Genomalt : Prédiction génomique du rendement et de la qualité brassicole chez l'orge d'hiver à 6 rangs

Gilles CHARMET [1], Pierre PIN [2], Marc SCHMITT [3], Nathalie LEROY [4], Bruno CLAUSTRES [4], Christopher BURT [4], Amélie GENTY [2]

1 - INRAE- UCA UMR GDEC, 5 chemin de Beaulieu 63000 Clermont-Ferrand – France
2 - SECOBRA Recherches SAS, Centre de Bois Henry, 78580 Maule – France
3 - IFBM, 7 rue du Bois de la Champelle, F-54500 Vandoeuvre les Nancy -France
4 - RAGT 2N, Place du bourg, 12510 Druelle - France

1 Introduction

L'orge (*Hordeum vulgare* L.) est une des deux espèces à l'origine de l'agriculture dans l'ancien monde. C'est probablement la première espèce qui fut domestiquée au moyen Orient, il y a 10 000 à 12 000 ans, à partir de l'espèce sauvage *Hordeum vulgare ssp. spontaneum* comme le suggèrent les grains fossiles retrouvés dans plusieurs sites archéologiques du croissant fertile (Zohary and Hopf, 1993). Badr *et al.* (2000) ont montré l'origine monophylétique de cette domestication, et émis l'hypothèse de la zone Israel-Jordanie comme berceau de sa culture, avec la région de l'Himalaya comme centre diversification secondaire.

L'ancêtre sauvage (*H. vulgare ssp. spontaneum*) possède un épi de type 2 rangs, avec les fleurs latérales demeurant vestigiales. Il est possible que les fermiers du néolithique aient sélectionné des mutants à 6 rangs, qui donnaient ainsi davantage de grains et par là un meilleur rendement. Le gène responsable du type 6, vrs1 (*six-rowed spike 1*), a été isolé par clonage positionnel (Komatsuda *et al.,* 2007). L'allèle sauvage *Vrs1* (type 2 rangs) code pour un facteur de transcription qui inclus un homéo domaine associé à un motif de type leucine zipper. La perte de fonction de *Vrs1* provoque la conversion des épillets latéraux rudimentaires chez les 2 rangs en épillets normaux et fertiles chez les 6 rangs. Les analyses phylogénétiques ont montré que le phénotype à 6 rangs est ainsi apparu de façon répétée dans le temps et dans différentes régions, par des mutations indépendantes de *Vrs1*.

Les orges à 6 rangs, (légèrement) plus productives, sont généralement préférées comme céréales fourragères, tandis que les orges à deux rangs sont privilégiées pour la production de malt et de bière. En effet, les orges à deux rangs ont des caractéristiques plus favorables pour la production de bière, notamment sa première étape qui est celle du malt. Le malt est en fait de l'orge germée, puis séchée à des températures plus ou moins élevées. La qualité du malt dépend de la taille des grains et de leur homogénéité, de sa friabilité et de son pouvoir diastasique, c'est à dire sa capacité à transformer l'amidon du grain en sucres fermentescibles, base de la fermentation alcoolique. Les orges brassicoles doivent avoir une teneur limitée en protéines, qui peut être responsable d'une bière trouble. Les orges à deux rangs, qui ont des grains plus gros et de tailles plus homogènes, sont donc préférées par les malteurs, par exemple en Allemagne et au Royaume-Uni. La France, premier producteur européen pour l'orge et premier exportateur Mondial de malt (1.2 Mt par an, 80 % de la production), utilise également des orges à 6 rangs, plus productives, généralement semées après un blé. C'est pourquoi les sélectionneurs français ont développé des programmes dédiés à l'amélioration variétale de l'orge à 6 rang brassicole. La plupart des schémas utilisent des haploïdes doubles pour accélérer les cycles de sélection.

L'évaluation de la valeur maltière s'effectue par un test de micro-maltage (ex. Haslemore *et al.,* 1982), qui demande du temps (> 4 jours) et une quantité de grains importante. Ce test n'est donc appliqué qu'à un nombre limité de lignées en fin de sélection, déjà triées pour des caractères agronomiques comme le rendement et les résistances aux maladies, juste avant les tests officiels pour l'inscription au catalogue. La pression de sélection pour la valeur maltière est donc assez faible et appliquée à des effectifs, ce qui se traduit par un progrès génétique assez lent.

Le terme sélection génomique (SG) a été proposé par Meuwissen *et al.* (2001), qui ont appliqué des méthodes de régression pénalisée ou bayésiennes à des populations animales, notamment aux taureaux de races laitières, pour estimer leur valeur génétique en descendance (breeding value). En SG, dans une première étape, les effets des marqueurs (couvrant tout le génome, c'est-à-dire non sélectionnés au préalable comme associés à des QTL) sont estimés à partir de données de génotypage et de phénotypage d'une population d'entraînement ou population de référence. Des méthodes adaptées au problème du sur-paramétrage doivent être utilisées, dans le cas général ou le nombre de marqueurs excède celui des individus génotypés. Puis ces effets à tous les marqueurs sont employés pour reconstruire la valeur génétique des individus d'une population pour lesquels on dispose seulement du génotypage. Ces valeurs sont appelées GEBV (Genomic Estimates of Breeding Values). La SG a été appliquée avec succès aux vaches laitières (Goddard and Hayes, 2007). En effet, dans les troupeaux laitiers, grâce à l'insémination artificielle, la sélection est beaucoup plus intense sur les taureaux. Mais ces derniers ne produisant pas de lait, l'estimation de leur valeur laitière se fait sur leur descendance femelle, ce qui est couteux en temps (au moins 8 ans) et en argent. L'avantage de la prédiction génomique des valeurs laitières par SG, qui peut être estimée dès la naissance des veaux pour quelques dizaines d'euros, permet donc une sélection plus intense et une mise en production (de sperme) plus précoce (3 ans). La SG

a permis de doubler le gain génétique par unité de temps, pour un coût réduit de 92 % (Shaeffer, 2006). Même si l'avantage de la SG est moins évident chez les plantes, il peut néanmoins s'avérer positif à condition que 1) le coût du génotypage soit significativement plus faible que celui du phénotypage (qui demande souvent de nombreuses répétitions dans des environnements différents pour être précis et 2) que la précision des prédictions génomiques soit comparable à celle des phénotypes, c'est-à-dire l'héritabilité des caractères (Bernardo & Yu, 2007; Crossa *et al.,* 2010; Heffner *et al.,* 2009; Jannink *et al.,* 2010; De los Campos *et al.,* 2013).

La condition 1 s'applique assez bien aux traits de qualité des céréales comme la qualité boulangère du blé ou la valeur brassicole de l'orge, dont le phénotypage est relativement cher et réalisé sur des générations tardives. Dans ce projet, nous avons analysé des caractères agronomiques et de qualité brassicole dans un ensemble de variétés inscrites et de lignées de sélection d'orge d'hiver à 6 rangs, et conduit une étude de faisabilité d'une sélection génomique intégrée dans des schémas d'amélioration variétale.

2 Matériel et méthode

▶ Matériel végétal

Deux sélectionneurs français, RAGT 2N et SECOBRA Recherche, anonymisés plus loin en breeder1 et breeder2, ont fourni chacun un set de lignées haploïdes doublées (HD) « propriétaires », qui avaient été préalablement triés pour des caractères adaptatifs comme la hauteur, résistance à la verse et aux maladies, date de floraison. Chaque set de lignées propriétaire a été évalué par son sélectionneur en 2 ou 3 lieux pendant deux campagnes, 2017/2018 et 2018/2019. Pour assurer une connectivité dans les données expérimentales, chaque sélectionneur a également évalué un set commun de variétés inscrites récemment au catalogue, plus loin appelées « fondateurs » (car fréquemment utilisées comme parents de croisement par les sélectionneurs).

▶ Données de génotypage

La « barley 50K iSelect SNP Array » (Bayer *et al.,* 2017) a été utilisée pour le génotypage pan-génomique des 679 lignées et cultivars. Des 44,040 SNP initialement présents sur la puce, 24 945 ont été retenu après filtrage sur les données manquantes (< 20 % par SNP), l'hétérozygotie des SNP (< 5 %) et la fréquence de l'allèle minoritaire (> 1 %), parmi lesquels 24,101 étaient cartographies sur la « barley physical map V2 » et utilisé dans la suite des analyses. Les données manquantes ont été imputés avec l'algorithme EM (expectation-maximization » (Poland *et al.,* 2012) implanté dans la fonction A.mat du package rrBLUP (Endelman, 2011).

Une matrice des relations additives génomique, K, a été calculée avec les 24 101 marqueurs selon van Raden (2008) avec la fonction A.mat :

$$K = WW^T / (2\sum(p_k\text{-}1)p_k)$$

Où W est une matrice centrée N × M dont la ligne $i\ Wik = Xik + 1 - 2p_k$ avec X_{ik} le génotype du *i-ème* individu pour le *k-ième* marqueur codé {− 1,0,1} et p_k la fréquence de l'allèle majoritaire au *k-ième* marqueur.

Une analyse en coordonnée principale (PCoA, commande "cmdscale" en R) a été appliquée à une matrice des distances de Rogers (Rogers 1972) calculée par la commande "dist" en R, pour illustrer les relations génétiques entre les lignées d'orge.

