	CF	LI	328	M2	49	9.2	0.91						
		GL	T	566	M2	27.1	7.9	0.87						
		RE	LI	131	M2	42.9	10.1	0.89						
		EM	LI / NT	1351	M1	-	-	-						
	2019	CF	LI	264	M2	38	12.6	0.86						
		GL	T	433	M2	40.1	16	0.83						
		RE	LI	113	M2	32.9	19.5	0.77						
		EM	LI / NT	1418	M1	-	-	-						
	Set3	2016	CP	T	95	M3	62.9	11.4	0.94	M3	8.2	0.79	0.97	
			HV		103		80.8	15.7	0.93		11	0.7	0.97	
			LC		141		151.5	25.2	0.94		23.59	1.3	0.98	
		2017	CP	T	77		25.9	7.1	0.91		7.8	0.46	0.98	
HV			82		24.7		5.8	0.93	17.9		0.3	0.99		
LC			94		23.7		17.3	0.79	15.1		1.3	0.97		
VB			104		23.3		5.1	0.93	8.1		1	0.96		
2018		CP	T	101	28.6		3.9	0.95	4.7		0.5	0.96		
		HV		102	45.2		9.0	0.93	11.1		0.49	0.98		
Set4		2018	ML	T	82		M2	42.1	23.7		0.78			
	EM		LI	63.6		16.2		0.89						
	ML			48.5		24.4		0.8						
	2019	EM	T	111	M2	25.5	8.3	0.86						
		LM		112		37.2	17.6	0.81						
		EM	LI	112		20.2	13.5	0.75						
		LM		112		31.3	10.2	0.86						

Table S5 Description of the four subsets constructed from the four initial datasets described in Table 1. **T**: treated. **LI**: low input. **NT**: Not treated. Genotype number referred to the number of lines in each environment, used for the analysis. Genotyping referred to the availability of molecular markers. NIRS N^{-1} referred to the use, or not, of the NIRS from the previous generation. Total nb NIRS count the number of NIRS available in the analysis. See Table S2& S3 for the description of the sites.

Subgroup	Panel	Year	Site	Treatment	Number of Lines	Lines Genotyped	NIRS N^{-1}	Total nb NIRS
SubSet SD	Set1	2016	HV	T	166	-	-	2
			LV					
		2017	HV	T	117	-	-	2
			LV					
		2018	HV	T	56	-	-	2
			LV					
	Set2	2018	CF	LI	305	-	-	2
			RE	LI				
			EM					
			GL	T				
			EM	NT				
			EM	NT				
		2019	CF	LI	259	-	-	2
			EM					
			RE	LI				
			EM	LI				
	Set3	2016	CP	T	87	-	-	2
			HV					
			HV	T				
			LC					
Set4	2017	LC	T	79	-	-	2	
		VB						
		CP	T					
	2018	HV	T	100	-	-	2	
		LM	T					
		LM	LI					
2019	EM	LI	82	-	-	3		
	EM	T						
	LM	LI						

Subgroup	Panel	Year	Site	Treatment	Number of Lines	Lines Genotyped	NIRS N^{-1}	Total nb NIRS
SubSet SDY	Set1	2017	HV	T	78	-	1	3
			LC					
		2018	HV	T	47	-	1	3
	LC							
	Set3	2017	LC	T	59	-	1	3
			VB					
		2018	CP	T	46	-	1	3
	HV							
	Set4	2018	CP	T	44	-	1	3
			HV					
2019		EM	T	76	-	1	5	
		LM	LI					
SubSet SDM	Set1	2016	HV	T	152	152	-	2
			LC					
		2017	HV	T	68	68	-	2
	LC							
	Set2	2019	GL	T	325	325	-	2
			EM	NT				
	Set3	2017	LC	T	57	57	-	2
			VB					
		2018	CP	T	69	69	-	2
			HV					
	Set4	2018	LM	T	71	71	-	3
			EM	LI				
		2019	EM	T	100	100	-	4
			LM	LI				
SubSet SYM	Set1	2017	HV	T	109	109	1	2
			LC	T				
		2017	CP	T	44	44	1	2
	HV		T					
	LC		T					
	Set3	2017	VB	T	57	57	1	2
			EM	T				
		2018	CP	T	42	42	1	2
	HV		T					
	Set4	2019	EM	T	76	76	1	2
			LM	T				
		2019	EM	LI	77	77	1	2
			LM	LI				

