



**HAL**  
open science

# Base génétique et potentiel d'évolution de la pathogénicité de *Fusarium graminearum*, bio-agresseur fongique des céréales

Benoit Laurent

► **To cite this version:**

Benoit Laurent. Base génétique et potentiel d'évolution de la pathogénicité de *Fusarium graminearum*, bio-agresseur fongique des céréales. Sciences du Vivant [q-bio]. Université de Bordeaux, 2016. Français. NNT: . tel-02801126

**HAL Id: tel-02801126**

**<https://hal.inrae.fr/tel-02801126>**

Submitted on 5 Jun 2020

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

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

Copyright



# Base génétique et potentiel d'évolution de la pathogénicité de *Fusarium graminearum*, bio-agresseur fongique des céréales

Benoit Laurent

## ► To cite this version:

Benoit Laurent. Base génétique et potentiel d'évolution de la pathogénicité de *Fusarium graminearum*, bio-agresseur fongique des céréales. Sciences agricoles. Université de Bordeaux, 2016. Français. <NNT : 2016BORD0317>. <tel-01502557>

**HAL Id: tel-01502557**

**<https://tel.archives-ouvertes.fr/tel-01502557>**

Submitted on 5 Apr 2017

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

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

## THESE

Présentée pour obtenir le grade de

**DOCTEUR DE L'UNIVERSITE DE BORDEAUX**

ECOLE DOCTORALE des Sciences de la vie et de la santé

SPECIALITE Microbiologie et Immunologie

Par

**Benoit LAURENT**

**« Base génétique et potentiel d'évolution de la pathogénicité  
de *Fusarium graminearum*, bio-agresseur fongique des  
céréales »**

Soutenue le 07 décembre 2016,

Membres du jury :

Pr. GALLUSCI Philippe  
Dr. DUPLESSIS Sébastien  
Dr. THARREAU Didier  
Pr. MALAGNAC Fabienne  
Dr. MARULLO Philippe  
Dr. FOULONGNE-ORIOU Marie

Professeur Université de Bordeaux  
Directeur de Recherche INRA  
Directeur de Recherche CIRAD  
Professeure Université Paris Sud  
Chargé de Recherche BioLaffort  
Chargée de Recherche INRA

Président du jury  
Rapporteur  
Rapporteur  
Examinatrice  
Examineur  
Directrice de thèse



## *Remerciements*

J'aimerais profiter de ces quelques lignes pour remercier les gens qui ont rendu cette aventure scientifique, et surtout humaine, incroyable.

Je voudrais tout d'abord remercier sincèrement Sébastien Duplessis et Didier Tharreau, Fabienne Malagnac et Philippe Marullo, pour me faire l'honneur de juger ces travaux. J'espère de tout cœur que cette lecture sera intéressante et utile. Merci à Pascal Frey, Anne Genissel, Thierry Langin et Sébastien Mongrand pour m'avoir conseillé et pour avoir échangé sur ces travaux à l'occasion de mon comité de thèse. Je voudrais aussi remercier Philippe Gallusci, qui en plus d'avoir accompagné mes premiers pas dans un laboratoire de recherche, a accepté de présider la soutenance de cette thèse.

Ces travaux, tels qu'ils sont présentés, n'auraient pas été possibles sans l'aide des différents collaborateurs. Merci tout d'abord aux organisateurs du réseau « Fusatox » et du consortium « Agreenium » pour leurs initiatives. Merci à Thierry Langin pour avoir permis la collaboration entre MycSA et le GDEC, ainsi que pour sa gentillesse. Merci à Pauline Lasserre-Zuber d'avoir initié et fait prospérer cette collaboration. Pauline, je te remercie énormément pour ton accueil en terre clermontoise et pour avoir rendu cet échange si chaleureux et sympathique. Merci pour ton aide lors des manip et pour toutes ces conversations, scientifiques ou non. Merci Richard pour ta bonne humeur et pour tout le travail que tu as fourni! Merci Cyrille, pour ton aide en l'absence de Pauline. Merci pour ta sincérité et tes précieux conseils. Merci à tous les autres membres de l'équipe maladie des céréales du GDEC pour votre accueil et votre sympathie.

I should probably thank the people that have declined my proposal of working in Zurich as it allows me to spend wonderful and refreshing time in Scotland (in many senses, including the literal one). Therefore, I'm grateful of Ross Houston to have welcomed me in his team. Thank you Ross for your honesty, your kindness and your help. Thank you Chris, our Greek librarian, for all the situations you help me with, your super reactivity and your friendship. I would also thank the other members of the Roslin who made the taste of all these beers even better. Thank you Alejandro, Saïf, Carolina, Diego, Justine, Emeric, Christelle for all the good time we spent, in Edinburgh and elsewhere in Scotland.

Merci à Jean Michel Savoie et Florence Forget pour leur accueil dans l'unité MycSA. Merci d'avoir soutenu la réalisation de ces travaux, merci pour votre enthousiasme et vos encouragements. Marie, merci pour ta proximité et ton amitié. Tu as tout fait pour que je m'épanouisse et que je puisse vivre cette thèse pleinement. Merci pour ta patience et ta pédagogie, qui auront su faire face, non sans mal à ma « tétu-tortu-osité ». Je suis fier du chemin parcouru en ta compagnie et je t'en remercie. Magalie, j'en perds mon clavier... Tu es une personne exceptionnelle, merci pour ton aide immense au quotidien et surtout merci pour ton amitié. Ces moments de complicité me manqueront... Cathy, quant à toi, tu as été un véritable repère dans cet environnement doryphoresque ! Merci pour tout le (sale) boulot que tu auras fait pour moi, merci pour ton sourire et ta bonne humeur qui à eux seuls peuvent changer un thésard grognon en thésard joyeux. Je ne pourrai parler plus longtemps de doryphore sans remercier les nordistes de l'aventure, Sylvain et Enric, qui auront participé activement à ces travaux de thèse et surtout à l'oxygénation masculine, dont j'avais parfois besoin, le temps d'un repas, d'une bière ou d'un café. Merci à vous les gars. Merci aux sist'r thésardes, Mathilde la matriarche, Léa et Ccori les cadettes, merci pour tout et surtout de m'avoir lâchement abandonné au moment où j'avais besoin de vous, bien que le kit du thésard en détresse m'aura été très utile ! Merci à l'ensemble des stagiaires 2013, notamment Claire et Stephan, avec qui j'ai commencé l'aventure à MycSA et dont l'amitié aura eu raison des kilomètres et des années. Une pensée à notre Padawan Fabien qui ne sait pas dans quoi il s'est engagé mais qui a l'air de se débrouiller comme un chef. Merci Fran pour tes travaux dans cette thèse, pour ton amitié et surtout pour tes conseils d'achat d'huile d'olive. Gégé, it was unexpected to meet you and I'm grateful from unexpectedness. Thank you for your transcontinental friendship, although it didn't help to win the Copa America. Merci Nadia pour ton aide et tes précieux conseils. Merci Christian pour votre honnêteté et votre sagesse. Merci Nathalie pour tes offrandes chocolatées et houblonnées, merci Marie-France pour m'avoir materné et merci à tous les autres membres de MycSA, permanents, retraités, stagiaires et ex-stagiaires. Merci pour toutes vos qualités, votre sympathie et votre bonne humeur!

J'aimerai ensuite remercier l'ensemble des gens du centre INRA de Bordeaux, qui ont rendu le lieu si plaisant à fréquenter. Merci notamment à Joanna, Elsa et David pour avoir partagé nos aventures de thésards rédactionnant et pour votre complicité. Merci à

Justine, qui a préféré sauter les classes qu'attendre les copains. Merci de nous avoir montré la voie (tu es un peu notre Moïse à nous), merci pour ton soutien, tes conseils et pour ton amitié aux saveurs houblonnées. Merci à l'ensemble des thésards SPE et Agreenium, autres thésards/postdocs/chercheurs, fusariés ou pas, que j'ai croisé sur la route, et avec qui j'ai partagé des conférences, des discussions et des rires. Je n'oublie pas de remercier Nicolas Fayamendy, professeur au lycée Victor Duruy, qui m'a transmis la passion des sciences. Félicitation Nicolas pour votre travail auprès des étudiants.

Me voilà arrivé au clan non-professionnel et mon Ctrl-F m'annonce déjà 32 mercis, encore un beau coup des gersois. Nous sommes loin des milles mercis qui vous reviennent à vous mes amis. Merci aux potes du rugby, Julien, Guigui, Aurel et les autres, merci de m'avoir laissé des intervalles le mardi soir pour que je puisse me défouler ! Merci aux amis Landais, Jcéou pour m'accompagner depuis mes 3 printemps, et toi Charlie pour ton soutien indéfectible, pour ton amitié en or, et pour avoir participé à la rédaction de cette thèse (j'espère que tu rajouteras ça sur ton CV). Antho, comme d'hab je suis emmerdé avec toi parce je ne sais pas si je dois te mettre dans la partie collaborateur, collègue, ami, ou famille. Je ne saurai jamais te remercier assez pour ta présence quotidienne et pour tous ces moments de partage. Comme quoi on peut-être Basque et génial. Bernie, merci de me supporter quand je viens à la colloc', de t'occuper d'Antho quand je ne suis pas là (ahah), pour ton amitié et pour tes encouragements, sportifs notamment. Merci à Erik, Dugue et Chat, merci à Matthieu et Yannick, merci aux autres montois, palois, bordelais qui ont rendu ces années universitaires si belles.

Je finirai par ces quelques lignes pour remercier tout d'abord mes beaux-parents pour leur gentillesse ; ma famille, pour leur amour au quotidien et leur soutien sans faille ; et bien sûr, Marine, mon oxygène, pour ce bonheur permanent.





*« Tu vois le point là-bas, c'est l'arrivée... »*

Julien Laurent, 13<sup>e</sup> Edition de la Junior des Sables.



# *Productions scientifiques*

## **Publications à comité de lecture :**

**B. Laurent, C. Palaiokostas, C. Spataro, M. Moinard, E. Zehraoui, R. D. Houston and M. Foulongne-Oriol.** High resolution mapping of the recombination landscape of the phytopathogen *Fusarium graminearum* suggests two-speed genome evolution. Accepted for publication with minor revisions in **Molecular Plant Pathology**

**B. Laurent, M. Moinard, C. Spataro, N. Ponts, C. Barreau and M. Foulongne-Oriol.** Landscape of genomic diversity and host adaptation in *Fusarium graminearum*. Submitted in **BMC genomics**.

## **Communications:**

**B. Laurent, M. Moinard, C. Spataro, S. Chéreau, E. Zehraoui, R. Blanc, P. Lasserre-Züber, C. Palaiokostas, R. D. Houston and M. Foulongne-Oriol (2016):** How the genomic era is putting quantitative genetics back under the spotlight: a case study of *Fusarium graminearum*. 13<sup>th</sup> European Fusarium Seminar, Paris, France. [Poster]

**B. Laurent, C. Palaiokostas, C. Spataro, M. Moinard, R. D. Houston and M. Foulongne-Oriol (2016):** A new linkage map of *Fusarium graminearum*. *Fusarium Workshop*; 13<sup>th</sup> European Fusarium Seminar, Paris, France. [Oral]

**B. Laurent, C. Palaiokostas, C. Spataro, M. Moinard, J.M. Savoie, R. D. Houston and M. Foulongne-Oriol (2016):** Using restriction site associated DNA-sequencing to construct the first high density genetic map of *Fusarium graminearum*. 11<sup>èmes</sup> Rencontres de Phytopathologie, Aussois, France. [Poster]

**B. Laurent, E. Zehraoui, S. Chéreau, M. Moinard, C. Spataro, N. Ponts, P. Lasserre-Zuber, R. Blanc, C. Palaiokostas, R. D. Houston et M. Foulongne-Oriol (2016) :** Variabilité du pouvoir pathogène de *F. graminearum* sur blé : quelques pistes? FusaTox meeting, Bordeaux, France. [Oral]

**B. Laurent, M. Moinard, C. Spataro, P. Lasserre-Zuber, N. Ponts and M. Foulongne-Oriol (2015):** The added value of -omic approach to dissect phytopathogenicity of *Fusarium*

*graminearum*: from genomic to genetic. 28<sup>th</sup> Fungal Genetics Conference, Asilomar, USA.  
[Poster]

**B. Laurent, N. Ponts, V. Atanasova, C. Barreau, M. Moinard, C. Spataro et M. Foulongne-Oriol (2014)** : Vers le décryptage d'un caractère complexe: Le potentiel toxigène de *Fusarium graminearum*. 6<sup>e</sup> Journées des doctorants SPE, Bordeaux, France. [Oral]

**B. Laurent, E. Zehraoui, N. Ponts, V. Atanasova, C. Barreau and M. Foulongne-Oriol (2013)** : *Fusarium* toxinogenic potential : A complex trait decrypted by QTL analysis. FusaTox meeting, Paris, France. [Oral]

**B. Laurent, N. Ponts, V. Atanasova-Penichon, C. Barreau and M. Foulongne-Oriol (2013)**: Disentangling mycotoxin regulatory pathways in *Fusarium graminearum* by quantitative genetics. 12<sup>th</sup> European Fusarium Seminar, Bordeaux, France [Poster]

# Tables des matières

## **Introduction Générale : page 1 – page 63**

<b>I. Les enjeux liés à l'étude.....</b>	<b>1</b>
<b>I.1. La sécurité alimentaire .....</b>	<b>1</b>
<b>I.2. Les céréales : une ressource alimentaire majeure.....</b>	<b>2</b>
<b>I.3. La santé des plantes, et des consommateurs.....</b>	<b>3</b>
<b>II. Les fusarioses des céréales.....</b>	<b>5</b>
<b>II.1. La fusariose des céréales et la contamination des grains en mycotoxines.....</b>	<b>5</b>
<b>II.2. L'apparition des fusarioses et des contaminations en mycotoxines : moyens de         lutttes.....</b>	<b>9</b>
<b>III. <i>Fusarium graminearum</i> et la fusariose des épis.....</b>	<b>14</b>
<b>III.1. Historique des recherches sur <i>Fusarium graminearum</i>.....</b>	<b>14</b>
<b>III.2. Précisions phylogénétiques.....</b>	<b>16</b>
<b>III.3. Le cycle de vie de <i>F. graminearum</i>.....</b>	<b>18</b>
<b>IV. La phase biotique des champignons phytopathogènes.....</b>	<b>21</b>
<b>IV.1. Bases génétiques et moléculaires de la relation hôte-pathogène.....</b>	<b>21</b>
<b>IV.2. Les effecteurs fongiques : le cas des métabolites secondaires.....</b>	<b>23</b>
<b>IV.3. Biosynthèse et rôle biologique des trichothécènes de types B produits par         <i>F. graminearum</i>.....</b>	<b>23</b>
<b>V. Les apports du génome de <i>F. graminearum</i> pour l'étude de la     pathogénicité.....</b>	<b>27</b>
<b>V.1. Un large catalogue d'effecteurs potentiels.....</b>	<b>28</b>
<b>V.2. Les différentes reprogrammations de l'expression du génome.....</b>	<b>28</b>
<b>V.3. La diversité génétique chez <i>F. graminearum</i>.....</b>	<b>29</b>
<b>VI. Vers une évolution du pathosystème <i>F. graminearum</i> x céréale ? .....</b>	<b>31</b>
<b>VI.1. L'évolution des relations hôte-pathogène.....</b>	<b>31</b>
<b>VI.2. Comment étudier les bases génétiques de l'agressivité ? .....</b>	<b>34</b>
<b>VI.3. L'agressivité chez <i>F. graminearum</i>? .....</b>	<b>58</b>
<b>VII. L'hypothèse et les objectifs de la thèse.....</b>	<b>61</b>