▶ Données phénotypiques

Breeder1 a fourni 259 lignées HD propriétaires et Breeder2, 315 HD. 105 « fondateurs », cad des variétés inscrites disponibles pour expérimentation selon le règlement UPOV ont été évaluées par chaque sélectionneur pour corriger les effets principaux lieux.

Breeder1 a évalué ses HD et 104 variétés dans 3 lieux en France, Thoiry, Auffay, Warmeriville, durant les 2 campagnes récolte 2018 (1872 parcelles) and 2019 (1327 parcelles). Dans chaque lieu était implanté un essai en 6 blocks incomplets, la majorité des lignées sans répétitions, avec quelques témoins répétés 15 à 20 fois (cv Etincel, Pixel and Visuel), et jusqu'à 50-60 fois (cv KWS-Tonic) en 2018. En 2019, un dispositif plus simple a été implanté avec seulement 3 blocs et un seul témoin (cv Pixel) répété 42 à 50 fois. Comme ces témoins étaient distribués au hasard dans la parcelle expérimentale, des modèles spatiaux ont été utilisés pour tenter de corriger les effets de l'hétérogénéité dans la parcelle. Les essais ont été conduits selon les pratiques agricoles locales, incluant un traitement fongicide.

Breeder2 a évalué ses 315 lignées HD et 91 des 105 variétés dans deux lieux, Cuperly et Premesques, durant les mêmes années 2018 (904 parcelles) and 2019. Dans chaque lieu était implanté un essai en 19 blocs incomplets avec les HD non répétés et des témoins répétés 10 fois (cv Amistar) ou 15-20 fois (cv Casino and Etincel) en 2018. Comme pour Breeder 1, des modèles spatiaux ont été employés pour corriger des hétérogénéités du champ.

Le jeu de variables commun disponible sur toutes les parcelles comprenait le rendement Yield (dt/ha), la teneur en protéines (%), le poids de 1000 grains (g), poids spécifique (Kg/hl), Calibration (% grains > 2.5 mm) et date d'épiaison (jour julien après le 1 janvier), ci-après dénommés variables agronomiques.

En outre, les traits suivants associés à la valeur maltière ont été mesurés par tests de micro-maltage sur un plus petit nombre de lieux (Cuperly et Warmeriville en 2018, Premesques and Warmeriville en 2019). Comme une seule répétition a été mesurée, y compris pour les témoins, aucune correction spatiale n'était possible pour ces 4 variables : friabilité du malt, extrait, viscosité et teneur en beta-glucanes.

La friabilité du malt a été estimée par la méthode (European Brewery Convention) 4.15

L'Extrait du malt a été déterminé par la méthode EBC 4.5.e1. Il définit le potentiel du malt à produire du moût soluble par un procédé standard de brassage.

La viscosité du moût est un paramètre important de la qualité du malt. Plus elle est faible, meilleure est la modification des grains durant la germination. La viscosité du moût est mesurée à 20°C à l'aide d'un viscosimètre calibré selon la EBC 8.4.

La viscosité du moût est liée à la teneur en beta-glucanes solubles. Un bon malt contient une quantité limitée de

ces polysaccharides pariétaux. Ils sont déterminés par la méthode EBC 4.16.2 (High molecular weight β-glucan content of malt and malt wort: fluorimetric method).

Les tests de micro maltage et les mesures des traits de valeur maltière ont été réalisés par l'IFBM.

▶ Analyse des données phénotypiques

Sur chaque essai les témoins répétés au hasard ont été utilisés pour une correction spatiale à l'aide du package SpATS de R (Rodriguez-Alverez *et al.,* 2018). Puis les données spatialement ajustées ont été exploitées par un modèle mixte linéaire (LMM) avec la librairie lme4 en R, avec les génotypes et leurs interactions comme facteurs aléatoires. Pour les traits de qualité du malt pour lesquels la correction spatiale n'était pas possible, les données brutes ont été utilisées.

L'utilisation du modèle mixte est justifiée car il est connu que dans les dispositifs très déséquilibrés comme ceux de ce projet, les prédictions des effets aléatoires sont mieux corrigées des effets fixes que les moyennes ajustées dans les modèles à effets fixes.

$$y_{ijk} = \mu + y_j + y{:}s_{jk} + g_i + g\,y_{ij} + gs_{ik} + \varepsilon_{ijk} \quad \textbf{(1)}$$

avec y_{ijkl} la variable phénotypique spatialement ajustée ou non du *i-ème* génotype la *j-ième* année dans le *k-ème* site, μ la moyenne générale, g_i l'effet *i-ème* génotype, y_j celui de la *j-ième* année, $y{:}s_k k$ est l'effet du *k-ième* lieu hierarchisé à la *j-ième* année, gy_{ij} is est l'interaction entre le *i-ème* génotype et la *j-ième* année, gs_{ik} l'interaction entre le *i-ème* génotype and le *k-ème* lieu et ε_{ijk} l'erreur résiduelle, qui contient donc l'interaction triple qui n'est pas estimable faute de répétitions. g_i et ses interactions ont été considérés comme des effets aléatoires en LMM.

L'équation (1) a été utilisée (commande VarComp dans la librairie R lme4) pour estimer les composantes de la variance σ^2_g, σ^2_{gy}, σ^2_{gs} and σ^2_e, et leurs intervalles de confiance (commande confint en R), qui permettent de calculer une héritabilité au sens large comme

$$h^2 = \sigma^2_g / (\sigma^2_g + \sigma^2_{gy}/ny + \sigma^2_{gs}/ns \text{ and } \sigma^2_e / nrep)$$

Avec *ny*, *ns* and *nrep* les nombres moyens d'années, de sites et de répétitions par génotype, respectivement. Les héritabilités pour chaque trait ont été calculées sur l'ensemble du dispositif, ou séparément avec les essais de chaque sélectionneur.

Les moyennes conditionnelles (cad. Corrigées des effets principaux environnement) de chaque génotype ont été extraites du LMM (commande ranef du package lme4 R) et utilisées pour la suite des analyses, en commençant pas la distribution des caractères et leurs corrélations 2 à 2.

Comme les variances génotypiques étaient en général plus grandes que les variances de leurs interactions, ces moyennes ajustées ont été utilisées par la suite pour tester la valeur prédictive des modèles de sélection génomique.

▶ Modèles de prédiction génomique

Le package R BWGS R (Charmet *et al.,* 2020) a été utilisé pour estimer la valeur prédictive de 4 modèles de sélection génomique : GBLUP, Bayes Cpi (Habier *et al.,* 2011), LASSO (Park & Casella, 2008), et EGBLUP. GBLUP est basé sur un contrôle génétique dit infinitésimal, où tous les marqueurs ont un effet tiré d'une distribution gaussienne, tandis de Bayes Cpi suppose une proportion pi de marqueurs ayant des effets nuls, et les autres des effets tirés d'une distribution de Student. LASSO suppose également une distribution des effets davantage centrée sur 0 que la loi Normale. EGBLUP est une extension de GBLUP avec un terme utilisant le carré de la matrice des relations addictives pour modéliser les interactions entre paires de marqueurs (Jiang and Reif, 2015).

▶ Validation des modèles

Plusieurs stratégies ont été utilisées pour comparer la valeur prédictive des modèles :

1. Validation croisée avec 10 volets tirés au hasard dans :

a. L'ensemble des lignées Breeder1 + Breeder2 + fondateurs (N = 679)

b. Les lignées du Breeder1 + fondateurs (N = 364)

c. Les lignées du Breeder2 + fondateurs (N = 420)

d. Les variétés inscrites (fondateurs) seulement (N = 105)

Chaque validation croisée (9/10-1/10) a été répétée 50 fois.

Les stratégies b et c donne une estimation de ce que chaque sélectionneur peut attendre en utilisant son propre matériel et les lignées inscrites, tandis que la stratégie a mesure l'avantage que pourraient avoir deux sélectionneurs à réunir leurs jeux de données pour construire des modèles de prédiction génomique. Enfin la stratégie d permet d'estimer la valeur prédictive d'une population d'entrainement de taille très réduite, ce que pourrait espérer par exemple un sélectionneur débutant n'ayant pas encore de matériel propre.

Pour mesurer si les valeurs prédictives des stratégies b-c comparées à a sont dues simplement à la taille de la population, on a appliqué la stratégie a) à des sous ensemble aléatoires de taille N dans {50, 100, 200, 300, 400, 500}, parmi les 679 lignées, avec 50 répétitions de chaque tirage. Nous illustrerons seulement avec les résultats pour le rendement et la friabilité du malt, les caractères qui présentent les extrêmes pour la valeur prédictive des modèles.

2. Une validation inter-sélectionneurs, en utilisant :

a. Breeder1 + variétés comme set d'entrainement et les lignées de Breeder 2 comme set de validation.

b. Breeder2 + variétés comme set d'entrainement et les lignées de Breeder 1 comme set de validation.