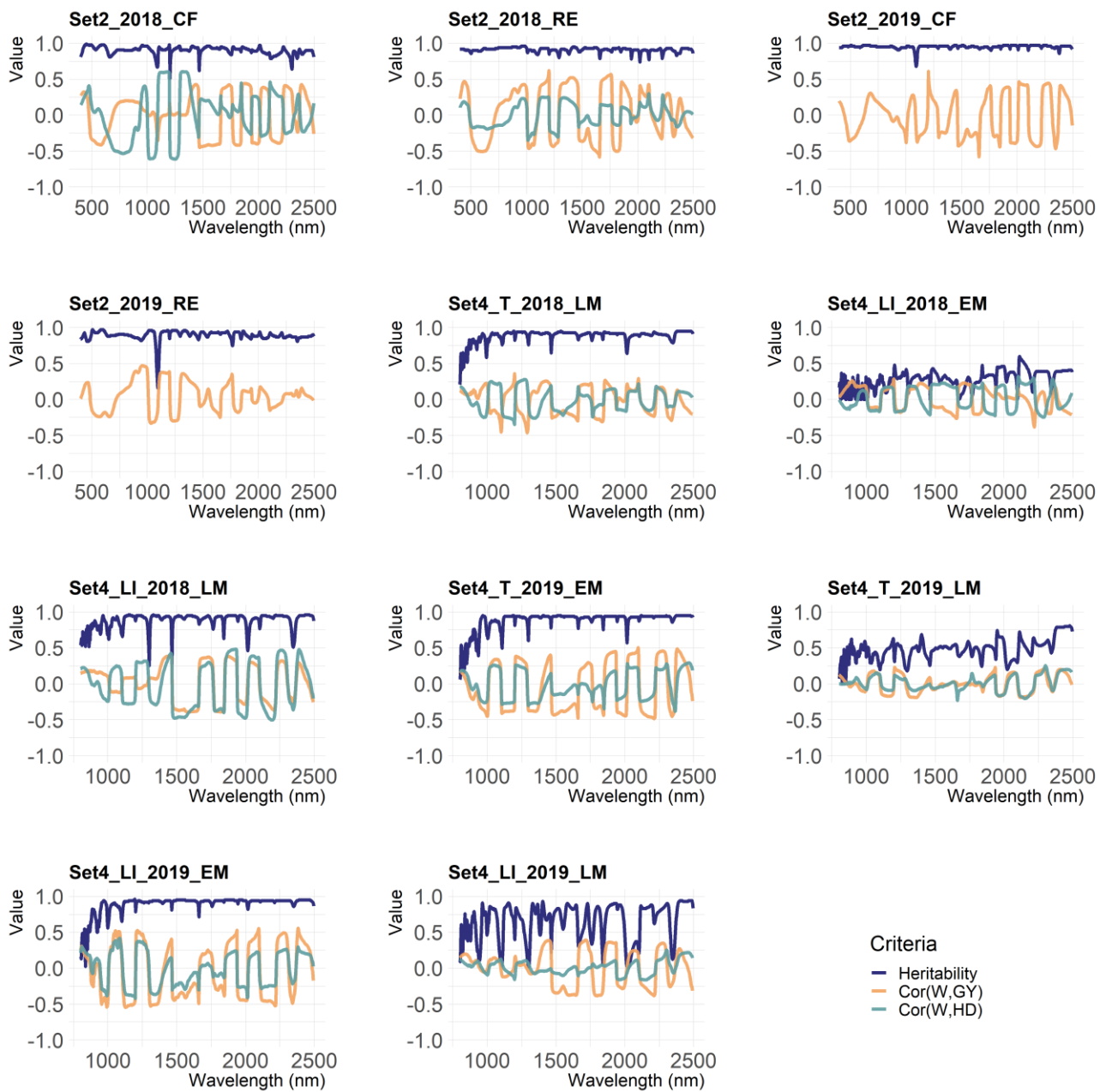


Fig. S1 Representation of the broad-sense Heritability (blue line) for each wavelength, the phenotypic correlation between adjusted absorbances and adjusted GY ($\text{cor}(W, GY)$, yellow line) and the phenotypic correlation between adjusted absorbances and adjusted HD ($\text{cor}(W, HD)$, green line) along spectra. All these criteria were calculated in nine environments from Set2 and Set4 (Table 3-1)

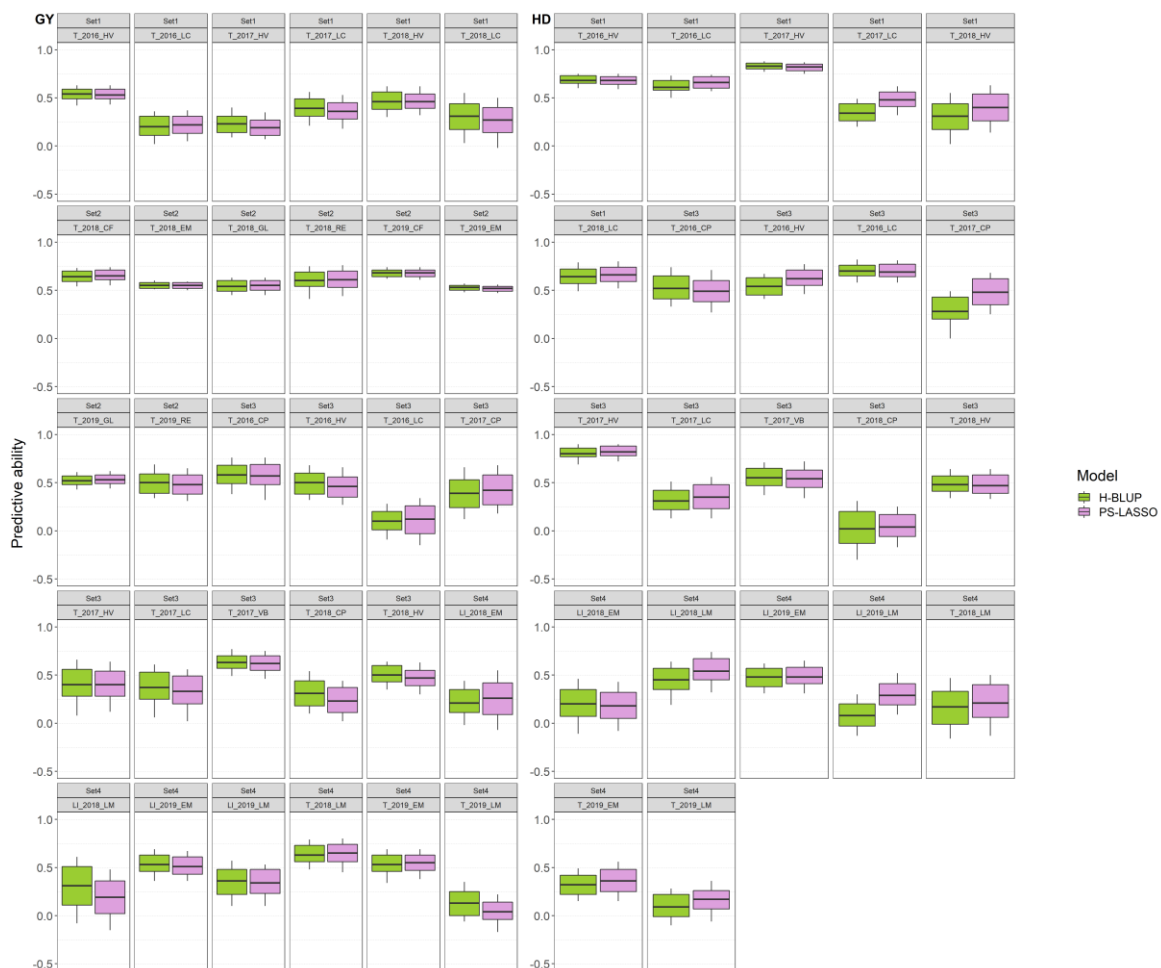


Figure S2. Boxplot of predictive abilities of phenomic selection for grain yield (GY) and heading date (HD) by panel-environment combinations of four different wheat breeding datasets (cf. Table1) with two different models (H-BLUP, PS-LASSO). Predictions were based on a 5-fold cross-validation with 25 repetitions. Each boxplot indicates the mean (bold horizontal line), the 1st to 3rd quartiles (box), the 1st to 9th deciles (whiskers).