# **Chapitre 1 : Caractérisation de la diversité génétique et phénotypique de souches de *F. graminearum* : page 64 – page 116**

<b>I. Etudes préliminaires de la collection de <i>F. graminearum</i></b> .....	<b>64</b>
<b>I.1.</b> Caractérisation génétique et phénotypique de <i>F. graminearum</i> .....	64
<b>I.2.</b> Vers la sélection de six souches de <i>F. graminearum</i> .....	67
<b>I.3.</b> Du modèle <i>in vitro</i> vers le modèle <i>in planta</i> .....	69
<b>I.4.</b> Matériels et méthodes.....	72
<b>II. Diversité génomique de <i>F. graminearum</i> et rôle pour l'adaptation à l'hôte; article soumis à « BMC genomics »</b> .....	<b>75</b>
<b>II.1.</b> Résumé.....	76
<b>II.2.</b> Introduction.....	77
<b>II.3.</b> Résultats.....	79
<b>II.3.1.</b> Identification de SNPs et InDels.....	79
<b>II.3.2.</b> Distribution génomique des variants.....	80
<b>II.3.3.</b> Annotation fonctionnelle des variants.....	83
<b>II.3.4.</b> Fonctions biologiques pouvant être affectées par les variants.....	85
<b>II.3.5.</b> Gènes montrant un excès de mutations non-synonymes.....	90
<b>II.4.</b> Discussion.....	92
<b>II.4.1.</b> Le polymorphisme génomique de <i>F. graminearum</i> est cohérent avec le style de vie d'un pathogène.....	92
<b>II.4.2.</b> Les multi-niveaux du polymorphisme dans le génome.....	93
<b>II.4.3.</b> Des preuves supplémentaires d'une organisation à deux-vitesses.....	95
<b>II.4.4.</b> La plus-value des données de génomique pour l'étude des gènes impliqués dans la pathogénicité.....	96
<b>II.5.</b> Conclusion.....	97
<b>II.6.</b> Procédures expérimentales.....	97
<b>II.7.</b> Figures et tableaux additionnels.....	101
<b>II.8.</b> Références bibliographiques.....	107
<b>III. Le séquençage des génomes : stratégies et enjeux</b> .....	<b>113</b>
<b>IV. Conclusion du chapitre 1</b> .....	<b>116</b>

**Chapitre 2 : La recombinaison méiotique chez *F. graminearum* : un mécanisme évolutif majeur, à la base de la cartographie génétique :  
page 117 – page 170**

<b>I. Construction de la population recombinante.....</b>	<b>117</b>
I.1. Le contrôle de la reproduction sexuée chez <i>F. graminearum</i> .....	117
I.2. Construction de souches hétérothalliques.....	119
I.3. Les tentatives de croisement.....	120
I.4. Matériels et méthodes.....	121
I.5. Une carte génétique : un outil unique pour des applications multiples.....	124
<b>II. La cartographie à haute résolution des profils de recombinaison chez le phytopathogène <i>F. graminearum</i> apporte de nouvelles preuves sur l'organisation à deux-vitesses de son génome ; article soumis à « Molecular Plant Pathology » .....</b>	<b>125</b>
II.1. Résumé.....	126
II.2. Introduction.....	127
II.3. Résultats.....	129
II.3.1. Souches et géotypages.....	129
II.3.2. Analyse de ségrégation, construction de la carte génétique et alignement sur le génom de référence.....	131
II.3.3. Profils de recombinaison.....	133
II.3.4. Enrichissement fonctionnel des séquences des régions hautement recombinantes.....	134
II.4. Discussion.....	137
II.4.1. La cartographie des crossovers révèle des profils de recombinaison chromosome-spécifiques.....	138
II.4.2. Caractérisation des régions génomiques recombinantes.....	139
II.4.3. Les régions chromosomiques recombinantes sont liées à plusieurs points chauds de recombinaison.....	140
II.4.4. Le rôle potentiel des profils de recombinaisons pour l'évolution du pathogène.....	140
II.5. Procédures expérimentales.....	142
II.6. Bibliographie.....	145
II.7. Informations complémentaires.....	151
<b>III. Apports technologiques et scientifiques.....</b>	<b>157</b>
III.1. Les mitochondries sont conservées.....	157
III.2. Validation de l'utilisation de souches $\Delta Mat$ .....	157
III.3. Les améliorations apportées sur la carte génétique.....	159
III.4. L'apport du géotypage à haut-débit.....	160
III.4.1. Le principe du RAD-seq.....	161
III.4.2. Quelques facteurs importants à considérer lors du géotypage par RAD- seq.....	163
III.4.3. Les implications pour la suite de l'analyse.....	168

IV. Matériels et méthodes.....	169
V. Conclusion du chapitre 2.....	170

**Chapitre 3 : Elucidation des bases génétiques de la variation de production de mycotoxines et d'agressivité chez *F. graminearum* :  
page 171 – page 212**

I. Une approche combinée de détection de QTL et de génomique comparative identifie une mutation sur le gène <i>FgVeA</i> avec un effet majeur sur l'agressivité de <i>F. graminearum</i> .....	173
I.1. Résumé.....	172
I.2. Introduction.....	173
I.3. Résultats.....	178
I.3.1. Phénotypage et distribution des caractères au sein de la population expérimentale.....	176
I.3.2. Corrélation entre caractères.....	179
I.3.3. Analyse de variance et estimation de l'héritabilité.....	179
I.3.4. Cartographie de QTL.....	181
I.3.5. Analyse <i>in silico</i> de la mutation causale.....	182
I.3.6. Validation fonctionnelle du gène <i>FgVeA</i> .....	184
I.4. Discussion.....	185
I.4.1. Vers une meilleure compréhension de l'agressivité.....	185
I.4.2. La variation d'agressivité est due à un QTL majeur à effet pléiotropique...187	
I.4.3. Du QTL au gène causal : la plus-value des approches NGS pour l'identification de gènes candidats.....	188
I.4.4. Implication du polymorphisme génétique du gène <i>FgVeA</i> pour la variation d'agressivité et le potentiel de production en TCTB.....	189
I.5. Conclusion.....	190
I.6. Matériels et méthodes.....	191
I.7. Références.....	196
II. Un caractère quantitatif et un gène majeur : un paradoxe ? .....	202
II.1. Quelle est la fréquence de l'allèle identifié au gène <i>FgVeA</i> dans les souches de <i>F. graminearum</i> . .....	202
II.2. Est-ce que d'autres QTL contrôlent la variation d'agressivité observée dans la population recombinante ? .....	203
I.1. D'autres polymorphismes affectant le gène <i>FgVeA</i> .....	204
III. Des indices supplémentaires sur le rôle du polymorphisme de <i>FgVeA</i> .....	206
I.2. L'allèle muté de <i>FgVeA</i> est lié à l'accumulation d'un métabolite intermédiaire des TCTB ? .....	207



I.3. Est-ce que la souche INRA-156 et INRA-171 sont compétitives ? .....	207
IV. Matériels et méthodes.....	209
V. Conclusion du chapitre 3.....	212

**Discussion générale et perspectives : page 213 – page 239**

I. Un premier bilan .....	213
II. Quels sont les éléments régulateurs de la recombinaison chez <i>F. graminearum</i> ?.....	216
III. La cartographie génétique de QTL, une approche pertinente pour l'étude de l'agressivité chez <i>F. graminearum</i> ? .....	220
IV. Quel est le rôle des technologies de séquençage dans le regain d'intérêt de la cartographie de QTL ? .....	223
V. L'étude des génomes : un résultat parallèle ou complémentaire? .....	224
VI. Quel est le rôle de <i>FgVeA</i> chez <i>F. graminearum</i> ?.....	225
VII. Le polymorphisme de <i>FgVeA</i> , quelles implications moléculaires ? .....	228
VIII. Un allèle rare de <i>FgVeA</i> , un impact sur le fitness des souches ? .....	230
IX. Quel est le potentiel d'évolution de <i>FgVeA</i> ? .....	232
X. La génétique de l'agressivité chez <i>F. graminearum</i> : un bilan actualisé des connaissances.....	233
XI. Les prochains défis pour une meilleure compréhension de l'agressivité et de son évolution .....	236

**Conclusion générale : page 239**

**Bibliographie : page 240 – page 261**

**Annexes: page 262– page 276**

- I. Rapport de Stage Fatima Z. Shailia: Etude de la diversité génétique et phénotypique de *F. graminearum*, producteur de toxines dans les céréales.....263
- II. Les scripts de bio-informatique .....264

# Liste des tableaux et figures

## Introduction générale

**Table 1** (Article): Liste non exhaustive des analyses de génétiques quantitatives chez les champignons : p44

**Tableau I\_1**: Liste de gènes précédemment associés à des variations du niveau d'agressivité : p60

**Figure I\_1** : Régime journalier moyen dans le monde par individu en 2011, calculé à partir des données de FAOSTAT : p2

**Photo I\_1**: un épi de blé (variété Roysac) infecté par *Fusarium graminearum*. © B. LAURENT : p5

**Figure I\_2** : principales familles et espèces de *Fusarium* productrices de fusariotoxines : p8

**Figure I\_3** : Grille d'évaluation du risque d'accumulation du déoxynivalénol (DON) dans le grain de blé tendre et d'aide au traitement contre la fusariose sur épi proposée par Arvalis : p12

**Figure I\_4** : Evolution du nombre de publications identifiées par le moteur de recherche spécialisé Web of Science comprenant « *Fusarium graminearum* » ou « *Gibberella zeae* » (forme sexuée de ce pathogène) dans le titre, et dates clés : p 15

**Figure I\_5** : Arbre phylogénétique du *Fusarium graminearum* species complex (FGSC) proposé par Yli-Mattila et al. 2009 (à gauche) et du genre *Fusarium* proposé par Aoki et al. 2014 (à droite) : p17

**Figure I\_6** : Cycle de vie de *F. graminearum*, associé au développement de la fusariose des épis chez le blé, inspiré de Trail et al. 2009 : p20

**Figure I\_7** : A] illustration des gènes *Tri* (localisés dans le cluster et hors du cluster) et B] schéma simplifié de la voie de biosynthèse des trichothécènes de type B : p 24

**Figure I\_8** : schéma simplifié de la stratégie d'infection adoptée par *F. graminearum*, mettant en lumière le rôle du déoxynivalénol dans la propagation du pathogène de l'épillet à l'épi chez le blé. Adapté de Brown et al. 2010 et Walter et al. 2010 : p26

**Figure I\_9**: Un bref historique sur les efforts apportés à la publication et la révision du génome de référence de *F. graminearum* : p27

**Figure I\_10** : Diagramme de Venn faisant état du transcriptome en fonction de l'hôte infecté : p29

**Figure I\_11** : Représentation de la densité en SNPs et du taux de recombinaison et de la localisation de certains types de gènes sur les quatre chromosomes de *F. graminearum* :p30

**Figure I\_12**: Mécanisme moléculaire du modèle en "zig-zag" : p32

**Encadré I\_1** : Les cinq forces évolutives bénéfiques à l'adaptation des pathogènes fongiques : p33

**Figure 1** (Article): Illustration schématique de la détection de QTL utilisant la stratégie classique de cartographie ou de GWA, adapté de Mackay et al. ,2009 : p43

**Figure I\_13**: Organisation des travaux de thèses : p63

## Chapitre 1

**Tableau CI\_1**: Polymorphisme estimé sur 20 marqueurs SSR génotypés : p69

**Table 1** (Article): Variant calling statistics, considering strain-specific reads and considering total reads produced. : p80

**Table 2** (Article): Genome-wide comparison of variants between pairs of isolates : p81

**Table 3** (Article): Significant ( $p$ -value < 0.01) gene ontology enrichment of the categories built from their variant contents and downstream coding-effect : p86

**Table 4** (Article): Putative effectors showing an excess of non-synonymous effect mutations : p90

**Table 5** (Article): Phenotypic information and geographical origin for the six strains : p91

**Figure CI\_1** : Arbre construit par Neighbor-Joining basé sur les distances euclidiennes calculées à partir de 20 marqueurs microsatellites génotypés sur 66 souches de *F. graminearum* isolées en France et 19 issues d'autres collections internationales : p66

**Figure CI\_2** : Concentration en TCTB ( $\mu$ g de toxines par g de biomasse fongique) pour les différentes souches INRA cultivées en milieu synthétique contenant du glucose : p67

**Figure CI\_3** : Notation des symptômes visuels, concentrations en TCTB, épis inoculés pour chaque souche, souches cultivées sur milieu solide de PDA, souches cultivées sur milieu Carrot agar : p70

**Figure CI\_4** : Graphique représentant les niveaux d'agressivités de la souche INRA-156 et INRA-171 observés chez 6 variétés de blé, contrastées pour leurs niveaux de résistance à la fusariose (échelle en bleue) : p71

**Figure 1** (Article): Average variant density by strain for the four chromosomes and the mitochondrial genome : p81

**Figure 2** (Article): Profiles of variant distribution by chromosome : p83

**Figure 3** (Article): Distribution of average variant content per gene per genome : p84

**Figure 4** (Article): Variant effect prediction and subsequent gene classification : p85

**Figure 5** (Article): Heatmap representation of variant and gene counts per 100 kb-long non-overlapping windows : p89

**Figure 6** (Article): Selected *F. graminearum*-specific gene content of each category of predicted variant effect : p89

**Figure CI\_5**: Les deux différentes stratégies pour identifier des variations génétiques à partir des séquences produites par les technologies NGS : p114

## Chapitre 2

**Tableau CII\_1**: Les différentes combinaisons testées pour la construction de la population recombinante : p120

**Tableau CII\_2**: Les amorces utilisées pour la délétion du gène *Mat 1-2-1* : p123

**Table 1** (Article): General features of genotyping procedure and linkage map construction: p132

**Tableau CII\_3**: Comparaison avec les cartes génétiques précédemment construites : p160

**Figure CII\_1**: base génétique de l'homothallisme chez *F. graminearum* et rôle supposé des différents gènes *Mat* : p118

