## RenKSeq: Développement d'un outil (RenKSeq) innovant de criblage de gènes à haut débit pour identifier rapidement de nouveaux gènes de résistance à la septoriose

Olivier Robert \*1, David Lopez 2, Andrea Garavito 2, Aurélie Evrard\*1, Ellen Goudemand-Dugue\*1, Lamia Aouini 3, Els Verstappen 3, Burkhard Steuernagel 4, Delphine Hourcade Marcolla 5, Brande Wulff 4, Justin Faris 6, Gert Kema 3, Cyrille Saintenac 2

- 1 FLORIMOND DESPREZ VEUVE & FILS, 59242 Cappelle-en-Pévèle, France
- 2 INRAE GDEC, UMR 1095, 5 chemin de Beaulieu 63100 Clermont-Ferrand, France
- 3 Wageningen University and Research (Wageningen Plant Research, Biointeractions and Plant Health), PO Box 16, 6700AA, Wageningen, The Netherlands
- 4 JOHN INNES CENTRE, Colney Lane NR4 7UH Norwich, UK
- 5 ARVALIS INSTITUT DU VEGETAL, 3 rue Joseph et Marie Hacquin-75116 PARIS, France
- 6 United States Department of Agriculture-Agricultural Research Service, Cereal Crops Research Unit, Edward T. Schafer Agricultural Research Center, Fargo, ND, 58102, USA

\*Coordinateur: Ellen Goudemand-Dugue, Ellen.GOUDEMAND-DUGUE@florimond-desprez.fr

#### Introduction

La septoriose est l'une des maladies les plus dommageables pour la culture du blé en France. Chaque année, les pertes estimées sont en moyenne entre 350 et 700 millions d'euros (Fones and Gurr, 2015). Cette maladie est causée par le champignon *Zymoseptoria tritici* dont l'une des principales caractéristiques est sa grande diversité génétique. Une étude récente suggère la présence de 50 millions d'haplotypes sur un hectare de champ de blé dans le Nord de la France lors de conditions épidémiques importantes (Siah *et al.*, 2018). Cette caractéristique est un frein majeur pour lutter efficacement et durablement contre cet agent pathogène car cela permet à *Z. tritici* de s'adapter rapidement aux stratégies de lutte déployées.

Le contrôle de cette maladie est principalement basé sur l'utilisation de fongicides et de variétés de blé résistantes à cet agent pathogène. De par la mise en place de directives Européennes et nationales, l'utilisation controversé des fongicides va être limité dans les années à venir et l'utilisation de variétés résistantes va devenir un des piliers de plus en plus important de la lutte contre cette maladie. Or, l'arsenal génétique à disposition des sélectionneurs pour lutter contre cette maladie reste limité et encore mal connu. Il est ainsi nécessaire de fournir plus d'outils et de ressources à disposition de la sélection comme des marqueurs diagnostiques pour réussir avec succès cette transition.

Au cours de ces dernières années, les progrès en génomique ont révolutionné les approches pour identifier de nouveaux gènes de résistance. En 2016, l'équipe de Brande Wulff (John Innes Centre) a mis au point une stratégie d'identification de gènes de résistance à la rouille chez le blé (Steuernagel et al., 2016). Cette méthodologie a permis d'identifier deux nouveaux gènes de résistance à la rouille noire, Sr22 et Sr45, en dix-huit mois alors que plusieurs années étaient nécessaires auparavant. Cette stratégie repose sur le prérequis majeur que le gène de résistance à identifier appartient à la famille des NBS-LRR. Cette stratégie comporte plusieurs étapes : (i) identification d'une accession présentant une résistance majeure à la rouille, (ii) production d'une population de tilling à partir de cette accession résistante, (iii) phénotypage de cette population et identification de familles sensibles, (iv)

séquençage de l'ensemble des NBS-LRR par capture de séquences chez l'accession résistante et chez les familles sensibles, (v) identification des mutations au sein des gènes NBS-LRR pour l'ensemble des familles sensibles. Le gène NBS-LRR présentant une mutation chez l'ensemble des familles sensibles représente le gène de résistance conférant la résistance chez l'accession étudiée.

La capture de gènes est une technique qui permet de réduire la complexité des séquences à étudier. Elle est principalement utilisée dans l'étude des grands génomes comme celui du blé. Cette méthodologie utilisée depuis longtemps présente plusieurs variantes. Cependant, la stratégie générale repose sur le design de séquences nucléotidiques appâts d'une longueur de 60 à 120 bp complémentaires des séquences à étudier. Une fois synthétisés, ces appâts sont en mis en solution avec le génome fragmenté d'intérêt. Les appâts hybridés avec les séquences complémentaires d'intérêt sont retirés de la solution, en général grâce à un système de billes magnétiques. Les séquences complémentaires sont ensuite éluées des appâts et séquencées à l'aide des dernières technologies de séquencage.

Dans le cas des résistances à la septoriose chez le blé, seulement 22 gènes majeurs de résistance et 89 méta-QTL ont été cartographiés génétiquement (Brown et al., 2015), un nombre limité si on le compare aux gènes identifiés pour d'autres maladies fongiques du blé telles que les rouilles ou l'oïdium. La résistance conférée par les gènes majeurs de résistance à la septoriose répond au système gène-à-gène dans lequel un gène majeur de résistance va reconnaitre un gène d'avirulence présent chez le champignon.

Il existe encore peu de connaissances sur la présence de ces gènes au sein des variétés françaises et très peu de marqueurs diagnostiques. Les gènes Stb4, Stb9, Stb11, Stb18 ont été identifié dans des variétés Françaises, ce qui suggèrent leur présence dans le germplasm français mais aucun marqueur n'est disponible pour un suivi fiable. Récemment, les gènes Stb6 (Saintenac et al., 2018) et Stb16q (Saintenac et al., 2021) ont été clonés. Ils codent des récepteurs à activité kinase (RLK). L'identification de ces gènes a permis le développement de marqueurs diagnostiques qui permettent actuellement de suivre ces gènes dans les programmes de sélection. Cela a permis



de montrer que le gène *Stb6* est présent dans environ 45 % des variétés françaises inscrites au catalogue alors que *Stb16q* est présent dans une très petite proportion.