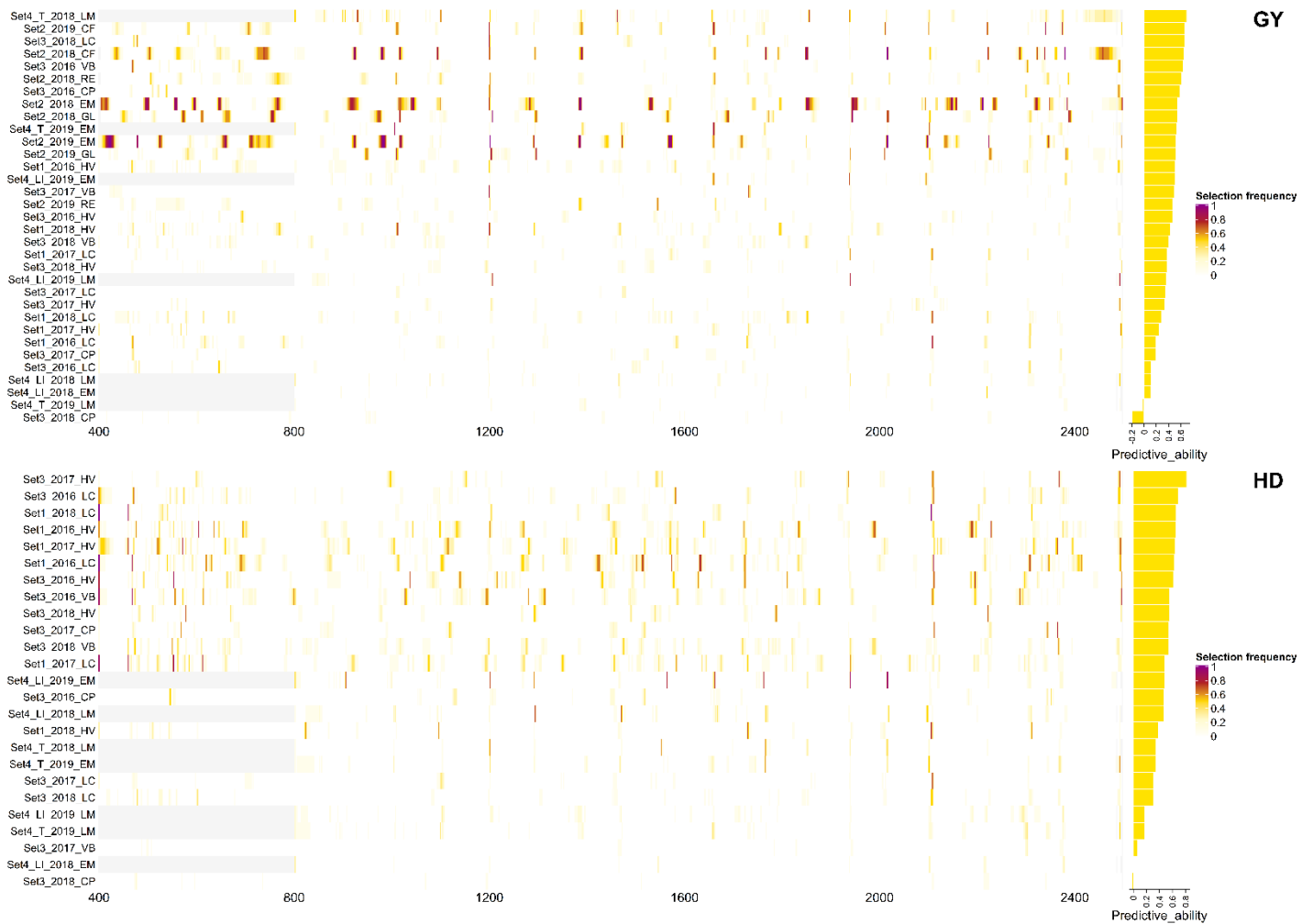


Figure S3. Representation of the selection frequency of each wavelength in each Set-Environment combination by the PS-LASSO model for Grain Yield (GY) and Heading Date (HD). One wavelength was considered as selected when its estimated effect was different from zero. For all the Set-Environment combinations, the predictive ability of the PS-LASSO was added to the selection frequency along spectra. The Set-Environment combinations were sorted in descending order of predictive ability. Selection frequency and predictive ability were calculated on a 5-fold cross-validation repeated 25 times.



Figure S5. Boxplot of predictive abilities as a function of the number of extra NIR spectra combined to spectra S for grain yield (GY) and heading date (HD) for all the panel-environment combinations. When an environment is defined as environment of prediction, it has its own spectra S. The other extra spectra correspond to spectra D or Y. S corresponds to model S without adding any extra spectra in the model. Multi-NIRS models are S+1 to S+4 which correspond to the addition of spectra S plus 1 to 4 extra NIR spectra D or Y. Predictions were based on a 5-fold cross-validation with 25 repetitions. Each boxplot indicates the mean (bold horizontal line), the 1st to 3rd quartiles (box), the 1st to 9th deciles (whiskers). Correlation can range -1 to 1 due to low number of individuals in the data (cf TableS5).