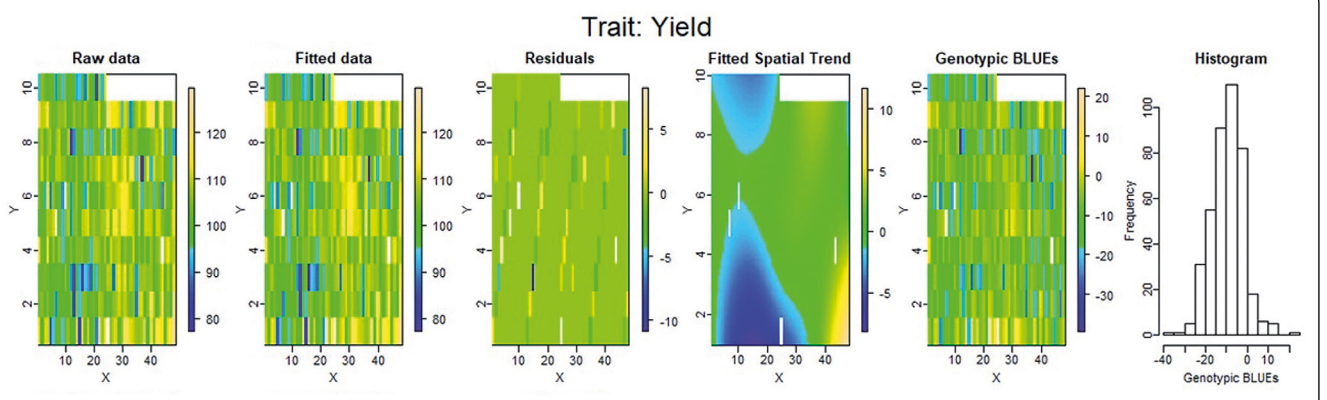
c. Les lignées de Breeder1 + Breeder2 lines pour l'entrainement et les variétés comme jeu de validation.

La valeur prédictive était calculée comme la corrélation entre les GEBV et les moyennes ajustées du LMM. Pour obtenir des intervalles de confiance de ces valeurs, nous avons utilisé une méthode de rééchantillonnage (bootstrap) comme décrite par Rutkoski *et al.* (2012).

3 Résultats

▶ Statistiques de base

L'ajustement spatial s'est révélé efficace pour réduire l'erreur résiduelle.



Le tableau 1 présente les composantes de la variance des effets aléatoires du modèle (1) appliqué aux données compète, et l'héritabilité des moyennes par génotype.

TRAIT	σ^2_g	σ^2_{gS}	σ^2_{gY}	σ^2_e	h^2
Yield	7.3 [5.08-9.20]	2.66 [0.93-2.25]	2.99 [0.85-2.15]	34.4 [31.8 -37.2]	**0.551**
Protein	0.067 [0.055-0.085]	0.015 [0-0.027]	0.010 [0-0.033]	0.305 [0.28-0.33]	**0.613**
TGW	7.22 [6.22-8.36]	0.52 [0.02-1.04]	0.68 [0.21-1.20]	8.59 [7.93-8.32]	**0.837**
TestW	1.90 [1.50-2.37]	0.21 [0.049-0.40]	0.20 [0.041-0.43]	1.94 [1.71-2.20]	**0.775**
Calibration	59.3 [51.2-68.4]	14.7 [10.9-18.7]	3.0 [0.13-6.23]	51.8 [47.8-56.4]	**0.853**
Heading	1.46 [1.18-2.28]	0.15 [0-0.54]	0.58 [0-068]	4.41 [3.99-5.11]	**0.652**
Friability	58.4 [51.2-66.6]	8.60 [6.47-11.98]	9.70 [7.56-18.2]	13.7 [10.16-19.22]	**0.895**
Extract	1.00 [0.43-1.02]	0.06 [0-0.36]	0.24 [0-0.88]	1.34 [0.94-2.12]	**0.753**
Viscosity	0.71 [0.60-0.80]	0.28 [0.16-0.31]	0.10 [0.06-0.18]	0.36 [0.207-0.55]	**0.769**
β-Glucan	163 [141.3-187.8]	18.6 [14.1-18.7]	15.7 [0-52.6]	80.6 [59.7-110.1]	**0.851**

Tableau 1 : *Composantes de la variance et leur intervalle de confiance pour les 10 caractères .*

La variance génotypique σ^2_g apparait plus grande que les deux composantes d'interaction σ^2_{gS} and σ^2_{gY}, conduisant à des héritabilités des moyennes génotypique variant 0.5 (rendement) à 0.9 (friabilité). Les héritabilités des caractères liés au malt sont toutes supérieures à 0.75, malgré un dispositif expérimental plus réduit que celui utilisé pour les caractères agronomiques.

▶ Corrélations et analyse en composantes principales

La figure 2 présente les distributions et corrélations 2 à 2 entre les moyennes ajustées des 10 caractères. Au sein des caractères agronomiques, la plus forte corrélation (0.77) est trouvée entre le PMG et la calibration, ce qui semble trivial. La teneur en protéines est corrélée négativement au rendement (-0.32), mais moins fortement que les valeurs rapportées pour le blé tendre (Oury *et al.,* 2003). De plus, pour l'orge brassicole, on ne cherche pas à augmenter systématiquement la teneur en protéines comme pour les blés panifiables, puisqu'un excès de protéines peut causer des problèmes lors de la filtration, comme illustré par la corrélation négative entre la teneur en protéines et le taux d'extraction (-0.35). Ainsi, c'est plutôt une stabilisation du taux de protéines qui est recherché, afin d'assurer une croissance optimale des levures, plutôt qu'un enrichissement continu.

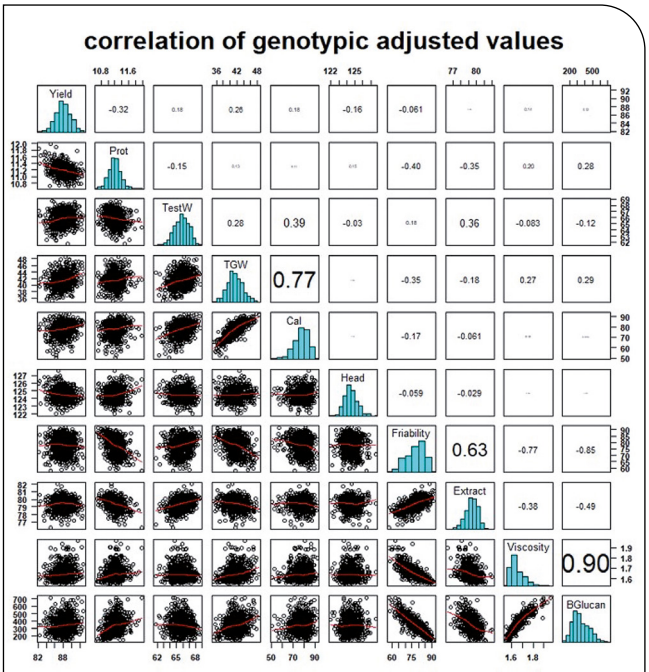


Figure 2

Au sein des caractères de valeur maltière, la plus forte corrélation est trouvée entre la viscosité et la teneur en β-Glucanes, et une autre entre la fiabilité et le taux d'extrait (0.64). Ces deux corrélations étaient attendues pour des raisons causales. La viscosité et la teneur en β-Glucanes sont négativement corrélées avec l'extrait, ce qui est favorable puisqu'on cherche à améliorer le taux d'extrait et à réduire la viscosité. Les caractères liés au malt sont faiblement corrélés aux caractères agronomiques, la plus forte en valeur absolue étant entre l'extrait et le taux de protéines (-0.42). Ces valeurs suggèrent que l'amélioration des caractères

agronomiques et de la valeur brassicole peut être obtenue de façon indépendante, et que les sélectionneurs d'orge ont la chance d'avoir entre leurs caractères cibles des corrélations majoritairement favorables.

Ces corrélations peuvent être illustrées par une analyse en composante principale. Le plan des axes 1-2 (Figure 3) montre clairement les deux groupes de variables de valeur du malt fortement corrélées entre elles et opposées sur l'axe 1, tandis que les caractères agronomiques sont principalement associés à l'axe 2, donc indépendants des traits brassicoles, en particulier les poids de 1000 grains et le calibrage, la teneur en protéines étant mal représentée dans ce plan, donc elle aussi indépendante des autres variables. La précocité d'épiaison, elle aussi mal représentée, n'est pas non plus corrélée aux caractères du malt.

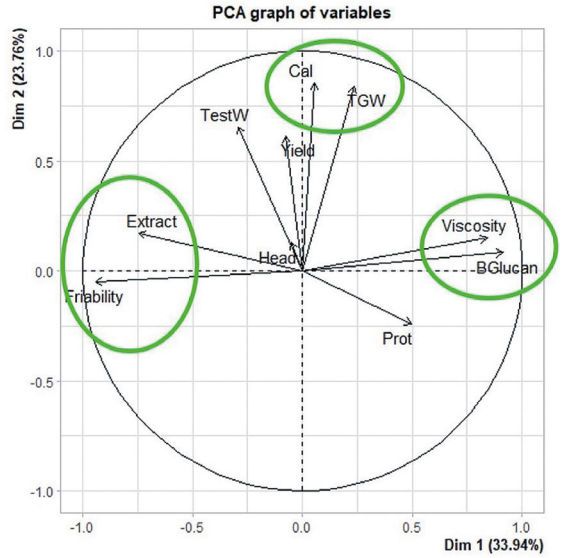


Figure 3

▶ Données moléculaires

La distribution des 24 101 marqueurs retenus après contrôle qualité est relativement homogène entre chromosomes, variant de 2 505 sur le chromosome 4H à 4 604 sur le chromosome 3H. La projection des 679 lignées et cultivars sur les axes 1 et 2 de l'analyse en coordonnées principales de la matrice de distance de Roger est présentée Figure 4.