Fort de ces découvertes récentes, le projet RenKSeq avait pour objectif de s'appuyer sur la stratégie MutRenSeq pour identifier de nouveaux gènes de résistance à la septoriose causée par *Zymoseptoria tritici* et Parastagonospora nodorum en ne se basant plus sur l'étude des gènes de type NBS LRR mais sur les gènes de type RLK pour fournir de nouvelles ressources à la sélection.

## 2

#### Matériel et méthode

#### ▶ Analyse bioinformatique des séquences de blé

Deux approches complémentaires ont été utilisées pour identifier les gènes de type RLK dans les séquences de blé. Pour les génomes annotés comme Chinese spring, les profils hmm de kinase (Lehti-Shiu and Shiu, 2012) ont été utilisés pour rechercher les kinases dans les protéomes prédits. Pour tous les autres génomes, une identification de novo a été réalisée en deux étapes avec le logiciel exonerate. A partir des séquences de génome dont les éléments transposables ont été masqués par RepeatMasker et 94,000 séquences de RLK présentes dans la littérature, une première identification a été réalisée. Ensuite, les régions ne portant pas de RLK lors de cette première approche ont été analysées avec une approche en mode sensible. Pour les génomes annotés, une comparaison des deux approches (Hmmscan et exonerate) a été réalisée et les résultats ont été mergés. Le design des séquences appâts a été réalisé grâce à un logiciel développé par le JIC dans le cadre du clonage des gènes de résistance à la rouille.

#### ▶ Production et phénotypage des populations de tilling

Chaque accession résistante étudiée a été multipliée pour disposer d'un stock de grains homogènes. Ces grains ont ensuite été traités avec différentes doses d'EMS pour identifier la dose entrainant environ 50 % de germination des grains. Plus de 1000 grains ont ensuite été traités avec cette dose. Chaque plante  $M_1$  issue de ce traitement a été multipliée et les graines  $M_2$  récoltées soit en batch pour chaque M1 soit épi par épi a constitué la population de tilling. Entre 8 et 12 plantes par famille  $M_2$  ont ensuite été phénotypées avec différentes souches de Z. tritici. Ce phénotypage a été effectué au stade plantule

(soit au stade une feuille soit deux feuilles) pour les populations Pocho, Pl41025 et Cadenza. Ces populations ont été inoculées par spray avec un inoculum à une concentration de 106 spores/ml. La population BG301 a été phénotypée par infiltration de la toxine SnTox2 pour identifier les plantes ne présentant pas de nécrose. La population M6 a été phénotypée au champ en utilisant un système de sprinkler pour maintenir une humidité importante tout au long de l'essai. Autour de 15 graines par famille  $\rm M_2$  ont été semées en ligne. Les plantes ont été inoculées au stade Z39 avec une solution de Z. tritici isolat T48 à  $\rm 10^6$  spores/ml. Le phénotypage de validation des familles  $\rm M_2$  identifiées comme sensibles a été réalisé avec une application de l'inoculum au pinceau.

#### ▶ Production et analyse des données de capture de séquence

L'ADN de chaque plante  $M_2$  ou  $M_3$  sensible a été extrait à l'aide du kit Qiagen DNeasy Plant Mini Kit. Ces ADN ont ensuite été dosés au qubit et envoyés à la société ArborBiosciences pour capture. Les fragments capturés ont été séquencés soit en utilisant la technologie MySeq (2x300 bp) soit la technologie NovaSeq (2x150 bp) afin d'attendre une couverture moyenne de 20X. L'assemblage de novo des RLK a été réalisé en utilisant le logiciel Mazurka. Les données ont été analysées en utilisant le pipeline MutantHunter, développé par l'équipe Anglaise (Steuernagel et al., 2016). Les contigs d'intérêt ont été annotés en utilisant le pipeline TriAnnot (Leroy et al., 2012).



#### Résultats

#### ▶ 1 - Développement d'un essai de capture des gènes de type RLK/RLP

Pour intégrer le plus grand nombre de gènes RLK dans l'essai de capture, un des prérequis essentiels au succès de l'approche MutRenSeq, l'ensemble des ressources génomiques à disposition chez le blé au moment de cette analyse a été pris en compte. Ainsi, 16 génomes de blé ont été utilisés. Cinq génomes de blé tendre (deux versions de Chinese spring, Cadenza, Claire, Paragon et Robigus), cinq génomes de blé tétraploïdes (Kronos, Svevo, Cappelli, Strongfield et Zavitan) et cinq génomes de blé diploïdes (*Triticum urartu, Triticum monococcum, Aegilops speltoides, Aegilops sharonensis and Aegilops tauschii*). L'analyse bioinformatique de ces génomes a permis d'identifier l'ensemble des RLK présents dans ces génomes (Table 1).

Assemblage	Taille de l'assemblage (pb)	Loci RLK (pb)	Nombre de loci RLK
Chinese Spring NVGSC	14 347 606 634	23 254 210	6 445
Chinese Spring Earlham Intitule	13 467 630 192	23 714 007	4 862
Zavitan	10 509 868 245	20 469 051	2 584
Cadenza	13 399 781 746	22 244 268	5 362
Claire	14 616 451 647	21 707 537	6 367
Kronos	10 540 467 196	14 940 058	3 706
Paragon	15 318 048 945	22 432 660	5 431
Robigus	14 791 409 785	21 874 636	5 281
Svevo	68 032 98 213	8 350 610	2 444
TGAC WGS_durum_v1	33 864 81 85	458 483	426
TGAC WGS_monococcum_v1	13 221 12 30	574 717	417
TGAC WGS_speltoides_v1	19 078 78 214	552 288	426
TGAC WGS_strongfield_v1	33 848 41 931	361 995	387
TGAC_WGS_tauschii_v1	12 002 21 237	665 015	441
TGAC WGS_urartu_v1	16 570 29 501	686 599	445
TSL_WGS_sharonensis_v1	17 308 99 803	611 984	445
Séquences issues de la bibligraphie		2 286 202	831
Total	128 Gb	185 Mb	45 290

Table 1: Ressources génomiques utilisées pour l'identification des gènes de type RI K chez le blé



Le nombre de RLK identifié par génome varie significativement de 6445 au sein du génome de référence produit par le consortium de séquençage jusqu'à 387 chez le blé dur Strongfield. Le catalogue complet des cibles RLK identifiés chez le blé correspond à 185 Mo pour 45 290 séquences. Ce catalogue comporte un ensemble de gènes redondants identifiés de façon indépendante chez chacune des accessions. Cette redondance a été prise en compte lors du design des séquences appâts.

