GIEC : sélection assistée par marqueurs et modèles du stade épi à 1 cm pour des blés tendres adaptés au changement climatique

Matthieu BOGARD^{1*}, Alison BENTLEY², Jacques LE GOUIS³, Jérôme AUZANEAU⁴, Simon GRIFFITHS⁵, Jérôme ENJALBERT⁶, Pascal GIRAUDEAU⁷, Patrice SENELLART⁸, Céline ZIMMERLI⁹, Céline DUQUE¹⁰, Philippe LEREBOUR¹¹, Sébastien CAIVEAU¹², Ellen GOUDEMAND¹³

- 1 Arvalis Institut du Végétal, 6 chemin de la côte vieille, 31450 Baziège, France
- 2 NIAB, Farm Office, Whitehouse Lane, Cambridge, CB3 0LF, United-Kingdom
- 3 INRAE, UMR GDEC, 5 chemin de Beaulieu, 63000 Clermont-Ferrand, France
- 4 Agri-Obtentions, chemin de la Petite Minière, 78280 Guyancourt, France
- 5 John Innes Centre, Norwich Research Park, Norwich, NR4 7UH, United-Kingdom
- 6 INRAE, UMR GQE, 12 route 128, 91190 Gif-sur-Yvette, France
- 7 Secobra recherche, Le Bois Henry, 78580 Maule, France
- 8 KWS-Momont, 7 rue de Martinval, 59246 Mons-en-Pévèle, France
- 9 BASF, 21 chemin de la Sauvegarde, 69130 Écully, France
- 10 Limagrain, Ferme de l'Etang, 77390 Verneuil-l'Étang, France
- 11 Unisigma, 2 Rue du Petit Sorri, 60480 Froissy, France
- 12 Syngenta France S.A.S., 2 avenue Gustave Eiffel, CS 40346, 28 008 Chartres Cedex
- 13 Florimond-Deprez, 3 rue Florimond Desprez, 59242 Cappelle-en-Pévèle, France

*Coordinateur : Matthieu BOGARD, m.bogard@arvalis.fr

1 Introduction

Il a été démontré que les changements climatiques ont un impact sur les productions agricoles et celle des céréales en particulier (Brisson *et al.*, 2010). L'augmentation des températures moyennes du fait des changements climatiques induit notamment une accélération de la croissance et du développement des plantes et l'on estime que la date d'épiaison du blé tendre a été avancée de 15 jours sur la période 1975-2020 (Figure 1).





Cette accélération de la vitesse de développement induit un raccourcissement du cycle de culture et une diminution de la durée d'interception du rayonnement ce qui est susceptible d'affecter le potentiel de rendement. Par ailleurs, du fait des changements climatiques et de l'augmentation des températures moyennes, il est attendu une augmentation de la fréquence et de l'intensité d'évènements « extrêmes », en particulier des épisodes de gel d'épis du fait de l'accélération du développement, d'épisodes caniculaires responsables de phénomènes d'échaudage des grains et d'épisodes de sécheresse estivales qui ont un impact négatif sur la productivité des blés (Trnka *et al.,* 2014).

Pour s'adapter à ces changements, les agriculteurs peuvent faire appel à deux leviers principaux : la modification de leurs pratiques agronomiques avec, par exemple, une modulation des dates de semis ou bien l'utilisation de variétés qui présentent une durée de cycle adaptée et/ou des facteurs génétiques de tolérance aux stress abiotiques. Une étude menée par ARVALIS Institut du végétal montre que l'avancement du stade épi à 1cm de 10 à 11 jours et l'avancement du stade épiaison de 5 à 6 jours permettraient de contrecarrer en partie l'occurrence accrue des épisodes de canicule (Gouache et al., 2012). Or, les hivers étant prédits plus doux à l'avenir, ces changements offrent l'opportunité d'avancer la date du stade épi à 1 cm sans augmenter les risques de gel, ce qui pourrait permettre de compenser l'effet de l'avancement du stade épiaison sur la durée de montaison. Cela semble envisageable du fait que le déterminisme génétique des stades « épi à 1 cm » et épiaison est partiellement indépendant.

Si la nécessité d'utiliser à l'avenir des variétés plus précoces afin d'esquiver les stress abiotiques de fin de cycle apparait clairement, il semble également nécessaire de tenir compte de la stabilité des variétés. En effet, d'autres travaux d'Arvalis ont permis de montrer que les variétés précoces à montaison réagissent d'avantage que les variétés tardives lorsque les conditions environnementales induisent un développement accéléré

1

des plantes (sous l'effet d'un automne-hiver doux par exemple). Ainsi, lors d'une année plus précoce de 10 jours en moyenne, les variétés les plus instables (qui sont généralement précoces) sont en avance de 18 jours en moyenne alors que les plus stables ne le sont que de quatre jours (données non publiées).

Dans le cadre du projet FSOV GIEC, des modèles basés sur des marqueurs moléculaires permettant de prédire la précocité chez le blé tendre ont été calibrés et une méthode permettant de définir des idéotypes de précocité à l'aide d'un modèle écophysiologique basé sur des marqueurs moléculaires a été proposée. Ces modèles peuvent potentiellement aider les sélectionneurs à identifier des lignées adaptées et stables en termes de développement, à évaluer si la variabilité génétique utilisée aujourd'hui en termes de précocité est suffisante pour faire face aux changements climatiques mais aussi aux agriculteurs afin d'identifier, parmi les variétés disponibles, celles les plus adaptées à leur contexte agro-climatique.

Matériel et méthode

Matériel végétal et expérimentations pour la calibration et la validation des modèles

Les modèles de prédiction issus du projet ont été calibrés sur des données générées dans le cadre des activités d'Arvalis Institut du Végétal. Ces données correspondent à des observations des dates des stades épi à 1 cm (GS30) et demi-épiaison (GS55) pour des variétés inscrites au catalogue officiel dans différents lieux en France sur la période 1997-2016 (Figure 2), essentiellement pour des dates de semis d'automne. La durée des phases semis - épi à 1 cm (Z30) et semis - épiaison (Z55) a ainsi pu être calculée pour chaque variété dans chaque essai en somme de degrés jours (base 0°C). Ensuite, la durée d'élongation des tiges (StEI) a été calculée comme la différence entre Z55 et Z30. Pour chaque caractère, seules les variétés présentent dans au moins 5 essais ont été sélectionnées pour la suite des analyses soit 438 variétés pour Z30, 718 pour Z55 et 404 pour StEl (Tableau 1).

