

Validation de marqueurs liés à des gènes d'intérêt en vue de l'établissement d'une base de données pour la sélection assistée chez le blé tendre.

Marion RANOUX¹, Claire NEDELLEC², Emmanuelle CARIOU-PHAM³, Robert BOSSY², Claude de VALLAVIEILLE-POPE⁴, Marc LECONTE⁴, Dialekti VALSAMOU², Delphine HOURCADE³, Laurent GUERREIRO^{3,5}, Pierre SOURDILLE^{1*}

1 – INRA – UMR 1095 GDEC (Génétique, Diversité et Ecophysiologie des Céréales) 5, Chemin de Beaulieu, 63039 Clermont-Ferrand cedex

2 – INRA – UR 1077 MIG (Mathématique, Informatique et Génome), Domaine de Vilvert, 78352 Jouy-en-Josas

3 – Arvalis Institut du Végétal 3, rue Joseph et Marie Hackin, 75116 Paris

4 – INRA – UR BIOGER-CPP (Biologie et Gestion des Risques en Agriculture - Champignons Pathogènes des Plantes), BP01, 78850 Thiverval Grignon

5 – Adresse actuelle : RAGT - Rue Emile Singla, Site de Bourran – BP 3336, 12033 Rodez Cedex 09

* Coordinateur : Pierre SOURDILLE, pierre.sourdille@clermont.inra.fr, Tél. : 04 73 62 44 37

1. Introduction

L'amélioration génétique du blé tendre (*Triticum aestivum* L.) est devenue une préoccupation majeure afin de pourvoir à la consommation mondiale qui ne cesse d'augmenter (environ 2% par an). Cependant, si les rendements ont régulièrement progressé ces dernières années, ils commencent à stagner alors qu'*a contrario*, ceux du maïs continuent de progresser avec régularité.

Si cette différence entre blé et maïs s'explique en grande partie par des conditions climatiques plus favorables à la culture du maïs (en particulier la hausse des températures ; Brisson et al. 2010), elle est due également à des changements radicaux de méthodes de sélection chez cette espèce. En particulier, lors des 20 dernières années, des efforts colossaux ont été entrepris au niveau des études génétiques et génomiques des caractères d'intérêt agronomique conduisant à l'identification de nombreuses liaisons entre marqueurs moléculaires et gènes impliqués dans l'expression de caractères d'intérêt aboutissant parfois au clonage de ceux-ci. Récemment, l'obtention de la séquence complète du génome du maïs (Schnable et al. 2009) a même contribué à amplifier le phénomène en augmentant à la fois le nombre de marqueurs disponibles (plusieurs dizaines de milliers de SNP disponibles) ainsi que le nombre de gènes clonés.

Chez le blé tendre, les efforts portés sur la génomique ont été beaucoup plus limités et tardifs en particulier du fait de la complexité du génome (polyploïde ($2n=6x=42$), grande taille (17 Gb) et fort pourcentage en séquences répétées (85%) ; Feuillet et al. 2012). Cependant, le retard a été partiellement rattrapé lors des 10 dernières années et même si la

séquence complète du génome n'est pas encore achevée, des travaux dans ce sens sont largement initiés en particulier au niveau du chromosome 3B qui vient d'être entièrement séquencé et annoté (Choulet et al. 2014) et des différents bras de chromosomes (IWGSC 2014). Les travaux visant à identifier des liaisons entre marqueurs et gènes d'intérêt ont été nombreux et la bibliographie à ce sujet est particulièrement riche (pour revue voir Neumann et al. 2011). Cependant, il existe peu de ressources regroupant toute cette information sous forme structurée (sites internet¹ MASWheat et CerealDB) et celle-ci est très dispersée dans de nombreuses revues rendant son exploitation pour la sélection particulièrement laborieuse. De plus, les résultats décrits dans ces articles sont parfois difficiles à reproduire ou n'ont été observés que pour des cultivars (voire des espèces) éloignés du matériel élite utilisé couramment dans les schémas de sélection européens. Enfin, la littérature sur le sujet des marqueurs liés à des gènes d'intérêt chez le blé est extrêmement vaste (plusieurs milliers d'articles) et extraire l'information pertinente est un travail laborieux et de longue haleine qui mérite d'être automatisé au moins partiellement.

L'objectif du projet était donc d'élaborer une base de données interrogeable aisément et répertoriant l'ensemble des informations disponibles dans la bibliographie décrivant l'existence d'une liaison entre un (ou plusieurs) marqueur(s) et des gènes d'intérêt agronomique chez le blé tendre (résistances aux maladies, gènes de développement, gènes impliqués dans la qualité boulangère, résistance aux stress abiotiques, etc.). Pour ce faire, nous avons développé

¹ Toutes les adresses URL des sites sont données dans la section « Références ».

une procédure automatisée d'analyse bibliographique. Les marqueurs ont ensuite été expertisés, c'est-à-dire que les conditions d'utilisation ont été mises au point, et validées sur du matériel de référence. Les marqueurs donnant les meilleurs résultats et pouvant être exploités dans des schémas de sélection ont été testés sur un panel élite étendu (335 variétés commerciales actuelles et 47 en cours de sélection) afin que leur pertinence puisse être mesurée. Finalement, une collection d'une soixantaine de variétés a été phénotypée pour les gènes de résistance aux rouilles jaune (*Yr*) et brune (*Lr*) ou d'autres maladies (septoriose, fusariose, oïdium). Une analyse d'association avec les marqueurs supposés liés à différents gènes impliqués dans les résistances a ainsi permis la validation de certains d'entre eux.