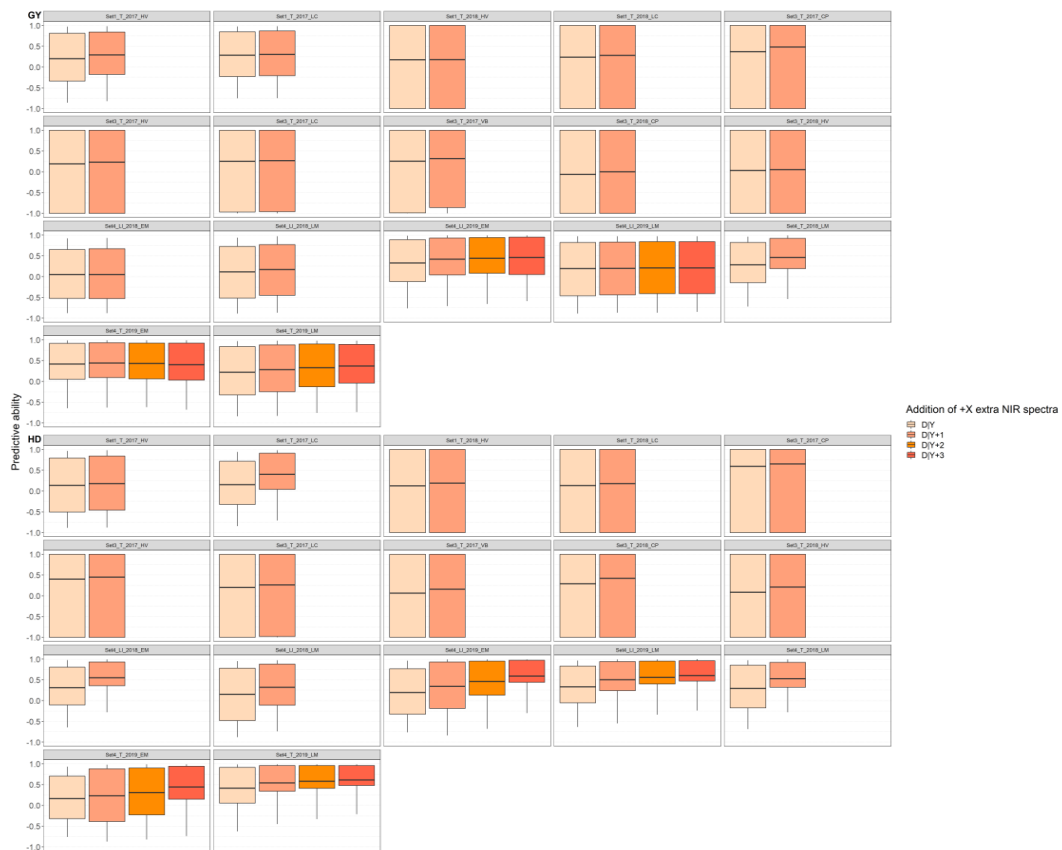


Figure S6. Boxplot of predictive abilities as a function of the number of extra NIR spectra combined excluding spectra S for grain yield (GY) and heading date (HD) for all the panel-environment combinations. When an environment is defined as environment of prediction, it has its own spectra S. The other extra spectra correspond to spectra D or Y. D|Y corresponds to model D or Y without adding any extra spectra in the model. Multi-NIRS models are D|Y +1 to D|Y +3 which correspond to the addition of spectra D or Y plus 1 to 3 extra NIR spectra excluding S. Predictions were based on a 5-fold cross-validation with 25 repetitions. Each boxplot indicates the mean (bold horizontal line), the 1st to 3rd quartiles (box), the 1st to 9th deciles (whiskers). Correlation can range -1 to 1 due to low number of individuals in the data (cf TableS5).

Annexe 2 : données supplémentaires du chapitre 4

Liste des tableaux

Table S1. Variance and standard error (SE) of the genotype (G), the environmental (E), the genotype x environment (GxE) and the residual effects for GY (**A**) and HD (**B**) for Set4-2018 and Set4-2019.

Table S2. Predictive abilities of the different genomic models, single NIRS models, multi-NIRS models and multi-omics models in scenario oGnE for GY and HD

Table S3. Predictive abilities of the different genomic models, single NIRS models, multi-NIRS models and multi-omics models in scenario nGnE for GY and HD

Table S4. Predictive abilities of the different genomic models, single NIRS models, multi-NIRS models and multi-omics models in scenario oGoEref for GY and HD

Table S5. Predictive abilities of the different genomic models, single NIRS models, multi-NIRS models and multi-omics models in scenario nGoEref for GY and HD

Liste des figures

Figure S1. Scheme of the nearest neighbour method to impute missing spectra of the validation set in the oGoEref scenario

Figure S2. Variance decomposition along NIR spectra

Figure S3. Averaged correlation along spectra between the predicted spectra and the acquired spectra of the validation set in the CVRandom_Ref scheme, for the prediction of GY

Figure S4. Averaged correlation along spectra between the predicted spectra and the acquired spectra of the validation set in the CVNewG_Ref scheme, for the prediction of GY

Table S1. Variance and standard error (SE) of the genotype (G), the environmental (E), the genotype x environment (GxE) and the residual effects for GY (**A**) and HD (**B**) for Set4-2018 and Set4-2019. (-) means that the variance effect and the SE associated could not be estimated.

A	Effect	Set4-2018		Set4-2019	
		Variance	SE	Variance	SE
	E	191.72	156.67	381.99	382.67
	G	16.77	2.96	12.76	5.65
	GxE	12.81	1.64	47.58	6.83
	Residual	15.56	0.97	27.11	2.25

B	Effect	Set4-2018		Set4-2019	
		Variance	SE	Variance	SE
	E	-	-	0.29	0.26
	G	-	-	12.18	1.73
	GxE	-	-	2.87	0.23
	Residual	-	-	0.08	0.01

Table S2. Predictive abilities of the different genomic models, single NIRS models, multi-NIRS models and multi-omics models in scenario **oGnE** for GY and HD. CVNewE is a leave-one environment out cross-validation scheme in which a new environment is added to the trial network. Predictive abilities are presented by set (**Set1-2016, Set2-2019, Set4-2018, Set4-2019**) and on average of all sets (**Average**). The best predictive ability by set and among category of models is in bold.

<i>oGnE scenario</i> (CVNewE)	<i>GY prediction</i>					<i>HD prediction</i>			
	Set1 2016	Set2 2019	Set4 2018	Set4 2019	Average	Set2 2019	Set4 2018	Set4 2019	Average
Genomic model									
EG	0.02	0.08	0.31	0.54	0.24	0.83	0.93	0.95	0.90
Single NIRS models									
EW	-0.04	0.19	0.25	0.35	0.19	0.71	0.64	0.95	0.77
EW_P	0.11	0.32	0.27	0.37	0.27	0.61	0.51	0.94	0.69
Multi-NIRS models									
EW _{cb}	-0.13	0.09	0.29	0.41	0.16	0.79	0.85	0.95	0.86
EW _{cb} -P	0.00	0.24	0.3	0.41	0.24	0.79	0.85	0.94	0.86
Multi-omics models									
EG_P	0.14	0.28	0.31	0.55	0.32	0.64	0.93	0.95	0.84
EGW _{cb}	-0.11	0.09	0.29	0.54	0.20	0.86	0.93	0.95	0.91
EGW _{cb} -P	-0.02	0.12	0.30	0.54	0.24	0.85	0.94	0.95	0.91

Table S3. Predictive abilities of the different genomic models, single NIRS models, multi-NIRS models and multi-omics models in scenario **nGnE** for GY and HD. CVNewGE is a leave-one environment out followed by a 5-fold cross-validation scheme in which new genotypes are unphenotyped in all the environment of the trial network and a new environment was added to the trial network. Predictive abilities are presented by set (**Set1-2016, Set2-2019, Set4-2018, Set4-2019**) and on average of all sets (**Average**). They were calculated based on 20 repetitions. The best predictive ability by set and among category of models is in bold.

<i>nGnE scenario</i> (CVNewGE)	<i>GY prediction</i>					<i>HD prediction</i>			
	Set1 2016	Set2 2019	Set4 2018	Set4 2019	Average	Set2 2019	Set4 2018	Set4 2019	Average
Genomic model									
EG	0.18	0.03	0.09	0.32	0.16	0.34	0.39	0.51	0.41
Single NIRS models									
EW	0.00	0.19	0.16	0.26	0.15	0.56	0.32	0.31	0.40
EW_P	0.13	0.32	0.19	0.27	0.23	0.41	0.23	0.31	0.32
Multi-NIRS models									
EW _{cb}	-0.09	0.09	0.18	0.29	0.12	0.68	0.42	0.4	0.5
EW _{cb_P}	0.04	0.26	0.20	0.28	0.20	0.64	0.43	0.4	0.49
Multi-omics models									
EG_P	0.26	0.27	0.16	0.38	0.27	0.24	0.35	0.53	0.37
EGW _{cb}	-0.03	0.06	0.11	0.39	0.13	0.70	0.57	0.59	0.62
EGW _{cb_P}	0.07	0.12	0.14	0.39	0.18	0.69	0.55	0.58	0.61

TableS4. Predictive abilities of the different genomic models, single NIRS models, multi-omics models and multi-NIRS models in scenario oGoEref for GY and HD prediction. CVRandom_Ref is a leave-one environment out followed by a 5-fold cross-validation schemes in which there are sparse testing in all the environment of the trial network which are not the environment of reference. Predictive abilities are presented by set (Set1-2016, Set2-2019, Set4-2018, Set4-2019) and on average of all sets (Average). They were calculated based on 50 repetitions. The optimum models are presented in red, and the best predictive ability by set an among category of models is in bold.