Caractère	# variétés	# essais (min - max)
Z30	438	5-186
Z55	718	5-825
StEl	404	5-164
b.Z30	144	6-9
b.Z55	212	30-50
h StEl	143	6-9

Tableau 1. Nombre de variétés et nombres minimaux et maximaux d'essais par variétés utilisés pour le calcul des moyennes génotypiques ajustées (Z30, Z55, StEl) et des coefficients de plasticité correspondants (b.Z30, b.Z55, b.StEl) pour le jeu de données de calibration.

Ces modèles ont ensuite été validés sur des données indépendantes acquises dans le cadre du projet sur un panel de 149 lignées issues de 9 populations (Tableau 2) testées dans 20 essais en France et en Angleterre en 2018 et 2019 (Figure 3). Les stades épi à 1 cm (GS30) et demiépiaison (GS55) ont été mesurés pour chaque variété



dans chaque essai. De la même manière que pour le jeu de données de calibration, la durée des phases semis – épi à 1 cm (Z30) et semis – épiaison (Z55) a été calculée pour chaque variété dans chaque essai en somme de degrés jours (base 0°C) et la durée d'élongation des tiges (StEl) a été calculée comme la différence entre Z55 et Z30.

Population	Partenaire	Туре	Nombre de lignées
HEROS x BELEPI	Florimond	DH	18
STADIUM x SOLEHIO	KWS-Momont	DH	18
FALADO X SHERLOCK	Secobra	F3	14
AUCKLAND x DESCARTES	Limagrain	DH	17
ACCROC x GALIBIER	Agri- Obtentions	F3	13
UN7-3060 x RUBISKO	Unisigma	DH, F4/F5	17
BUENNO x TRIOMPH	syngenta	DH	17
MAGIC_inra	INRA GQE	SSD	17
MAGIC_niab	NIAB	SSD	18
TOTAL	149		

Tableau 2. Constitution du panel de validation testé dansle cadre du projet FSOV GIEC. Le panel était constitué de149 lignées issues de 9 populations.

Calcul de moyennes ajustées et de coefficients de plasticité

Une valeur moyenne par génotype ajustée des effets essais a été calculée à l'aide du modèle mixte suivant :

$$Y_{ij} = \mu + g_i + e_j + \varepsilon_{ij}$$

Avec Y_{ij} le phénotype du génotype *i* dans l'essai *j*, μ la moyenne générale, g_i l'effet du génotype *i*, e_i l'effet aléatoire de l'essai *j* et ε_{ij} l'erreur résiduelle.

Ensuite, pour chacun des caractères étudiés, un coefficient de plasticité a été calculé selon la méthode de Finlay and Wilkinson (1963). Les données de départ étant fortement déséquilibrées, elles ont d'abord été filtrées de manière à ne garder qu'un ensemble de variétés bien représentées dans un sous-ensemble d'essais. Ainsi, pour Z30, 9 essais ont été retenus avec 144 variétés présentes dans 6 à 9 essais (Tableau 1). Pour Z55, 50 essais ont été retenus avec 212 variétés présentes dans 30 à 50 essais (Tableau 1). Enfin, pour StEl, 9 essais ont été retenus avec 143 variétés présentes dans 6 à 9 essais (Tableau 1). Le modèle utilisé pour le calcul des coefficients de plasticité était le suivant :

$$Y_{ij} = \mu + g_i + e_j + b_i e_j + \varepsilon_{ij}$$

Avec Y_{ij} le phénotype du génotype *i* dans l'essai *j*, μ la moyenne générale, g_i l'effet du génotype *i*, e_j l'effet de l'essai *j*, b_i l'écart au modèle additif de la variété *i* dans l'essai *j* et ε_{ij} l'erreur résiduelle. Selon ce modèle, $1+b_i$ peut être défini comme la variation attendue de la précocité de la lignée i en fonction de la précocité moyenne dans les différents essais (norme de réaction génotypique). Ainsi, les génotypes avec $1+b_i$ proche de 1 se comportent comme la moyenne des variétés tandis que celles avec $1+b_i$ élevé réagissent de manière accentuée aux fluctuations environnementales (plasticité élevée) et celle avec $1+b_i$ faible sont plus stables (leur précocité est proche d'un environnement à l'autre). Ces coefficients de plasticité calculés pour Z30, Z55 et StEl sont dénommés b.Z30, b.Z55, b.StEl respectivement.

L'ensemble de ce matériel a été caractérisé sur le plan génotypique à l'aide des marqueurs SNP des puces TaBW420K ou TaBW35K (Paux *et al.,* 2022). Dans ce



dernier cas, une imputation des données des marqueurs TaBW420K a été réalisée à l'aide du logiciel Beagle v5.2. Le matériel végétal étudié a également été caractérisé à l'aide de marqueurs permettant de quantifier le nombre de copies des gènes Ppd-B1, Vrn-A1 et Vrn-A3 (Díaz *et al.*, 2012).

Modèle de génétique d'association

Une analyse de génétique d'association a été réalisée pour les différents caractères étudiés à l'aide du modèle :

$$Y_i = \mu + \sum_{1}^{p} \alpha_p + \beta_m X_{im} + G_i + \varepsilon_i$$

Avec Y_{ij} le phénotype de la variété *i*, μ la moyenne générale, α_p l'effet du $p^{ième}$ axe de l'analyse en coordonnées principales de la matrice d'apparentement, β_m l'effet du marqueur *m*, X_{im} le génotype de la lignée *i* au marqueur *m*, G_i un effet génotypique aléatoire de la lignée *i* et ε_{ij} l'erreur résiduelle. Ce modèle a été ajusté avec le package MM4LMM (Laporte and Mary-Huard, 2018) pour chaque marqueur. Un seuil de $-log_{10}$ (*p*) = 4 a été fixé arbitrairement afin de déclarer un marqueur comme étant significativement associé au caractère d'intérêt. L'assignation chromosomique et la position des marqueurs a été obtenue à partir du génome de référence RefSeqv1 (IWGSC *et al.*, 2018).