Les nuages des lignées des deux sélectionneurs présentent à la fois une zone de recouvrement et une zone « privative ». Les cultivars sont davantage dispersés sur l'ensemble du plan, avec une plus forte densité dans la zone médiane qui correspond à la zone de superposition des deux breeders. Ceci peut s'expliquer par l'utilisation par les deux sélectionneurs de variétés inscrites comme parent de croisement, mais aussi par l'existence d'un germplasme propre à chacun, qui explique un début de divergence entre les deux sets de lignées avancées de sélection. Toutefois, le chevauchement des deux sélectionneurs semble assez large pour anticiper la possibilité d'une prédiction croisée, en utilisant le matériel d'un sélectionneur pour construire le modèle et celui de l'autre pour valider les prédictions. L'indice de différenciation Fst calculé entre les deux populations, 0.03, est en effet largement inférieur à celui rapporté pour des programmes de sélection de blé tendre entre différents états des USA, 0.09 à 0.15 (Sneller *et al* 2021).

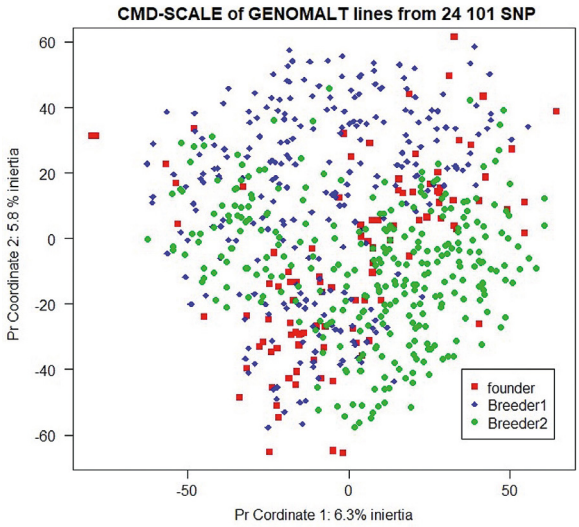


Figure 4

▶ Prédictions génomiques

Les résultats de valeurs prédictives obtenues par validation croisée sur les différents jeux de lignées sont présentés dans le tableau 2.

TRAIT	randomCV	BRE1 +FO.CV	BRE2 +FO.CV	FounderCV
Yield	0.446 / 0.022	0.451 / 0.028	0.395 / 0.025	0.489 / 0.056
Prot	0.517 / 0.016	0.659 / 0.020	0.282 / 0.036	0.387 / 0.095
TGW	0.667 /0.012	0.798 / 0.010	0.518/ 0.021	0.560 / 0.080
TestW	0.662 / 0.012	0.717 / 0.016	0.612 / 0.017	0.672 / 0.080
Cal	0.693 / 0.013	0.685 / 0.020	0.598 / 0.020	0.310 / 0.064
Head	0.522 / 0.022	0.533 / 0.028	0.371 / 0.034	0.097 / 0.110
Friability	0.805 / 0.009	0.824 / 0.009	0.775 / 0.008	0.749/ 0.030
Extract	0.654 / 0.009	0.778 / 0.014	0.635 / 0.014	0.802 / 0.030
Viscosity	0.690 / 0.011	0.729 / 0.016	0.662 / 0.012	0.694 / 0.036
β-Glucan	0.753 / 0.009	0.789 / 0.011	0.733 / 0.016	0.729 / 0.042

Tableau 2

La validation croisée au hasard utilisant l'ensemble des lignées (N = 679) donne des valeurs prédictives moyennes pour le rendement et la teneur en protéines (0.45-0.50), et bonnes à très bonnes pour les caractères de qualité liés au malt, toujours supérieures 0.65, et jusqu'à 0.80 pour la friabilité. Notons que la prise en compte d'un QTL majeur détecté par GWAS ($r^2 = 0.25$, résultat non montré) comme effet fixe n'améliore que très marginalement la valeur prédictive du modèle (de 0.806 à 0.814). Ces résultats sont très encourageants quant à la possibilité de trier précocement et à moindre coût un plus grand nombre de lignées candidates pour la valeur maltière, assurant ainsi un progrès génétique plus rapide pour ces caractères.

Les colonnes 2 et 3 montrent les valeurs prédictives que pourraient obtenir chaque sélectionneur en utilisant ses lignées propriétaires + les variétés inscrites, sans avoir

à partager ses données avec un autre sélectionneur. Comparé au dispositif complet de la colonne 1, la taille de la population d'entrainement serait dont plus petite (N = 359 et N = 410, respectivement), ce qui devrait conduire à de plus faibles valeurs prédictives, mais le phénotypage est issu d'un dispositif plus équilibré, ce qui devrait donner une meilleure répétabilité des données. Ces deux effets s'équilibrent sans doute l'un par l'autre, car les valeurs prédictives obtenues par chaque sélectionneur sont très proches de celles obtenues avec les données complètes. La colonne 4 présente les valeurs prédictives en validation croisée obtenues avec les seules variétés évaluées par les deux sélectionneurs (N = 95. Bien que ces valeurs soient plus variables que celles obtenues avec les données complètes (écart-type 2-4 fois plus grands), elles sont presque aussi élevées, en particulier pour les caractères du malt, ce qui est assez inattendu.

Pour déterminer si la valeur prédictive des sous-ensembles de lignées était déterminée par la taille de l'échantillon, nous avons procédé à des tirages au hasard parmi la population totale. La figure 5 présente les valeurs prédictives pour le rendement et la friabilité obtenu avec des populations d'entrainement tirées au hasard vs les populations de chaque sélectionneur ou les variétés inscrites.

Comme attendu par la théorie, la valeur prédictive des échantillons aléatoires décroit avec sa taille, tandis que la variabilité des prédictions augmente. En utilisant les lignées d'un seul sélectionneur et/ou les variétés, les résultats sont contrastés : pour les deux caractères, les prédictions obtenues avec les lignées de Breeder 2 + variétés est proche de celle obtenues avec l'échantillon aléatoire de même taille, tandis que la valeur prédictive de Breeder 1 + variétés et des variétés seules se situent bien au-dessus des valeurs obtenues avec les échantillons aléatoires. La valeur prédictive des variétés inscrites, malgré la faible taille de la population (N=95), est particulièrement élevée pour le rendement, elle est même supérieure à celle obtenue par validation croisée en utilisant la population totale (N=679).

Ces différences de valeur prédictive entre populations d'entrainement de même taille peuvent difficilement être attribuées à des différences d'apparentement moyens entre les populations d'entrainement et de validation. En effet, les coefficients de parentés estimés (par normalisation de la matrice K issue de A.mat) au sein de chaque sous-population ne sont pas très différents en moyenne (0.195, 0.200 and 0.201 pour breeder 2, breeder 1 et variétés, respectivement). Il n'y a pas de structure nette entre les groupes de lignées et aucun qui corresponde aux germplasms de chaque sélectionneur.

Et en effet, les héritabilités au sens large estimées à partir des 95 variétés évaluées par les deux sélectionneurs sont aussi élevées, voire supérieures (rendement) à celles estimées sur l'ensemble des lignées. De plus, les héritabilités estimées dans le matériel de Breeder2 sont toujours inférieures à celles estimées dans les lignées de Breeder 1, ce qui est cohérent avec les valeurs prédictives présentées au-dessus.