2. Matériel et méthode

a- Matériel végétal

Pour le test des marqueurs moléculaires, le matériel végétal analysé est constitué de deux panels : (1) le panel de référence contient différentes lignées connues comme portant les allèles intéressants de gènes d'intérêt, d'après la bibliographie. Cinq graines de chaque lignée ont été collectées auprès du centre de ressources biologiques (CRB) de Clermont-Ferrand ou auprès des auteurs des publications ; (2) le panel élite uniquement génotypé avec les marqueurs moléculaires validés, est constitué de 335 variétés (Annexe 1) actuellement cultivées et de 47 lignées en cours de sélection.

Dans les deux cas, l'ADN est extrait à partir d'un mélange de feuilles issues de trois plantes selon un protocole utilisant le bisulfite de sodium en vigueur au laboratoire.

b- Analyse automatisée de la littérature

La méthode automatique d'extraction d'information utilise une analyse profonde du texte et l'apprentissage automatique de la chaîne AlvisNLP (Nedellec *et al.* 2009). L'apprentissage automatique supervisé s'applique à des exemples annotés par 13 experts du domaine, principalement des sélectionneurs de semences et les deux auteurs de l'article, biologistes du GDEC. Le corpus annoté est composé de 72 articles sélectionnés pour leur intérêt et leur représentativité et principalement publiés par le journal *Theoretical and Applied Genetics* (Springer). Ce journal a été choisi car il mentionne un grand

nombre de marqueurs et il est accessible en ligne. L'importance de la source d'information (articles) est essentielle et par exemple, l'éditeur Springer permet un tri efficace des articles. Chaque document est annoté deux fois, en double aveugle. Les annotateurs utilisent l'éditeur d'annotation AlvisAE auquel ils ont été formés (Papazian *et al.* 2012). Le document de consignes fourni aux annotateurs décrit le modèle, définit les entités et relations et donne de nombreux exemples (Nedellec *et al.* 2013).

c- Développement et validation des marqueurs

Certains sites internet proposent déjà quelques marqueurs utilisables en sélection (46 dans MASWheat). Ceux-ci ont été validés en priorité car ils sont déjà régulièrement exploités et ont servi à évaluer l'efficacité des méthodes d'extraction d'information. Les recherches bibliographiques ont été orientées autour de quatre grandes thématiques : (1) résistances aux stress biotiques (maladies (rouilles, oïdium, septoriose, fusariose...) et ravageurs (nématodes, cécidomyie, mouche de Hess...)) ; (2) résistances aux stress abiotiques (sécheresse, température, salinité, verse...) ; (3) développement des plantes (hauteur, photopériode, précocité, vernalisation, floraison, tallage, caractéristiques de l'épi...) ; (4) qualité boulangère (rhéologie de la pâte, dureté du grain, taux de protéines, qualité de l'amidon...). Des gènes peuvent être associés à chaque thématique, et des données de liaison avec des marqueurs moléculaires existent pour certains d'entre eux. L'information initiale sur ces gènes a été enrichie en particulier avec le marqueur le plus proche, ses conditions de révélation et une liste de variétés de référence possédant les allèles favorables. Les marqueurs facilement automatisables (SSR, SNP, STS) ont été sélectionnés en priorité. Pour chaque marqueur lié à un gène d'intérêt, les amorces permettant son amplification ont été récupérées de même que le protocole permettant sa révélation. Dans le cas où il s'agissait d'un marqueur révélé par hybridation (RFLP, EST), les séquences correspondantes ont été recherchées dans les bases de données afin de dériver des amorces utilisables en PCR. Ces amorces ont été utilisées pour amplifier des fragments chez les lignées de références et des lignées témoins (Chinese Spring, Renan, Courtot). Les produits d'amplification ont été séquencés afin d'identifier des mutations (SNP) permettant un génotypage des lignées de référence. Lorsque les séquences n'étaient pas disponibles ou si le marqueur était localisé relativement loin du gène

(>10cM), des marqueurs plus proches automatisables ont été recherchés dans les bases de données internationales de marqueurs (GrainGenes). Dans tous les cas, les marqueurs ont été testés dans les conditions initiales sur le panel de référence dérivé de la bibliographie. Un marqueur a été considéré comme validé lorsque les profils observés correspondaient à ceux décrits dans la littérature. Dans le cas où les marqueurs utilisés n'étaient pas ceux issus de la bibliographie, un marqueur a été considéré comme validé s'il présente un profil permettant de différencier des lignées possédant le gène considéré, de celles ne l'ayant pas. Au final, la base de données a été incrémentée des données de validation : amorces, conditions de révélation, variétés de référence, profils (allèles).

d- Phénotypage des résistances

Des données de phénotypage pour les différents caractères ont été collectées afin de confirmer la pertinence des marqueurs validés précédemment. Dans un premier temps, chaque partenaire a évalué une collection de 30 variétés (15 variétés commerciales proposées par ARVALIS extraites du panel élite (Annexe 1) et 15 issues du matériel propre à chaque sélectionneur). De façon à optimiser et maximiser le nombre de caractères évalués, ceux-ci ont été répartis au début du projet de telle manière qu'ils soient relevés chez au minimum deux partenaires différents pour la robustesse des données.

Deux types d'essais ont été conduits : des essais en conditions contrôlées sur la rouille jaune sur une soixantaine de lignées au stade plantule en collaboration avec l'INRA de Grignon et des essais en plein champs sur la campagne 2012-2013.