Modèles de prédiction basés sur des marqueurs génétiques

Afin de prédire les caractères phénotypiques étudiés, différents modèles statistiques basés sur des marqueurs moléculaires ont été calibrés :

1 - Un modèle de régression linéaire multiple (GWAS) avec pour variables explicatives des marqueurs identifiés par génétique d'association :

$$Y_i = \mu + \sum_{1}^{m} \beta_m X_{im} + \varepsilon_i$$

2 - Un modèle linéaire mixte de prédiction génomique (GBLUP) :

$$Y_i = \mu + G_i + \varepsilon_i$$



3 - Un modèle linéaire mixte avec pour variables explicatives des marqueurs identifiés par génétique d'association et un effet génotypique aléatoire (GSGWAS) :

$$Y_i = \mu + \sum_{1}^{m} \beta_m X_{im} + G_i + \varepsilon_i$$

Avec Y_i la moyenne génotypique pour Z30, Z55, StEl, ou le coefficient de plasticité $(1+b_i)$ du genotype *i*, μ la moyenne générale, β_m l'effet du marqueur m, X_{im} le génotype de la variété i au marqueur m, G_i un effet génotypique aléatoire du génotype i calculé sur la base d'une matrice d'apparentement et ε_{ii} l'erreur résiduelle. Ces modèles ont été ajustés avec le package BGLR (Pérez and de los Campos, 2013). Les marqueurs inclus dans les modèles GWAS et GSGWAS ont été sélectionnés sur la base des résultats d'études de génétique d'association et de l'analyse du déséquilibre de liaison afin de ne garder qu'un marqueur représentatif de chaque région associée. Les capacités prédictives de ces différents modèles ont été évaluées après validation croisée de type « k-folds » et sur la base de la qualité des prédictions obtenues pour les 149 lignées évaluées au cours du projet.

Modèle écophysiologique basé sur des marqueurs génétiques et identification d'idéotypes

Un modèle écophysiologique basé sur Weir et al. (1984) a été calibré à l'aide de marqueurs génétiques pour 408 variétés. Ce modèle de prédiction de la précocité est basé sur le concept d'accumulation de temps thermique modifié par la sensibilité génotypique à la photopériode et les besoins en vernalisation. Dans ce modèle, la durée des phases de développement est simulée de façon séquentielle et deux paramètres (GDDpv et GDDp) déterminent la durée des phases levée-épi à 1 cm et épi à 1 cm - épiaison. Ces deux paramètres ont été optimisés pour chacune des 408 variétés de manière à minimiser l'erreur quadratique moyenne entre stades observées et stades simulés. Pour chaque paramètre (GDDpv et GDDp), une analyse de génétique d'association et les modèles de prédictions basés sur des marqueurs génétiques décrits précédemment ont été calibrés puis validés sur les données du projet GIEC.

En parallèle, une analyse des risques de gel et de canicule a été réalisée pour 72 lieux en France à l'aide de données météorologiques pour la période 1986-2015. Pour chaque lieu × année, le dernier jour de gel où la température minimale journalière était inférieure à -5°C et le premier jour de canicule où la température maximale journalière était supérieure à 30°C ont été calculés. Pour chaque lieu, la période située entre le 90^{ème} percentile du premier jour de gel et le 30^{ème} percentile du premier jour de canicule a été définie et les lieux ont été regroupés en fonction des niveaux de risque de gel et de canicule par la méthode k-means.

Enfin, des idéotypes de précocité pour chaque lieu et trois dates de semis ont été définis par simulations à l'aide du modèle écophysiologique basé sur des marqueurs. Les paramètres GDDpv et GDDp pour 1019 variétés ont été prédits à partir des données de génotypage. Pour chaque lieu × variété × date de semis, les stades GS30 et GS65 (demi-floraison) ont été simulés. La date du stade GS65 a été calculée en ajoutant 124°C jours à la date de GS55 simulée (données non publiées). Un

coefficient d'évitement génotypique a été calculé comme le pourcentage d'années où les dates d'occurrence des stades GS30 et GS65 étaient situés entre le 90^{ème} percentile du premier jour de gel et le 30^{ème} percentile du premier jour de canicule. Des valeurs moyennes ont été calculées pour chaque variété × date de semis × groupe d'environnements et les idéotypes ont été définis comme les variétés (ou les combinaisons alléliques à différents loci) permettant de maximiser l'évitement des épisodes de gel et de canicule. Ce travail méthodologique sur l'identification d'idéotypes de précocité par simulations écophysiologiques basées sur des marqueurs génétiques a été publié dans Bogard *et al.*, (2021).

3 Résultats

Statistiques descriptives pour les moyennes génotypiques et les coefficients de plasticité

Les moyennes ajustées pour le jeu de données de calibration variaient de 843 à 1136°C jours pour Z30, de 1443 à 1861°C jours pour Z55 et de 520 à 720°C jours pour StEl entre les différents génotypes (Figure 4). Dans ce même jeu de données, les coefficients de plasticité pour Z30, Z55 et la durée d'élongation des tiges variaient respectivement de 0.75 à 1.33, 0.89 à 1.15 et de 0.18 à 1.92 (Figure 4).

Les moyennes ajustées pour le jeu de données de validation correspondant aux données phénotypiques acquises durant le projet sur 149 lignées testées dans 20 essais en France et en Angleterre pendant deux ans présentaient une gamme de variation similaire au jeu de données de calibration. Les moyennes génotypiques ajustées des effets essais variaient de 916 à 1166°C jours pour Z30, de 1478 à 1819°C jours pour Z55 et de 528 à 688°C jours pour StEl (Figure 4). Les coefficients de plasticité pour Z30, Z55 et la durée d'élongation des tiges variaient respectivement de 0.78 à 1.19, 0.73 à 1.22 et de 0.53 à 1.52 (Figure 4).



Figure 4 : (A) Distributions des moyennes génotypiques des durées des phases semis-épi à 1cm (Z30), semis-épiaison (Z55), de la durée d'élongation des tiges (StEI) et (B) des coefficients de plasticité correspondant (b.Z30, b.Z55, b.StEI) pour les données de calibration et de validation.