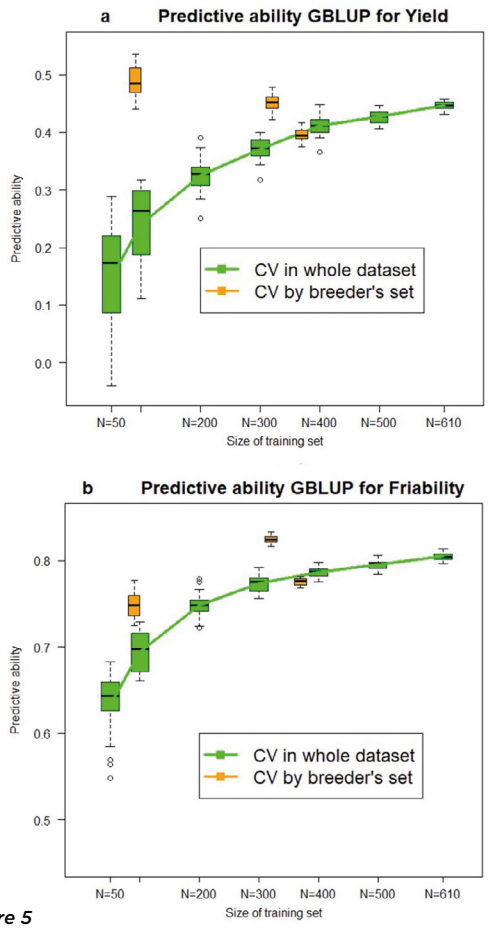
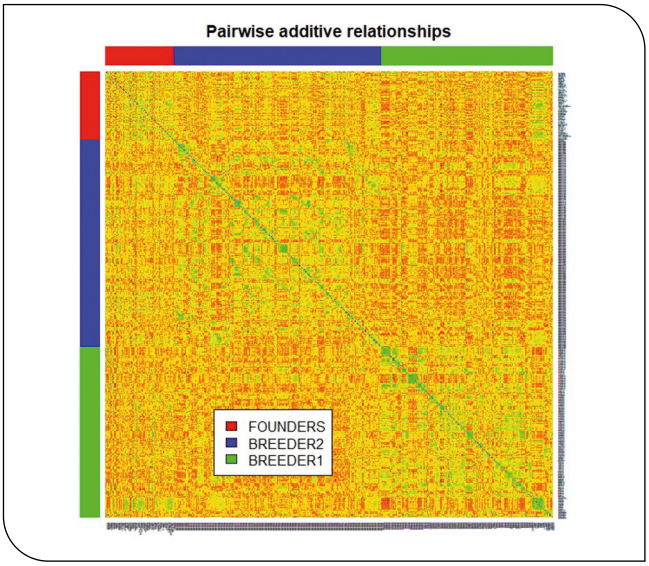
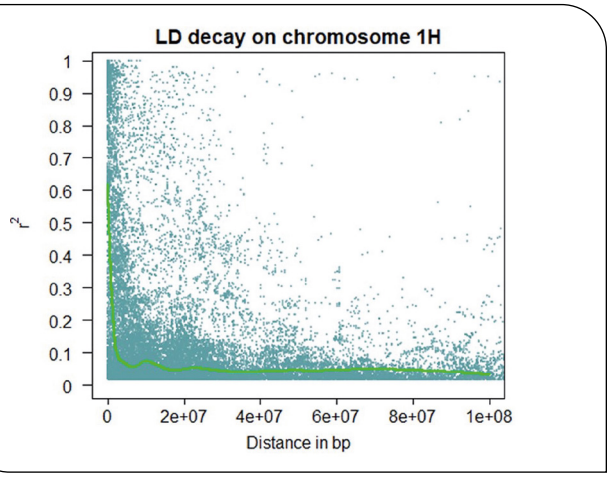


Figure 5



TRAIT / h^2	All	Breeder1	Breeder2	Founders
Yield	0.551	0.479	0.459	0.600
Protein	0.613	0.493	0.575	0.670
TGW	0.837	0.834	0.680	0.840
TestW	0.775	0.727	0.695	0.810
Calibration	0.853	0.786	0.768	0.806
Heading	0.652	0.575	0.580	0.450
Friability	0.895	0.850	0.840	0.879
Extract	0.753	0.706	0.680	0.777
Viscosity	0.769	0.748	0.624	0.759
β-Glucan	0.851	0.708	0.638	0.779

La figure 4 présente l'effet du nombre de marqueurs choisis au hasard sur la valeur prédictive du GBLUP pour les deux traits extrêmes : rendement et friabilité. Comme attendu, la valeur prédictive moyenne augmente et son écart-type diminue avec le nombre de marqueurs, jusqu'à un plateau obtenu avec seulement 2000 marqueurs. L'explication la plus vraisemblable est que l'étendue du déséquilibre de liaison est assez grande entre chacun des 2000 marqueurs et ses voisins pour être capable de capturer l'information de tous les QTL localisés dans l'intervalle. Pour tester cette hypothèse, la figure 5 présente la décroissance du DL avec la distance physique pour le chromosome 1H.



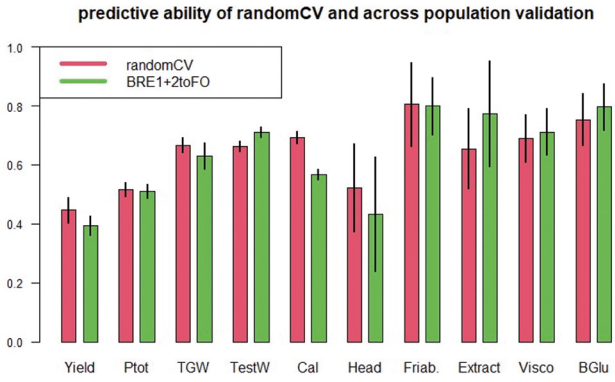
Bien que le DL semble décroitre très rapidement à l'échelle du chromosome entier (courbe verte), il reste supérieur à 0.4 sur une distance d'environ 2 Mb. Compte tenu de la taille du génome de l'orge 4,250 Mb, environ 2,100 markers (4,200/2) régulièrement espacés devraient être suffisant pour assurer une couverture du génome avec un DL minimum de 0.3. Cette valeur correspond à celle déterminée empiriquement pour la valeur prédictive qui plafonne avec M = 2000 marqueurs.

Le tableau 3 présente les valeurs prédictives du GBLUP obtenues entre populations, c'est-à-dire en utilisant un sous-ensemble prédéfini de lignées pour l'entrainement et un autre pour la validation. La taille des jeux d'entrainement est en gros décroissante de gauche à droite. Comme attendu les valeurs prédictives décroissent avec la taille du jeu d'entrainement, plus rapidement que les valeurs obtenues par validation croisée, particulièrement pour le rendement et la teneur en protéines. Toutefois elle demeure dans une gamme permettant une utilisation pratique pour les traits de valeur en malterie.

TRAIT	BRE1+2 to FO	BRE1 to FO	BRE2 to FO
Yield	0.394 / 0.075	0.254 / 0.101	0.248 / 0.107
Prot	0.509 / 0.097	0.575 / 0.086	0.250 / 0.095
TGW	0.630 / 0.071	0.595 / 0.064	0.547 / 0.096
TestW	0.710 / 0.049	0.562 / 0.083	0.556 / 0.066
Cal	0.568 / 0.068	0.529 / 0.071	0.472 / 0.101
Head	0.432 / 0.090	0.408 / 0.095	0.332 / 0.111
Friability	0.799 / 0.040	0.718 / 0.052	0.677 / 0.050
Extract	0.773 / 0.039	0.653 / 0.058	0.726 / 0.040
Viscosity	0.712 / 0.044	0.644 / 0.061	0.572 / 0.072
β-Glucan	0.796 / 0.040	0.745 / 0.048	0.706 /0.060

Tableau 3

Ainsi les valeurs en colonnes 1 sont très proches de celles de la colonne 1 du tableau 2, pour des tailles similaires du jeu d'entrainement (N = 612 en validation croisée au hasard 10 feuilles, N = 569 avec le subset BRE1+BRE2). La Figure 5 illustre les valeurs prédictives pour les 10 caractères obtenues par validation croisée au hasard ou par validation entre populations. Cette dernière donne des valeurs prédictives légèrement inférieures pour les caractères agronomiques, mais très proches pour les caractères liés au malt, et même supérieures pour l'extrait.



Pour explorer pourquoi les caractères liés au malt sont prédits de façon plus précise et plus robuste que les caractères agronomiques, nous avons testé d'autres modèles qui ne s'appuient pas sur le modèle génétique infinitésimal comme le GBLUP. En effet, LASSO et Bayes Cpi estiment les effets additifs d'un nombre limité de marqueurs, en autorisant un plus grand nombre à avoir des effets nuls, tandis que EGBLUP estiment également les effets des interactions entre paires de marqueurs. Les résultats sont présentés dans le tableau 4.

TRAIT	h^2	h	GBLUP	Bayes Cpi	LASSO	EGBLUP
Yield	0.551	0.742	0.446 / 0.022	0.443 / 0.026	0.338 / 0.030	0.463 / 0.016
Prot	0.613	0.783	0.517 / 0.016	0.514 / 0.013	0.482 / 0.020	0.513 / 0.012
TGW	0.837	0.915	0.667 / 0.012	0.669 / 0.015	0.627 / 0.011	0.671 /0.010
TestW	0.775	0.880	0.662 / 0.012	0.666 / 0.012	0.647 / 0.017	0.677 / 0.009
Cal	0.853	0.923	0.693 / 0.013	0.690 / 0.013	0.639 / 0.021	0.705 / 0.013
Head	0.652	0.807	0.522 / 0.022	0.519 / 0.018	0.511 / 0.019	0.518 / 0.017
Friability	0.895	0.946	0.805 / 0.009	0.806 / 0.006	0.789 / 0.008	0.805 / 0.011
Extract	0.753	0.868	0.654/ 0.009	0.658/ 0.009	0.650/ 0.010	0.669/ 0.009
Viscosity	0.769	0.876	0.690 / 0.011	0.697 / 0.007	0.657 / 0.015	0.700 / 0.009
β-Glucan	0.851	0.922	0.753 / 0.009	0.754 / 0.009	0.722 / 0.015	0.761 / 0.009

Tableau 4 : *héritabilités au sens large et sa racine carrée estimées dans le dispositive complet (2 années x 5 lieux) et les valeurs prédictives de 4 modèles statistiques (voir MM) obtenues par validation croisée dans la population totale de lignées (moyenne et écart-type de 50 tirages.*

Comme attendu, compte-tenu des limites du dispositif, le rendement et la teneur en protéines ont des valeurs modérées d'héritabilité. C'est également le cas pour la date d'épiaison, un trait généralement considéré comme très héritable chez les céréales. Ceci est sans doute dû à une faible gamme de variation dans le matériel étudié, constitué de lignées d'orge d'hiver à 6 rangs adaptées à l'Europe de l'ouest. Il convient de noter que la colonne 1 est l'héritabilité des moyennes parcellaires, quelquefois appelée répétabilité, qui dépend donc du dispositif, et dont la racine carrée est un proxi de la limite supérieure que peut atteindre la valeur prédictive d'un modèle génomique.