Cet ensemble de séquences a ensuite été utilisé pour le design des appâts, des séquences de 120 pb qui sont la base de l'essai de capture. Pour réaliser le design de l'essai ne pouvant dépasser 200 000 appâts, nous avons modifié plusieurs paramètres dont l'espacement entre les différents appâts sur la séquence des RLK mais également le pourcentage de similarité entre les séquences appâts. Un premier lot de 236 591 appâts, d'une longueur de 120 pb, a été conçu. Cette valeur a été obtenue en sélectionnant des appâts très spécifiques (90 % d'identité) tout en augmentant l'espace entre les appâts (35 pb). Après cette conception initiale des appâts, des tests *in silico* ont été réalisés pour prédire l'efficacité de capture.

La partie la plus difficile de ce processus consistait à capturer le plus grand nombre possible de cibles RLK tout en éliminant les régions répétées. Tandis que le premier était facilement obtenu via la recherche Blast de nos appâts contre le RLKome, la nature intrinsèquement répétée de certains candidats (par exemple, kinases riches en leucine répétées, domaines de facteur de croissance d'élongation trouvés dans les WAKs) rendait difficile la distinction entre répétitions biologiques utiles ou non. Nous avons donc identifié des régions annotées en tant que répétitions dans le génome de Chinese Spring qui appartenaient au moins partiellement aux RLK supposées. L'annotation manuelle des WAK réalisée par nos collègues Anglais de Rothamsted Research s'est avérée une ressource précieuse à cet égard (The International Wheat Genome Sequencing Consortium (IWGSC) et al., 2018). Nous avons ensuite utilisé la fréquence d'apparition des appâts parmi les régions d'intérêts extraites (c'est-à-dire les RLKomes) et leurs fréquences dans des assemblages non masqués. Cette méthode a permis d'éliminer quelques centaines d'appâts dont les fréquences étaient trop différentes entre les sources génomiques et les sources ciblées. Ce processus itératif a généré 206 000 appâts qui ont été soumis à la société Arbor Sciences pour la synthèse. Selon les contrôles qualité de cette société, 26 000 appâts présentaient une forte propension secondaire. Un travail de repositionnement/redesign de ces appâts a été réalisé et nous a permis de finaliser l'essai de capture comprenant 199 733 séquences appât.

#### 2 - Création et phénotypage de populations de tilling à partir de variétés résistantes

Au total, cinq accessions ont été étudiées dans ce projet :

- L'accession d'origine sud-américaine Pocho qui présente un large spectre de résistance à *Zymoseptoria tritici*.
- L'accession de blé synthétique M6 qui porte le gène majeur *Stb8*.
- Une lignée recombinante BG301 provenant de l'accession BR34 et Grandin qui porte le gène de résistance à *Parastagonospora nodorum Snn2*.

- L'accession de blé dur (*Triticum dicoccum*) PI41025 portant un gène majeur de résistance à Z. tritici sur le chromosome 3A.
- L'accession Cadenza portant le gène de résistance Sth6

<u>Pocho</u>: 911 familles M<sub>2</sub> indépendantes ont été produites et phénotypées avec l'isolat IPO9415 de Z. tritici avirulent sur cette accession. Au total, 47 familles ont présenté des symptômes et ont été sélectionnées pour validation par inoculation au pinceau. Les validations ont été effectuées sur environ 30 plantes par famille M2. Seul neuf familles mutantes indépendantes présentaient au moins une plante sensible (plus de 20 % des régions porteuses de pycnides, 50 % de nécroses). De manière inattendue, le phénotype n'a pas été hérité en tant que trait mendélien pour toutes les familles. Les plantes sensibles portant le plus probablement la mutation causale à l'état homozygote ont été cultivées pour la multiplication des graines. Cependant, ces dernières ont montré une condition physique réduite et une capacité limitée à produire des graines. Nous avons réussi à isoler les semences de la famille 349. Tous les descendants de M3 étaient sensibles. Pour les autres familles, nous avons eu des difficultés à obtenir des semences M3 de plantes sensibles et, pour certaines, des plantes sont même mortes après l'inoculation. Le génotypage à partir de la puce SNP 35K des mutants sensibles des 9 familles ne montrant pas de délétion de larges fragments de chromosome, une plante sensible par famille a été choisie pour la capture des RLK.

BG301: Un total de 478 plants M<sub>1</sub> a été généré après traitement à l'EMS. Dix mutants M<sub>2</sub> indépendants, insensibles à SnTox2 ont été identifiés au cours de la première année. La deuxième année, les graines M<sub>3</sub> de ces plantes ont été récoltées et replantées. Les plantes M3 ont été réévaluées avec des cultures contenant SnTox2 pour vérifier leurs insensibilités à SnTox2, et elles ont été génotypées avec des marqueurs moléculaires connus pour être liés au locus Snn2 afin de déterminer si l'un des mutants présentait des délétions chromosomiques. L'analyse d'infiltration a indiqué que les dix plantes étaient insensibles à la culture contenant SnTox2, mais une analyse génotypique a indiqué que trois des dix hébergeaient d'importantes délétions chromosomiques et n'étaient donc pas utiles pour RenKSeq. L'ADN des sept plantes ne présentant pas de délétions a été extrait pour réaliser la capture des RLK.