**Photo CII\_1**: A gauche, INRA-156 $\Delta$ *mat* et à droite INRA-156 WT, cultivés dans les mêmes conditions sur milieu Carot Agar induisant le développement de périthèces : p120

**Figure CII\_2**: Stratégie expérimentale pour la délétion du gène *Mat 1-2-1* : p123

**Figure 1** (Article): Circos plot of the distribution of several genomic attributes and genotyping features along the four chromosomes of *F. graminearum* : p130

**Figure 2** (Article): Alignment of the linkage map (left side) to the reference genome RRES v4.0 (right side) : p132

**Figure 3** (Article): Scatterplot showing the linkage map position (y-axis) and physical position (x-axis) : p133

**Figure 4** (Article): Circos plot of the distribution of several genomic sequence features along the four chromosomes of *F. graminearum* : p136

**Figure 5** (Article): A. Recombination rate distribution according to chromosomal section. B. The percentage of total size of chromosome allocated to recombinant-active sections is given in red, and the chromosome average recombination rate (cM/Mb) in grey : p137

**Figure CII\_3**: Comparaisons phénotypiques entre la souche INRA-156 sauvage, la souche INRA-156 $\Delta$ *mat* délétée pour le gène *mat1-2-1* et la souche INRA-171 : p159

**Figure CII\_4**: la stratégie de génotypage par séquençage (illustrée et proposée par Integrated DNA technologies®) : p162

**Figure CII\_5**: Distribution de la taille des fragments après la digestion simulée des génomes de la souche INRA-156 et INRA-171 avec l'enzyme de digestion *Pst*I : p164

**Figure CII\_6** : Exemple de profils retrouvés après alignement des séquences issues du RAD-seq de la population recombinante sur le génome de référence (RRES v4.0) : p165

**Figure CII\_7** : Le nombre de marqueurs génotypés observé dans l'intervalle entourant les sites de coupures *PstI* prédits et le nombre théorique de marqueurs : p165

**Figure CII\_8** : Relation entre la taille des fragments de restriction et la profondeur de séquençage observée à partir d'une expérience de RAD-seq utilisant *PstI* chez *C. elegans* : p166

## **Chapitre 3**

**Table 1** (Article): Minimal, maximal, median and mean values within the progeny and for the parents and Spearman rank order correlation between the experiments : p178

**Table 2** (Article): Spearman rank order correlation values between the different phenotypes, for the different trials : p179

**Table 3** (Article): Analyses of variance and heritability estimates ( $h^2$ ) for the aggressiveness-related traits measured *in planta* and *in vitro* from combined experiment data : p180

**Table 4** (Article): QTL associated with phenotype variation under experiment condition and mapping of the morphological Mendelian trait : p182

**Tableau CIII\_1**: Les amorces utilisées pour le génotypage Kasp™ : p210

**Tableau CIII\_2** : amorces utilisées pour tester la présence de souche dans l'épi : p211

**Figure 1** (Article) : Phenotypic characteristics of parental strains : p177

**Figure 2** (Article) : Distribution of the different phenotypes assayed in the offspring : p178

**Figure 3** (Article) : Analytical process of the *in silico* analysis of the genomic region of interest : p183

**Figure 4** (Article) : Identification of the causal mutation : p184

**Figure 5** (Article) : Phenotyping data of INRA-156 strain, INRA-171 strain and INRA-171cFgVeA strain, of which the *FgVeA* allele has been replaced by the allelic version of INRA-156 strain : p185

**Figure CIII\_1** : Représentation en boîtes à moustache de la production de TCTB (en µg/mL) en fonction de l'allèle au locus *FgVeA*. « A » est l'allèle sauvage et « H » est l'allèle alternatif présenté par la souche INRA-195 : p205

**Figure CIII\_2** : Illustration du pic inconnu identifié en spectrométrie de masse (détection UV) : p206

**Figure CIII\_3** : Symptômes visuels sur épis de blés (3 plantes x 3 épis / souche) en inoculation souche seule ou en co-inoculation : p207

**Figure CIII\_4** : Confirmation de la présence de chaque souche dans les épis après amplification par PCR de marqueurs spécifiques : p208

## **Discussion générale et perspectives**

**Figure D\_1:** Les différences entre la formation des crossovers et les non-crossovers (<http://www.microbiology.columbia.edu/symington/resolvases.html>) : p217

**Figure D\_2 :** Représentation du chromosome I. A] le nombre de CO par fenêtre différentes de 100kb observé durant notre analyse B] Résultat (en séquence par kb) du séquençage après immunoprécipitation de la chromatine (ChIP-seq) ciblant les marques H<sub>3</sub>K<sub>27</sub>me<sub>3</sub> (en orange) et les marques H<sub>3</sub>K<sub>4</sub>me<sub>2</sub> (en vert), proposé par Conolly et al. (2013) : p219

**Figure D\_3 :** modèle de fonctionnement des protéines Velvet et du complexe Velvet chez *A. nidulans* (Bayram et al. 2008) : p227

**Figure D\_4 :** Localisation du locus *FgVeA* sur le chromosome I (étoile noire, 4 786 462 – 4 788 117 pb), en rouge le nombre de variants/kb moyen calculé par fenêtre de 100kb, en bleu le nombre de CO (en dizaine) détecté tous les 100 kb : p232

**Figure D\_5 :** Localisation des différents gènes candidats sur les chromosomes associés à l'agressivité de *F. graminearum* identifiés lors des études précédentes et dans cette étude (échelle en Mb) : p235

## Tables des abréviations

15-ADON :	15-acetyl-deoxinivalenol
3-ADON :	3-acetyl-deoxynivalenol
ADN :	Acide désoxyribonucléique
AFLP :	Amplified fragment length polymorphism
ANOVA :	Analysis of variance
ARN :	Acide ribonucléique
CAPS :	Cleaved amplified polymorphic sequence
CIM :	Composite interval mapping
cM :	CentiMorgan
CO :	Crossover
DON :	Deoxynivalenol
DSB :	Double strand break
EFSA :	European Food Safety Authority
FANCM :	Fanconi anemia, complementation group M
FAO :	Food and Agriculture Organization of the United Nations
FHB :	Fusarium Head Blight
Fx :	Fusarenone X
GWA :	Genome wide association
GWAS :	Genome wide association studies
KASP™ :	Kompetitive Allele Specific PCR
INRA :	Institut national de la recherche agronomique
InDel :	Insertion/Deletion
Mb :	Mégabases
mRNA :	Messenger ribonucleic acid
rRNA :	ribosomal ribonucleic acid
MycSA :	Mycologie et sécurité des aliments
NIV :	Nivalenol
pb :	Paire de base
PCR :	Polymerase chain reaction
QTL :	Quantitative trait <i>locus/loci</i>
QTN :	Quantitative trait nucleotide
ROS :	Reactive oxygen species
SIM :	Simple interval mapping
SNP :	Single-nucleotide polymorphism
SSR :	Simple sequence repeats
TCTB :	Type-B trichothecenes
TCTA :	Type-A trichothecenes



# Préface

Les céréales sont une composante majeure de l'alimentation humaine et animale. Il est donc important d'en assurer et d'en pérenniser la production. *Fusarium graminearum* est un des agents majoritaires responsables de la fusariose des épis des céréales. L'infection des céréales par ces pathogènes amène à des pertes de rendement au champ ainsi que des contaminations des grains par des métabolites fongiques toxiques appelés les mycotoxines. L'occurrence de *F. graminearum*, et ses conséquences, sont problématiques : les méthodes de lutttes actuelles ne protègent pas contre l'apparition de ce pathogène au champ et contre les contaminations des grains en mycotoxines. Une meilleure connaissance de la biologie du pathogène est donc nécessaire pour développer des stratégies de lutte plus efficace et durable. Les enjeux liés aux recherches sur ce pathogène, ainsi qu'un état actuel des connaissances en lien avec ces travaux, seront présentés dans une première partie introductive.

Ces travaux de thèses ont été initialement conçus pour étudier le déterminisme génétique de la production de mycotoxines par *F. graminearum*, testée dans des conditions *in vitro*. Les collaborations réalisées au cours de cette période doctorale ont permis d'étendre ce sujet à un modèle végétal. Les travaux présentés dans la suite du manuscrit proposent d'étudier le déterminisme génétique de la pathogénicité de *F. graminearum* lors de son interaction avec le blé. Ces travaux sont motivés par l'acquisition de nouvelles connaissances sur les éléments régissant l'interaction entre le pathogène et son hôte, afin notamment de mieux en comprendre son fonctionnement et son évolution.

Pour ce faire, une démarche de cartographie génétique des facteurs de pathogénicité chez *F. graminearum* a été initiée. Ces travaux sont développés selon trois chapitres principaux, proposant les résultats en lien aux trois étapes clés classiquement développées pour mener à bien ce genre de stratégie expérimentale : i) la sélection de souches parentales potentielles ii) l'obtention d'une population recombinante pour la construction d'une carte génétique iii) la cartographie des *loci* (QTL, pour quantitative

trait *loci*) en lien à la variation d'agressivité observée dans la population expérimentale. Pour chaque chapitre, les résultats principaux sont présentés sous forme de publication, soumise ou en cours de rédaction. Certains résultats complémentaires sont également décrits dans des sous-chapitres. Pour faciliter la lecture, les procédures expérimentales sont présentées à la fin de chaque sous-chapitre.

Des données obtenues durant le développement de cette approche ont apporté des indications supplémentaires sur la dynamique du génome de *F. graminearum*. L'étude du déterminisme génétique de l'agressivité, et celle de la dynamique des génomes, apportent toutes deux des informations intéressantes sur le potentiel d'évolution de ce pathogène. Ce dernier point sera notamment discuté en fin de manuscrit.

# **Introduction générale**



## I. Les enjeux liés à l'étude

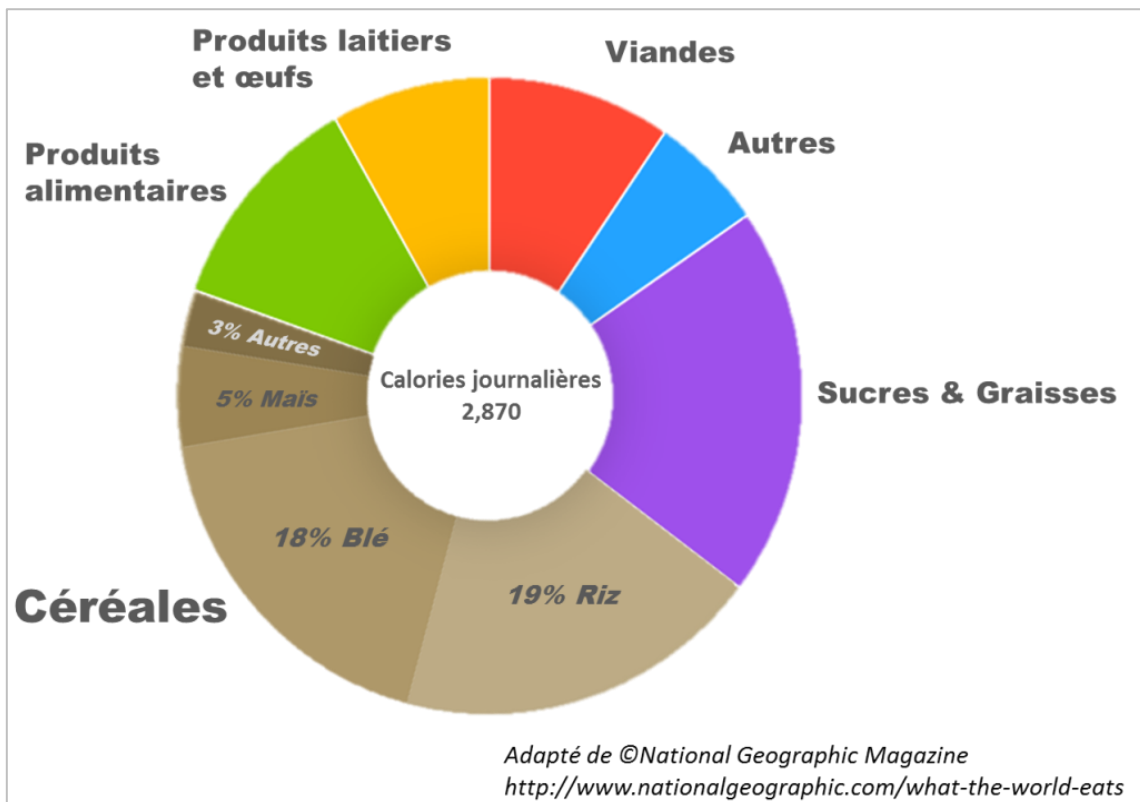
### I.1. *La sécurité alimentaire*

Selon l'Organisation des Nations Unies pour l'Alimentation et l'Agriculture (Food and Agriculture Organization of the United Nations, FAO) : « La sécurité alimentaire existe lorsque tous les êtres humains ont, à tout moment, un accès physique et économique à une nourriture suffisante, saine et nutritive leur permettant de satisfaire leurs besoins énergétiques et leurs préférences alimentaires pour mener une vie saine et active » (Sommet mondial de l'alimentation, 13 - 17 Novembre 1996 Rome). Suivant cette définition : deux composantes de la sécurité alimentaire sont distinguables : la sous-nutrition et la malnutrition (Baudin 2014). Malgré des améliorations observées depuis les dix dernières années, environ 795 millions de personnes sont toujours sous-alimentées dans le monde. Plus du double (~2 milliards) souffriraient de malnutrition (FAO 2015). Des efforts sont donc encore nécessaires pour assurer la sécurité alimentaire à l'échelle de la planète. Cet enjeu est confronté à de nouveaux objectifs: i) l'augmentation des productions alimentaires de ~70% pour l'horizon 2050 afin de répondre à la croissance démographique et l'urbanisation des populations (FAO 2009) ii) limiter l'impact écologique du développement agricole sur l'environnement (FCCC/CP/2015/L.9).

Dans le monde, de nombreux *consortia* et instances politiques font du développement d'une agriculture saine, durable et suffisante, leur priorité. L'union européenne a ainsi débloqué, pour l'année 2016-2017, ~141 millions d'euros pour la recherche sur ce sujet (Programme de travail 2016-2017, Décision de la Commission Européenne C(2015)6776 du 13 Octobre 2015). Ces programmes se traduisent plus localement par l'action des différents instituts nationaux. En France, l'Institut National de la Recherche Agronomique (INRA) propose comme une de ses quatre priorités d' « assurer des systèmes alimentaires sains et durable » (Document d'orientation INRA 2010 - 2020). Ces travaux de thèse s'inscrivent dans cette démarche, en proposant des connaissances dont le but est d'augmenter et pérenniser les rendements des productions céréalières, ainsi que d'en assurer la qualité sanitaire.