Pour les essais en conditions contrôlées sur la rouille jaune, les graines d'une soixantaine de variétés ont été fournies à l'INRA de Grignon par les sélectionneurs partenaires du projet. Les tests ont été réalisés à l'aide de 12 races de *Puccinia striiformis* permettant de révéler la présence de 10 gènes *Yr* s'exprimant au stade plantule. Le postulat sur les gènes *Yr* présents dans les différentes variétés est présenté dans l'Annexe 2. Dans l'Annexe 3, nous avons rajouté les gènes de résistance *Yr* dont l'absence ne peut être certifiée du fait de l'absence en France d'isolats virulents de *P. striiformis* correspondants. Certaines variétés sont résistantes à toutes les races et il n'est pas possible de connaître les gènes impliqués. Beaucoup de variétés possèdent un gène majeur, mais leur niveau de résistance au

stade adulte suggère qu'elles possèdent d'autres gènes ne s'exprimant que lors de stades tardifs sans qu'il soit possible de déterminer lequel au vu des résultats des tests.

Pour les essais en plein champ, les tests ont été réalisés à l'aide d'inoculum de rouilles jaune et brune et de septoriose fournis par INRA de Grignon. Pour la fusariose et la mosaïque, la contamination a été conduite naturellement respectivement sur des précédents favorables (maïs) ou dans des parcelles connues pour être infestées.

Les semences des variétés à planter ont été collectées puis redistribuées par Arvalis auprès des sélectionneurs partenaires du projet.

e- Etude d'association

L'ensemble des données de génotypage et de phénotypage sur tous les critères suivis au cours du projet et obtenus par les partenaires a finalement été centralisé afin de mener des études d'association. L'analyse d'association n'a pas été faite sur tout le génome, mais seulement sur les marqueurs validés afin d'évaluer la force de la liaison entre les marqueurs et les gènes considérés ainsi que la fréquence des gènes d'intérêt au sein du matériel élite français et européen.

L'analyse d'association a été conduite en croisant simplement les données de phénotypage avec celles de génotypage et en analysant la différence des moyennes au marqueur. Des *box plots* par marqueur et par phénotype ont été réalisés systématiquement.

3. Résultats

a- Analyse automatisée

Le moteur de recherche sémantique AlvisIR a été développé sur une base bibliographique pertinente de 3170 articles. Il utilise comme index de recherche les résultats de prédiction de gènes, de marqueurs, de méthodes et de variétés à partir de nomenclatures définies initialement. Les phénotypes et caractères sont prédits en utilisant l'ontologie *WheatPhenotype* de 624 concepts et synonymes, développée pour le projet. Pour l'entraînement des modules d'apprentissage et afin d'améliorer les prédictions, nous avons utilisé comme exemples les annotations manuelles des articles. Elles contiennent 1551 relations binaires et ternaires entre 5848 entités comme par exemple des relations marqueur/allèle, allèle/taille, variété/gène... Certaines relations sont moins représentées comme

par exemple la relation variété/allèle dans la mesure où ce genre d'information est rare dans les articles et figure en général dans les matériels supplémentaires présentés sous forme de tableaux difficiles à exploiter. En ce qui concerne la reconnaissance automatique des entités, deux cas sont possibles : (1) des entités « figées » (gènes, allèles...) qui varient marginalement. Ce cas est le plus favorable et seuls subsistent quelques problèmes d'ambiguïté traités par une analyse du contexte (par exemple des noms de variétés proches ou identiques à d'autres noms usuels comme Brigadier, Oratorio, Racine...). La méthode développée obtient un taux élevé de réussite de 60% selon la métrique de F-mesure utilisée en extraction d'information ; (2) des entités terminologiques (caractères par exemple) qui montrent une très grande diversité et constituent un cinquième des entités. L'essentiel de l'activité du groupe est dirigé vers cette partie et une première version d'une méthode en cours de développement (ToMap (Ratkovic *et al.* 2011)) a été appliquée sur un corpus spécifique avec une précision encourageante de 90%.

b- Validation des marqueurs

Initialement, une liste de 107 gènes d'intérêt couvrant quatre thématiques (résistances aux stress biotiques et abiotiques, développement de la plante

et qualité boulangère) avait été établie : 10 pour le développement de la plante, 80 pour les stress biotiques, 5 pour les stress abiotiques et 12 pour la qualité boulangère. Pour ces 107 gènes, une information de liaison avec des marqueurs était disponible sur le site internet MASWheat pour 70 d'entre eux. Pour les autres une recherche bibliographique a permis d'identifier les marqueurs les plus proches. Pour chaque gène, entre un et six marqueurs sont décrits dans la littérature ou sur le site. Dix-neuf gènes supplémentaires ont été relevés dans la littérature conduisant à un total de 126 gènes évalués. Pour ces 126 gènes, 298 marqueurs ont été testés sur le panel de référence de 16 lignées et 235 ont donné un résultat satisfaisant (qualité d'amplification optimale) ce qui représente au moins un marqueur efficace pour 75 des 126 gènes. Par exemple, le gène *VRN-A1* possède un profil complexe avec 10 allèles (cinq allèles hiver et cinq printemps) distribués sur toute la longueur du gène. Il y a cinq allèles différents dans son promoteur (3 allèles hiver et deux allèles printemps), trois allèles dans l'intron 1 (1 hiver et deux printemps) et deux allèles dans l'exon 7 (1 hiver et 1 printemps). Au total, sept marqueurs SNP ont pu être validés pour permettre d'identifier les différents allèles printemps ou hiver présents dans les variétés testées.