<i>oGoEref scenario</i> (<i>CVRandom_Ref</i>)	<i>GY prediction</i>					<i>HD prediction</i>			
	Set1 2016	Set2 2019	Set4 2018	Set4 2019	Average	Set1 2016	Set4 2018	Set4 2019	Average
Genomic models									
EG	0.12	0.31	0.32	0.55	0.32	0.84	0.93	0.95	0.91
EG_GxE	0.17	0.54	0.46	0.6	0.44	0.80	0.94	0.94	0.89
Single NIRS models									
EW	0.11	0.27	0.27	0.38	0.26	0.70	0.91	0.94	0.85
EW_WxE	0.24	0.35	0.24	0.4	0.31	0.68	0.88	0.94	0.83
Multi-NIRS models									
<i>Optimum</i>									
EW _{cb}	0.17	0.40	0.36	0.45	0.35	0.80	0.91	0.94	0.88
EW _{cb} -W _{cb} xE	0.38	0.60	0.33	0.49	0.45	0.78	0.92	0.94	0.88
EW _{cb} -W _{diag} xE	0.37	0.60	0.37	0.50	0.46	0.78	0.92	0.94	0.88
EW _{cb} -P	0.34	0.55	0.34	0.48	0.43	0.79	0.92	0.94	0.88
<i>Nearest Neighbour</i>									
EW _{cb}	0.07	0.19	0.28	0.4	0.24	0.65	0.88	0.94	0.82
EW _{cb} -W _{cb} xE	0.18	0.20	0.21	0.38	0.24	0.60	0.89	0.93	0.81
EW _{cb} -W _{diag} xE	0.12	0.16	0.14	0.35	0.19	0.59	0.88	0.93	0.80
EW _{cb} -P	0.06	0.16	0.21	0.33	0.19	0.62	0.88	0.93	0.81
<i>Weighted Multivariate Ridge Regression</i>									
EW _{cb}	0.13	0.28	0.27	0.38	0.26	0.69	0.75	0.89	0.78
EW _{cb} -W _{cb} xE	0.28	0.40	0.24	0.41	0.33	0.66	0.72	0.87	0.75
EW _{cb} -W _{diag} xE	0.27	0.41	0.24	0.41	0.33	0.66	0.73	0.87	0.75
EW _{cb} -P	0.22	0.33	0.25	0.40	0.30	0.69	0.74	0.87	0.77

Multi-omics models*Optimum*

EGW _{cb}	0.14	0.39	0.34	0.57	0.36	0.86	0.93	0.95	0.91
EGW _{cb} _W _{cb} xE	0.36	0.60	0.36	0.64	0.49	0.84	0.94	0.94	0.91
EGW _{cb} _W _{diag} xE	0.35	0.60	0.40	0.64	0.50	0.85	0.94	0.94	0.91
EGW _{cb} _P	0.32	0.54	0.35	0.61	0.46	0.85	0.94	0.94	0.91
EGW _{cb} _GxE	0.19	0.59	0.46	0.62	0.46	0.84	0.94	0.95	0.91
EGW _{cb} _GxE_W _{cb} xE	0.39	0.67	0.44	0.66	0.54	0.83	0.94	0.94	0.90
EGW _{cb} _GxE_W _{diag} xE	0.37	0.68	0.45	0.66	0.54	0.83	0.94	0.94	0.90
EGW _{cb} _GxE_P	0.33	0.65	0.44	0.64	0.52	0.84	0.94	0.94	0.91

Nearest Neighbour

EGW _{cb}	0.08	0.24	0.32	0.55	0.30	0.82	0.93	0.94	0.90
EGW _{cb} _W _{cb} xE	0.17	0.23	0.30	0.58	0.32	0.80	0.93	0.94	0.89
EGW _{cb} _W _{diag} xE	0.13	0.19	0.25	0.55	0.28	0.79	0.92	0.94	0.88
EGW _{cb} _P	0.08	0.19	0.30	0.52	0.27	0.80	0.92	0.94	0.89
EGW _{cb} _GxE	0.13	0.52	0.45	0.61	0.43	0.78	0.93	0.94	0.88
EGW _{cb} _GxE_W _{cb} xE	0.21	0.41	0.400	0.61	0.41	0.77	0.93	0.94	0.88
EGW _{cb} _GxE_W _{diag} xE	0.18	0.38	0.35	0.57	0.37	0.76	0.93	0.94	0.88
EGW _{cb} _GxE_P	0.12	0.42	0.42	0.57	0.38	0.77	0.93	0.94	0.88

Weighted Multivariate Ridge Regression

EGW _{cb}	0.11	0.33	0.33	0.56	0.33	0.84	0.93	0.94	0.90
EGW _{cb} _W _{cb} xE	0.26	0.44	0.35	0.62	0.42	0.83	0.94	0.94	0.90
EGW _{cb} _W _{diag} xE	0.25	0.45	0.33	0.62	0.41	0.83	0.93	0.94	0.90
EGW _{cb} _P	0.19	0.38	0.32	0.59	0.37	0.84	0.93	0.94	0.90
EGW _{cb} _GxE	0.15	0.55	0.45	0.61	0.44	0.81	0.94	0.94	0.90
EGW _{cb} _GxE_W _{cb} xE	0.29	0.57	0.44	0.64	0.48	0.81	0.94	0.94	0.90
EGW _{cb} _GxE_W _{diag} xE	0.28	0.57	0.42	0.64	0.48	0.81	0.94	0.94	0.90
EGW _{cb} _GxE_P	0.21	0.56	0.43	0.62	0.46	0.81	0.94	0.94	0.90

TableS5. Predictive abilities of the different genomic models, single NIRS models, multi-omics models and multi-NIRS models in scenario nGoEref for GY and HD prediction. CVNewG_Ref is a leave-one environment out followed by 5-fold cross-validation schemes in which in which new genotypes are unphenotyped in all the environment of the trial network which are not the environment of reference. Predictive abilities are presented by set (Set1-2016, Set2-2019, Set4-2018, Set4-2019) and on average of all sets (Average). They were calculated based on 50 repetitions. The optimum models are presented in red, and the best predictive ability by set an among category of models is in bold.