Impact des variations du nombre de copies de gènes de précocité

Le nombre de copies des gènes Ppd-B1, Vrn-A3 et Vrn-A1 pour les génotypes des jeux de données de calibration et de validation variaient respectivement de 2 à 8, 0 à 8 et de 2 à 6. L'analyse des corrélations entre le nombre de copies de ces gènes et les caractères Z30, Z55, StEL ainsi que les coefficients de plasticité associés ont permis de mettre en évidence des corrélations faibles mais significatives entre Z55 et le nombre de copies de Ppd-B1 et Vrn-A1, entre StEl et le nombre de copies de Ppd-B1, Vrn-A1 et Vrn-A3 et entre le coefficient de plasticité de Z55 et Vrn-A3 (Tableau 3). Les corrélations avec le nombre de copies de Ppd-B1 étaient toujours négatives (les variétés précoces possédaient en moyenne un nombre de copies de Ppd-B1 élevé) tandis que les corrélations avec le nombre de copies de Vrn-A1 ou Vrn-A3 étaient toujours positives (les variétés tardives possédaient en moyenne un nombre de copies de Vrn-A1 ou Vrn-A3 élevé).

Trait	Ppd-B1	Vrn-A1	Vrn-A3
Z30	-0.01	-0.09	0.03
Z55	-0.14	0.11	-0.01
StEl	-0.20	0.19	0.13
b.Z30	0.04	0.00	0.07
b.Z55	-0.11	0.08	0.12
b.StEl	-0.08	0.10	0.02

Tableau 3. Corrélations entre nombre de copies de Ppd-B1, Vrn-A1, Vrn-A3 et les durées des phases semis-épi à 1 cm (Z30), semis-épiaison (Z55), la durée d'élongation des tiges (StEl) et les coefficients de plasticité correspondant (b.Z30, b.Z55, b.StEl). Les corrélations significatives au seuil α =0.05 sont indiquées en gras.

Génétique d'association pour les moyennes génotypiques et les coefficients de plasticité

L'analyse de génétique d'association a permis de mettre en évidence l'importance de Ppd-D1 sur les variations phénotypiques pour Z30, Z55, StEl et b.Z55 (Figure 5A). Un marqueur pour ce gène a ensuite été utilisé comme cofacteur dans le modèle de génétique d'association afin d'identifier des loci supplémentaires. Un second QTL majeur a ainsi été identifié dans la région de Ppd-B1 pour Z30, Z55, StEl et b.Z55 (Figure 5A). D'autres QTL ont également été identifiés à proximité de gènes de précocité connus tels que Ft3-B1 pour b.Z55, Elf3-B1 pour Z30 tous deux situés sur le chromosome 1B, Elf3-D1 sur le 1D pour Z55, Ppd-A1 situé sur le 2A pour Z30, une zone à proximité de Lumini Dependens (LD) sur le 3A pour Z30, une zone proche de Vrn-A1 sur le 5A pour Z30 et Z55 et une zone proche de Vrn-B1 sur le 5B pour Z30 et le coefficient de plasticité correspondant (Figure 5A).

Un marqueur représentatif de chaque QTL a été sélectionné après analyse du déséquilibre de liaison entre marqueurs situés dans la même région. Le nombre de marqueurs sélectionnés était de 18 pour Z30, 17 pour Z55, 13 pour StEl, quatre pour b.Z30 et cinq pour b.Z55 et b.StEl. Seul le marqueur BGABW004965 dans la région de Ppd-D1 était commun aux différents caractères. Néanmoins, il y avait clairement d'autres régions communes à différents caractères et représentés par différents marqueurs. Par exemple, les marqueurs BWS2256 et BGABW020552 sont situés dans la région de Ppd-B1 et ont été sélectionnés pour Z55 et StEl. Les marqueurs sélectionnés ont ensuite été utilisés pour calibrer des modèles de prédiction.



Figure 5 : Résultats de génétique d'association pour (A) les moyennes génotypiques des durées des phases semis-épi à 1 cm (Z30), semisépiaison (Z55), de la durée d'élongation des tiges (StEI) et (B) des coefficients de plasticité correspondants (b.Z30, b.Z55, b.StEI) et (B) pour les paramètres du modèle écophysiologique (GDDpv et GDDp). Chaque point représente un marqueur positionné selon la carte physique du génome de référence RefSeqv1. L'ordonnée représente le degré de significativité des marqueurs. Les points en couleurs représentent les marqueurs considérés comme significatifs pour les caractères correspondants. L'échelle de l'axe des ordonnées varie en fonction du chromosome considéré.



Génétique d'association pour les moyennes génotypiques et les coefficients de plasticité

L'analyse de génétique d'association a permis de mettre en évidence l'importance de Ppd-D1 sur les variations phénotypiques pour Z30, Z55, StEl et b.Z55 (Figure 5A). Un marqueur pour ce gène a ensuite été utilisé comme cofacteur dans le modèle de génétique d'association afin d'identifier des loci supplémentaires. Un second QTL majeur a ainsi été identifié dans la région de Ppd-B1 pour Z30, Z55, StEl et b.Z55 (Figure 5A). D'autres QTL ont également été identifiés à proximité de gènes de précocité connus tels que Ft3-B1 pour b.Z55, Elf3-B1 pour Z30 tous deux situés sur le chromosome 1B, Elf3-D1 sur le 1D pour Z55, Ppd-A1 situé sur le 2A pour Z30, une zone à proximité de Lumini Dependens (LD) sur le 3A pour Z30, une zone proche de Vrn-A1 sur le 5A pour Z30 et Z55 et une zone proche de Vrn-B1 sur le 5B pour Z30 et le coefficient de plasticité correspondant (Figure 5A).

Un marqueur représentatif de chaque QTL a été sélectionné après analyse du déséquilibre de liaison entre marqueurs situés dans la même région. Le nombre de marqueurs sélectionnés était de 18 pour Z30, 17 pour Z55, 13 pour StEl, quatre pour b.Z30 et cinq pour b.Z55 et b.StEl. Seul le marqueur BGABW004965 dans la région de *Ppd-D1* était commun aux différents

caractères. Néanmoins, il y avait clairement d'autres régions communes à différents caractères et représentés par différents marqueurs. Par exemple, les marqueurs BWS2256 et BGABW020552 sont situés dans la région de Ppd-B1 et ont été sélectionnés pour Z55 et StEl. Les marqueurs sélectionnés ont ensuite été utilisés pour calibrer des modèles de prédiction.