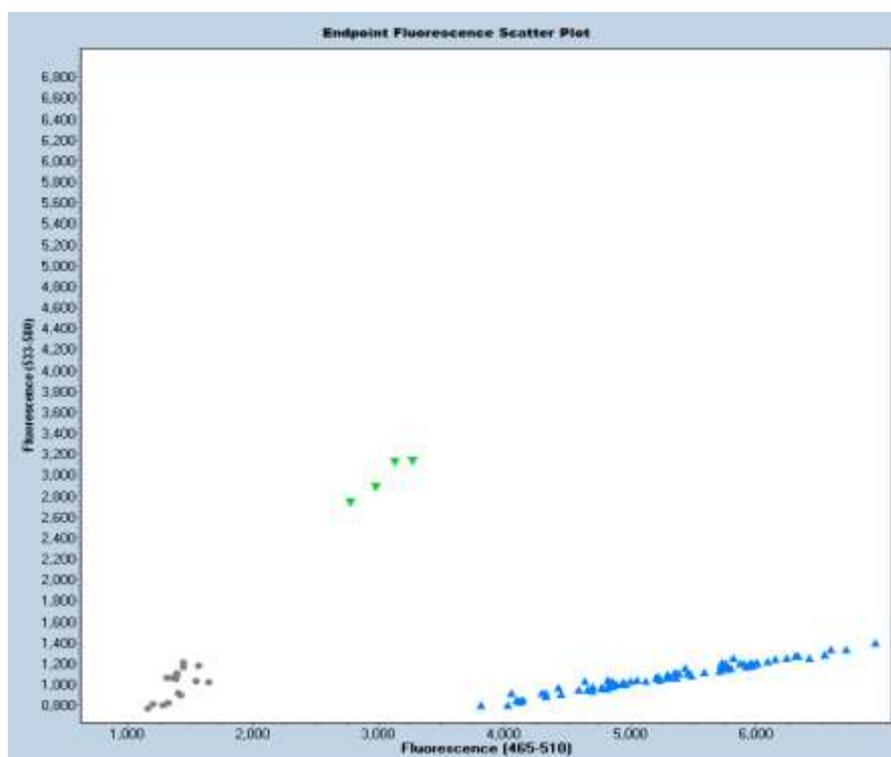


Figure 1 : Exemple de l'évaluation du marqueur KASPar 86 lié à un allèle printemps de l'intron 1 du gène de vernalisation *VRN-A1* sur un panel de lignées témoins. Les triangles verts figurent les lignées avec l'allèle printemps, les bleus l'allèle hiver et en gris les témoins négatifs (eau).

A titre d'exemple, le marqueur 86 (Figure 1) permet de distinguer efficacement l'un des deux allèles printemps du promoteur. En revanche, aucun marqueur pertinent n'a pu être trouvé pour les 51 autres gènes.

Pour 125 de ces 235 marqueurs, des essais haut-débit (KASPar) ont été développés ce qui représente un total de 75 gènes. Pour tous ceux-ci, le ou les marqueurs les plus pertinents ont été choisis et 64 d'entre eux ont été testés sur le panel élite (382 variétés élites cultivées actuellement ou en cours de sélection). A titre d'exemple, la Figure 2 montre le

résultat du test avec le marqueur KASPar 86 précédent lié au gène *VRN-A1*. Le résultat montre que seulement deux variétés possèdent cet allèle dans le panel testé et seulement une possède un allèle original inconnu. Pour ce gène *VRN-A1*, l'analyse complète des sept marqueurs KASPar disponibles permet de conclure que le panel est composé de 368 lignées de type hiver et sept lignées de type printemps. Il n'est pas possible de conclure pour sept lignées qui possèdent des allèles originaux non encore classés.

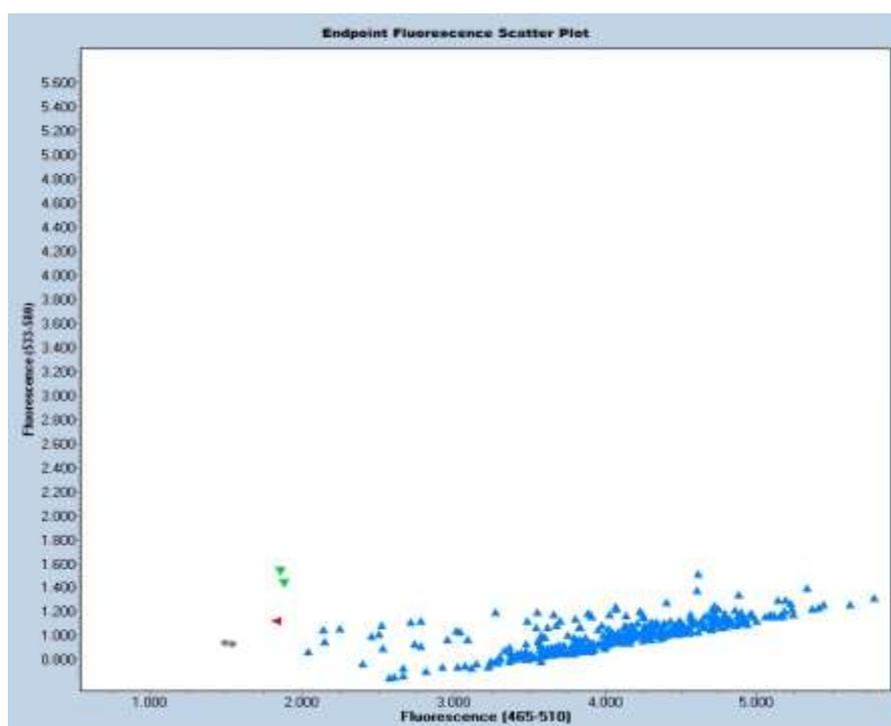


Figure 2: Exemple de l'évaluation du marqueur KASPar 86 lié à un allèle printemps de l'intron 1 du gène de vernalisation *VRN-A1* sur un panel de 382 lignées élites cultivées ou en cours de sélection. Les triangles verts figurent les lignées avec l'allèle printemps, les bleus l'allèle hiver, en rouge une lignée avec un allèle original inconnu et en gris les témoins négatifs (eau).