## 1.2. *Les céréales : une ressource alimentaire majeure*

La culture des céréales a joué un rôle important dans l'histoire et l'évolution de nos sociétés. Elle aurait notamment permis la sédentarisation des populations et leur expansion (Garel 2015). Du fait de leur importance et de leur richesse nutritive, les céréales sont considérées comme la ressource alimentaire actuelle la plus importante au monde (Sarwar et al. 2013; Passions Céréales 2012). Sur une moyenne de 2 870 calories journalières, environ 45% sont apportées par la consommation de céréales ; parmi lesquelles le blé et le riz sont les céréales les plus consommées (Figure I\_1).



**Figure I\_1** : Régime journalier moyen dans le monde par individu en 2011, calculé à partir des données de FAOSTAT (<http://faostat3.fao.org>).

En 2013, 2,8 milliards de tonnes de céréales ont été produites et ont généré, via leur commerce, ~126 milliards de dollars américains (USD, FAOSTAT). Il existe de nombreuses espèces de céréales cultivées, dont les plus produites sont, en poids, le maïs (36.9% de la production totale), le riz (26,7%), le blé (25,8%) et l'orge (5,2%). Ces

chiffres, en décalage avec les données de consommation humaine (Figure I\_1), traduisent l'importance de l'utilisation du maïs et du blé pour l'alimentation animale. Alors que la majorité de la production céréalière est destinée à l'alimentation humaine dans les pays en voie de développement, une grande partie de la production dans les pays développés est destinée à l'alimentation animale, et répond à l'évolution de la consommation en viande dans ces pays.

En France, les enjeux liés à la production céréalière sont aussi économiques. Avec 33 millions de tonnes exportées en 2013, notre pays est le deuxième exportateur mondial après les Etats-Unis (FAOSTAT). La culture du blé, qui constitue à elle seule 50% des exportations céréalières françaises, génère annuellement ~5 milliards de dollars (USD) et représente le deuxième marché agricole national après la vente du vin (FAOSTAT).

### *1.3. La santé des plantes, et des consommateurs*

La santé des plantes est un facteur déterminant pour les rendements agricoles (Flood 2010). En effet, les cultures sont soumises à de nombreuses pressions biotiques et/ou abiotiques, qui peuvent affecter grandement la productivité agricole mondiale. En moyenne, 10% à 16% des récoltes mondiales seraient perdues à cause des pressions biotiques (Chakraborty & Newton 2011) et ces estimations cachent des disparités importantes en fonction de la plante cultivée et de la région considérée (Chakraborty & Newton 2011). Alors qu'une éventuelle hausse de productivité rendrait les cultures plus sensibles aux attaques d'agents pathogènes (appelés « pathogènes » dans la suite du manuscrit) (Oerke 2006; Bebbber et al. 2014), les changements climatiques engendreraient une augmentation globale des pressions parasitaires (Chakraborty & Newton 2011). Au-delà des pertes directes au champ, celles cumulées de la récolte jusqu'à la consommation peuvent varier de 25 à 55% (Flood 2010). Parmi ces pertes indirectes, on distingue celles liées à la dégradation de la qualité sanitaire des aliments. Ce cas concerne notamment certains pathogènes fongiques qui contaminent les aliments avec des métabolites toxiques, de faible poids moléculaires, physiquement et chimiquement stables: les mycotoxines (Reddy et al. 2010). On estime que 25% de la production alimentaire mondiale issue des cultures végétales contiendraient des

mycotoxines (Reddy et al. 2010). Les mycotoxines considérées comme les plus problématiques sont les aflatoxines, l'ochratoxine A, la patuline, le déoxynivalénol, la zéaralénone, les fumonisines et les toxines T-2 et HT-2 (Reddy et al. 2010; Charmet et al. 2016). Ces métabolites présentent des toxicités variées : hépatotoxique, cancérigène, neurotoxique, perturbateur endocrinien, immunotoxique ou néphrotoxique (Reddy et al. 2010). Elles sont produites majoritairement par les espèces fongiques du genre *Aspergillus* et *Penicillium* pour les trois premières, et par les espèces du genre *Fusarium* pour les autres. Chez l'homme, si les cas de toxicité aiguë due à une ingestion d'aliments très fortement contaminés sont relativement rares, les problèmes liés à une exposition chronique demeurent préoccupants (Charmet et al. 2016). En ce sens, un guide pour limiter l'occurrence des mycotoxines a été proposé par la FAO (World Health Organization 2012), et des législations ont été mises en place dans le monde pour définir, pour certaines, leurs teneurs maximales admissibles dans l'alimentation humaine (Régulation EC No 1881/2006 en Europe). Pour l'alimentation animale, seule la présence d'aflatoxine est réglementée en Europe (2006/576/CE) et des recommandations existent pour les autres (2002/32/CE).

Les cultures céréalières subissent de nombreuses pressions biotiques. Par exemple, en France, les céréales sont notamment soumises à la septoriose, l'oïdium, les piétins et rhizoctones, les rouilles (brune ou jaune), les charbons, l'helminthosporiose ou encore, et faisant l'objet des travaux de cette thèse, les fusarioses (Charmet et al. 2016). Toutes ces maladies sont provoquées par des agents fongiques dont certains produisent aussi des mycotoxines (Reddy et al. 2010). Ainsi, l'attaque des céréales par les espèces du genre *Fusarium* est particulièrement préoccupante puisque qu'elle va générer à la fois des pertes de rendements directes au champ, par le développement de fusarioses ; mais aussi indirectes via les contaminations en mycotoxines, parmi lesquelles cinq des mycotoxines majeures citées précédemment.



# II. Les fusarioses des céréales

## II.1. *La fusariose des céréales et la contamination des grains en mycotoxines*

La fusariose peut affecter tous les compartiments de la plante et peut se développer chez la majorité des céréales cultivées. Le terme générique de fusariose peut s'accorder en fonction des compartiments affectés, et sa traduction anglophone peut varier en fonction de la plante infectée. Ainsi, on distingue par exemple la fusariose des épis ; traduit par *Fusarium* Head Blight (FHB) pour le blé (Photo I\_1) et *Gibberella* Ear Rot ou *Fusarium* Ear Rot pour le maïs (GER/FER) ; de la fusariose du collet (*Gibberella* stalk rot), des racines ou des semences (Seedling blight/root rot). La fusariose des épis, dont l'étude est le cœur du sujet de cette thèse, induit un flétrissement et un blanchissement des épis (Photo I\_1). Les grains sont atrophiés et ridés et présentent une couleur grisée, rosée ou brune.



**Photo I\_1 :** Un épi de blé (variété Royssac) infecté par *Fusarium graminearum*. © B. LAURENT

Alors que toutes les formes de fusarioses engendrent des pertes de rendements au champ, seule la fusariose des épis semble être associée à la présence de mycotoxines dans les grains. Ces molécules toxiques s'accumulent majoritairement pendant la croissance des grains mais aussi occasionnellement lors d'accidents durant leur stockage (grains humides lors de la mise en silo ou autres conditions favorisant le développement du pathogène au cours du stockage). De fait, la fusariose des épis est généralement considérée comme la forme de la maladie la plus problématique (Dean et al. 2012).

La fusariose des épis résulte de l'infection des céréales par des espèces fongiques du genre *Fusarium* ou *Microdochium*, qui peuvent agir seules ou de concert. Contrairement aux espèces du genre *Microdochium*, seules les espèces fusariennes semblent produire des mycotoxines (Bottalico & Perrone 2002; Ferrigo et al. 2016). Les mycotoxines produites par les espèces fusariennes, aussi appelées fusariotoxines (Figure I\_2), appartiennent à plusieurs familles de molécules : les trichothécènes, majoritairement de type B comme le déoxynivalénol (DON) et formes acétylées, le nivalénol (NIV) et sa forme acétylée ; parmi les trichothécènes de type A, citons les toxines T-2 et HT-2 ; les fumonisines et la zéaralénone (ZEA). Il serait pourtant réducteur de limiter les fusariotoxines à ces seules molécules puisque d'autres composés sont occasionnellement identifiés, comme par exemple en Europe les enniatines, la beauvericine, la moniliformine et la fusarine C (Charmet et al. 2016). Finalement, pas moins d'une vingtaine de mycotoxines produites par les espèces du genre *Fusarium* ont été identifiées et pour lesquelles l'exposition pour l'homme et l'animal est dangereuse (Glenn 2007; Bottalico & Perrone 2002). Si chaque espèce présente son propre spectre de production (Glenn 2007), dans certains cas, des différences sont observées au niveau même de l'espèce et définissant leur chémotype. Par exemple, chez *F. graminearum*, trois chémotypes sont classiquement distingués, et semblent différencier des populations différentes (Van der Lee et al. 2015; Ward et al. 2008). En plus de la ZEA, commune à l'ensemble des souches (Morgavi & Riley 2007; Reddy et al. 2010), les souches produisent majoritairement du DON ainsi que du 15-acétyldéoxynivalénol (chémotype 15-ADON), ou majoritairement du DON ainsi que du

3-acétyldéoxynivalénol (chémotype 3-ADON), ou du nivalénol au lieu du déoxynivalénol (chémotypes NIV).

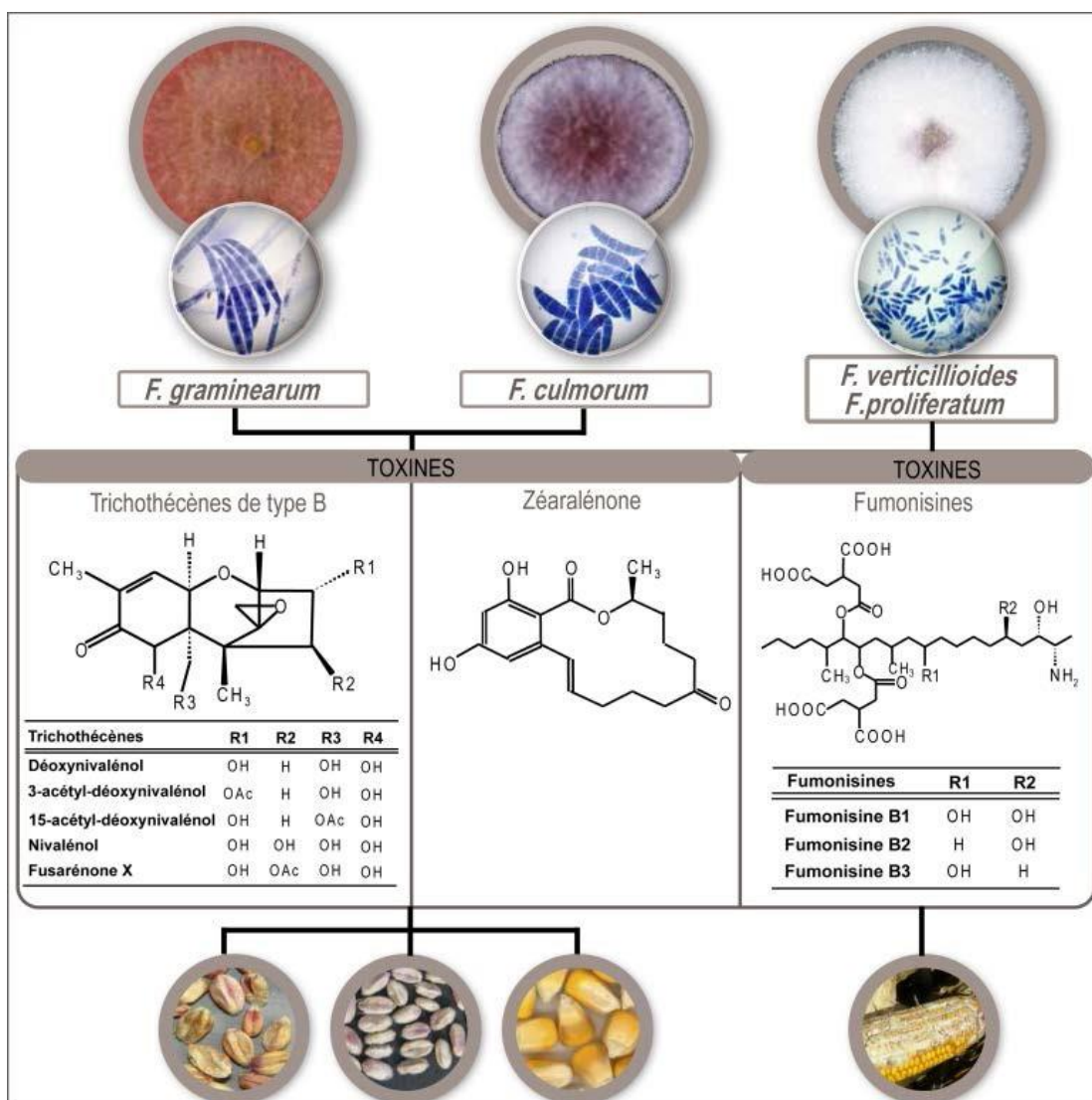
*F. graminearum* s'avère être l'une des espèces retrouvées majoritairement associées à la fusariose des épis dans le monde (Boutigny et al. 2014; Basler 2016; Backhouse 2014; Van der Lee et al. 2015; Osborne & Stein 2007; Bottalico & Perrone 2002). Des répartitions géographiques variées sont observées pour chaque chémotype et peuvent donc entraîner des problèmes sanitaires différents (Van der Lee et al. 2015). En France, les souches de *F. graminearum* de chémotype 15-ADON semblent être prédominantes (Boutigny et al. 2014). Cependant, il est important de noter que *Fusarium culmorum*, *Fusarium avenaceum* et *Fusarium poae* sont fréquemment retrouvés sur les céréales à pailles en France et en Europe (Figure I\_2). Ces fluctuations amènent à des contaminations occasionnelles mais préoccupante de T-2 et HT-2, en beauvericine ou en enniatines (Reddy et al. 2010; Glenn 2007). *Fusarium pseudograminearum* est lui majoritaire en Australie alors que *Fusarium asiaticum* est très présent en Chine (Bottalico & Perrone 2002; Becher et al. 2013). *Fusarium verticilloides* et *Fusarium proliferatum* sont très couramment retrouvés, à l'échelle mondiale, sur les cultures de maïs (en plus de *F. graminearum* et *F. culmorum*), amenant des contaminations en fumonisines (Figure I\_2) (Ferrigo et al. 2016).

Si l'occurrence et la toxicité du DON, des fumonisines (B<sub>1</sub>), de la ZEA, des T-2 et HT-2 ont abouti à une réglementation en Europe, des études sur les autres fusariotoxines sont en cours par l'autorité européenne de sécurité alimentaire et pourraient amener à l'élaboration de nouvelles réglementations (ex : enquête EFSA, 2013 pour les formes acétylées du DON). Il est aussi important de noter que les mycotoxines coexistent très souvent entre-elles, avec d'autres mycotoxines et/ou avec d'autres familles de contaminants (plomb, cadmium, résidus insecticides et fongicides, etc.). Les effets dits cocktails, bien que préoccupants, restent peu caractérisés (Smith et al. 2016; Charmet et al. 2016).