<i>nGoEref scenario</i> (CVNewG_Ref)	<i>GY prediction</i>					<i>HD prediction</i>			
	Set1 2016	Set2 2019	Set4 2018	Set4 2019	Average	Set1 2016	Set4 2018	Set4 2019	Average
Genomic models									
EG	0.25	0.39	0.24	0.38	0.32	0.35	0.39	0.40	0.38
EG_GxE	0.23	0.54	0.34	0.44	0.39	0.35	0.39	0.42	0.39
Single NIRS models									
EW	0.16	0.29	0.19	0.30	0.23	0.61	0.31	0.28	0.40
EW_WxE	0.25	0.35	0.20	0.37	0.29	0.61	0.32	0.28	0.40
Multi-NIRS models									
<i>Optimum</i>									
EW _{cb}	0.25	0.44	0.27	0.36	0.33	0.72	0.36	0.38	0.49
EW _{cb} _W _{cb} xE	0.40	0.60	0.29	0.43	0.43	0.72	0.35	0.39	0.49
EW _{cb} _W _{diag} xE	0.39	0.60	0.32	0.42	0.43	0.72	0.36	0.39	0.49
EW _{cb} _P	0.36	0.55	0.3	0.38	0.40	0.73	0.36	0.38	0.49
<i>Nearest neighbour spectra</i>									
EW _{cb}	0.11	0.18	0.11	0.17	0.14	0.43	0.14	0.09	0.22
EW _{cb} _W _{cb} xE	0.19	0.20	0.09	0.19	0.17	0.44	0.14	0.08	0.22
EW _{cb} _W _{diag} xE	0.15	0.16	0.06	0.19	0.14	0.43	0.13	0.09	0.22
EW _{cb} _P	0.08	0.15	0.08	0.16	0.12	0.43	0.14	0.09	0.22
<i>Weighted Multivariate Ridge Regression</i>									
EW _{cb}	0.19	0.31	0.17	0.30	0.24	0.59	0.31	0.28	0.39
EW _{cb} _W _{cb} xE	0.29	0.40	0.19	0.37	0.31	0.60	0.32	0.27	0.40
EW _{cb} _W _{diag} xE	0.29	0.41	0.18	0.36	0.31	0.60	0.32	0.28	0.40
EW _{cb} _P	0.25	0.34	0.14	0.32	0.26	0.60	0.32	0.28	0.40

Multi-omics models*Optimum*

EGW _{cb}	0.29	0.48	0.25	0.47	0.37	0.74	0.59	0.51	0.61
EGW _{cb} -W _{cb} xE	0.42	0.63	0.28	0.53	0.47	0.74	0.58	0.50	0.61
EGW _{cb} -W _{diag} xE	0.41	0.63	0.31	0.53	0.47	0.74	0.57	0.51	0.61
EGW _{cb} -P	0.39	0.58	0.30	0.50	0.44	0.74	0.57	0.50	0.60
EGW _{cb} -GxE	0.29	0.61	0.33	0.52	0.44	0.73	0.57	0.53	0.61
EGW _{cb} -GxE-W _{cb} xE	0.43	0.67	0.33	0.56	0.50	0.74	0.57	0.52	0.61
EGW _{cb} -GxE-W _{diag} xE	0.41	0.68	0.35	0.55	0.50	0.74	0.57	0.51	0.61
EGW _{cb} -GxE-P	0.38	0.65	0.36	0.53	0.48	0.74	0.56	0.51	0.60

Nearest neighbour spectra

EGW _{cb}	0.20	0.27	0.18	0.35	0.25	0.49	0.40	0.45	0.45
EGW _{cb} -W _{cb} xE	0.23	0.26	0.16	0.37	0.26	0.50	0.39	0.43	0.44
EGW _{cb} -W _{diag} xE	0.20	0.21	0.11	0.37	0.22	0.50	0.39	0.46	0.45
EGW _{cb} -P	0.15	0.21	0.14	0.35	0.21	0.50	0.37	0.45	0.44
EGW _{cb} -GxE	0.21	0.51	0.30	0.42	0.36	0.49	0.39	0.44	0.44
EGW _{cb} -GxE-W _{cb} xE	0.26	0.40	0.27	0.41	0.34	0.50	0.39	0.42	0.44
EGW _{cb} -GxE-W _{diag} xE	0.22	0.37	0.23	0.41	0.31	0.50	0.39	0.43	0.44
EGW _{cb} -GxE-P	0.17	0.42	0.27	0.40	0.32	0.50	0.39	0.42	0.44

Weighted Multivariate Ridge Regression

EGW _{cb}	0.24	0.40	0.22	0.43	0.32	0.65	0.48	0.49	0.54
EGW _{cb} -W _{cb} xE	0.32	0.48	0.24	0.48	0.38	0.66	0.49	0.48	0.54
EGW _{cb} -W _{diag} xE	0.32	0.48	0.23	0.48	0.38	0.66	0.50	0.48	0.55
EGW _{cb} -P	0.28	0.44	0.21	0.45	0.35	0.66	0.49	0.49	0.55
EGW _{cb} -GxE	0.24	0.56	0.33	0.48	0.40	0.65	0.49	0.49	0.54
EGW _{cb} -GxE-W _{cb} xE	0.33	0.57	0.31	0.51	0.43	0.66	0.49	0.47	0.54
EGW _{cb} -GxE-W _{diag} xE	0.32	0.58	0.31	0.51	0.43	0.66	0.50	0.46	0.54
EGW _{cb} -GxE-P	0.28	0.57	0.31	0.49	0.41	0.66	0.49	0.48	0.54