c- Analyse d'association

Afin d'estimer la corrélation entre la présence d'allèles particuliers aux locus marqueurs et d'allèles favorables pour les gènes d'intérêt suivis, des essais sur différentes maladies (rouilles brune et jaune, septoriose, oïdium, fusariose) ont été réalisés en 2013. Ceux-ci sont relativement hétérogènes dans la mesure où des effets peuvent être masqués par l'attaque d'autres pathogènes ou par des infestations par des souches naturelles. Soixante marqueurs validés sur les témoins et liés à différents gènes de résistance ont été testés sur un panel de 57 variétés commerciales phénotypées aux stades plantule et adulte. La faible taille de ce panel fait qu'une

majorité (49/60) des marqueurs est monomorphe sur les variétés. Néanmoins, certains résultats sont intéressants. Par exemple, les marqueurs SCOAB, WMC44 et WMC170 sont associés à une résistance accrue aux rouilles brune et jaune. Ainsi, pour le microsatellite WMC170, 18 des 19 variétés pour lesquelles la présence du gène de résistance *Yr32* est suspectée, possèdent l'allèle à 208 paires de bases à ce locus ce qui confirme la liaison de ce marqueur avec ce gène. De même, pour le gène de résistance à la fusariose *FHB3* (Figure 3), l'allèle 3 du microsatellite GWM639 est systématiquement associé à une résistance améliorée et 10 des 57 variétés testées possèdent l'allèle favorable au locus.

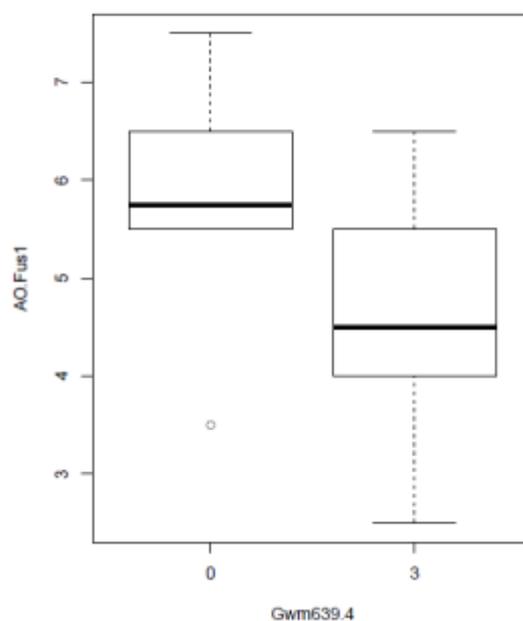


Figure 3 : Exemple de l'évaluation du marqueur microsatellite GWM639 lié à un allèle de résistance à la fusariose au gène *FHB3*. Dans ce box plot, les variétés qui ont l'allèle 3 montrent une résistance au pathogène plus élevée.

4. Discussion et Conclusion

L'analyse automatisée d'articles va devenir un outil incontournable dans la mesure où le nombre de revues s'est multiplié ces dernières années et que les outils de génomique ont gagné en puissance de telle sorte qu'il est devenu aisé de générer de grandes masses de données de génotypage sur des collections ou des populations en ségrégation à un moindre coût. De ce point de vue, les outils développés dans le cadre de ce projet ont montré leur efficacité et leurs limites. Les résultats peuvent encore être améliorés en particulier pour le rappel des caractères et phénotypes qui est un problème difficile à résoudre. Pour cela, l'idéal serait de revoir et d'améliorer les annotations manuelles réalisées dans ~280 zones de texte de 55 articles. Il sera sûrement nécessaire de compléter et augmenter les annotations déjà faites de façon à perfectionner encore les détections automatiques. En termes d'amélioration de l'outil d'analyse, il reste à compléter les résultats d'annotation et les résultats

expérimentaux présents dans la base de données WheatPhenotypeDB par la description des protocoles pour qu'ils puissent être interprétés efficacement. Celle-ci est d'ores et déjà accessible aux sélectionneurs partenaires. Elle contient une information relative à plus de 120 gènes d'intérêt avec une information validée et pertinente pour 89 gènes. Les différents projets de développement de puces de SNP (Wang *et al.* 2014; CerealDB; BreedWheat (E. Paux communication personnelle)) vont contribuer à augmenter encore le nombre de liaisons. En particulier, cela devrait permettre d'étoffer le panel des lignées phénotypées et confirmer certains résultats que nous n'avons pas pu valider dans la mesure où notre panel de 57 variétés était trop faible et toutes les lignées étaient monomorphes pour une majorité des marqueurs que nous avons évalués. Nous pourrions ainsi espérer identifier des allèles causaux aux locus d'intérêt. Cette base devra donc poursuivre une veille active afin d'enrichir régulièrement son contenu et rester un outil attractif et pertinent pour les sélectionneurs.