 $\underline{\text{M6}}$ : Une population de tilling comprenant 513 plantes  $\mathrm{M}_2$  séparées en 1216 épis a été développée. Dans un premier temps, une partie de cette population a été phénotypée au stade plantule avec la souche T48 Malheureusement, l'accession résistante (M6) ainsi que la population était sensible.



Figure 1: Phenotypage au stade adulte de la population de tilling M6.



Avant de reconduire un phénotypage de la population entière, nous avons phénotypé 50 génotypes de la population ITMI (haploïdes doublés entre la variété sensible Opata et M6) avec les isolats T48, IPO98032 et IPO98050 pour déterminer les bases génétiques de la résistance à la septoriose chez l'accession M6. Une analyse QTL a été effectuée en utilisant ces phénotypes et les données de génotypage disponibles dans la littérature. Pour la souche T48, nous avons observé quatre pics sur le 2B, 6B, 1D et 2D. Avec la souche IPO98032, 3 QTL sont observés dont un sur le 6B et 2 sur le 1D pour les nécroses, et un QTL majeur expliquant les picnides sur le 6B. Ces résultats sont assez bien corrélés avec Simon et al. où 3 QTL au stade plantule ont été identifiés sur les chromosomes 1D, 2D et 6B mais avec un faible score LOD. Néanmoins, il semble que le gène Stb8 sur le chromosome 7B (la cible initiale pour M6) ne soit pas révélé au stade plantule. L'isolat utilisé pour cartographier le gène Stb8 était T48 mais sur des plantes adultes. D'après ces données, Stb8 ne semble s'exprimer ainsi qu'au stade adulte. Avec notre test de semis au stade plantule, nous ciblons donc d'autres QTL que celui contenant Stb8. Un essai au stade adulte sur la population de mutant M6 en conditions semi contrôlées a été réalisé au printemps/ été 2019 (Figure 1). Malheureusement, cette année de sécheresse a tout d'abord impacté la germination et le développement des mutants EMS de cette population. De plus, la sécheresse a limité la propagation de la septoriose et malgré une inoculation artificielle et un arrosage important, très peu de symptômes ont été observé sur les génotypes sensibles et aucun mutant qui avait atteint le stade adulte n'a montré de symptômes de septoriose.

PI41025 : Des graines pour 803 M2 ont été produites. Ces mutants ont été phénotypés mais le phénotypage n'a pas révélé de différence entre le parent sensible (ben) et résistant (PI41025) à partir desquels une population bi-parentale avait été utilisée pour identifier le gène de résistance chez PI41025. De nombreux tests de conditions de culture, de feuille pour l'inoculation, de différents isolats et de méthodes d'inoculation ont été testés sur les graines des deux parents de cette population mais aucun n'a pu reproduire les résultats originaux. L'évaluation du stock original du génotype PI41025 a permis de reproduire le phénotypage de résistance mais pas les graines qui ont servi à faire la population mutante. Cette expérience a montré que le stock de grains utilisé pour réaliser la population n'était pas celui de l'accession PI41025.

<u>Cadenza</u>: Nous avons utilisé sept familles de mutants portant une mutation dans le gène *Stb6* et identifiés lors du clonage de ce gène pour réaliser la preuve de concept de cette stratégie pour cloner le gène *Stb6*.

Au final, 23 familles ont été identifiées comme sensibles à la septoriose à partir de trois populations de tilling (Pocho, BG301 et Cadenza).

#### 3 - Capture des RLK chez les parents des populations de tilling et les familles identifiées comme sensibles

Une fois les mutants sensibles identifiés, la stratégie MutRenSeq nécessite de réaliser un assemblage de novo de l'ensemble des RLK présent chez l'accession résistante et ensuite d'utiliser cette séquence comme référence pour aligner les séquences obtenues à partir de chacun des mutants sensibles. L'identification des contigs de la référence qui contient le plus de mutations et notamment une mutation par famille sensible doit permettre d'identifier le gène de résistance.

Nous avons séquencé les RLK de quatre accessions utilisées dans le projet (Pocho, M6; BG301 et PI41025) ainsi que du génome de référence Chinese spring avec la technologie Illumina MiSeq qui génèrent des lectures plus longues pour permettre l'assemblage d'une meilleure séquence de référence des RLK. Cette expérimentation a généré ~10 Go de séquence pour chaque accession. Chacun des 5 génotypes représentait 2,3 à 4,3 millions de paires de lecture 2x300 pb (Table 2).

Pour notre étude, nous visions une profondeur de 20x. Pour Chinese Spring, les 3,5 millions de paires de lecture ont donné une profondeur médiane de 18x. Parmi ces séquences, 284 avaient une profondeur moyenne <2x et 1358 sont au-dessous de 10x. Les séquences en dessous de la profondeur médiane (17x) ont été communiquées à Arbor Science. L'analyse a révélé que les appâts correspondants étaient dominés par des appâts dont les scores de propension secondaire étaient extrêmement élevés. Arbor Science a synthétisé un lot de nouveaux appâts pour améliorer l'efficacité de la capture. Cette partie a permis de valider l'efficacité de la capture avec le cultivar Chinese Spring pour la plupart des RLK présentes dans son génome. De plus, en moyenne, 85 % des séquences des quatre autres génotypes ont été correctement cartographiées sur l'assemblage Chinese Spring montrant l'efficacité de l'enrichissement.

	Nombre de paires de lecture	% de lecture mappées sur la ref	Paires correctement mappées sur le RLKome	% lectures mappées sur le RLKome	Profondeur sur le RLKome (X)
Chinese spring	3 445 819	84,84 %	1 308 929	37,99	3 262
Pocho	3 436 686	79,20 %	1 290 079	37,54	3 063
P141025	4 015 234	74,96 %	1 548 081	38,56	2 559
BG301	4 385 093	76,91 %	1708 645	38,96	1 931
W7984	2 354 737	78,80 %	861 155	36,57	1 025
Total paires lecture	17 637 569		6 716 889		
Total en bases	10 582 541 400				

Table 2 : Lectures issues de Miseq à partir des cinq génotypes. Nombre brut de lectures 2X300 pb produites.