De plus, l'importance des mycotoxines est à prendre en compte en l'état actuel des connaissances, et ne sont étudiées/légiférées que celles qui sont recherchées. Pour illustrer ce problème, prenons l'exemple d'une mycotoxine produite par *F.*

# Introduction Générale

*graminearum*, dénommée NX-2, qui a été caractérisée pendant la réalisation de ces travaux de thèse (Varga et al. 2015). L'occurrence et la toxicité de cette molécule restent encore inconnues et ne fait l'objet d'aucune législation.



**Figure I\_2 :** Principales familles et espèces de *Fusarium* productrices de fusariotoxines

Une connaissance approfondie des facteurs influençant le développement de la maladie et la production de mycotoxines, ainsi qu'un suivi des communautés fusariennes, semblent donc essentiels pour anticiper au mieux les risques (Van der Lee et al. 2015).

## II.2. *L'apparition des fusarioses et des contaminations en mycotoxines : moyens de luttés*

L'intensité de la fusariose des épis est très souvent corrélée à des contaminations plus sévères en mycotoxines (Wegulo 2012; Paul et al. 2005).

- Facteurs abiotiques et biotiques (+++<sup>1</sup>)

Le développement de la maladie est très sensible aux conditions climatiques. De plus, les conditions favorables à l'infection peuvent varier en fonction du pathosystème considéré. La température et l'humidité sont les facteurs les plus importants (Ferrigo et al. 2016). Par exemple, pour le blé, des températures modérées et une forte humidité au moment de la floraison sont favorables au développement de *F. graminearum*, alors que des conditions plus froides semblent être favorables au développement de *F. poae* (Ferrigo et al. 2016). Les années plus sèches permettraient le développement de *F. culmorum* (Beyer et al. 2014). Dans le cas du maïs, l'infection par *F. verticillioides* est particulièrement problématique dans des conditions plus sèches et lors de la présence d'insectes. En effet, les blessures qu'ils engendrent sur la plante favorisent le développement de ce pathogène (Ferrigo et al. 2016).

Afin d'anticiper et d'ajuster au mieux les moyens de lutte, une attention particulière a été apportée au développement de modèles de prévisions de risques en lien avec le climat (Prandini et al. 2009). Une limite majeure de ces modèles réside cependant dans leurs niveaux d'incertitudes qui s'avèrent encore être importants et dépendants notamment de facteurs non-modélisables. C'est le cas par exemple des décisions des acteurs agricoles et de leur choix de pratiques culturales (Terzi et al. 2014; Prandini et al. 2009).

- Pratiques culturales (+++<sup>1</sup>)

Un deuxième facteur influençant majoritairement l'établissement de la maladie est directement lié aux pratiques culturales. Sur ce volet, le travail du sol et les rotations de cultures semblent être les facteurs les plus critiques (Ferrigo et al. 2016). Ainsi il est possible de limiter la présence d'inoculum d'une année sur l'autre, et donc le développement d'épidémies, par une gestion des débris végétaux (par le labour par

<sup>1</sup> : importance du facteur. '+++' : facteur majoritaire ; '++' : facteur important.

exemple) ou en évitant des rotations de cultures dites « à risque » (culture de maïs suivi de blé par exemple) (Wegulo et al., 2015). Le désherbage peut s'avérer aussi être utile puisque de nombreuses espèces fusariennes, et particulièrement *F. graminearum*, ont été identifiées sur les adventices (Postic et al. 2012). Au vu du transport à longue distance de l'inoculum (Keller et al. 2014), les pratiques culturales doivent cependant être considérées sur une échelle régionale pour espérer une efficacité optimale.

Bien que primordiales, les pratiques culturales sont mises en place selon la libre volonté des acteurs des filières, qui peuvent parfois être confrontés à d'autres problématiques, entraînant des compromis et des prises de risques. Par exemple, afin de protéger les sols, les débris végétaux sont parfois laissés à la surface, favorisant le développement de la maladie. Similairement, la culture de certaines céréales ou de certaines variétés sont préférées pour leur performance agronomique, au dépend de leur comportement face au développement de fusarioses (Ferrigo et al. 2016).

- Traitement chimique (++<sup>1</sup>)

Lorsque les risques sont avérés (conditions climatiques et/ou pratiques culturales dangereuses), l'utilisation de fongicides est parfois préconisée (Charmet et al. 2016; Ferrigo et al. 2016). Toutefois, ces molécules présentent des résultats variables et leur application doit être effectuée dans des fenêtres de temps très courtes pour être efficace (Ferrigo et al. 2016; Terzi et al. 2014). Dans le cas du pathosystème *F. verticillioides* x maïs, les traitements insecticides peuvent s'avérer particulièrement efficaces, puisque les insectes facilitent le développement de la maladie (Bowers et al. 2014). Cependant, l'utilisation de fongicides et insecticides n'apparaît pas comme une méthode de choix afin de respecter les différents engagements environnementaux (voir le plan Ecophyto et le plan Ambition Bio 2017 par exemple en France). Suivant cette problématique, l'utilisation de moyens de luttés biologiques est encouragée, bien que leurs efficacités actuelles restent limitées. L'utilisation de produits biologiques, couplée à des doses raisonnées de fongicides présente des efficacités intéressantes et prometteuses (Terzi et al. 2014). D'un point de vue sanitaire, l'utilisation de fongicides est là aussi sujette à controverse, indépendamment de la résilience et de la toxicité de la molécule. En effet, l'utilisation de certains fongicides, en modifiant la flore

microbienne présente sur l'épi, favoriserait le développement de pathogènes toxigènes (Simpson et al. 2001). Dans certains cas, il semblerait que l'application de fongicides éliciterait directement la production de mycotoxines par le champignon (Audenaert et al. 2010).

- Sélection variétale (++<sup>1</sup>)





Contrairement aux fongicides, l'utilisation de variétés de céréales résistantes aux fusarioses apparaît comme une solution prometteuse pour respecter les engagements environnementaux (Terzi et al. 2014). Alors que les législations en Europe et plus particulièrement en France, interdisent l'utilisation d'ingénieries moléculaires pour la conception et la vente de plantes résistantes, ces approches de sélection variétale consistent à identifier des sources naturelles de résistances à partir de collections de diversité et à les introduire, par le biais de croisement et de sélection génétique, dans des variétés cultivées (Buerstmayr et al. 2009).

La génétique de la résistance à la fusariose est quantitative : aucune variété à résistance totale n'a été identifiée. A ce jour, différents types de résistance peuvent être distingués : la résistance à l'infection initiale (type I), la résistance à la propagation du champignon dans les tissus végétaux (type II), la capacité de dégrader les mycotoxines (type III), la résistance à l'accumulation en mycotoxine (type IV) et la résistance à l'infection du grain (type V) (Ferrigo et al. 2016). Le patrimoine génétique des céréales, et plus particulièrement du blé, démontre une grande variabilité génétique pour la résistance à la fusariose des épis (Buerstmayr et al. 2009). Actuellement, plus d'une centaine de régions chromosomiques, appelées QTL pour « quantitative trait *loci* », ont été associées à la résistance à la fusariose chez le blé (Buerstmayr et al. 2009), et un peu moins d'une trentaine chez l'orge (Bai et Shaner 2004; Massman et al. 2011). Chez le blé, les sources les plus efficaces de résistance ont été identifiées chez la variété chinoise Sumai3, avec notamment l'identification du QTL *Fhb1* qui permet la réduction des symptômes (résistance de type II) de 20-30% en moyenne (Buerstmayr et al. 2009; Buerstmayr & Lemmens 2015). Les mécanismes moléculaires liés à ces résistances demeurent moins bien caractérisés (Buerstmayr et al. 2009). De nouvelles approches sont aussi à l'étude et visent à identifier directement les métabolites végétaux

impliqués dans la résistance, notamment chez le maïs, pour lequel la sélection génétique est moins développée (Gauthier et al. 2015). Cette méthode de sélection reste toutefois compliquée à mettre en œuvre du fait de la complexité des voies de biosynthèse et du rôle physiologique associé à ces molécules (Charmet et al. 2016).

- Aide à la décision (++)

Fort des connaissances actuelles, de nombreux instituts publics et privés proposent des outils d'aide à la décision afin d'aider les acteurs des filières céréalières à gérer les problématiques en lien avec les fusarioses (Figure I\_3 : exemple de grille proposée par Arvalis, un institut de recherche appliquée agricole).

Gestion des résidus*		Sensibilité variétale	Risque	Pluie à la floraison		
				<10	10-40	>40
 Céréales à paille, colza, lin, pois, féverole, tournesol	Labour ou résidus enfouis	Peu sensibles	1			
		Moyennement sensibles	2			
		Sensibles	3			T**
	Techniques sans labour ou résidus en surface	Peu sensibles	2			
		Moyennement sensibles	3			
		Sensibles	3			T
 Betteraves, pomme de terre, soja, autres	Labour ou résidus enfouis	Peu sensibles	2			
		Moyennement sensibles	3			
		Sensibles	3			T
	Techniques sans labour ou résidus en surface	Peu sensibles	2			
		Moyennement sensibles	4			
		Sensibles	4		T	T
 Maïs et sorgho fourrages	Labour ou résidus enfouis	Peu sensibles	2			
		Moyennement sensibles	4			
		Sensibles	4			
	Techniques sans labour ou résidus en surface	Peu sensibles	5		T	T
		Moyennement sensibles	5		T	T
		Sensibles	6	T	T	T
 Maïs et sorgho grains	Labour ou résidus enfouis	Peu sensibles	2			
		Moyennement sensibles	3			
		Sensibles	4			
	Techniques sans labour ou résidus en surface	Peu sensibles	5		T	T
		Moyennement sensibles	6		T	T
		Sensibles	6	T	T	T
		Sensibles	7	T	T	T

ARVALIS-Institut du végétal 2011

**Figure 3 :** Grille d'évaluation du risque d'accumulation du déoxynivalénol (DON) dans le grain de blé tendre et d'aide au traitement contre la fusariose sur épi proposée par Arvalis (<http://www.arvalis-infos.fr/qualite-sanitaire-des-cereales-comment-gerer-le-risque-don-facteurs-de-risque-et-grilles-d-evaluation--@/view-18629-arvarticle.html>). L'échelle de risque s'étend de 1 à 7 ; 1 étant le risque le plus faible et 7 le risque le plus fort. « T » : un traitement fongicide est préconisé.



# Introduction Générale

En France, bien que les données de contamination en mycotoxine restent confidentielles, notamment pour limiter leurs effets négatifs sur les exportations, entre 10 et 20% des lots peuvent dépasser les normes réglementaires certaines années (Charmet et al. 2016). Les différents leviers d'actions présentés ci-dessus restent donc insuffisants pour parfaitement prévenir l'apparition de maladies et ne garantissent pas des seuils en mycotoxines compatibles avec les différentes réglementations.

Afin de proposer de nouvelles solutions, il devient donc important de mieux caractériser **la relation entre ce pathogène et son hôte**, ainsi que les **facteurs amenant à la production de mycotoxines**.

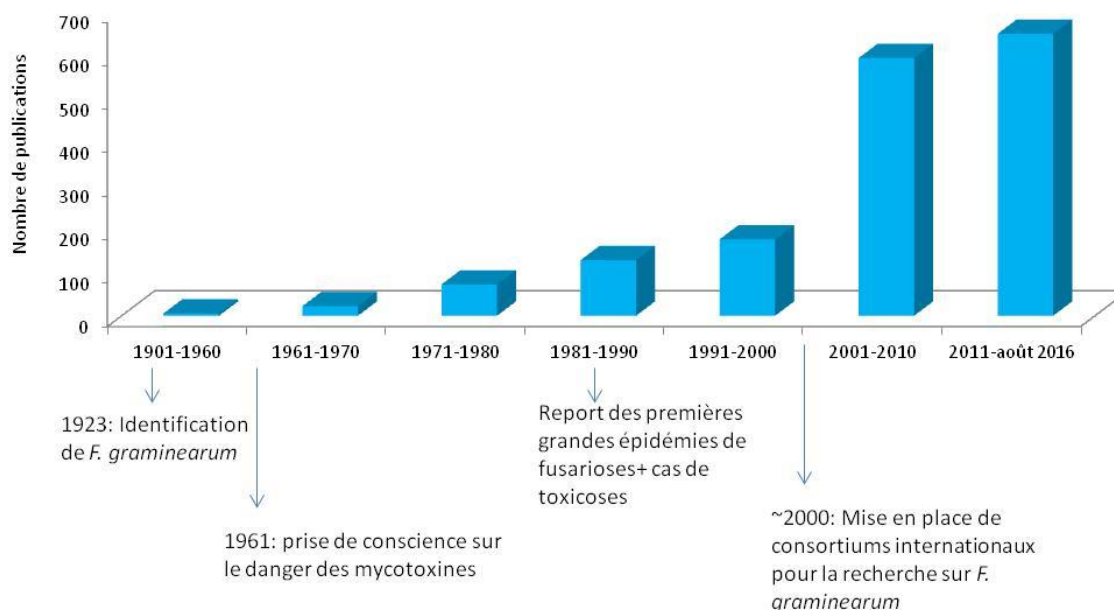
Les travaux développés dans la suite de ce manuscrit s'inscrivent dans cet axe stratégique de recherche. Ils se concentrent principalement sur l'étude de la fusariose des épis et plus précisément du pathosystème *F. graminearum* x blé (*Triticum aestivum*).