Références bibliographiques

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

CerealDB : URL

<http://www.cerealsdb.uk.net/cerealgenomics/CerealsDB/indexNEW.php>

CRB de Clermont-Ferrand : URL

<https://www6.clermont.inra.fr/umr1095/Equipes/Infrastructures-experimentales/Centre-de-Ressources-Biologiques>

Choulet F, Alberti A, Theil S, Glover N, Barbe V et al. (2014) Structural and functional partitioning of bread wheat chromosome 3B. *Science* 345 DOI: 10.1126/science.1249721

Feuillet C., Stein N., Rossini L., Praud S., Mayer K., Schulman A., Eversole K., Appels R. (2009) Integrating cereal genomics to support innovation in the Triticeae. *Functional & Integrative Genomics* 12: 573-583

GrainGenes : URL

<http://wheat.pw.usda.gov/GG2/index.shtml>

IWGSC (The International Wheat Genome Sequencing Consortium) (2014) A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science* 345: DOI: 10.1126/science.1251788

MASWheat : URL <http://maswheat.ucdavis.edu/>

Nédellec C., Nazarenko A., Bossy R. (2009) "Information Extraction" in *Ontology Handbook*, S. Staab, R. Studer

(eds.), Springer Verlag, Berlin (DEU): Springer Science - Business Media Deutschland GmbH (International Handbooks on Information Systems), 2nd edition revised 663-686

Nédellec C., Bossy R., Ranoux M., Valsamou D., Sourdille P. (2013) Consignes d'annotation d'articles sur la sélection du blé par marqueurs génétiques. *Projet FSOV Sam Blé*. avril 2013.

Neumann K., Kobiljski B., Denčić S., Varshney R.K., Börner A. (2011) Genome-wide association mapping: a case study in bread wheat (*Triticum aestivum* L.). *Molecular Breeding* 27: 37-58

Papazian F., Bossy R., Nédellec C. (2012) AlvisAE: a collaborative Web text annotation editor for knowledge acquisition. *Proc 6th Linguistic Annotation Workshop (The LAW VI)* 149-52.

Ratkovic Z., Golik W., Warnier P. (2012) Event extraction of bacteria biotopes: a knowledge-intensive NLP-based approach". *BMC Bioinformatics* 2012, 13(Suppl 11):S8, 26.

Schnable P.S., Ware D., Fulton R.S., Stein J.C., Wei F.S et al. (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science* 326: DOI: 10.1126/science.1178534

Wang S., Wong D., Forrest K., Allen A., Chao S. et al. (2014) Characterization of polyploid wheat genomic diversity using a high-density 90,000 single nucleotide polymorphism array. *Plant Biotechnol J.* 12: 787-796

WheatPhenotypeDB : URL

<http://genome.jouy.inra.fr/~rboassy/cgi-bin/FSOV/SAM.cgi>

Annexe 1 : Liste du panel élite (335 variétés).

ACCOR	ACCROC	ACIENDA	ACIENTO	ACOUSTIC	ADAGIO	ADEQUAT
ADHOC	AEROBIC	AFRICA	AGRESTIS	AGUILA	ALCAZAR	ALCEDO
ALCHEMY	ALDRIC	ALEZAN	ALFA	ALIGATOR	ALIGRE	ALIXAN
ALIZEO	ALLEZ-Y	ALLISTER	ALTAMIRA	ALTEO	ALTIGO	ALTRIA
AMADOR	AMBELLO	AMBITION	AMERIGO	AMUNDSSEN	ANDALOU	ANDINO
ANTONIUS	APACHE	APRILIO	ARACK	ARAMIS	ARDELOR	AREZZO
ARISTOTE	ARKEOS	ARLEQUIN	AS-DE-COEUR	ASANO	ASTRAKAN	ASTUCE
ATHLON	ATTITUDE	ATTLASS	AUBUSSON	AUDI	AUTAN	AVALON
AVANTAGE	AXIMACK	AZIMUT	AZTEC	AZZERTI	AZZURO	BAGATELLE 7
BAGOU	BAGUETTE 10	BAGUETTE 11	BALANCE	BALTHAZAR	BALTIMOR	BANDERA
BAROK	BAROUDEUR	BASTIDE	BATTANT	BERMUDE	BISCAY	BOBINO
BOISSEAU	BOKARO	BOLOGNA	BOREGAR	BOSTON	BOTTICELLI	BRANDO
BRIGADIER	BROMPTON	BUENNO	CADENZA	CALISTO	CAMP REMY	CAMPARI
CAMPERO	CAPHORN	CAPNOR	CAPO	CARNAVAL	CARRE	CATALAN
CELESTIN	CENTENAIRE	CEZANNE	CHARGER	CHEVALIER	CHINESE-SPRING	CIGALO
CLAIRE	COMODOR	CONTREFOR	CORDIALE	CORVUS	COURTOT	CRAKLIN
CROISADE	CROUSTY	DENMAN	DIALOG	DIEGO	DINOSOR	DUXFORD
EINSTEIN	EMERALD	ENESCO	ENORM	EPHOROS	EPIDOC	EQUILIBRE
ESPERIA	ESTICA	ETECHO	EUCLIDE	EVEIL	EXELCIOR	EXOTIC
EXPERT	EXSEPT	FARANDOLE	FARINELLI	FIorenZO	FIORETTO	FLAIR
FLAMENKO	FLUOR	FOLIO	FOLKLOR	FORBAN	FOURMI	FRELON
GALACTIC	GALAXIE	GALOPAIN	GANDHI	GARANTUS	GARCIA	GATSBY
GAUGAIN	GLASGOW	GONCOURT	GRAINDOR	GRAVITAS	GRETHEL	GRISBY
GUADALUPE	GULLIVER	HAMAC	HATTRICK	HAUSSMANN	HEKTO	HEREWARD
HERMANN	HOURRA	HUMBER	HYBERY	HYPERION	HYSORE	HYSUN
HYXO	ILLICO	INCISIF	INGENIO	INOUI	INSPIRATION	INSTINCT
INTACT	INTENSE	INTERET	IRIDIUM	ISENGRAIN	ISIDOR	ISTABRAQ
JOSELIN	JULIUS	KALANGO	KALYSTAR	KARILLON	KORELI	KUMBERRI
LANCELOT	LEAR	LIMES	LORD	MALACCA	MANAGER	MARCELIN
MATTIS	MAXWELL	MAXYL	MEISTER	MELKIOR	MENDEL	MENESTREL
MERCATO	MESSAGE	MESSIDOR	METEOR	MINOTOR	MIROIR	MULAN
MUSIK	NIRVANA	NOGAL	NOGENT	NUAGE	NUCLEO	OAKLEY
OCTET	ÆDIPE	OPATA	ORATORIO	ORDEAL	ORNICAR	ORQUAL
ORVANTIS	OXEBO	PACTOLE	PAINDOR	PAJERO	PAKITO	PALADAIN
PALEDOR	PARADOR	PAROLI	PEPIDOR	PERFECTOR	PERICLES	PHARE
PIERROT	PIKO	PIRENEO	PLAYER	PLETHORE	POTENT	PREFERENCE
PREMIO	PRESTIDOR	PREVERT	PROLOG	PULSAR	PYTAGOR	QUALITY
QUEBON	RACINE	RAGLAN	RAISON	RAZZANO	RECITAL	RENAN
RESSOR	RIALTO	RICHEPAIN	RITMO	ROBIGUS	RODRIGO	ROSARIO
ROYSSAC	RUNAL	RUSTIC	SAINT-EX	SAMURAI	SANKARA	SANTIAGO
SATURNUS	SAVANNAH	SCENARIO	SCIPION	SCOR	SCORPION-25	SEBASTO
SELEKT	SEQUOIA	SEYRAC	SHANGO	SIRTAKI	SOGOOD	SOISSONS
SOKAL	SOKRATES	SOLEHIO	SOLLARIO	SOLSTICE	SOLUTION	SOPHYTRA
SORRIAL	SPARK	SPONSOR	STIGG	SUMO	SWEET	SWINGGY
TABASCO	TAMARO	TAPIDOR	TERRON	TIAGO	TILBURI	TIMBER
TIMING	TOISONDOR	TOREADOR	TORKA	TRAPEZ	TREMIE	TROCADERO
TULIP	TUXEDO	USKI	VALODOR	VERLAINE	VICTO	VISCOUNT
VIVANT	VOLONTAIRE	W7984	WARRIOR			