Les traits ayant les plus fortes héritabilités sont donc logiquement ceux qui présentent les plus fortes valeurs prédictives. Globalement, il y a très peu de différences entre les 4 modèles, avec LASSO qui montrent des valeurs légèrement inférieures, particulièrement pour les traits à faible héritabilité, rendement et teneur en protéines. Bayes Cpi donne des valeurs prédictives très proches de celles du GBLUP, la différence étant souvent sur la troisième décimale. En comparaison, EGBLUP, qui est supposé modéliser les interactions épistatiques, montre des valeurs prédictives supérieures au GBLUP qui ne modélise que les effets additifs, avec parfois une amélioration significative bien que faible, par exemple pour le rendement.

4 Discussion

Malgré un dispositif expérimental de taille relativement réduite et fortement déséquilibré (2 années, 5 lieux, mais seulement 2 ou 3 pour le matériel propriétaire de chaque sélectionneur) les données ont permis d'obtenir de bonnes valeurs de répétabilité pour la majorité des caractères. Ceci nous a permis d'utiliser les moyennes ajustées par LMM pour les prédictions génomiques.

Bien que la sélection assistée par marqueurs ait été proposée il y a plus de 20 ans (e.g. Han *et al.,* 1997), seuls quelques caractères contrôlés par un petit nombre de QTL à effet fort comme le pouvoir diastasique ou la teneur en β-glucanes content étaient concernés (Li *et al.,* 2009; Fang *et al.,* 2019). Le développement d'outils de marquage dans haute densité a été développé depuis une dizaine d'années et a ouvert la voie à l'utilisation des approches modernes de génétique quantitative comme la sélection génomique. Du fait de cette disponibilité récente des outils et de son importance secondaire, les publications sur la sélection génomique de l'orge sont encore plus récentes.

Etant donné leur coût financier et en temps, les caractères de qualité du malt ont fait l'objet des premières études de sélection génomique. Un des premiers rapports fut Schmidt *et al.* (2015), qui ont exploré les possibilités de la sélection génomique pour la valeur maltière dans deux programmes de sélection, orge d'hiver et orge de printemps. Les auteurs ont étudié davantage de traits que nous, incluant les activités enzymatiques (α-amylase and β-glucanase), mais nos 4 traits liés au malt étaient également inclus. Ils ont utilisé la puce Illumina-9K SNP et retenu 4359 marqueurs pour l'orge d'hiver, qui leur ont donné des valeurs prédictives variant de 0.625 (Extrait) à 0.798 (teneur en β-glucane), donc des valeurs très proches de celles de notre étude, malgré une taille de population d'entrainement très faible (N = 102). Il est notable que dans l'étude de Schmidt *et al.* (2015), les valeurs prédictives étaient plus faibles chez l'orge de printemps que chez l'orge d'hiver, de 0.16 en moyenne, malgré des tailles de populations plus grandes. Les auteurs expliquent ce résultat par une plus grande homogénéité génétique des orges d'hiver, comme nous l'avons-nous même rapporté.

Nielsen *et al.* (2016) ont rapporté des valeurs prédictives du G-BLUP dans une population peu structurée d'orges de printemps avec 3540 marqueurs SNP. Par la « leave-one-out » (un cas extrême de validation croisée), ils ont obtenu des valeurs prédictives variant de 0.40 (teneur en protéine) à 0.68 (poids du grain),et même 0.83 pour la teneur en ergostérol, un trait que nous n'avons pas mesuré. Comme dans notre étude, le "leave set out" (validation inter-populations) donne des valeurs prédictives plus faibles allant de 0.31 (protéines) à 0.52 (PMG), et 0.72 pour l'ergostérol.

Il est généralement reconnu qu'augmenter la taille de la population d'entrainement améliore la valeur prédictive des modèles génomique, comme le prédisent la théorie (e.g. Daetwyler *et al.* 2008; Goddard 2009) ou des études par simulation (e.g. Iwata & Jannink, 2011). Nos résultats sont conformes aux attentes théoriques lorsque la population d'entrainement est tirée aléatoirement dans l'ensemble des lignées. De tels résultats ont été rapportés par Nielsen *et al.* (2016). Néanmoins cette relation est loin d'être une règle d'or. Par exemple, Edwards *et al.* (2019), ont montré que, pour une taille fixée, il est préférable d'augmenter le nombre de croisements (descendances) plutôt que la taille des descendances de chaque croisement. Ceci pourrait expliquer pourquoi la petite population de 95 variétés inscrites donne une valeur prédictive plus grande qu'attendue. Il est en effet probable que le nombre de croisements à l'origine de ces 95 variétés du catalogue est plus grand que celui dont dérivent les lignées avancées d'un seul sélectionneur ?

L'avantage de regrouper les lignées avancées de deux sélectionneurs pour constituer une plus grande population de référence ne conduit pas, dans notre étude, à de meilleures valeurs prédictives, du moins en validation croisée, en particulier pour les caractères de valeur maltière. Un résultat similaire avait déjà été rapporté par Lorenz & Smith (2015). En utilisant des lignées d'orge de deux programmes universitaires aux USA (MN and ND), ils montraient qu'ajouter des lignées génétiquement distantes provenant d'un autre programme à la population d'entrainement n'améliorait pas, voire diminuait la valeur prédictive des modèles. Toutefois, le matériel de sélection de ces deux programmes était nettement plus divergent génétiquement que ne le sont les matériels des deux sélectionneurs de notre étude. Une très nette structuration est visible dans leur Figure 1 (heatmap), ce qui n'est pas le cas dans notre matériel.

Bien que l'échelle pour les relations génétiques additives soient différente de la nôtre, l'apparentement moyen entre les deux programmes était significativement inférieur à l'apparentement moyen au sein de chaque programme,

ce qui à nouveau n'est pas le cas de notre matériel, qui apparait génétiquement beaucoup plus homogène. Toutefois, quand les modèles sont validés sur un set de validation indépendant (variétés), la valeur prédictive obtenue (tableau 3), les valeurs prédictives obtenues en assemblant les lignées des deux sélectionneurs dans la population d'entrainement sont supérieures à celle obtenues avec les lignées d'un seul sélectionneur. D'un point de vue pratique, cela signifie que les programmes d'amélioration ne présentent pas une forte divergence génétique, on peut gagner à rassembler les données pour constituer une plus grande population de référence. Nos résultats montrent également, comme Nielsen *et al.* (2016), qu'un nombre aussi faible que 200 marqueurs est suffisant pour obtenir les meilleures prédictions, ce qui est dû à l'étendue du DL dans le matériel d'orge d'hiver à 6 rangs. Ceci reflète une taille effective assez limitée dans ce germplasm, également visible dans les coefficients d'apparentements relativement élevés, que ce soit au sein ou entre les matériels des deux sélectionneurs.

Comme déjà et souvent rapporté (e.g. Heslot *et al.,* 2012), nous n'avons pas trouvé de grandes différences dans la valeur prédictive des différents modèles. Souvent le bon vieux GBLUP, qui remplace les apparentements « pedigree » par des apparentements « génomiques », apparait comme la meilleure méthode. Des résultats similaires ont été rapporté par Wang *et al.* (2015). En utilisant des données simulées, ils montrent que Bayes Cpi a une meilleure valeur prédictive uniquement dans les scénarios à 20 QTL. Pour les autres architectures génétiques, qu'elles soient simulées ou réelles, le RR-BLUP (équivalent au GVLUP) dépasse légèrement les autres méthodes. Comme dans notre précédente étude sur le blé tendre (Charmet *et al.,* 2020), le modèle qui est supposé capturer les interactions additive-additive entre marqueurs présente effectivement des valeurs prédictives légèrement supérieures à celles du GBLUP.

Dans cette étude, nous en sommes restés à des méthodes de prédiction génomique mono caractère. Bien qu'une étude récente (Bhatta *et al.,* 2020) présente des valeurs prédictives significativement supérieures en utilisant des prédictions génomiques multi caractère, nous ne pensons pas que ça puisse être le cas avec nos données. En effet, les corrélations génétiques rapportées dans Bhatta *et al.* sont bien plus élevées que les nôtres. Les seuls caractères fortement corrélés sont PMG/calibrage, friabilité/extrait et viscosité/β-glucane. Or ces traits sont déjà très bien prédits (0.6-0.8), il est donc peu probable que des modèles multi caractère montrent une forte amélioration des valeurs prédictives.