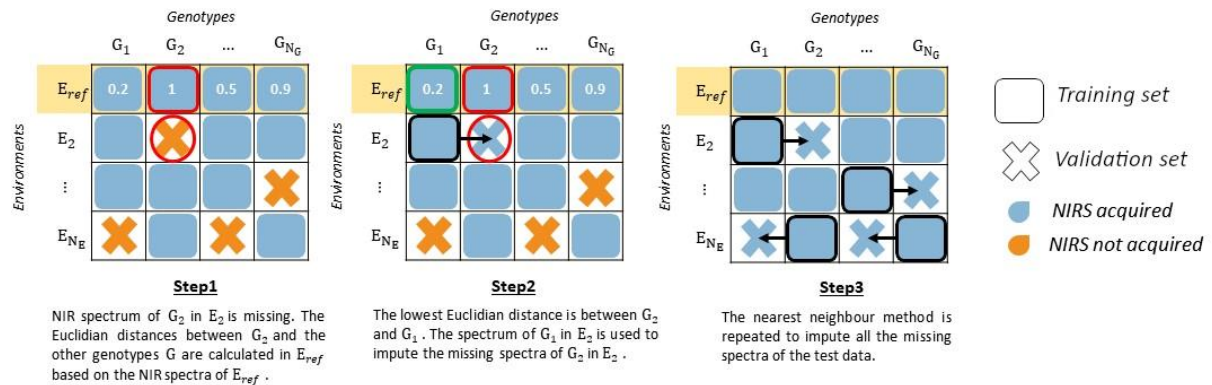


Figure S1. Scheme of the nearest neighbour method to impute missing spectra of the validation set in the **oGoEref** scenario. The square represents the data constituting the training set and the cross represents the data constituting the validation set. The training set was always phenotyped for the target trait whereas the validation set was not phenotyped for the target trait. The blue colour indicates that the NIR spectrum was acquired on the genotype i in the environment j whereas the orange colour indicates that the NIR spectrum was not acquired. Finally, the yellow colour indicates the environment of reference.

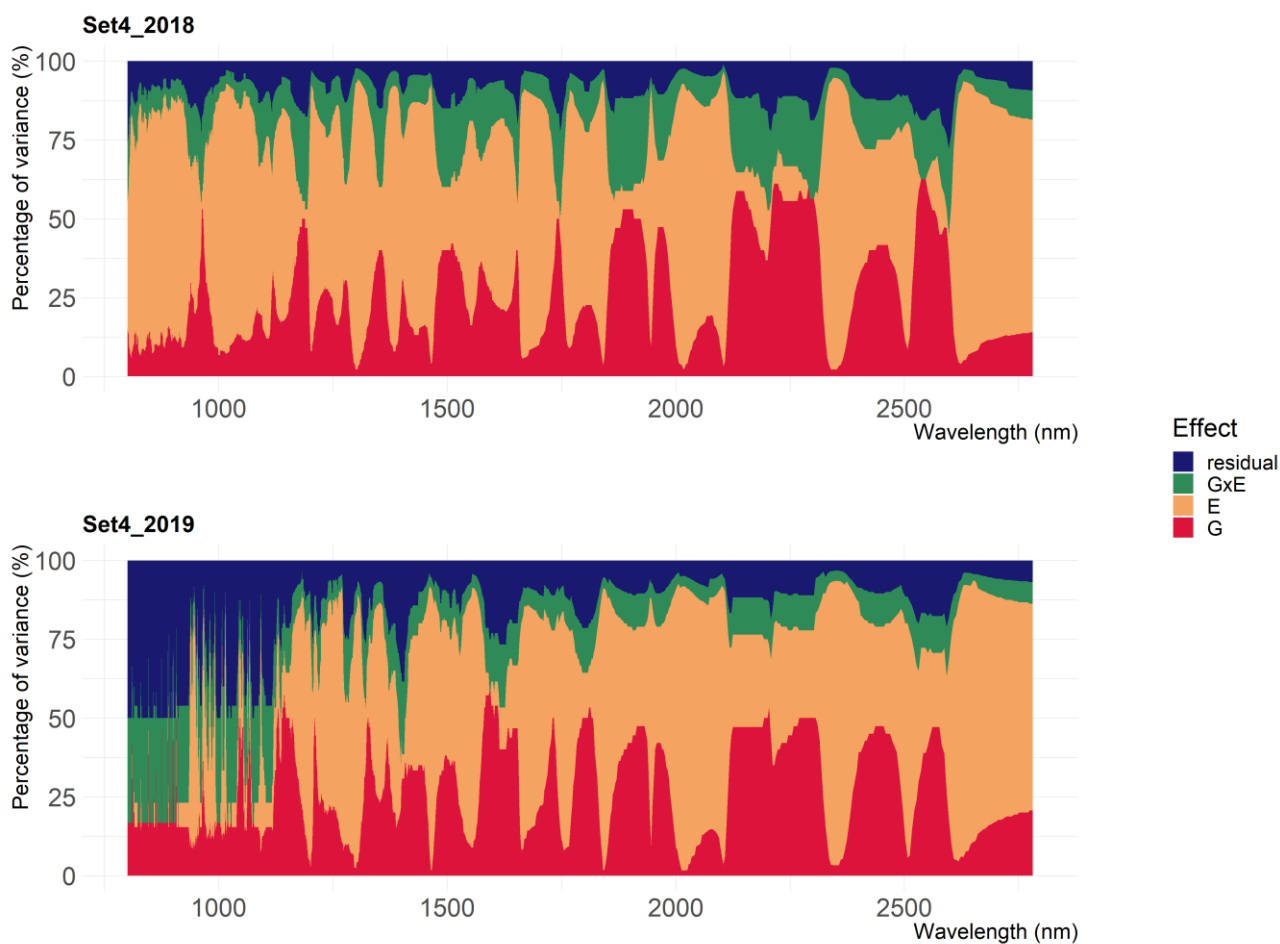


Figure S2. Proportion of the genotypic (G), the environmental (E), the genotype x environment (GxE) and residual variance along the spectrum for Set4-2018 and Set4-2019.