Capacités prédictives des modèles basés sur des marqueurs pour les moyennes génotypiques et les coefficients de plasticité

Les capacités prédictives estimées par validation croisée des modèles statistiques basés sur des marqueurs apparaissaient élevées pour Z30, Z55 et StEl mais faibles pour tous les coefficients de plasticité. Le modèle le plus prédictif était généralement GSGWAS (Figure 6).

Les différents caractères étudiés ont ensuite été prédits avec le modèle GSGWAS pour les 149 lignées testées dans le cadre du projet et ont été comparées aux valeurs observées (Figure 7). La corrélation entre valeurs prédites et observées était élevée pour Z30, Z55 et StEl (respectivement 0.76, 0.86 et 0.69), modérée pour b.Z55 (0.45) et faible voire nulle pour b.Z30 et b.StEl (respectivement -0.09 et 0.11).



GIEC : sélection assistée par marqueurs et modèles du stade épi à 1 cm pour des blés tendres adaptés au changement climatique



Calibration et évaluation d'un modèle écophysiologique basé sur des marqueurs génétiques pour la précocité

L'optimisation des paramètres GDDpv et GDDp du modèle écophysiologique pour 408 variétés a été réalisée. Les médianes des erreurs de prédiction par génotype pour GS30 et GS55 avec les paramètres optimisés s'élevaient à 5.5 et 3.3 jours. L'analyse de génétique d'association a permis de mettre en évidence la forte association de Ppd-D1 avec les variations des paramètres GDDpv et GDDp dans le panel d'étude (Figure 5B). Un marqueur proche de Ppd-D1 a ensuite été utilisé comme cofacteur dans le modèle de génétique d'association et d'autres régions ont pu être identifiées, notamment à proximité de gènes de précocité connus. Ainsi, en plus de Ppd-D1, GDDpv est apparu associé à trois autres régions sur les chromosomes 2D, 3A (à 38Mbp de LD) et 5B à 46Mbp de Vrn-B1 (Figure 5B) tandis que GDDp est apparu associé à 4 autres régions dont une sur le chromosome 1A (à 5Mbp de Ft3-A1), deux régions sur le 1D dont une proche de Elf3-D1 (à 5Mbp) et quatre associations sur le 2B dans la région de Ppd-B1 (Figure 5B). D'autres loci proches de Hd-1 sur les chromosomes 6A et 6B présentaient un niveau de significativité proche du seuil choisi dans cette étude (Figure 5B).

Les modèles GWAS, GBLUP et GSGWAS ont été ajustés et leurs capacités prédictives ont été évaluées par validation croisée. Les meilleures capacités prédictives ont été obtenues avec le modèle GSGWAS. Ce modèle a été utilisé afin de prédire les paramètres GDDpv et GDDp pour chacune des 149 lignées évaluées dans le cadre du projet FSOV GIEC. Les stades GS30 et GS55 de ces lignées ont ensuite été simulés pour les 20 essais du projet et comparés aux stades observés. Les corrélations intra-essais entre valeurs prédites et observées pour GS30 et GS55 variaient de 0.37 à 0.71 (médiane 0.60) et de 0.74 à 0.90 (médiane 0.86) (Figure 8A). Les médianes des erreurs de prédiction par génotype pour GS30 et GS55 avec les paramètres prédits sur la base des données de génotypage pour ce jeu de données indépendant s'élevaient à 8.5 et 4.2 jours respectivement (Figure 8B).





Analyse des risques de gel et de canicule

L'analyse des risques de gel et canicule dans les régions productrices de blé tendre en France a permis de mettre en évidence des situations contrastées. Ainsi, Le 90^{ème} percentile du dernier jour de gel et le 30^{ème} percentile du premier jour de canicule variaient du 30^{ème} au 100^{ème} jour calendaire (Figure 9A) et du 136^{ème} au 182^{ème} jour calendaire (Figure 9B). Les différentes régions ont ensuite été classées en trois groupes avec des niveaux de risques contrastés (Figure 9C). Le groupe 1 rassemblait 27 lieux avec des niveaux de risques gel et canicule faibles (dernier jour de gel variant du 31^{ème} au 68^{ème} jour de l'année et premier jour de canicule variant du 158^{ème} au 182^{ème} jour de l'année ; Figure 9C). Le groupe 2 (24 lieux) présentait un risque de gel modéré et un risque de canicule élevé (dernier jour de gel variant du 45^{ème} au 75^{ème} jour de l'année et premier jour de canicule variant du 136^{ème} au 160^{ème} jour de l'année ; Figure 9C). Enfin, le groupe 3 rassemblant 21 lieux était caractérisé par un risque de gel élevé et un risque de canicule modéré (dernier jour de gel variant du 72^{ème} au 100^{ème} jour de l'année et premier jour de canicule variant du 149^{ème} au 177^{ème} jour de l'année ; Figure 9C).



Figure 9: Analyse des risques (A) de gel et (B) de canicule pour 72 lieux de référence. Pour chaque lieu, une analyse basée sur 30 années de données météorologiques (période 1986-2015) a permis de calculer le 90^{eme} percentile du dernier jour de gel (température minimum < -5°C) et le 30^{eme} percentile du premier jour de canicule (température maximum > 30°C) exprimés en jour de l'année. (C) L'ensemble des lieux ont ensuite été classés selon 3 groupes à l'aide de la méthode des k-means. Les régions en blanc n'ont pas été considérées dans l'analyse.



Identification d'idéotypes par simulation pour maximiser l'échappement des stress gel et canicule

Les stades GS30 et GS65 de 1079 variétés pour trois dates de semis ont été simulés dans le réseau de 72 lieux à l'aide des données climatiques historiques (période 1986-2015). Les fréquences d'échappement génotypiques ont ensuite été calculées pour chaque lieu × date de semis puis des valeurs moyennes ont été obtenues pour chacun des groupes d'environnements. Les résultats indiquent qu'il est possible d'atteindre la fréquence d'échappement maximale quelle que soit la date de semis pour les lieux du groupe 1 (Figure 10A). Dans les lieux des groupes 2 et 3, l'utilisation de variétés précoces et d'une date de semis tardive semble avantageuse puisqu'elle offre une plus grande flexibilité dans le choix variétal. Dans ces régions, il apparait nécessaire d'optimiser le couple variété × date de semis afin d'atteindre la fréquence d'échappement maximale (Figures 10B et 10C).