Il est bon de noter que les corrélations génétiques trouvées dans ce matériel d'orge d'hiver à 6 rangs sont toutes favorables aux objectifs de sélection. En effet, friabilité et extrait, que l'on cherche à augmenter, sont corrélées négativement à la viscosité et β-glucanes, qui doivent être réduits. PMG et calibrage sont mécaniquement corrélés au rendement, et la teneur en protéines ne lui est pas fortement et négativement corrélées comme rapporté pour le blé tendre (e.g. Oury *et al.,* 2003). De plus, les sélectionneurs cherchent à stabiliser la teneur en protéines plutôt qu'à l'augmenter. Finalement, ces traits semblent indépendants de la précocité d'épiaison, ce qui permet d'envisager l'amélioration de la valeur brassicole aussi bien dans des variétés précoces que tardives afin de les adapter aux climats locaux, présents et futurs.

5 Conclusion

Cette étude, basée sur un matériel génétique représentatif de deux programmes appliqués d'amélioration variétale de l'orge d'hiver à 6 rangs, montre des résultats très encourageants quant aux possibilités d'utiliser la sélection génomique pour accélérer er rendre plus efficace les schémas de sélection, particulièrement pour les traits de valeur maltière, par exemple en permettant une sélection pour ces traits à des stades plus précoces, permettant ainsi d'augmenter l'intensité de sélection. Les ressources génétiques d'un seul sélectionneur et du matériel librement utilisable sont suffisantes pour obtenir de bonnes prédictions, mais rassembler des données de plusieurs breeders peut permettre quelques améliorations.

6 Remerciements

Les auteurs tiennent à remercier G Cresté et PM Leroux de SECOBRA Recherches (Maule, France), R Dupont, M Tison et G Touzy de RAGT 2N (Rodez France), S Schwebel & C. Colin et leur équipe de l'IFBM (Vandœuvre-lès-Nancy, France) pour la fourniture du matériel, la conduite des expérimentations et les analyses, incluant les tests de micro maltage, la sous-traitance du génotypage et le nettoyage des données.

Références bibliographiques

Badr, K. M, Sch, R., El Rabey, H., Effgen, S., Ibrahim, H.H., Pozzi, C., Rohde,,W., Salamini, F. (2000). On the Origin and Domestication History of Barley (Hordeum vulgare), Molecular Biology and Evolution, 17, 499–510, https://doi.org/10.1093/oxfordjournals.molbev.a026330

Bhatta, M., Gutierrez, L., Cammarota, L., Cardozo, F., Germán, S., Gómez-Guerrero, B., *et al.* (2020) Multi-trait Genomic Prediction Model Increased the Predictive Ability for Agronomic and Malting Quality Traits in Barley (Hordeum vulgare L.), G3 Genes|Genomes|Genetics, 10, 1113–1124, https://doi.org/10.1534/g3.119.400968

Bayer, M.M., Rapazote-Flores, P., Ganal, M., Hedley, P.E., Macaulay, M., Plieske, J., Ramsay, L., Russell, J., Shaw, P.D., Thomas, W., Waugh, R. (2017). Development and Evaluation of a Barley 50k iSelect SNP Array. Frontiers in Plant Science 8, 1792. doi: 10.3389/fpls.2017.01792

Bernardo, R., and Yu, J.M. (2007) Prospects for genomewide selection for quantitative traits in maize. Crop Science 47, 1082-1090.

Breiman L. (2001). Random Forests. Machine Learning 45, 5–32.

Charmet, G., Tran, L.G., Auzanneau, J., Rincent, R., Bouchet, S. (2020). BWGS: A R package for genomic selection and its application to a wheat breeding programme. PLoS ONE 15(4): e0222733. https://doi.org/10.1371/journal.pone.0222733

Crossa J., de los Campos G., Perez P., Gianola D., Burgueño J., Araus J.L., Makumbi D., Singh R.P., Dreisigacker S., Yan J, Arief V., Banziger M. and Braun H.J. (2010). Prediction of Genetic Values of Quantitative Traits in Plant Breeding Using Pedigree and Molecular Markers. Genetics 186, 713-724

Cullis, B. R., Smith, A.B., Coombes, N.E. (2006). On the designof early generation variety trials with correlated data. J. Agric. Biol. Environ. Stat. 11: 381–393. https://doi.org/10.1198/108571106X154443

Daetwyler, H.D., Villanueav, B., Wooliams, J.A. (2008). Accuracy of Predicting the genetic risk of disease using a genome-wide approach. PLos ONE 3(10)e3395

De los Campos, G., Gianola, D., Rosa, G.J.M., Weigel, K.A., Crossa,J. (2010). Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert spaces methods. Genetics Research 92, 295-308.

De los Campos, G., Hickey, J. M., Pong-Wong, R., Daetwyler, H. D., & Calus, M. P. L. (2013). Whole-Genome Regression and Prediction Methods Applied to Plant and Animal Breeding. Genetics, 193(2), 327–345. http://doi.org/10.1534/genetics.112.143313

Edwards, S.M., Buntjer, J.B., Jackson, R. *et al.* (2019). The effects of training population design on genomic prediction accuracy in wheat. Theor. Appl. Genet. 132, 1943–1952 (2019). https://doi.org/10.1007/s00122-019-03327-y

Endelman, J.B. (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. Plant Genome 4,250-255. doi: 10.3835/plantgenome2011.08.0024

Endelman, J.B., and Jannink, J.L. (2012). Shrinkage estimation of the realized relationship matrix. G3: Genes Genom Genet 2, 1405-1413. doi: 10.1534/g3.112.004259

Fang, Y., Zhang ,X., Xue, D. (2019). Genetic Analysis and Molecular Breeding Applications of Malting Quality QTLs in Barley. Front. Genet. 10, 352. doi: 10.3389/fgene.2019.00352

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

Goddard, M. (2009). Genomic selection: prediction of accuracy and maximization of long term response. Genetica 136(2), 245-257.

Habier, DRL Fernando RL, Kizilkaya K and Garrick DJ (2011) Extension of the bayesian alphabet for genomic selection. BMC Bioinformatics2011. 12:186.

Han, F., Romagosa, I., Ullrich, S. E., Jones, B. L., Hayes, P. M., and Wesenberg, D. M. (1997). Molecular marker-assisted selection for maltingquality traits in barley. Mol. Breed. 3, 427–437. doi: 10.1023/A:1009608312385

Haslemore, R. M., Slack, C.R., Brodrick,K.N. (1982). Assessment of malting quality of lines from a barley breeding programme, New Zealand Journal of Agricultural Research, 25:4, 497-502, DOI: 10.1080/00288233.1982.10425212

Heffner, E.L., Sorrells, M.E., Jannink, J.L. (2009). Genomic Selection for Crop Improvement. Crop Science 49, 1-12.

Heslot, N., Yang, H.P., Sorrells M.E., Jannink, J.L. (2012). Genomic Selection in Plant Breeding: A Comparison of Models. Crop Sci. 52, 146–160.

Iwata H., Jannink J. L. (2011) Accuracy of genomic selection prediction in barley breeding programs: a simulation study based on the real single nucleotide polymorphism data of barley breeding lines. Crop Sci 2011,51: 1915-1927.

Jannink J. L., Lorenz A. J., Iwata H. (2010). Genomic selection in plant breeding: from theory to practice. Briefings in Functional Genomics & Proteomics 9, 166-177.

Komatsuda, T., Pourkheirandish, M., He, C., Azhaguvel, P., Kanamori, H., Perovic,D., Stein, N., Graner, A., Wicker, T., Tagiri, A., Lundqvist, U., Fujimura, T., Matsuoka, M., Matsumoto,T., Masahiro Yano M. (2007). Six-rowed barley originated from a mutation in a homeodomain-leucine zipper I-class homeobox gene. Proceedings of the National Academy of Sciences 104 (4) 1424-1429, DOI: 10.1073/pnas.0608580104

Li C.D., Cakir M., Lance R. (2009) Genetic Improvement of Malting Quality through Conventional Breeding and Marker-assisted Selection. In: Zhang G., Li C. (eds) Genetics and Improvement of Barley Malt Quality. Advanced Topics in Science and Technology in China. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-01279-2_9

Lorenz, A., and Smith, K. P. (2015). Adding Genetically Distant Individuals to Training Populations Reduces Genomic Prediction Accuracy in Barley. Crop Sci. 55, 2567-2667 doi :10.2135/cropsci2014.12.0827

Meuwissen, T.H.E., Hayes, B., Goddard, M.E. (2001) Prediction of total genetic value using genome-wide dense marker maps, Genetics 157, 1819-1829.

Nielsen NH, Jahoor A, Jensen JD, Orabi J, Cericola F, Edriss V, *et al.* (2016) Genomic Prediction of Seed Quality Traits Using Advanced Barley Breeding Lines. PLoS ONE 11(10): e0164494. doi:10.1371/journal.pone.0164494

Oury FX, Berard P, Brancourt-Hulmel M, Depatureaux C, Doussinaults G, Galic N, Heumez E, Lecomte C, Pluchard P, Rolland B, Rousset M, Trottet M (2003) Yield and grain protein concentration in bread wheat : a review and a study of multi-annual data from a French breeding program. J. Genet. & Breeding 57 :59-68.

Park, T., Casella, G. (2008). The bayesian lasso. Journal of the American Statistical Association. 103, 681–686.