### **III. *Fusarium graminearum* et la fusariose des épis**

#### **III.1. *Historique des recherches sur F. graminearum***

*F. graminearum* a été identifié pour la première fois en 1923 par Dickson, sur des grains de maïs nécrosés. Il tient son nom de l'aspect fusiforme de ses spores (Figure I\_2). En 1940, Mitchell et Beadles démontrent que l'infection de grains de maïs par ces « micro-organismes » est à l'origine de pathologies chez les animaux consommant ces grains (Kadis et al. 1971). En 1961, le rôle de l'aflatoxine a été mis en évidence sur la santé animale et humaine (Blount 1961). Bien que ce composé soit produit par d'autres espèces fongiques que les espèces du genre *Fusarium*, c'est cette découverte qui amena à une réelle prise de conscience du danger inhérent à la présence des composés d'origines fongiques dans l'alimentation (Kadis et al. 1971). A la fin du siècle dernier, des épidémies importantes de fusarioses affectèrent la production céréalière mondiale (Dubin et al. 1997). Une estimation proposa notamment que 2,6 milliards de dollars avaient été perdus de 1999 à 2000 dans le nord Dakota et le Minnesota pour le blé et l'orge uniquement (Nganje et al. 2002). En Chine et en Inde, ces épidémies s'accompagnèrent d'importantes toxicoses humaines conduisant à des vomissements, des diarrhées et d'autres symptômes gastroentéritiques (Li et al. 1999; Bhat et al. 1989; Hussein et Brasel 2001). Ces toxicoses ont été rapidement associées à la présence de fortes doses de mycotoxines produites par *F. graminearum* dans les grains (Li et al. 1999; Bhat et al. 1989; Hussein et Brasel 2001). Cette recrudescence d'événements a permis d'importantes levées de fond à partir des années 1990-2000, et l'émergence de *consortia* internationaux pour la recherche sur ce pathogène et sur les mycotoxines (Stack 2000). Cette période d'investissement se reflète d'ailleurs très bien par le nombre croissant de travaux accordés à ce pathogène dans les années 2000 (Figure I\_4). Depuis cette date, ce chiffre continue de croître. L'attention particulière portée à ce pathogène a permis notamment la publication de son génome en 2007 (disponible à partir de 2005), et faisant de cet organisme un des premiers champignons dont le génome a été séquencé (Cuomo et al. 2007). La publication de ces données a constitué

une étape clé de la recherche sur ce pathogène, dont l'apport a été, et reste considérable pour la compréhension des mécanismes moléculaires sous-jacents à ses différents stades de vie (Kazan et al. 2012).



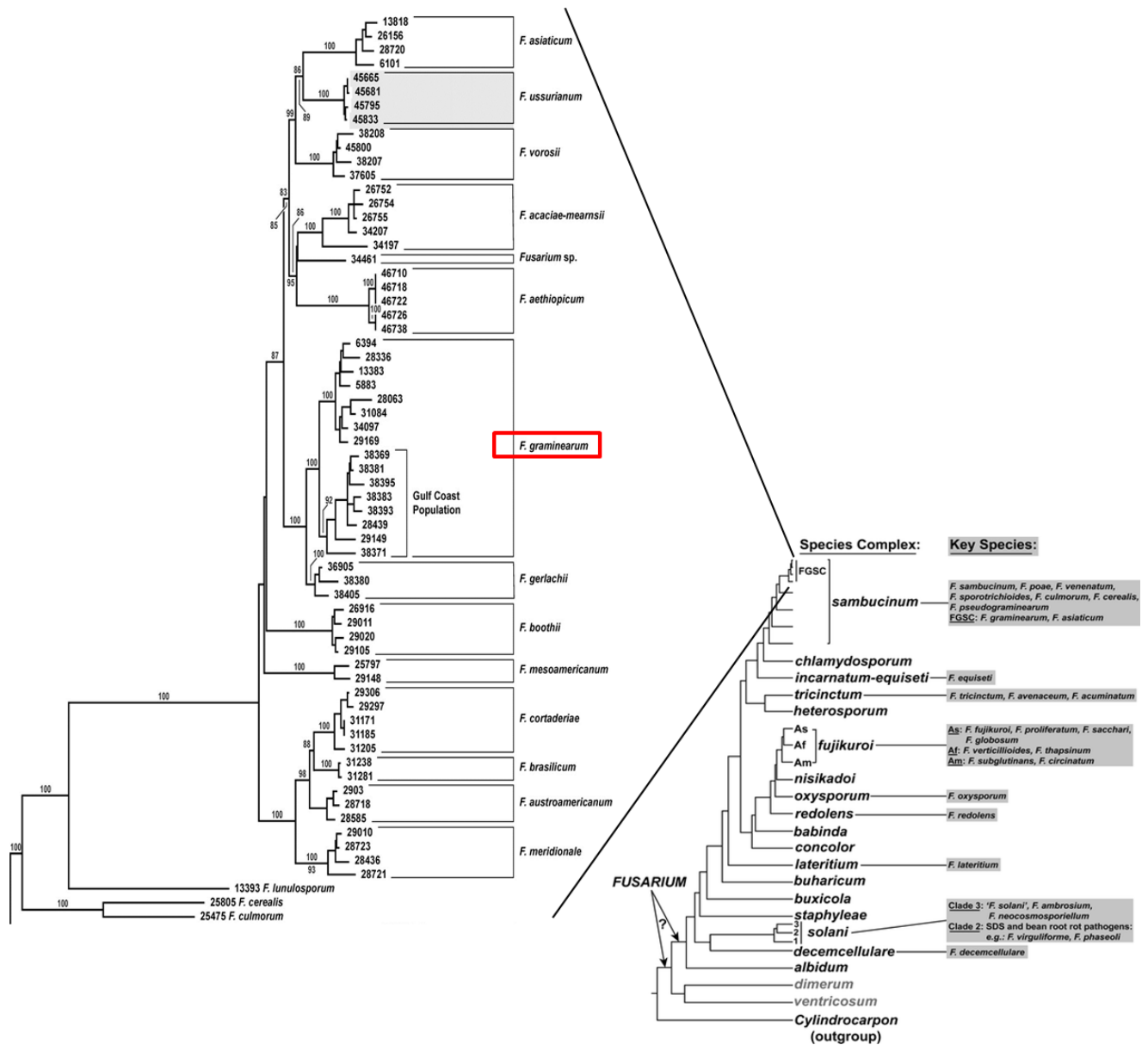
**Figure I\_4 :** Evolution du nombre de publications identifiées par le moteur de recherche spécialisé Web of Science comprenant « *Fusarium graminearum* » ou « *Gibberella zeae* » (forme sexuée de ce pathogène) dans le titre, et dates clés.

La bibliographie associée à la biologie de ce pathogène est aujourd'hui riche. Ce dernier point, associé à des facilités de cultures et de manipulations au laboratoire, font apparaître de plus en plus *F. graminearum* comme un modèle fondamental pour l'étude des relations hôte-pathogène (Galazka & Freitag 2014).

## III.2. *Précisions phylogénétiques*

*F. graminearum* appartient à la division des ascomycètes et à la classe des sordariomycètes (Goswami & Kistler 2004). Il appartient au groupe *Sambucinum* des espèces fusariennes (Aoki et al. 2014, Figure I\_5). La ressemblance morphologique de *F. graminearum* avec d'autres espèces fusariennes a toujours été source de confusions et de débats. L'avènement de la biologie moléculaire a permis de clarifier la phylogénie de cette espèce, alimentant paradoxalement les controverses (O'Donnell et al. 2004; Leslie et al. 2007; Aoki et al. Geiser 2014). Ainsi, on peut parfois retrouver le nom de *F. graminearum* pour le complexe d'espèces qui porte le même nom (FGSC pour *Fusarium graminearum* species complex, Figure I\_5). Les organismes composant ce complexe sont tous génétiquement proches et difficilement distinguables sur des critères morphologiques (Aoki et al., 2014). Il comprendrait au moins 16 espèces différentes (Starkey et al. 2007; Van der Lee et al. 2015; Aoki et al. 2014; Yli-Mattila et al. 2009). Bien que des croisements entre espèces soient possibles dans certains cas, l'éloignement génétique est assez important pour générer des réarrangements chromosomiques et autres biais de ségrégation dans la descendance (Lee et al. 2008). Ce résultat conforte des études phylogénétiques récentes, qui défendent un/des phénomène(s) récent(s) de spéciation entre ces isolats (Aoki, O'Donnell & D. Geiser 2014). Tout au long de ce manuscrit la mention de *F. graminearum* fera référence à *F. graminearum sensu stricto* et non au complexe d'espèces.

# Introduction Générale



**Figure I\_5 :** Arbre phylogénétique du *Fusarium graminearum* species complex (FGSC) proposé par Yli-Mattila et al. 2009 (à gauche) et du genre *Fusarium* proposé par Aoki et al. 2014 (à droite).

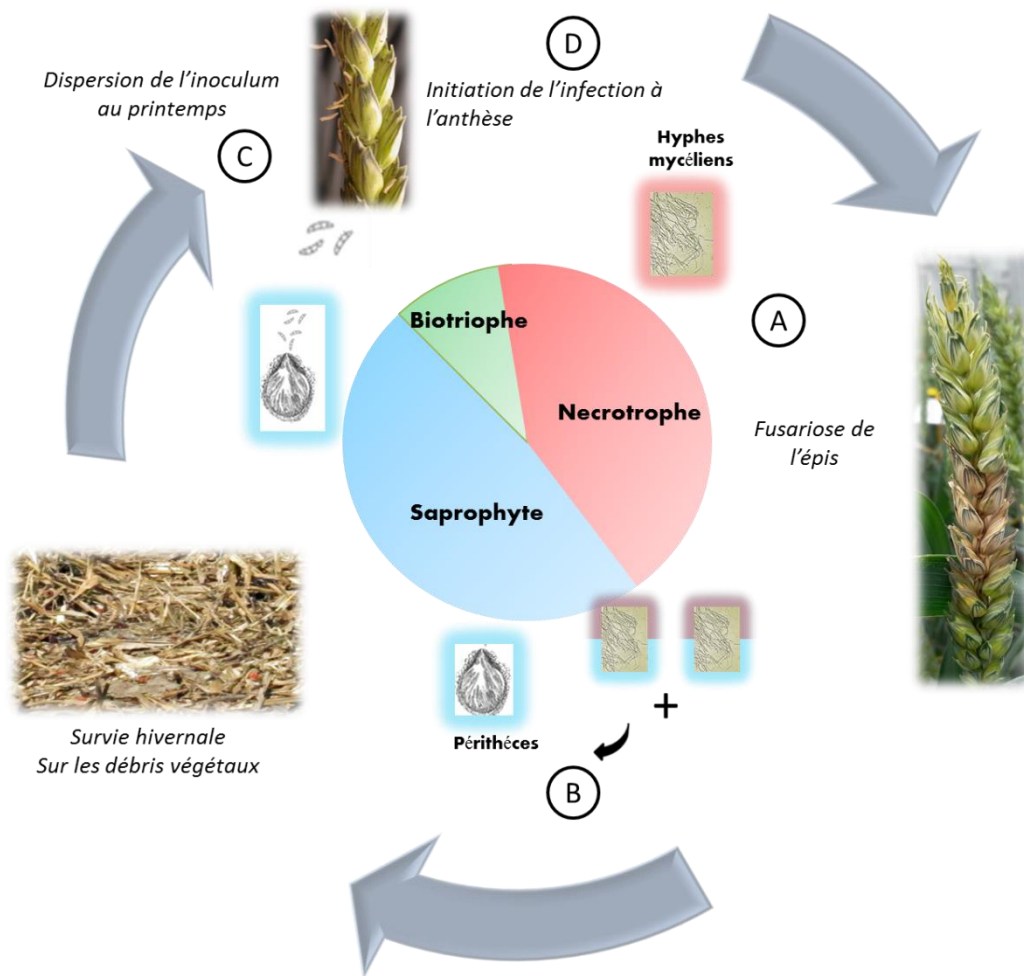
## III.3. *Le cycle de vie de F. graminearum*

La croissance végétative de *F. graminearum* se fait sous la forme de filaments haploïdes, appelé hyphes (Figure I\_6A). Cette particularité lui vaut son appartenance au groupe plus généraliste des champignons filamenteux, ou moisissures. Tout au long de son cycle, *F. graminearum* alterne des modes trophiques et des cycles de reproduction variables (Trail 2009). Il est donc possible de retrouver associé au nom de *F. graminearum* le nom *Gibberella zeae*, correspondant à sa forme sexuée (*i.e.* téléomorphe). Cette forme sexuée est produite à la sortie de l'été et correspond à un organe fructifère, appelé périthèce, visible sur les débris végétaux laissés en décomposition sur les champs (Figure I\_6B, Trail, 2009). A ce stade, son mode de vie est saprophyte, puisqu'il se nourrit de matière morte (Goswami & Kistler 2004) La formation des périthèces résulte de la rencontre et de la fusion de deux hyphes, constituant les seules structure diploïdes de son cycle de vie (Trail & Common 2000). *F. graminearum* est homothallique et la fusion d'hyphes peut se faire entre individus différents (« fécondation croisée ») ou entre hyphes issus du même individu (« autofécondation », Yun et al. 2000). Le périthèce est une structure très importante pour le cycle de *F. graminearum* : il constitue une forme de résistance, en plus d'être le lieu de production des ascospores (Figure I\_6C). Ces ascospores sont formées suite aux événements de méiose affectant les cellules diploïdes contenues dans les périthèces (Trail & Common 2000) (Figure I\_6C).

Les ascospores représentent l'inoculum primaire dont l'expulsion et le transport dans l'air, par la pluie et les insectes, de manière concomitante à la floraison des blés, engendrent l'établissement d'épidémies sévères de fusariose des épis (Goswami et Kistler, 2004). De fait, le développement de la cette maladie résulte de la germination et de la pénétration d'hyphes haploïdes dans la plante (Figure I\_6D). Cette pénétration se produit soit par les ouvertures naturelles, soit au travers des tissus végétaux par le biais de structures spécifiques (Boenisch et al. 2011; Walter et al. 2010). De plus, il semblerait que le rôle des spores asexuées, formées pendant l'hiver mais aussi durant l'infection des céréales, puisse aussi être une source importante d'inoculum (Osborne et Stein, 2007). *In fine*, ces sources d'inoculum (ascospores et conidies) permettent une dispersion à courte et à longue distance du pathogène

(Maldonado-Ramirez et al. 2005; Schmale Iii et al. 2006; Prussin et al. 2013; Osborne & Stein 2007). La fusariose des épis est considérée cependant comme monocyclique, puisque la production et la germination des conidies, durant l'infection de la plante, est mineure sur l'incidence de la maladie (Wegulo et al. 2015).

Une fois dans la plante, le champignon semble adopter une stratégie d'infection hémibiotrophe (Brown et al. 2010; Kazan et al. 2012). Cette observation est elle aussi source de discussion au sein de la communauté scientifique. Les partisans de cette définition se justifient notamment par le développement du pathogène, aux prémices de l'infection, de manière asymptomatique dans l'espace extracellulaire des tissus végétaux et donc de manière biotrophe (Figure I\_6D) (Brown et al. 2010). Cette période, plus ou moins longue, précède l'apparition de nécroses sur les épis, traduisant la croissance du pathogène dans les cellules végétales. Cette phase s'accompagne de la mort des cellules végétales et donc l'adoption d'une relation nécrotrophique par le pathogène (Brown et al. 2010; Kazan et al. 2012) (Figure I\_6A). Elle permet au champignon d'accumuler des réserves importantes, notamment lipidiques, consommées plus tard pendant la mise en place des périthèces à la sortie de l'été et à l'hiver (Guenther et al. 2009).



**Figure I\_6 :** Cycle de vie de *F. graminearum*, associé au développement de la fusariose des épis chez le blé, inspiré de Trail et al. 2009.