Pour disposer de la meilleure séquence de référence, différents assembleurs de novo ont été testés avec les séquences des cinq lignées. Les tests ont révélé que la longueur assemblée variait de 12 à 17 Mb (13 Mb ont été trouvés avec CLC). Cependant, Masurca était le meilleur assembleur en terme de métriques d'assemblage, avec une L50 inférieure et une N50 supérieure à toutes les autres et même surpassait CLC. De loin, Masurca a produit la plus petite quantité de contigs et, bien que les auteurs recommandent l'utilisation de bibliothèques brutes, nous avons constaté que le prétraitement avec Trimmomatic donnait des statistiques encore meilleures. Les assemblages de novo RLKome ont ensuite tous été réalisés avec Masurca et Trimmomatic (Table 3).

Assemblage	Taille	Nombre de contigs	L50	N50	Contigs avec une couverture d'appâts > 60%
Cinnose spring	25 Mb	12 907	2 616	3 148	7 858
Pocha	30 Mb	15 621	2 971	3 210	8 577
P141025	24 Mb	12 698	2 218	3 339	5 612
BG301	31 Mb	15 637	2 938	3 345	8 254
W7984	31 Mb	15 125	2 948	3 276	8 693

**Table 3 :** Assemblage de novo des RLK chez les parents des populations de tilling.

<u>Cadenza</u>: Pour valider cette stratégie de clonage de gènes de résistance à la septoriose, nous avons séquencé les RLK chez l'accession Cadenza (2x150 pb) et chez les sept familles identifiées comme sensibles à la souche avirulente *Stb6* IPO323 (Figure 2). L'analyse des séquences de Cadenza a permis l'assemblage de 85 171 contigs avec une N50 de 1,8 kb. La recherche de contigs présentant des mutations a révélé la présence d'un seul contig portant sept mutations provenant chacune d'une des sept familles de mutants sensibles. L'analyse de ce contig a montré qu'il portait le gène *Stb6*. Cette expérience a permis de valider cette stratégie pour le clonage des gènes de résistance à la septoriose.

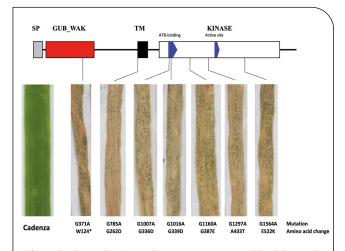


Figure 2 : Caractéristiques des sept mutants sensibles à la souche IPO323, utilisés pour la capture des RLK. En haut est représenté les exons du gène Stb6. Le phénotype des feuilles infectées par la souche IPO323 à 21 dpi ainsi que la position des mutations sont représentées.

Pocho: Pour la population Pocho, aucun contig n'a été identifié comme présentant une mutation dans les neuf familles sensibles identifiées lors du phénotypage. L'analyse par MutantHunter a révélé la présence de 6 contigs comportant une mutation pour 6 des 9 familles et un contig portant 7 mutations. Un seul de ces contigs montre la présence d'un gène (par l'annotation TriAnnot) mais les mutations n'étaient pas présentes dans les exons de ce gène. Ainsi, cette analyse ne nous a pas permis d'identifier de gènes de résistance candidat. Pour tenter d'expliquer ces résultats et regarder si un des 7 contigs identifiés cartographie sur le gène ou les gènes de résistance présents chez l'accession Pocho, nous avons décidé de réaliser une analyse du déterminisme génétique de la résistance à la septoriose chez Pocho. Nous avons ainsi développé une population SSD de 78 individus issue du croisement entre Pocho et l'accession de blé sensible Obelisk. Une analyse QTL a ensuite été réalisée suite au phénotypage de la population par trois souches portant des virulences différentes et grâce au génotypage de la population avec la puce SNP 35K. L'analyse des données issues de la première souche montre la présence de cinq QTL (2A, 3A, 5B, 7B et 7D) de résistance chez l'accession Pocho (Figure 3). La présence de ces cinq QTL vis-à-vis des souches utilisées pour le phénotypage de la population de tilling Pocho explique vraisemblablement pourquoi nous n'avons pas pu identifier un gène candidat car l'effet de la mutation au niveau d'un de ces QTL aurait un faible impact sur le phénotype.

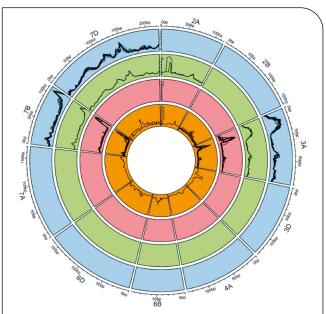


Figure 3 : Analyse QTL des régions impliquées dans la résistance à la souche IPO9415 dans l'accession Pocho. Chaque couleur représente un phénotype. Un QTL majeur sur le chromosome 3A et sur le chromosome 7B a été identifié.

<u>BG301</u>: 27 contigs portant une mutation chez les 7 mutants analysés ont été trouvés. Chacun de ces contigs a été ancré sur la séquence de référence de CS mais aucun n'a été positionné sur le chromosome 2D où le gène *Snn2* est cartographié. L'annotation des gènes présents sur ces scaffolds a permis d'identifier 11 scaffolds portant un gène dont 6 étaient des kinases.