Poland, J., Endelman, J., Dawson, J., Rutkoski, J., Wu, S., Manes, Y., Dreisigacker, S., Crossa, J., Sánchez-Villeda, H., Sorrells, M. and Jannink, J.L. (2012). Genomic Selection in Wheat Breeding using Genotyping-by-Sequencing. The Plant Genome 5, 103-113, https://doi.org/10.3835/plantgenome2012.06.0006

Rodriguez-Álvarez, M.X., Boer, M.P., van Eeuwijk, F.A.,, Eilers, P.H. (2018). Correcting for spatial heterogeneity in plant breeding experiments with P-splines. Spatial Statistics 23, 52-71. https://doi.org/10.1016/j.spasta.2017.10.003.

Schaeffer, L. R. (2006) Strategy for applying genome-wide selection in dairy cattle. Journal of Animal Breeding and Genetics 123, 218-223 DOI: 10.1111/j.1439-0388.2006.00595.x

Schmidt, M., Kollers, S., Maasberg-Prelle, A. *et al.* (2016). Prediction of malting quality traits in barley based on genome-wide marker data to assess the potential of genomic selection. Theor Appl Genet 129, 203–213. https://doi.org/10.1007/s00122-015-2639-1

Sneller, C.; Ignacio, C.; Ward, B.; Rutkoski, J.; Mohammadi, M. (2021) Using Genomic Selection to Leverage Resources among Breeding Programs: Consortium-Based Breeding. Agronomy, 11, 1555. https://doi.org/10.3390/agronomy11081555

Wang, X., Yang, Z.F., Xu, C.W. (2015). A comparison of genomic selection methods for breeding value prediction. Sci. Bull., 60, 925-935. https://doi.org/10.1007/s11434-015-0791-2

Zohary, D., and M. Hopf. (1993) Domestication of plants in the Old World. The origin and spread of cultivated plants in West Asia, Europe and the Nile Valley. Clarendon Press, Oxford, England.

Genomalt : Prédiction génomique du rendement et de la qualité brassicole chez l'orge d'hiver à 6 rangs

Gilles CHARMET [1], Pierre PIN [2], Marc SCHMITT [3], Nathalie LEROY [4], Bruno CLAUSTRES [4], Christopher BURT [4], Amélie GENTY [2]

1 - INRAE- UCA UMR GDEC, 5 chemin de Beaulieu 63000 Clermont-Ferrand – France
2 - SECOBRA Recherches SAS, Centre de Bois Henry, 78580 Maule – France
3 - IFBM, 7 rue du Bois de la Champelle, F-54500 Vandoeuvre les Nancy -France
4 - RAGT 2N, Place du bourg, 12510 Druelle - France

La France est le 2ième exportateur d'orge brassicole dans le monde

1,97 million d'hectares d'orge en 2021, soit 21 % de la surface céréalière sur 120 000 exploitations. **10,4 M de tonnes** produits, dont 5,7 M exportées. Les orges de brasserie **représentent plus de 4 millions de tonnes** soit plus d'un tiers du total
La culture de **l'orge d'hiver 6 rangs brassicoles** est une **particularité française qui est grandement responsable de la compétitivité** de la filière Orge-Malt-Bière

 # 30 % **# <1 %** **# 70 %**

De l'orge au malt...

La malterie française compte 14 unités de production regroupées en 3 groupes faisant parti des 5 premiers mondiaux.
En 2020/2021, 1,6 million de tonnes d'orges de brasserie ont été transformées en 1,4 million de tonnes de malt. Depuis 1967, la malterie française est le 1er exportateur mondial de malt. La France exporte 80 % de sa production de malt soit 1,2 million de tonnes par an.

Du malt à la bière...

Des nombreuses brasseries ont été créées en France ces dernières années. Aujourd'hui, avec plus de 2 300 brasseries la France est le 1ier pays européen en nombre de brasseries. La brasserie française emploie 7 900 personnes.
La consommation française de bière est de 22 millions d'hectolitres en 2020, soit une consommation de 33 litres par habitant et par an. Cette consommation place les Français dans les plus faibles consommateurs de bière européens, juste avant l'Italie.

Les critères de qualité brassicole

Qualité brassicole de l'orge:
- Calibrage: pour assurer une germination rapide et homogène: % grains >2.5mm
- Protéines: suffisamment pour "nourrir" les levures mais pas trop (filtration): 9.5 à 11.5%
- beta-glucanes (fibres solubles): colmatage des filtres, trouble de la bière (mais recherchées pour nutrition humaine)

Qualité du malt:
- Taux d'extrait: rendement en malt / tonne d'orge
- Désagrégation: capacité de friabilité du grain de malt
- Pouvoir diastasique: activité enzymatique des alpha- et beta-amylases

⇨ Les tests de micromaltage sont long/coûteux/exigeants en grains
Intérêt des **prédictions génomique** en (pré)sélection

Matériel & Méthodes

Matériel végétal:
- 574 lignées avancées des programmes de SECOBRA (Breeder1) et RAGT (Breeder2)
- 105 variétés "fondatrices" inscrites au catalogue officiel

Phénotypage:
- parcelles observation sur 5 lieux, 2 années (2018-2019)
- Mesure du rendement grain, teneur protéine, PMG, PS, calibrage, épiaison et hauteur
- Micromaltage réalisé sur deux lieux par an
- Caractères brassicole mesurés: friabilité, extrait, viscosité du mout et teneur en beta-glucanes

Génotypage:
- 50k Illumina Infinium iSelect genotyping array disponible chez SGS TraitGenetics GmbH

Analyses:
- Ajustement spatial des données phénotypiques
- Analyse GxE
- QC sur données de marquage (24K marqueurs retenus)
- Diversité génétique (multidimensional scaling)
- Modèle de sélection génomique (GBLUP, Bayes Cpi, LASSO, EGBLUP)
- Validation croisées aléatoires et indépendantes

Résultats

- Le pourcentage de la somme des carrés suggère que l'effet du génotype est majeur pour les caractères brassicoles étudiés (Fig1).
- L'ACP des variables phénotypiques suggère des corrélations positives ou négatives entre caractères brassicoles (Fig2).
- L'analyse de diversité montre un large chevauchement du matériel des deux sélectionneurs, avec l'absence de structure apparente (Fig3), malgré un début de divergence génétique (Fst 0.03).
- Les aptitudes prédictives (*r*) des validations croisées réalisées sur les différents modèles varient de 0.4 à 0.8 (Fig4). Les caractères brassicoles sont très bien prédits avec des aptitudes de 0.77 en moyenne.
- Les validations croisées obtenues avec différents sous-échantillonnages de marqueurs suggèrent qu'un petit set de 1000-2000 marqueurs est suffisant (Fig5) pour construire des modèles prédictifs efficaces.

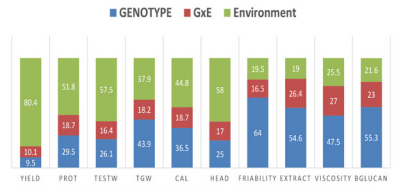


***Fig1**. Sommes des carrés issus des anova réalisées sur les différents caractères étudiés*

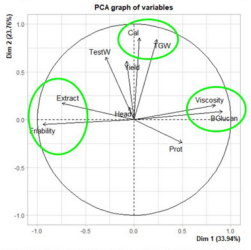


***Fig2**. Analyse en composantes principales des caractères phénotypiques*

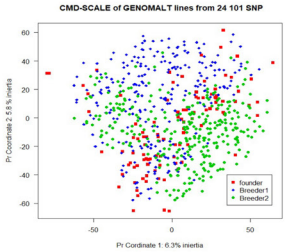


***Fig3**. Analyse en coordonnées principales du matériel végétal*

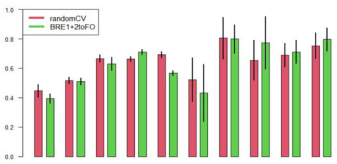


***Fig4**. Résultats des validations croisées des modèles prédictifs (GBLUP)*

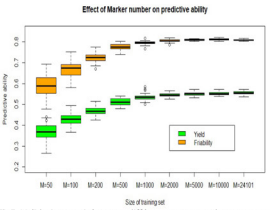


***Fig5**. Validations croisées avec différents sous-sets de marqueurs*

Conclusions

- Les caractères brassicoles observés sont peu sujets aux interactions GxE et sont corrélés favorablement pour atteindre les objectifs en sélection (cad fort extrait, forte friabilité, faible viscosité et faible teneur en béta-glucanes)
- Faible divergence entre les germplasmes des deux sélectionneurs (Fst 0.03)
- Aptitude prédictive pour les caractères brassicoles très élevée avec la méthode GBLUP et quelque soit la méthode de validation croisée
- Sous-échantillonnage de 1000-2000 marqueurs semble suffisant et optimal

⇨ Résultats extrêmement encourageants pour envisager un déploiement de la sélection génomique dans les programmes d'orge d'hiver 6 rangs pour assister la sélection des caractères brassicoles

INRAE: Gilles CHARMET – IFBM: Marc SCHMITT – RAGT 2n: Nathalie LEROY, Bruno CLAUSTRES, Christopher BURT – SECOBRA Recherches: Amélie GENTY, Pierre PIN