La fusariose des épis chez le blé constitue le pathosystème le plus étudié du fait de son importance économique (Brown et al. 2010). Les connaissances de ces mécanismes d'infection chez d'autres céréales et pour d'autres compartiments de la plante sont plus succinctes. De plus, la phase hémibiotrophe est très certainement la mieux caractérisée de toutes car la plus dommageable pour les rendements et la qualité sanitaire des grains. Les stades de vie hivernaux, qui semblent pourtant jouer un rôle important dans le cycle épidémiologique de la fusariose sont bien moins caractérisés (Trail 2009).



# IV. La phase biotique des champignons phytopathogènes

La phase biotique est le stade de vie du champignon en interaction avec la plante vivante. Pour les champignons phytopathogènes, le succès de la colonisation et de l'infection des tissus végétaux dépend de leurs capacités à modifier les cellules végétales pour contourner les mécanismes de reconnaissance et pour l'acquisition des nutriments (Selin et al. 2016). Cette interaction est sous le contrôle d'effecteurs fongiques, majoritairement des petites protéines sécrétées et des métabolites secondaires (Rep 2005; Pusztahelyi et al. 2015; Selin et al. 2016).

### IV.1. *Bases génétiques et moléculaires de la relation hôte-pathogène.*

- La relation qualitative :

Dans les années 1930, Harold Henry Flor fut le premier à définir les bases génétiques de la relation hôte-pathogène chez *Melampsora lini*, l'agent responsable de la rouille du lin (*Linum usitatissimum*). Ces travaux ont permis la conception du modèle « gène pour gène » qui propose qu'un seul gène de plante (nommé gène *R*) et le gène complémentaire du pathogène (nommé gène *Avr*) permettent la reconnaissance par la plante du pathogène et la mise en place de résistance (Flor 1971). Dans le cas de cette relation, l'acquisition de l'allèle de résistance « *R* », supposée dominante, ou de l'allèle de sensibilité « *r* », supposée récessive, amène à une variation qualitative de la résistance. Chez le pathogène, la forme allélique reconnue par la plante sera appelée allèle d'avirulence, alors que la forme non-reconnue sera appelée allèle de virulence. Cette relation est en théorie monogénique, et est supposée être variété-spécifique (Flor 1971). Les mécanismes moléculaires liés à cette relation ont été bien étudiés : les gènes *R* codent généralement des protéines qui reconnaissent soit des effecteurs de pathogènes soit des cibles d'effecteurs de pathogènes. Les protéines NB-LRR (pour nucleotide binding-leucine rich repeat) sont peut-être les plus connues et forment une grande famille de récepteurs d'effecteurs de pathogènes (Nimchuk et al. 2003). Chez la

plante, cette reconnaissance induit notamment la mise en place d'un mécanisme hautement efficace de défense de type hypersensible (Zhang et al. 2013). Ce type de relation est très fréquemment retrouvé lors de l'infection des plantes par des pathogènes biotrophes (Zhang et al. 2013).

- La relation quantitative :

En plus de cette composante qualitative de la pathogénicité une composante quantitative existe et décrit l'intensité de la relation parasitique (Van der Plank 1963). Pour la plante, cette composante fait référence aux résistances quantitatives et pour les pathogènes, cette composante fait référence à l'agressivité. Le terme de virulence est parfois utilisé, notamment par les pathologistes animaux, même si pour d'autres la virulence fait référence à la composante qualitative de la pathogénicité (Van der Plank 1963). Afin d'éviter des confusions sémantiques, notons que le terme d'agressivité sera utilisé dans ce manuscrit et fera référence à la composante quantitative de la pathogénicité.

Chez la plante, les résistances quantitatives sont gouvernées par les *loci* quantitatifs de résistance (QRL) et peuvent-être liées, par exemple, à des régulations morphologiques ou développementales, des transporteurs ou des voies de signalisations cellulaires, des réactions basales de défenses, des systèmes de détoxification de phytotoxines ou des allèles de gènes *R* à effet plus faible (Zhang et al. 2013; Poland et al. 2009).

L'agressivité d'un pathogène s'exprime de différentes manières et peut donc être quantifiée de différentes façons (Pariaud et al. 2009). Les caractères les plus communément mesurés sont l'efficacité d'infection, la période de latence, le taux de production d'inoculum, la durée infectieuse, la taille des lésions et pour certains pathogènes, la production de mycotoxines (Pariaud et al. 2009; Lannou 2012). D'une manière générale, il est important de considérer tous les traits d'histoire de vies (caractéristiques observables d'un organisme en lien avec sa croissance, sa survie, sa morphologie, sa physiologie ou son comportement) qui pourraient être reliés à la dynamique épidémiologique et à l'évolution du pathogène (Lannou 2012). Bien que très sensible à l'environnement, l'agressivité des champignons est généralement sous

contrôle génétique (Pariaud et al. 2009). Les bases génétiques de l'agressivité des champignons restent cependant bien moins caractérisées que les résistances quantitatives des plantes.

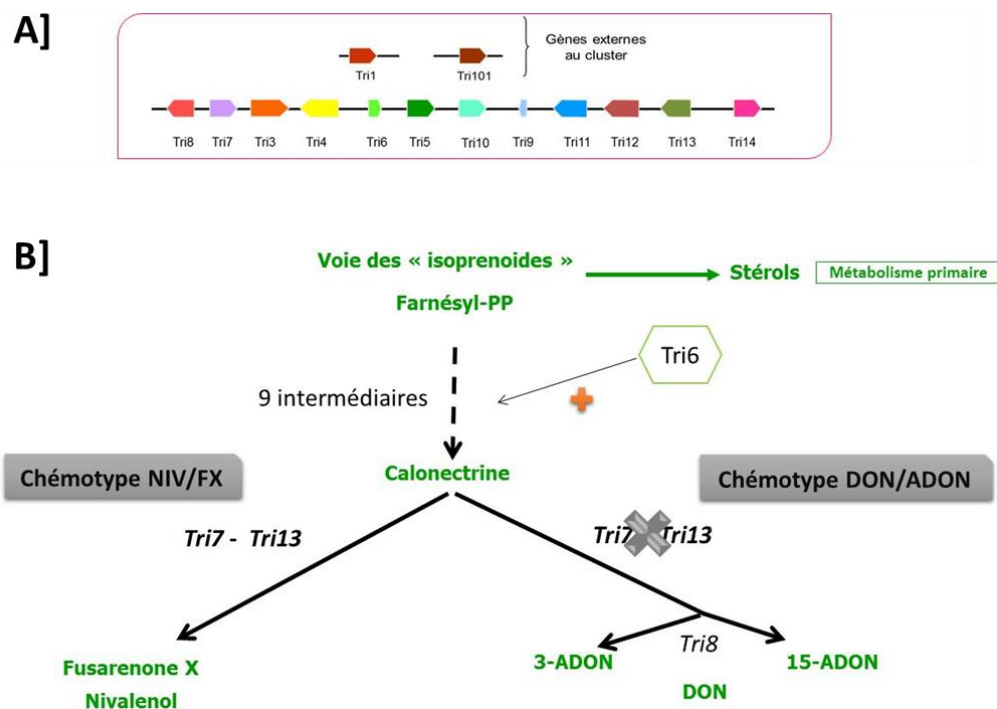
## **IV.2. *Les effecteurs fongiques : le cas des métabolites secondaires.***

Les métabolites secondaires correspondent à des molécules bioactives, souvent de faibles poids moléculaires (< 10 000 daltons), et dont la définition fait référence au caractère non-essentiel de ces molécules l'organisme (Keller et al. 2005). Bien que leurs rôles ne soient pas exclusifs à l'établissement d'interaction biotique, les métabolites secondaires y jouent un rôle important (Pusztahelyi et al. 2015). Ils participent donc activement à la survie des espèces. Le panel de métabolites secondaires peut varier entre espèces fongiques. Il traduit souvent le mode de vie adopté par le champignon (biotrophie stricte, nécrotrophie stricte ou hémibiotrophie ; Pusztahelyi *et al.*, 2015). Leurs synthèses dérivent pour la plupart de précurseurs produits par les voies métaboliques dites primaires. De fait, la production de métabolites secondaires est coûteuse pour le pathogène : ils ne sont pas produits constitutivement mais en réponse à de nombreux signaux environnementaux (Brakhage 2013). Les gènes impliqués dans leur production ou liés à leurs régulations sont d'ailleurs très souvent regroupés spatialement sur les chromosomes, formant des clusters. Cette organisation permettrait notamment une reprogrammation rapide et coordonnée de la production de métabolites secondaires (Osbourn 2010). Les métabolites secondaires fongiques peuvent être classés en quatre grands groupes principaux: les terpènes, les peptides non-ribosomiaux, les alcaloïdes et les polyketides (Keller et al. 2005). Les mycotoxines appartiennent toutes à l'une de ces quatre familles de molécules (Keller et al. 2005).

## **IV.3. *Biosynthèse et rôle biologique des trichothécènes de types B produits par *F. graminearum****

Les trichothécènes de type B (TCTB) correspondent à des terpènes. Ils sont produits à partir de la transformation du farnésyl-pyrophosphate en trichodiène par l'action de la trichodiène synthase (Figure I\_7). Cette enzyme est codée par le gène

*Tri5*. Dix des treize autres gènes participant à la biosynthèse des trichothécènes de type B, tous dénommés gènes *Tri*, sont regroupés en un cluster de 25 kilobases (kb, Figure I\_7) (Kimura et al. 2003). L'expression de tous les gènes *Tri* (dans le cluster et hors du cluster) est régulée par le facteur de transcription général de la voie, codé par le gène *Tri6* (Seong et al. 2009). Ce dernier est lui aussi localisé dans le cluster. Des variations alléliques affectant les gènes *Tri* sont à l'origine des chémotypes. Elles concernent le gène *Tri8* et détermineront le chémotype 3-ADON ou 15-ADON (Alexander et al. 2011). Ce cluster est aussi composé de deux pseudo-gènes chez les souches productrices de DON (*Tri7* et *Tri13*), dont l'activité dans d'autres souches détourne la production de DON vers la production de NIV (Figure I\_7) (Lee et al. 2002).



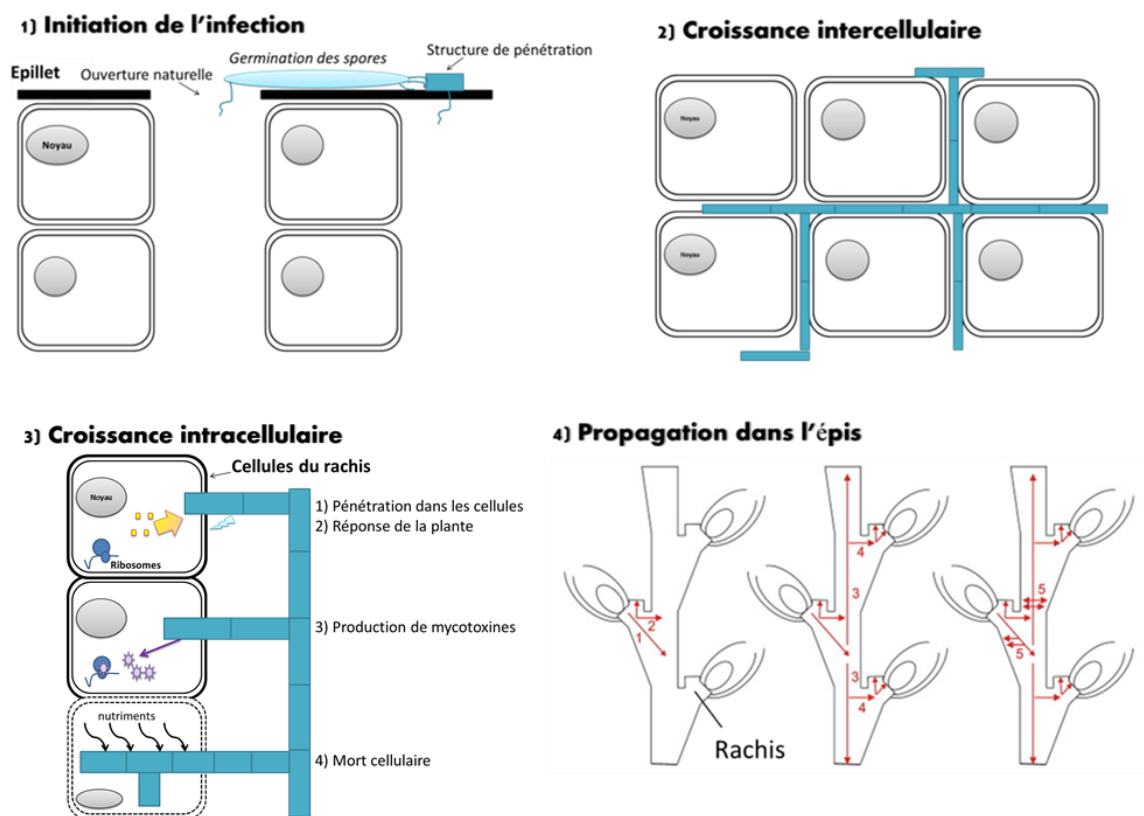
**Figure I\_7** A] Illustration des gènes *Tri* (localisé dans le cluster et hors du cluster) et B] schéma simplifié de la voie de biosynthèse des trichothécènes de type B.

La cytotoxicité des trichothécènes de type B résulte de leur affinité avec les ribosomes, pouvant conduire à un arrêt de la synthèse protéique et à la mort des cellules (Pestka, 2007; Ueno, 1983). Pour se protéger de la toxicité de ces molécules, le champignon semble internaliser la biosynthèse des trichothécènes de type B dans des structures spécifiques, appelées « toxisomes » (Menke et al. 2013). La production de

TCTB fait suite à la perception par *F. graminearum* de nombreux signaux extérieurs, comme la lumière, le pH, les sources carbonées ou azotées... (Merhej et al. 2011 pour revue). Lors de l'interaction avec la plante, la production de TCTB par le champignon semble s'effectuer en réponse à la mise en place des mécanismes de défense de la plante, et notamment en réponse à la production des espèces réactives de l'oxygène (ROS) (Montibus et al. 2013; Ponts 2015; Ponts et al. 2009) (Figure I\_8).

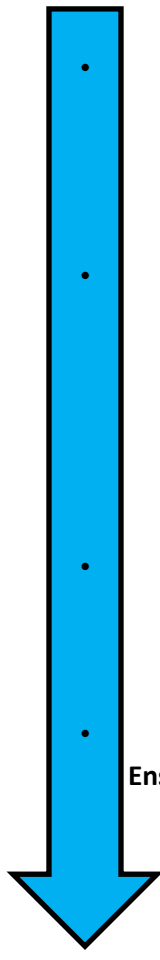
Dans le pathosystème *F. graminearum* x blé, la production de DON n'est pas indispensable pour l'initiation de l'infection mais permet la propagation de la maladie (Bai et al. 2002). Dans ce sens, le DON peut être assimilé à un facteur d'agressivité. Il interviendrait notamment juste avant le passage du rachis des épis et empêcherait la mise en place de barrière physique au niveau de ces tissus par la plante. Ceci permettrait donc la libre propagation des hyphes mycéliens dans les autres épillets composant l'épi (Brown et al. 2010; Bai et al. 2002; Hallen-Adams et al. 2011; Walter et al. 2010 Figure I\_8).