De nombreuses mutations étaient présentes sur ces scaffolds et au niveau de ces gènes. Nous avons donc appliqué une étape de filtration pour éliminer toutes les mutations non induites par l'EMS, option non incluse dans le pipeline MutantHunter. Suite à ce filtre, aucun des 27 scaffolds n'a été retenu par le pipeline MutantHunter, ce qui suggère que ces scaffolds étaient vraisemblablement des faux positifs. L'échec d'identification d'un gène peut être lié au fait que celui-ci soit mal assemblé ou assemblé en plusieurs contigs. Ainsi, nous avons analysé en particulier les 16 gènes présents dans l'intervalle de confiance (dont 11 kinases) défini pour le gène *Snn2* sur la séquence de référence internationale CS. Nous avons reconstruit ces gènes sur la séquence de référence du parent BG301 en utilisant ces données de capture pour vérifier que le gène n'était pas fragmenté dans la référence. Nous avons ensuite utilisé le pipeline MuntantHunter sur cette nouvelle séguence de référence mais les résultats ont été identiques. Nous avons également fait varier les paramètres d'analyse du pipeline MutantHunter ainsi que la référence du parent BG301 (assemblage Novaseq ou Miseq) mais aucun de ces changements n'a permis d'identifier un gène candidat.

▶ 4 - Identification et clonage des gènes effecteurs chez Z. tritici qui interagissent avec les gènes cibles chez les blés (gènes Stb) et qui interagissent dans la spécificité blé tendre / blé dur afin d'améliorer le phénotypage et la sélection pour la résistance des blés

Le gène AvrStb6 a été cloné récemment ce qui permet d'envisager la production de la protéine Stb6. Cette dernière pourrait être utilisée comme preuve de concept pour infiltrer et phénotyper le germplasm Français. Nous avons planifié la production de la protéine AvrStb6 pour l'infiltration dans les tissus de blé afin de vérifier si une réaction se produisait. Les expériences préliminaires n'ont pas abouti et la production en système hétérologue semble compliquée dû à la présence de nombreuses cystéines dans cette protéine.

Le clonage de l'effecteur AvrStb6 chez IPO323 a montré un nombre considérable de SNP par rapport à l'haplotype AvrStb6 présent chez la souche virulente IPO95052. En combinant le séquençage du génome entier d'une collection mondiale de 136 isolats de Z. tritici et des essais de phénotypage sur des cvs. Volcani 447 (blé dur), Bulgarie, Taichung 29 et TE9111, nous avons observé une diversité génétique étendue avec une différenciation génétique modérée entre les groupes géographiques. Les isolats originaires du Moyen-Orient, centre d'origine du blé, présentaient une diversité génétique accrue par rapport aux populations d'autres zones géographiques. Nous avons montré de nombreuses variations du génome structurel à grande échelle et confirmons la présence de 13 chromosomes principaux hébergeant la plupart des gènes effecteurs identifiés et des chromosomes accessoires structurellement variables avec des taux plus élevés de SNP et de polymorphismes de présence / absence de gènes effecteurs. Le profilage de ces effecteurs a révélé un groupe de 183 effecteurs conservés et un groupe de 88 effecteurs présentant des polymorphismes de présence / absence significativement plus proches des

éléments transposables que des conservés. Lorsque nous avons phénotypé 118 isolats de Z. tritici sur les cultivars mentionnés, nous avons constaté que la majorité (87) infectent l'une ou l'autre des espèces de blé (tendre ou dur), tandis qu'une minorité (12 isolats) causait la maladie dans les deux cas. L'évaluation des différences génomiques entre les isolats de blé tendre et de blé dur a révélé quatre gènes effecteurs affectés par du polymorphisme SNP non synonymes dans des isolats de blé tendre. De futures analyses fonctionnelles sont nécessaires pour comprendre le rôle des candidats effecteurs identifiés sur les cultivars de blé tendre et de blé dur. Cette étude contribue à une meilleure compréhension de la spécificité de l'hôte dans les isolats mondiaux de *Z. tritici* et contribue ainsi à une gestion améliorée de la maladie de cet important agent pathogène des céréales.

## 4

#### Discussion

Ce projet a permis d'identifier le catalogue des RLK chez le blé, une famille de gènes importante qui intervient dans de nombreuses fonctions biologiques dont la résistance à la septoriose. Cette ressource va permettre une meilleure compréhension de leur diversité, leur localisation chromosomique et de leur fonction à plus long terme.

Nous avons rencontré plusieurs problèmes lors du développement et du phénotypage des populations de tilling. L'expression de certains gènes de résistance qui n'opère qu'à un stade précis du développement de la plante et des conditions climatiques défavorables pour le phénotypage des populations au stade adulte ont limité l'analyse de certaines populations.

Alors que cette stratégie s'est avérée fructueuse pour le clonage de Stb6 (identification d'un seul contig portant le gène), nous n'avons pas pu identifier les gènes de résistance ciblés chez les accessions résistantes étudiées. Les raisons de cet échec peuvent être multiples et il est difficile de conclure. La première hypothèse est tout simplement que le gène de résistance n'est pas un RLK. Une autre hypothèse probable pour la population Pocho est la présence de plusieurs QTL de résistance. Les mutants sensibles ainsi identifiés peuvent porter des mutations dans un de ces différents QTL et il est impossible d'identifier une région ou un gène commun aux différents mutants. Une autre hypothèse est basée sur l'absence de capture du RLK candidat par notre essai mais cela serait surprenant car les appâts peuvent s'hybrider avec des séquences divergentes de 20 %. La dernière hypothèse est liée aux problèmes d'analyse rencontrés (assemblage et alignement) de ces données pour cette famille de gènes qui est l'une des plus complexe du génome de blé avec autour de 6000 à 7000 copies par génome. Ce qui représente environ deux fois plus de copies que la famille des NBS-LRR.

Les gènes de résistance présents dans les accessions de ce projet sont en cours de clonage en utilisant d'autres approches en parallèle. L'identification de ces gènes nous permettra d'identifier les étapes bloquantes de cette méthodologie afin de l'améliorer. L'identification très récente d'un premier gène candidat de résistance dans une de ces populations montre que ce gène a



été capturé pour moitié dans l'essai de capture. Seul la partie 5' du gène est présent dans les séquences et ainsi toutes les mutations présentes dans les différentes familles sensibles n'ont pas pu être retrouvées par notre approche. De plus, l'accession résistante porte un cluster de cinq gènes RLK alors qu'une seule copie est présente chez la référence Chinese spring. Ce cluster présente notamment une copie avec une similarité de 98 % avec notre gène candidat. Lors de l'alignement des séquences de mutants sur notre séquence de référence, ces deux copies avec 98 % de similarité s'alignent sur la référence ne permettant pas de détecter des mutations fixées chez les mutants sensibles. Ce cas de figure montre

la complexité d'analyse de cette famille multigénique. L'une des solutions serait de privilégier la capture et le séquençage de long fragments pour éviter l'alignement de copies multiples.