Annexe 2 : Liste des variétés utilisées pour l'analyse en conditions contrôlées des gènes de résistance à la rouille jaune. Année : année de diffusion de la variété. Hypothèse gènes (Yr) : postulat sur les gènes de résistance au stade plantule déduits d'après les types d'infection observés après inoculation par 12 pathotypes de *Puccinia striiformis*. Lorsque la variété est résistante à tous les pathotypes, elle est indiquée Résistant. Marqueur : marqueur trouvé le plus proche des gènes considérés (ordre des marqueurs conservé avec l'ordre des gènes de la colonne précédente) d'après la littérature. ? : pas de marqueur identifié comme lié au(x) gène(s).

Variété	Année	Hypothèse gènes (Yr)	Marqueur	Variété	Année	Hypothèse gènes (Yr)	Marqueur
ACCROC	10	7	GWM526	FARINELLI	11	7	GWM526
ACOUSTIC	11	7	GWM526	FIGARO	12	7	GWM526
ADHOC	11	1	STM673	FLAMENKO	11	7	GWM526
ALIGATOR	10	17	Ventriup-LN2	FLUOR	11	Résistant	?
ALIXAN	5	32	WMC170	FOLKLOR	11	32	WMC170
ALIZEO	11	6 et 32	WMC76 et WMC170	GONCOURT	9	6 et 17	WMC76 et Ventriup-LN2
ALLEZ-Y	11	3, 25 ou/et Sd	WMC356, GDM33 ou/et ?	GRAINDOR	6	7 et (9 ou 17)	GWM526 et (GWM582 ou Ventriup-LN2)
ALTAMIRA	9	17	Ventriup-LN2	HYBERY	11	1	STM673
ALTEO	11	Résistant	?	HYPSTAR	8	7 et (17, 32)	GWM526 et (Ventriup-LN2, WMC170)
ALTIGO	7	17	Ventriup-LN2	HYSUN	4	2 et (3 ou Sd)	WMC364 et (WMC356 ou ?)
AMBELLO	11	17	Ventriup-LN2	KARILLON	11	2	WMC364
APACHE	98	7 et 17	GWM526 et Ventriup-LN2	KORELI	6	17	Ventriup-LN2
AREZZO	8	17	Ventriup-LN2	MATTIS	11	17	Ventriup-LN2
ARKEOS	11	6 et 32	WMC76 et WMC170	MIROIR	11	7	GWM526
AUBUSSON	2	17	Ventriup-LN2	MUSIK	11	7	GWM526
AZZERTI	10	17	Ventriup-LN2	ORVANTIS	0	17	Ventriup-LN2
BAROK	9	17	Ventriup-LN2	PAKITO	11	Résistant	?
BERMUDE	7	17 et 6	Ventriup-LN2 et WMC76	PALEDOR	5	2 et (3 ou Sd)	WMC364 et (WMC356 ou ?)
BOISSEAU	7	Résistant	?	PREMIO	7	7	GWM526
BOREGAR	8	32	WMC170	RACINE	11	6 et 17	WMC76 et Ventriup-LN2
CAPHORN	1	6 et 17	WMC76 et Ventriup-LN2	SAINT EX	11	Résistant	?
CARRE	11	32	WMC170	SCENARIO	11	17	Ventriup-LN2
CHEVRON	9	6 et 17	WMC76 et Ventriup-LN2	SCOR	9	32	WMC170
COMPIL	10	17	Ventriup-LN2	SELEKT	7	6	WMC76
CONTREFOR	11	17	Ventriup-LN2	SOISSONS	88	2	WMC364
CROISADE	11	7	GWM526	SOKAL	11	Résistant	?
DINOSOR	5	17 ?	Ventriup-LN2	SWEET	11	6 et 17	WMC76 et Ventriup-LN2
EPHOROS	4	3, 25 et Sd	WMC356, GDM33 et ?	TRAPEZ	9	32	WMC170
EUCLIDE	7	32	WMC170	TULIP	11	32	WMC170
EXPERT	8	6	WMC76				