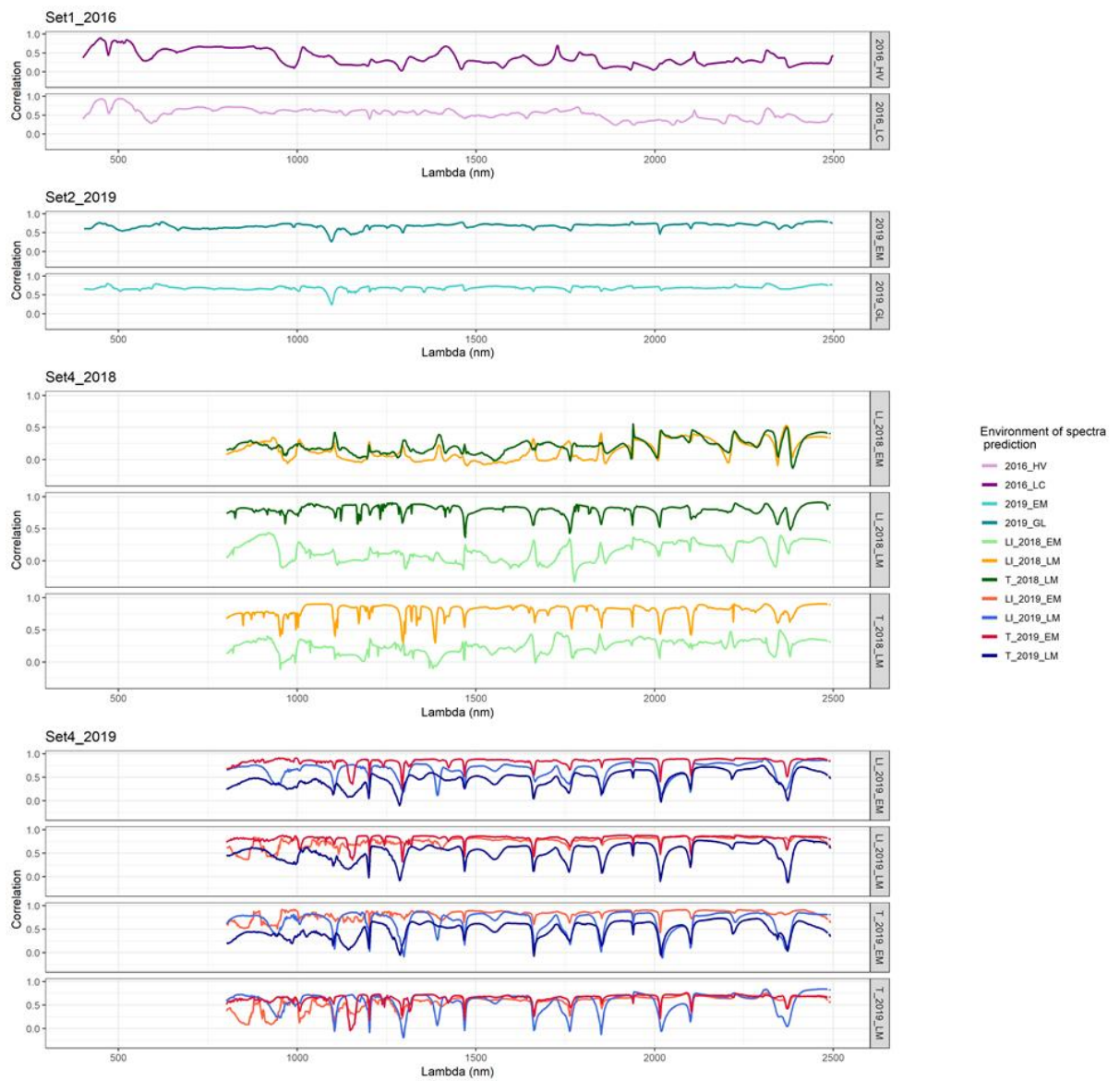


Figure S3. Averaged correlation along spectra between the predicted spectra and the acquired spectra of the validation set in the CVRandom_Ref scheme, for the prediction of GY. In each set, the title of the correlation box is the reference environment and the coloured lines the environments for which spectra are predicted.

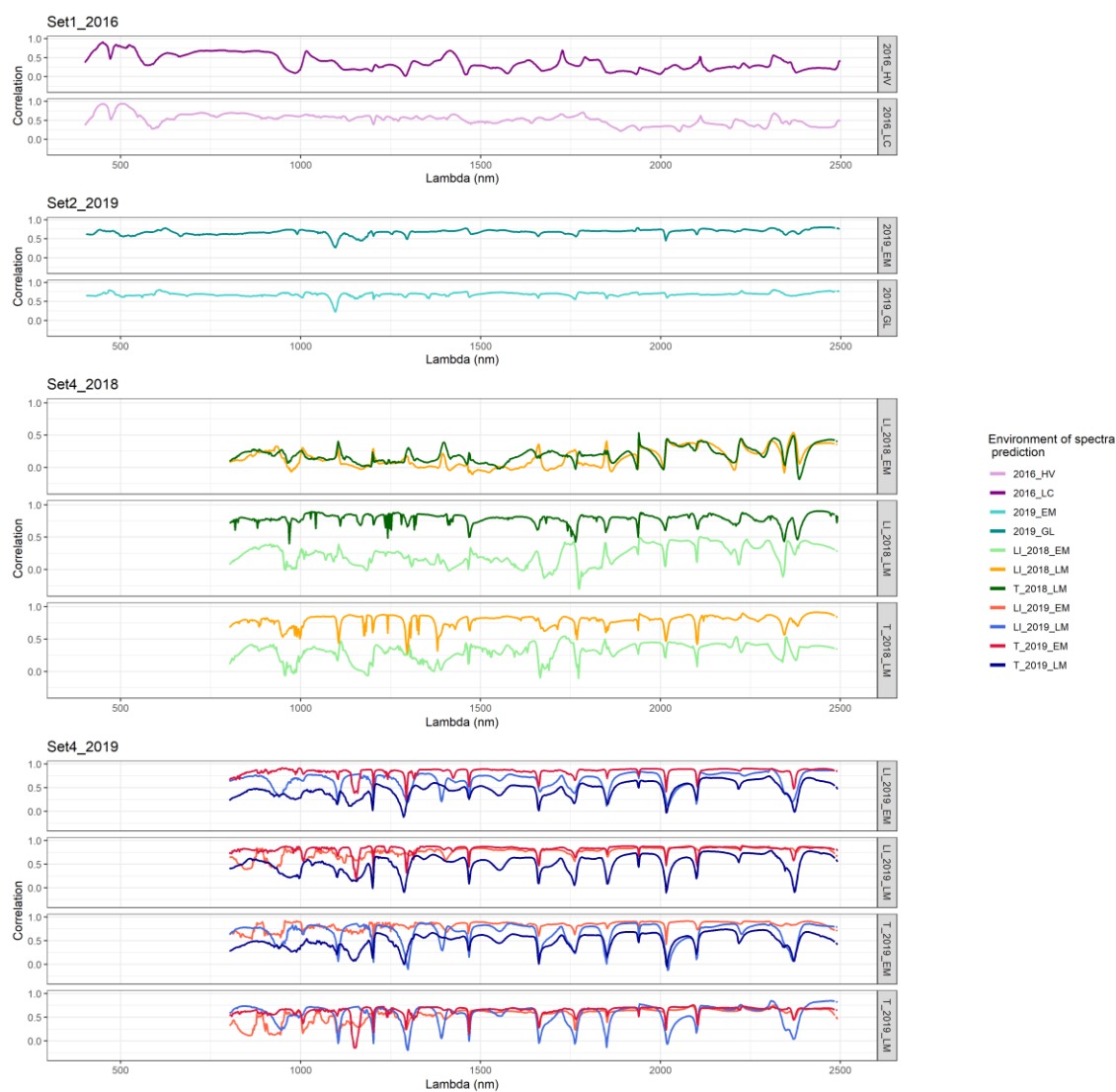


Figure S4. Averaged correlation along spectra between the predicted spectra and the acquired spectra of the validation set in the CVNewG_Ref scheme, for the prediction of GY. In each set, the title of the correlation box is the reference environment and the coloured lines the environments for which spectra are predicted.