Figure 10 : Moyenne des fréquences d'échappement génotypiques pour 1019 variétés × 3 dates de semis × trois groupes de lieux. (A) Les lieux du groupe 1 sont caractérisés par des risques gel et canicule faibles, (B) le groupe 2 par un risque de gel modéré et un risque de canicule élevé et (C) le groupe 3 par un risque de gel élevé et un risque de canicule modéré. Des simulations des stades épi à 1cm (GS30) et floraison (GS65) à l'aide d'un modèle écophysiologique basé sur des marqueurs génétiques ont été réalisées pour chaque variété représentant une combinaison des paramètres GDDpv et GDDp du modèle dans 72 lieux pour 30 années climatiques (période 1986-2015) et trois dates de semis (1^{er} octobre, 30 octobre et 30 novembre). La fréquence d'échappement génotypique a été calculée pour chaque variété × lieu × date de semis comme la fréquence où les stades GS30 et GS65 étaient simulés entre le 90^{ème} percentile du dernier jour de gel et le 30^{ème} percentile du premier jour de canicule. Une valeur moyenne a été calculée pour chaque variété × groupe de lieux et représentée à l'aide d'une échelle de couleur allant du rouge (fréquence d'échappement égale à zéro) au vert (fréquence d'échappement égale à un).

4 Discussion

Identifier le déterminisme génétique et être en mesure de prédire la phénologie revêt une importance cruciale dans la sélection et la production des blés. En effet, du point de vue des sélectionneurs, il est capital de pouvoir proposer une variabilité génétique suffisante afin de couvrir les besoins du marché avec des variétés adaptées aux différents contexte culturaux rencontrés (température, photopériode, dates de semis). De même il est important de pouvoir anticiper si cette variabilité génétique sera suffisante pour faire face aux changements climatiques qui vont induire des effets importants sur la précocité mais aussi sur les dates d'occurrence, la fréquence et l'intensité des stress abiotiques de fin de cycle notamment. Du côté des agriculteurs, cette capacité à prédire la phénologie est également cruciale. En effet, les outils d'aide à la décision utilisés par les agriculteurs pour optimiser différents aspects de la culture des blés (choix variétal, fertilisation, protection des cultures) reposent largement sur des modèles de prédiction de la précocité.

Les analyses de génétiques d'association réalisées dans le cadre du projet ont mis en évidence le rôle majeur des gènes de sensibilité à la photopériode du groupe 2 (Ppd-D1, Ppd-B1 et dans une moindre mesure Ppd-A1) dans le matériel inscrit en France. Ces régions étaient associées à de nombreux caractères, que ce soit à la précocité moyenne ou aux paramètres du modèle écophysiologique. D'autres études ont pu montrer l'importance de ces régions sur le déterminisme de la précocité chez les céréales (Turner et al., 2005 ; Beales et al., 2007 ; Le Gouis et al., 2012). D'autres régions du génome identifiées par génétique d'association ont pu être reliées à des gènes de précocité connus tels que Ft3-A1, Elf3-B1 et Elf3-D1 sur les chromosomes du groupe 1 (Bonnin et al., 2008; Alvarez et al., 2016; Wang et al., 2016), LD sur le chromosome 3B (Rousset et al., 2011), ou à proximité des gènes Vrn-A1 et Vrn-B1 sur le groupe 5 (Rousset et al., 2011). Dans ces derniers cas, même si ces gènes sont connus pour fortement structurer la variabilité au niveau mondial (en séparant les types hiver/printemps), le panel étudié étant exclusivement de type hiver et les degrés d'association relativement modérés, ces associations pourraient s'expliquer par la présence d'autres allèles différenciant les types hivers entre eux (Rousset et al., 2011). Par ailleurs, l'analyse des corrélations entre durées de phases et polymorphisme de nombre de copies pour les gènes Ppd-B1, Vrn-A1 et Vrn-A3 ont permis de mettre en évidence des effets faibles mais significatifs sur différents caractères liés à la précocité et à sa plasticité, ce qui a pu être démontré par ailleurs (Díaz et al., 2012 ; Würschum et al., 2015, 2019).

Les travaux réalisés au cours du projet ont permis d'une part de travailler sur la précocité moyenne des variétés mais aussi sur la stabilité de la précocité à travers l'étude de coefficients de plasticité. Néanmoins, les résultats obtenus n'ont pas permis d'aboutir à des capacités prédictives satisfaisantes (sauf peut-être pour b.Z55). Cette incapacité à produire des modèles pour ces caractères de stabilité peut s'expliquer en partie du fait d'une quantité de données plus faible en nombre de variétés étudiées et en nombre d'environnements par variété mais aussi parce que le jeu de données sélectionné était encore déséquilibré (toutes les variétés n'étaient pas présentes dans tous les environnements). Or, compte tenu de la méthode de calcul des coefficients de plasticité, cela est susceptible d'affecter la qualité d'estimation des coefficients.

En plus des travaux effectués sur la précocité moyenne et les coefficients de plasticité, ce projet a permis d'aborder l'identification d'idéotypes de précocité adaptés à différents lieux × dates de semis afin de maximiser l'échappement

GIEC : sélection assistée par marqueurs et modèles du stade épi à 1 cm pour des blés tendres adaptés au changement climatique



aux stress abiotiques. Depuis que le concept d'idéotype a été proposé par Donald, (1968), les approches développées pour parvenir à des idéotypes adaptés à différents types d'environnement ont évoluées et impliquent largement des modèles écophysiologiques (Semenov and Stratonovitch, 2013 ; Semenov et al., 2014; Martre et al., 2015; Rötter et al., 2015). Ainsi, la recherche dans ce domaine est passé de l'expertise sur les caractères morpho-physiologiques désirables dans un environnement donné à la simulation des performances dans des réseaux d'essais virtuels et à l'identification des combinaisons de paramètres ou des combinaisons d'allèles les plus favorables pour un type d'environnement donné. L'approche suivie dans le cadre du projet était largement inspirée de Zheng et al. (2012). Néanmoins, les capacités prédictives du modèles écophysiologique basé sur des marqueurs semblent clairement supérieures à ce qui a pu être publié à ce jour pour la France (Bogard et al., 2014, 2015). Par ailleurs, la méthode d'identification d'idéotype qui repose sur le calcul d'une fréquence d'évitement génotypique est nouvelle et offre une certaine flexibilité. Ainsi, ces travaux à visée méthodologique devraient pouvoir être étendus à d'autres séries climatiques, notamment des séries futures afin d'évaluer si la variabilité génétique actuelle est suffisante pour faire face aux évolutions induites par le changement

Références bibliographiques

Alvarez MA, Tranquilli G, Lewis S, Kippes N, Dubcovsky J. 2016. Genetic and physical mapping of the earliness per se locus Eps-Am1 in *Triticum monococcum* identifies EARLY FLOWERING 3 (ELF3) as a candidate gene. Functional & Integrative Genomics 16, 365–382.