Annexe 3 : Liste des variétés utilisées pour l'analyse en conditions contrôlées des gènes de résistance à la rouille jaune. Année : année de diffusion de la variété. Hypothèse gènes (Yr) : postulat des gènes de résistance Yr dont l'absence ne peut être avérée d'après les types d'infection observés pour 12 pathotypes. Lorsque la variété est résistante à tous les pathotypes, elle est indiquée Résistant.

Variété	Année	Hypothèse gènes (Yr)	Variété	Année	Hypothèse gènes (Yr)
ACCROC	10	7 et (2, 3 ou 25)	FARINELLI	11	7 et (1, 2, 3, 4, 6, 9, 17, 32 ou 25)
ACOUSTIC	11	7 et (2, 3, 4 ou 25)	FIGARO	12	7 et (2, 3, 4 ou 25)
ADHOC	11	1 et (2, 3 ou 25)	FLAMENKO	11	7 et (2, 3 ou 25)
ALIGATOR	10	17 et (1, 2, 3, 4, 9 ou 25)	FLUOR	11	Résistant
ALIXAN	5	32 et (1, 2, 3, 4, 9, 17 ou 25)	FOLKLOR	11	32 et (1, 2, 3, 4, 9, 17 ou 25)
ALIZEO	11	(6 et 32) et (1, 2, 3, 4, 9, 17 ou 25)	GONCOURT	9	(6 et 17) et (1, 2, 3, 4, 9 ou 25)
ALLEZ-Y	11	2 et 3 et 25	GRAINDOR	6	7 et (1, 2, 3, 4, 6, 9, 17, 32 ou 25)
ALTAMIRA	9	17 et (1, 2, 3, 9 ou 25)	HYBERY	11	1 et (2, 3 ou 25)
ALTEO	11	Résistant	HYSTAR	8	7 et (1, 2, 3, 4, 6, 9, 17, 32 ou 25)
ALTIGO	7	17 et (1, 2, 3, 9 ou 25)	HYSUN	4	2 et 3 et 25
AMBELLO	11	17 et (1, 2, 3, 4, 9 ou 25)	KARILLON	11	2 ou sensible
APACHE	98	(7 et 17) et (1, 2, 3 ou 25)	KORELI	6	17 et (1, 2, 3, 9 ou 25)
AREZZO	8	17 et (1, 2, 3, 4, 9 ou 25)	MATTIS	11	17 et (1, 2, 3, 4, 9 ou 25)
ARKEOS	11	(6 et 32) et (1, 2, 3, 4, 9, 17 ou 25)	MIROIR	11	7 et (2, 3 ou 25)
AUBUSSON	2	17 et (1, 2, 3, 4, 9 ou 25)	MUSIK	11	7 et (2, 3 ou 25)
AZZERTI	10	17 et (1, 2, 3, 4, 9 ou 25)	ORVANTIS	0	17 et (1, 2, 3, 4, 9 ou 25)
BAROK	9	17 et (1, 2, 3, 4, 9 ou 25)	PAKITO	11	Résistant
BERMUDE	7	(6 et 17) et (1, 2, 3, 4, 9 ou 25)	PALEDOR	5	2 et 3 et 25
BOISSEAU	7	Résistant	PREMIO	7	7 et (1, 2, 3, 4, 6, 9, 17, 32 ou 25)
BOREGAR	8	32 et (1, 2, 3, 4, 9, 17 ou 25)	RACINE	11	(6 et 17) et (1, 2, 3, 4, 9 ou 25)
CAPHORN	1	(6 et 17) et (1, 2, 3, 4, 9 ou 25)	SAINT EX	11	Résistant
CARRE	11	32 et (1, 2, 3, 4, 9, 17 ou 25)	SCENARIO	11	17 et (1, 2, 3, 4, 9 ou 25)
CHEVRON	9	(6 et 17) et (1, 2, 3, 4, 9 ou 25)	SCOR	9	32 et (1, 2, 3, 4, 9, 17 ou 25)
COMPIL	10	17 et (1, 2, 3, 4, 9 ou 25)	SELEKT	7	6 et (1, 2, 3 ou 25)
CONTREFOR	11	17 et (1, 2, 3, 4, 9 ou 25)	SOISSONS	88	2 et (3 ou 25)
CROISADE	11	7 et (1, 2, 3, 4, 6, 9, 17, 32 ou 25)	SOKAL	11	Résistant
DINOSOR	5	17 et (1, 2, 3, 4, 9 ou 25)	SWEET	11	(6 et 17) et (1, 2, 3, 4, 9 ou 25)
EPHOROS	4	2 et 3 et 25	TRAPEZ	9	32 et (1, 2, 3 ou 25)
EUCLIDE	7	32 et (1, 2, 3, 4, 9, 17 ou 25)	TULIP	11	32 et (1, 2, 3, 4, 9, 17 ou 25)
EXPERT	8	6 et (2, 3 ou 25)			