Ce projet a permis de mettre en avant les QTL de résistance présents dans l'accession Pocho qui pourront être introgréssés dans le matériel français. De plus, les mutants sensibles identifiés dans ce projet sont une ressource importante pour l'identification des gènes de résistance. Finalement, cet essai en cours d'amélioration est également utilisé par d'autres collaborateurs pour l'identification de gènes de résistance à d'autres maladies du blé.

### Références bibliographiques

Brown, J.K.M., Chartrain, L., Lasserre-Zuber, P., and Saintenac, C. (2015). Genetics of resistance to Zymoseptoria tritici and applications to wheat breeding. Fungal Genet. Biol. 79, 33-41.

Fones, H., and Gurr, S. (2015). The impact of Septoria tritici Blotch disease on wheat: An EU perspective. Fungal Genet. Biol. 79, 3-7.

**Lehti-Shiu, M.D., and Shiu, S.-H.** (2012). Diversity, classification and function of the plant protein kinase superfamily. Philos. Trans. R. Soc. B Biol. Sci. *367*, 2619–2639.

Leroy, P., Guilhot, N., Sakai, H., Bernard, A., Choulet, F., Theil, S., Reboux, S., Amano, N., Flutre, T., Pelegrin, C., *et al.* (2012). TriAnnot: A Versatile and High Performance Pipeline for the Automated Annotation of Plant Genomes. Front. Plant Sci. 3.

Saintenac, C., Lee, W.-S., Cambon, F., Rudd, J.J., King, R.C., Marande, W., Powers, S.J., Bergès, H., Phillips, A.L., Uauy, C., et al. (2018). Wheat receptor-kinase-like protein Stb6 controls gene-for-gene resistance to fungal pathogen Zymoseptoria tritici. Nat. Genet. 50, 368–374.

Saintenac, C., Cambon, F., Aouini, L., Verstappen, E., Ghaffary, S.M.T., Poucet, T., Marande, W., Berges, H., Xu, S., Jaouannet, M., *et al.* (2021). A wheat cysteine-rich receptor-like kinase confers broad-spectrum resistance against Septoria tritici blotch. Nat. Commun. *12*, 433.

Siah, A., Bomble, M., Tisserant, B., Cadalen, T., Holvoet, M., Hilbert, J.-L., Halama, P., and Reignault, P. (2018). Genetic Structure of *Zymoseptoria tritici* in Northern France at Region, Field, Plant, and Leaf Layer Scales. Phytopathology\* 108, 1114-1123.

Steuernagel, B., Periyannan, S.K., Hernández-Pinzón, I., Witek, K., Rouse, M.N., Yu, G., Hatta, A., Ayliffe, M., Bariana, H., Jones, J.D.G., et al. (2016). Rapid cloning of disease-resistance genes in plants using mutagenesis and sequence capture. Nat. Biotechnol. 34, 652-655.

The International Wheat Genome Sequencing Consortium (IWGSC), Appels, R., Eversole, K., Stein, N., Feuillet, C., Keller, B., Rogers, J., Pozniak, C.J., Choulet, F., Distelfeld, A., et al. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science 361, eaar7191.



# RenKSeq : Développement d'un outil innovant de criblage de gènes à haut débit pour identifier rapidement de nouveaux gènes de résistance à la septoriose

Olivier Robert \*1, David Lopez 2, Andrea Garavito 2, Aurélie Evrard\*1, Ellen Goudemand-Dugue\*1, Lamia Aouini 3, Els Verstappen 3, Burkhard Steuernagel 4, Delphine Hourcade Marcolla 5, Brande Wulff 4, Justin Faris 6, Gert Kema 3, Cyrille Saintenac 2

- 1 FLORIMOND DESPREZ VEUVE & FILS, 59242 Cappelle-en-Pévèle, France
- 2 INRAE GDEC, UMR 1095, 5 chemin de Beaulieu 63100 Clermont-Ferrand, France
- 3 Wageningen University and Research (Wageningen Plant Research, Biointeractions and Plant Health), PO Box 16, 6700AA, Wageningen, The Netherlands
- 4 JOHN INNES CENTRE, Colney Lane NR4 7UH Norwich, UK
- 5 ARVALIS INSTITUT DU VEGETAL, 3 rue Joseph et Marie Hacquin-75116 PARIS, France
- 6 United States Department of Agriculture-Agricultural Research Service, Cereal Crops Research Unit, Edward T. Schafer Agricultural Research Center, Fargo, ND, 58102, USA

 ${\bf `Coordinateur: Ellen\ Goudemand-Dugue, \it Ellen. GOUDEMAND-DUGUE@florimond-desprez. fr}$ 

**Résumé:** Le projet avait pour objectif de mettre au point une stratégie innovante pour accélérer le clonage de nouveaux gènes de résistance à la septoriose et fournir un ensemble de marqueurs diagnostiques associés à ces gènes. Cette stratégie basée sur l'approche MutRenSeq (Figure 1) développée pour le clonage de gènes de résistance à la rouille se focalisaient sur les gènes de type récepteurs kinase ou RLK, une famille de gènes impliquée dans la résistance à la septoriose. Le projet a permis de développer un essai de capture des gènes de type RLK chez le blé afin de mieux comprendre leur diversité dans les variétés. Bien que la stratégie développée ait été validée pour le clonage du gène Stb6, celle-ci n'a pas permis le clonage de nouveaux gènes de résistance. Ce projet a cependant permis l'identification de mutants sensibles, une ressource indispensable pour l'identification de gènes de résistance et également l'identification de nouveaux QTL de résistance à la septoriose chez la variété résistante Pocho.