Beales J, Turner A, Griffiths S, Snape JW, Laurie DA. 2007. A Pseudo-Response Regulator is misexpressed in the photoperiod insensitive Ppd-D1a mutant of wheat (*Triticum aestivum* L.). Theoretical and Applied Genetics **115**, 721-733.

Bogard M, Hourcade D, Piquemal B, Gouache D, Deswartes J-C, Throude M, Cohan J-P. 2021. Marker-based crop model-assisted ideotype design to improve avoidance of abiotic stress in bread wheat. Journal of Experimental Botany **72**, 1085–1103.

Bogard M, Pierre J-B, Huguenin-Bizot B, Hourcade D, Paux E, Le Bris X, Gouache D. 2015. A simple approach to predict growth stages in winter wheat (*Triticum aestivum* L.) combining prediction of a crop model and marker based prediction of the deviation to a reference cultivar: A case study in France. European Journal of Agronomy **68**, 57-68.

Bogard M, Ravel C, Paux E, Bordes J, Balfourier F, Chapman SC, Le Gouis J, Allard V. 2014. Predictions of heading date in bread wheat (*Triticum aestivum* L.) using QTL-based parameters of an ecophysiological model. Journal of Experimental Botany **65**, 5849–5865.

Bonnin I, Rousset M, Madur D, Sourdille P, Dupuits C, Brunel D, Goldringer I. 2008. FT genome A and D polymorphisms are associated with the variation of earliness components in hexaploid wheat. Theoretical and Applied Genetics **116**, 383–394.

Brisson N, Gate P, Gouache D, Charmet G, Oury F-X, Huard F. 2010. Why are wheat yields stagnating in Europe? A comprehensive data analysis for France. Field Crops Research 119, 201-212. climatique, mais également à d'autres stress abiotiques comme les dégâts de gel sur épi qui sont susceptibles de devenir plus fréquent à l'avenir du fait d'un développement accéléré.

5 Conclusions

Le projet FSOV GIEC a ainsi permis de produire des modèles avec des capacités prédictives satisfaisantes pour la durée des phases semis-épi à 1cm (Z30), semisépiaison (Z55) et pour la durée d'élongation des tiges (StEl). En revanche, les modèles pour les coefficients de plasticité correspondants présentaient des capacités prédictives moindre, en particulier pour Z30 et StEl. Le travail réalisé dans le cadre de ce projet a également permis de calibrer un modèle écophysiologique à l'aide de marqueurs génétiques avec de bonnes capacités prédictives, en particulier pour prédire le stade épiaison. Enfin, une méthode d'identification d'idéotypes de précocité visant à maximiser l'échappement aux stress abiotiques (gel et canicule) a été proposée. Cette méthode présente une certaine flexibilité et pourrait être appliquée à d'autres stress abiotiques et pour des séries climatiques futures.

Díaz A, Zikhali M, Turner AS, Isaac P, Laurie DA. 2012. Copy number variation affecting the photoperiod-B1 and vernalization-A1 genes is associated with altered flowering time in wheat (*Triticum aestivum*). PLOS ONE **7**, e33234.

Donald CM. 1968. The breeding of crop ideotypes. Euphytica **17**, 385-403.

Finlay KW, Wilkinson GN. 1963. The analysis of adaptation in a plant-breeding programme. Australian Journal of Agricultural Research **14**, 742–754.

Gouache D, Le Bris X, Bogard M, Deudon O, Pagé C, Gate P. 2012. Evaluating agronomic adaptation options to increasing heat stress under climate change during wheat grain filling in France. European Journal of Agronomy **39**, 62–70.

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.

Laporte F, Mary-Huard T. 2018. MM4LMM: inference of linear mixed models through MM algorithm. R package version 1.0.5.

Le Gouis J, Bordes J, Ravel C, *et al.* 2012. Genome-wide association analysis to identify chromosomal regions determining components of earliness in wheat. Theoretical and Applied Genetics **124**, 597-611.

Martre P, Quilot-Turion B, Luquet D, Memmah M-MO-S, Chenu K, Debaeke P. 2015. Chapter 14 - Model-assisted phenotyping and ideotype design. Editors: Sadras VO, Calderini DF. Crop Physiology (Second Edition). San Diego: Academic Press, 349–373.

Paux, E.; The Breedwheat Consortium; Lafarge, S.; Balfourier, F.; Derory, J.; Charmet, G.; Alaux, M.; Perchet, G.; Bondoux, M.; Baret, F.; *et al.* Breeding for economically and environmentally sustainable wheat varieties: an integrated approach from genomics to selection. *Biology* **2022**, *11*, 149, doi:10.3390/biology11010149.

GIEC : sélection assistée par marqueurs et modèles du stade épi à 1 cm pour des blés tendres adaptés au changement climatique



Pérez P, de los Campos G. 2013. BGLR: A Statistical Package for Whole Genome Regression and Prediction. R package version.

Rötter RP, Tao F, Höhn JG, Palosuo T. 2015. Use of crop simulation modelling to aid ideotype design of future cereal cultivars. Journal of Experimental Botany **66**, 3463–3476.

Rousset M, Bonnin I, Remoué C, et al. 2011. Deciphering the genetics of flowering time by an association study on candidate genes in bread wheat (*Triticum aestivum L.*). Theoretical and Applied Genetics **123**, 907–926.

Semenov MA, Stratonovitch P. 2013. Designing high-yielding wheat ideotypes for a changing climate. Food and Energy Security **2**, 185–196.