Introduction: La septoriose est l'une des maladies les plus dommageables pour la culture du blé en France. Chaque année, les pertes sont estimées entre 350 et 700 millions d'euros pour la France. Seulement 22 gènes majeurs de résistance et 89 méta-QTL ont été cartographiés génétiquement (Brown et al., 2015). Même si certains comme Stb4, Stb9, Stb11, Stb18 ont été identifié dans des variétés Françaises, il existe encore peu de connaissances sur la présence de ces gènes au sein des variétés françaises et très peu de marqueurs diagnostiques. L'identification des gènes Stb6 et Stb16q comme des RLK suggère que cette grande famille a un rôle prépondérant dans la résistance du blé vis-à-vis de la septoriose. L'analyse de cette grande famille pourrait permettre de mieux connaitre leur diversité et également faciliter le clonage de nouveaux gènes de résistance et l'identification de marqueurs.

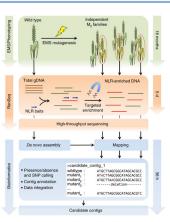


Figure 1: Stratégie MutRenSeq (Steuernagel et al. 2016)

## WP1 - Développement d'un outil de capture des gènes RLK

Afin d'être le plus exhaustif, 16 génomes de blé ont été utilisés pour rechercher les gènes de type RLK, soit en utilisant les RLK annotés par la communauté, soit en utilisant le pipeline présenté dans la figure 2. Cette analyse a permis l'identification d'un total de 45 290 RLK (Table 1) à partir desquels un design de 199 733 séquences appâts a été réalisé en collaboration avec la société Daicel Arbor.



Figure 2: Pipeline bioinformatique utilisé pour l'identification des RLK chez le blé

Assembly	Assembly size (pb)	RLK loci (pb)	RLK Loci	
Chinese Spring IWGSC	14 347 606 634	23 254 210	6 445	
Chinese Spring Earlham Institute	13 467 630 192	23 714 007	4 862	
Zavitan	10 509 868 245	20 469 051	2 584	
Cadenza	13 399 781 746	22 244 268	5 362	
Claire	14 616 451 647	21 707 537	5 357	
Kronos	10 540 467 196	14 940 058	3 706	
Paragon	15 318 048 945	22 432 660	5 431	
Robigus	14 791 409 785	21 874 636	5 281	
Svevo	6 803 298 213	8 350 510	2 444	
TGAC_WGS_durum_v1	3 386 481 085	458 483	426	
TGAC_WGS_monococcum_v1	1 322 112 030	574 717	417	
TGAC_WGS_speltoides_v1	1 907 878 214	552 288	426	
TGAC_WGS_strongfield_v1	3 384 841 931	361 995	387	
TGAC_WGS_tauschii_v1	1 200 221 237	665 015	441	
TGAC_WGS_urartu_v1	1 657 029 501	686 599	445	
TSL_WGS_sharonensis_v1	1 730 899 803	611 984	445	
Bibliography sequences		2 286 202	831	
Total	128 Gb	185 Mb	45 290	

Table 1: RLK identifié pour chacun des 16 génomes étudiés

## WP2 – Développement et phénotypage de 5 populations de tilling réalisées à partir d'accessions résistantes

- L'accession d'origine sud-américaine Pocho qui présente un large spectre de résistance à  $Zymoseptoria\ tritici$  (911  $M_2$ , 9 familles sensibles, Figure 3)
- L'accession de blé synthétique M6 qui porte le gène majeur Stb8. (513  $\mathrm{M}_2$ , 0 sensible)
- Une lignée recombinante BG301 qui porte le gène de résistance à *Parastagonospora nodorum Snn2*. (478 M<sub>2</sub>, 7 familles sensibles)
- L'accession de blé dur (*Triticum dicoccum*) PI41025 portant un gène majeur de résistance à *Z. tritici* sur le chromosome 3A. (803 M2, 0 sensible)
- L'accession Cadenza portant le gène de résistance  $\mathit{Stb6}$ . (1200  $M_5$ , 7 familles sensibles)

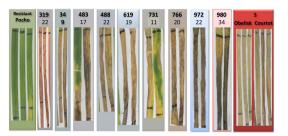


Figure 3: Phénotypes des 9 plantes  $M_2$  issues de la population de tilling Pocho à 21 dpi après inoculation avec une souche avrulente.

#### WP3 – Identification des gènes de résistance présent dans les accessions d'intérêt

L'essai de capture des RLK (WP1) a été utilisé pour séquencer les RLK des 23 familles sensibles identifiées dans le WP2 ainsi que les parents de ces populations. L'ensemble des mutations présentes dans ces RLK a été identifié ainsi que les RLK portant une mutation dans plusieurs familles sensibles de la même population par le pipeline MutantHunter. Cette analyse a permis d'identifier un seul contig portant sept mutations, une par famille sensibles et issues de la population Cadenza. Ce contig porte le gène Stb6, ce qui a permis de valider la stratégie de clonage. Cependant, aucun gène portant une mutation pour l'ensemble des familles sensibles pour les populations Pocho ou BG301 n'a été identifié. Certaines optimisations sont encore nécessaires pour lever les difficultés liées à l'étude d'une des plus grandes familles multigéniques du blé et pallier à ce manque d'efficacité.

En parallèle, une approche d'analyse QTL sur une population RIL issue de Pocho a permis d'identifier 5 QTL (2A, 3A, 5B, 7B et 7D) impliqués dans la résistance suite au phénotypage avec trois isolats de *Z. tritici* génétiquement différentes (Figure 4).

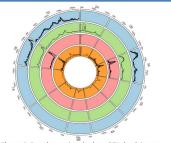


Figure 4: Représentation de deux QTL de résistance sur le chromosome 3A et 7B identifiés chez l'accession Pocho suite au phénotypage avec l'isolat IPO9415.

