Semenov MA, Stratonovitch P, Alghabari F, Gooding MJ. 2014. Adapting wheat in Europe for climate change. Journal of Cereal Science **59**, 245–256.

Trnka M, Rötter RP, Ruiz-Ramos M, Kersebaum KC, Olesen JE, Žalud Z, Semenov MA. 2014. Adverse weather conditions for European wheat production will become more frequent with climate change. Nature Climate Change **4**, 637-643.

Turner A, Beales J, Faure S, Dunford RP, Laurie DA. 2005. The Pseudo-Response Regulator Ppd-H1 Provides Adaptation to Photoperiod in Barley. Science **310**, 1031-1034.

Wang J, Wen W, Hanif M, Xia X, Wang H, Liu S, Liu J, Yang L, Cao S, He Z. 2016. TaELF3-1DL, a homolog of ELF3, is associated with heading date in bread wheat. Molecular Breeding **36**, 161.

Weir AH, Bragg PL, Porter JR, Rayner JH. 1984. A winter wheat crop simulation model without water or nutrient limitations. The Journal of Agricultural Science **102**, 371-382.

Würschum T, Boeven PHG, Langer SM, Longin CFH, Leiser WL. 2015. Multiply to conquer: Copy number variations at Ppd-B1 and Vrn-A1 facilitate global adaptation in wheat. BMC Genetics **16**.

Würschum T, Rapp M, Miedaner T, Longin CFH, Leiser WL. 2019. Copy number variation of Ppd-B1 is the major determinant of heading time in durum wheat. BMC Genetics **20**, 64.

Zheng B, Chenu K, Dreccer MF, Chapman SC. 2012. Breeding for the future: what are the potential impacts of future frost and heat events on sowing and flowering time requirements for Australian bread wheat (*Triticum aestivum*) varieties? Global Change Biology **18**, 2899-2914.



GIEC : sélection assistée par marqueurs et modèles du stade épi à 1 cm pour des blés tendres adaptés au changement climatique

Matthieu BOGARD¹*, Alison BENTLEY², Jacques LE GOUIS³, Jérôme AUZANEAU⁴, Simon GRIFFITHS⁵, Jérôme ENJALBERT⁶, Pascal GIRAUDEAU⁷, Patrice SENELLART⁸, Céline ZIMMERLI⁹, Céline DUQUE¹⁰, Philippe LEREBOUR¹¹, Sébastien CAIVEAU¹², Ellen GOUDEMAND¹³ **Coordinateur : Matthieu BOGARD*, m.bogard@arvalis.fr

CONTEXTE ET OBJECTIFS

L'augmentation des températures moyennes du fait des changements climatiques induit une accélération de la croissance et du développement des blés et une augmentation de la fréquence et de l'intensité des stress abiotiques de fin de cycle. Ce projet vise à proposer des modèles de prédictions basés sur des marqueurs génétiques pour des caractères de précocité ainsi qu'une méthode d'identification d'idéotypes afin de maximiser l'échappement aux stress abiotiques (Figure 1).



Figure 1: stratégie de l'étude pour la calibration et la

validation de modèles basés sur des marqueurs génétiques et l'identification d'idéotypes de précocité par simulations

afin de maximiser l'échappement aux stress abiotiques.

MATERIELS ET METHODES

Les modèles de prédiction ont été calibrés sur des données d'Arvalis Institut du Végétal puis validés sur un panel de 149 lignées (Tableau 1) testées dans 20 essais (Figure 2). Ces données correspondaient à des observations des stades épi à 1cm et épiaison qui ont permis de calculer la durée des phases semis-épi à 1cm, semisépiaison, la durée d'élongation des tiges et les coefficients de plasticité correspondants. Un modèle écophysiologique paramétré avec des marqueurs génétiques a été utilisé afin d'identifier des idéotypes de précocité permettant de maximiser l'échappement au gel et à la canicule après analyse de ces risques climatiques en France.



Figure 2: Réseau d'essai du panel FSOV GIEC.

I	Tableau 1: Constitution du panel de validation testé dans le
	cadre du projet FSOV GIEC. Le panel était constitué de 149
	lignées issues de 9 populations.

Population	Partenaire	Туре	Nombre de lignées
HEROS x BELEPI	Florimond	DH	18
STADIUM x SOLEHIO	KWS-Momont	DH	18
FALADO x SHERLOCK	Secobra	F3	14
AUCKLAND x DESCARTES	Limagrain	DH	17
ACCROC x GALIBIER	Agri-Obtentions	F3	13
UN7-3060 x RUBISKO	Unisigma	DH, F4/F5	17
BUENNO x TRIOMPH	syngenta	DH	17
MAGIC_inra	INRA GQE	SSD	17
MAGIC_niab	NIAB	SSD	18
TOTAL			149

RESULTATS

Des modèles statistiques de prédiction basés sur des marqueurs génétiques de la durée des phases semis-épi à 1cm (Z30), semis-épiaison (Z55), de la durée d'élongation des tiges (StEI) et les coefficients de plasticité correspondant (b.Z30, b.Z55, b.StEI) ont été testés sur un panel indépendant et montrent de bonnes capacités prédictives sauf pour b.Z30 et b.StEI (Figure 3).

Une méthode d'identification d'idéotypes de précocité basée sur des simulations avec un modèle écophysiologique paramétré avec des marqueurs génétiques permettant de maximiser l'échappement aux stress abiotiques a été développée. Elle a été appliquée pour des environnements contrastés en termes de risque de gel et de canicule (Figure 4).





CONCLUSIONS ET PERSPECTIVES

Figure 4: Identification d'idéotypes pour différents groupes d'environnements avec des risques gel et canicule contrastés et trois dates de semis. Chaque point représente une variété, l'échelle de couleur représente le niveau d'échappement aux stress.

11

Le projet FSOV GIEC a permis de produire des modèles avec des capacités prédictives satisfaisantes pour la durée des phases semis-épi à 1cm, semis-épiaison, la durée d'élongation des tiges et la plasticité à épiaison. Les travaux ont également permis de proposer une méthode d'identification d'idéotypes de précocité visant à maximiser l'échappement aux stress abiotiques. Cette méthode pourrait être appliquée à d'autres stress et pour des séries climatiques futures.



GIEC : sélection assistée par marqueurs et modèles du stade épi à 1 cm pour des blés tendres adaptés au changement climatique