Des différences d'agressivité entre souches produisant soit du DON, soit du NIV ont été observées (Maier et al. 2006) et semblent être directement liées à la capacité à produire soit le DON soit le NIV (Cumagun 2004; Eudes et al. 2000). Les différences d'agressivité entre souches de chémotypes 3-ADON et 15-ADON sont beaucoup plus controversées (Spolti et al. 2013). L'action des mycotoxines ne se limite pas non plus uniquement à la colonisation des tissus végétaux. Elles semblent jouer des rôles importants à d'autres moments du cycle de vie de ce pathogène, notamment pour la compétition avec d'autres microorganismes évoluant dans la même niche écologique (Reverberi et al. 2010; Audenaert et al. 2014). De plus, les mycotoxines ne seraient pas les seuls effecteurs dont dispose ce pathogène pour compléter son cycle (Miedaner et al. 2000), comme en atteste les travaux étudiant la fonction des gènes présents dans le génome de *F. graminearum* (Cuomo et al. 2007).



**Figure I\_8 :** schéma simplifié de la stratégie d'infection adoptée par *F. graminearum*, mettant en lumière le rôle du déoxynivalénol dans la propagation du pathogène de l'épillet à l'épi chez le blé. Adapté de Brown et al. 2010 et Walter et al. 2010.

## V. Les apports du génome de *F. graminearum* pour l'étude de la pathogénicité

- 
- **2005 : Mise à disposition du génome de la souche de référence PH-1 par le Broad Institute :**
    - ✓ 36 Megabases (Mb)
    - ✓ 43 supercontigs, peu de séquences répétées
    - ✓ 4 chromosomes (Gale et al. 2005)
  - **2007 : Publication officielle du génome** (Cuomo et al. 2007)
    - ✓ Annotation automatique de **11,640** gènes codants pour des protéines
    - ✓ Séquençage d'une deuxième souche à 0.4X et identification d'environ 10,000 nucléotides polymorphiques (SNPs pour single nucleotide polymorphism)
    - ✓ Mise en évidence du « Repeat Induced Point mutation » (RIP, système de mutation fongique induit par les séquences répétées et limitant la mobilité et l'accumulation de transposons)
  - **2011 : Révision et ré-annotation du génome proposées sur la *Fusarium Graminearum* DataBase (FGDB):**
    - ✓ **13,718** gènes codant pour des protéines
    - ✓ 2,461 gènes ré-annotés (Wong et al. 2011)
  - **2015 : Nouveau séquençage de la souche PH-1 par le Rothamsted Institute et mise en ligne sur EnsemblFungi** (King et al. 2015) :
    - ✓ **38 Mb**
    - ✓ Annotation automatique de **14,164** gènes codant pour des protéines
    - ✓ 4 scaffolds alignés sur les 4 chromosomes,
    - ✓ Prédiction de la position des centromères

**Figure I\_9** : Un bref historique sur les efforts apportés à la publication et la révision du génome de référence de *F. graminearum*.

Le génome de référence de *F. graminearum* a été publié pour la première fois en 2007, et a été révisé plusieurs fois par la suite (King et al. 2015; Cuomo et al. 2007; Wong et al. 2011) (Figure I\_9). L'annotation automatique du génome a révélé un répertoire de gènes varié (Figure I\_9), codant notamment une part importante de facteurs de transcription, d'enzymes hydrolytiques, ainsi que de nombreux transporteurs (Cuomo et al. 2007). Cependant, la fonction de beaucoup des gènes identifiés reste inconnue.

## V.1. *Un large catalogue d'effecteurs potentiels*

Pour mieux comprendre le rôle des gènes de *F. graminearum*, les séquences génomiques ont fait l'objet d'études plus ciblées d'extraction de données (*i.e.* genome mining), qui consistent à chercher, grâce aux connaissances acquises chez des espèces modèles, des séquences ou motifs caractéristiques dont la présence permettrait d'attribuer une fonction aux gènes.

Une attention particulière a été portée sur les gènes impliqués dans la voie de biosynthèses de métabolites secondaires. L'existence de 20 non-ribosomale peptides synthases, 15 polykétides synthases et 17 terpenoïdes synthases a ainsi été prédit (Cuomo et al. 2007). Ces enzymes catalysent des réactions carrefours, réorientant la production de métabolites issus du métabolisme primaire pour la production de métabolites secondaires (Keller et al. 2005). Une autre analyse plus récente a prédit l'existence de 67 clusters de gènes représentant des voies de biosynthèse différentes (Sieber et al. 2014). Bien que la plupart des produits finaux associés à ces clusters restent inconnus, ce résultat suggère que le répertoire de métabolites secondaires produits par *F. graminearum* est certainement large (Sieber et al. 2014). La prédiction d'effecteurs protéiques potentiels basée sur l'étude des signatures nucléotidiques spécifiques a été particulièrement étudiée. Ces signatures correspondent notamment à des signaux de sécrétion, suggérant une action extracellulaire de la protéine codée. Chez *F. graminearum*, 616 gènes semblent ainsi coder des protéines sécrétées (Brown et al. 2012; King et al. 2015). Ce dernier résultat, appuyé notamment par des études de protéomiques (Rampitsch et al. 2013; Lowe et al. 2015), laisse penser que *F. graminearum* dispose aussi d'un nombre important d'effecteurs protéiques.

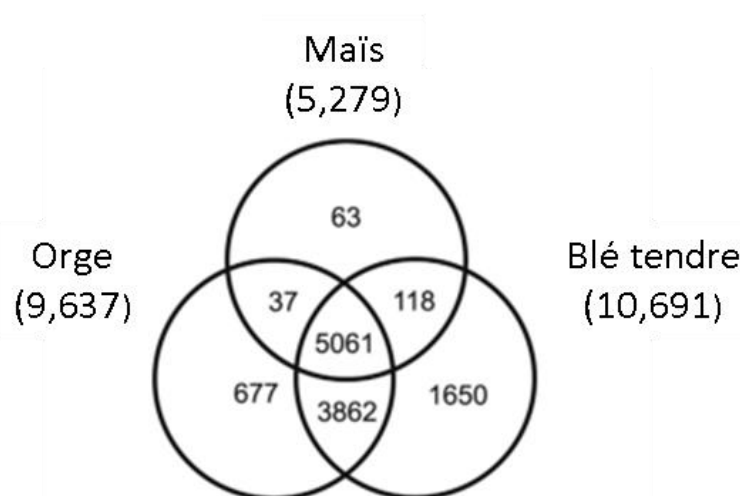
La réelle implication de la majorité de ces effecteurs (métabolites secondaires et protéines sécrétées) reste à déterminer (Jia & Tang 2015).

## V.2. *Les différentes reprogrammations de l'expression du génome*

Au-delà de l'étude du génome, l'étude des mécanismes régulant son expression reste primordiale. Des données d'expression globale des gènes suggèrent des



reprogrammations rapides et importantes de l'ensemble des gènes durant les différents stades de vie du pathogène (Carapito et al. 2008; Carapito et al. 2013; Harris et al. 2016; Lee et al. 2014; Lysoe et al. 2011). Cette régulation d'expression dans le temps semble aussi vraie dans l'espace. Ainsi, il apparait que malgré l'existence d'un bloc commun de ~5000 gènes exprimés constitutivement lors de l'infection des céréales (un peu plus d'un tiers des gènes de *F. graminearum*), ~ 6300 autres gènes montrent des spécificités d'expression en fonction de l'hôte infecté (Lysoe et al. 2011; Harris et al. 2016). Ce résultat suggère une spécialisation importante du transcriptome en réponse à l'hôte infecté (Figure I\_10).



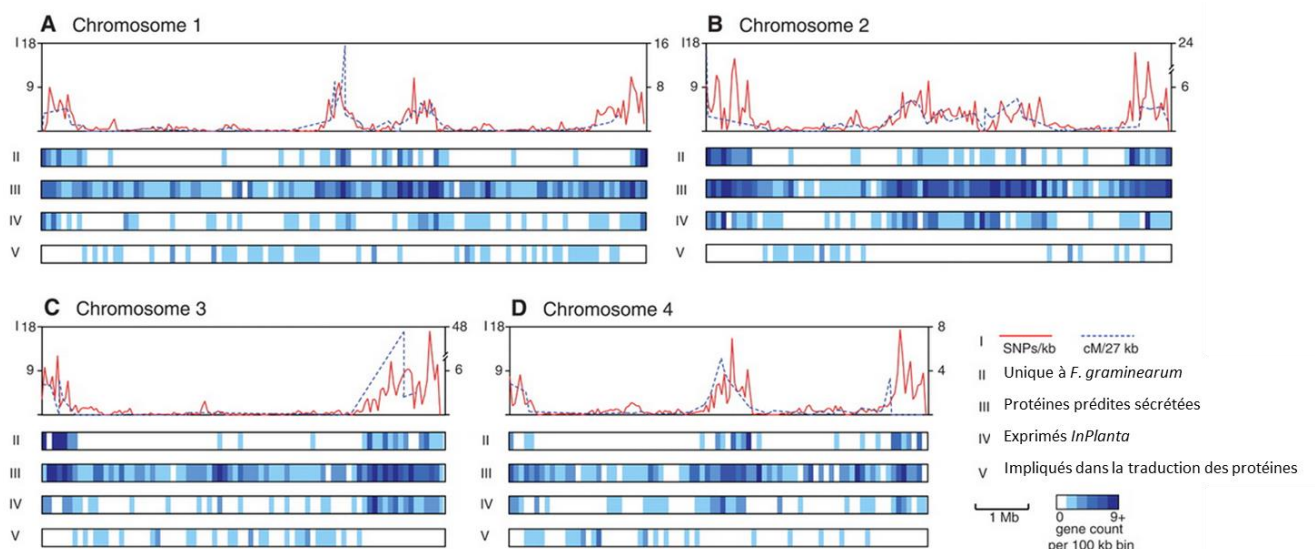
**Figure I\_10 :** Diagramme de Venn faisant état du transcriptome en fonction de l'hôte infecté. Les chiffres correspondent au nombre de gènes exprimés par *F. graminearum*. Adapté des travaux de Harris et al. 2016.

### V.3. *La diversité génétique chez F. graminearum*

Des données issues de la génétique des populations ont montré des niveaux de diversité génétique et d'échange génétique importants entre isolats de *F. graminearum*, et ce à l'échelle de l'épi, du champ ou du continent (Talas & McDonald 2015; Pan et al. 2016; Talas et al. 2011; Ward et al. 2008; Zeller et al. 2004). Ces résultats appuient l'existence d'une haute fréquence de croisement sexué entre souches. A l'échelle du génome, le séquençage d'une deuxième souche de *F. graminearum*, bien que partiel (0,4 X), a permis la localisation de ~10 000 nucléotides polymorphiques sur le génome (SNP pour « single nucleotide polymorphism »), répartis non-aléatoirement. Les

régions les plus polymorphiques semblent être localisées aux niveaux des télomères et subtélomères, mais aussi à l'intérieur des chromosomes (Cuomo et al. 2007). Alors que le polymorphisme des télomères et subtélomères est couramment observée chez les génomes eucaryotes (Mefford & Trask 2002; Kuo et al. 2006; Petes 2001), celui observé au niveau des régions internes aux chromosomes est plus inhabituel. Ce dernier point serait le reflet d'événements ancestraux de fusions chromosomiques, et ces régions correspondraient donc à d'anciennes régions télomériques/subtélomériques (Ma et al. 2010; Cuomo et al. 2007). Ces phénomènes expliqueraient notamment le faible nombre de chromosomes actuellement observés dans cette espèce (n=4, Gale et al. 2005) comparé à d'autres espèces fongiques et incluant des espèces fusariennes proches (Ma et al. 2010). A titre d'exemple, *Fusarium oxysporum*, une autre espèce fusarienne phytopathogène, est composée de 15 chromosomes (Ma et al. 2010).

De manière intéressante, les gènes exprimés spécifiquement en fonction de la plante infectée, ainsi que les gènes codant des effecteurs putatifs, semblent s'accumuler dans les régions chromosomiques polymorphiques (Cuomo et al. 2007, Figure I\_11). Une hypothèse est donc que ces gènes évoluent plus rapidement. De nouvelles études sont nécessaires pour appuyer ce résultat, dont la profondeur de séquençage associée à l'expérience en est la limitation majeure.



**Figure I\_11** : Représentation de la densité en SNPs et du taux de recombinaison (exprimé en centiMorgan tous les 27 kilobases), ainsi que de la localisation de certains types de gènes sur les quatre chromosomes de *F. graminearum*. Adapté de Cuomo et al. 2007.

# VI. Vers une évolution du pathosystème *F. graminearum* x céréale ?

## VI.1. *L'évolution des relations hôte-pathogène.*

Les relations entre les champignons et les plantes sont parmi les plus anciennes : elles perdurent depuis plus de ~240 millions d'années (Schwendemann et al. 2011). Dans le cas des relations parasitiques, celles-ci sont contrôlées par deux besoins antagonistes qui sont de type nutritif et/ou de reproduction pour le pathogène et de résistance pour la plante (Poland et al. 2009). Ce phénomène pousse donc à une constante évolution des deux partenaires, communément appelée « course à l'armement » (Anderson et al. 2010).

- Pour la relation qualitative :

L'évolution des relations qualitatives de la pathogénicité est très bien documentée et fait l'objet du modèle en « zig-zag » (Figure I\_12) (Jones & Dangl 2006). La première phase consiste à la reconnaissance par la plante des molécules du pathogène (les PAMPs pour pathogen associated molecular pattern) ; celui qui met en place un système de défense basal : l'immunité élicitée par les PAMPs (PTI pour PAMPs-triggered immunity) (Figure I\_12). Certains pathogènes échappent à la PTI en produisant des effecteurs interférant avec la PTI, amenant à la susceptibilité élicitée par les effecteurs (ETS pour effector-triggered susceptibility) (Figure I\_12). Les effecteurs des pathogènes peuvent être reconnus par des récepteurs de plantes (protéines NB-LRR par exemple) qui mettent en place la réaction de type réponse hypersensible (HR) et l'immunité élicitée par les effecteurs (ETI pour effector-triggered immunity) (Figure I\_12). Les pathogènes produisant un effecteur différent (non-reconnu) sont sélectionnés, suppriment l'ETI et engendrent une ETS (Figure I\_12). L'évolution du pathosystème suit ce cycle en « dent de scie » (en anglais le « Boom and Bust »).

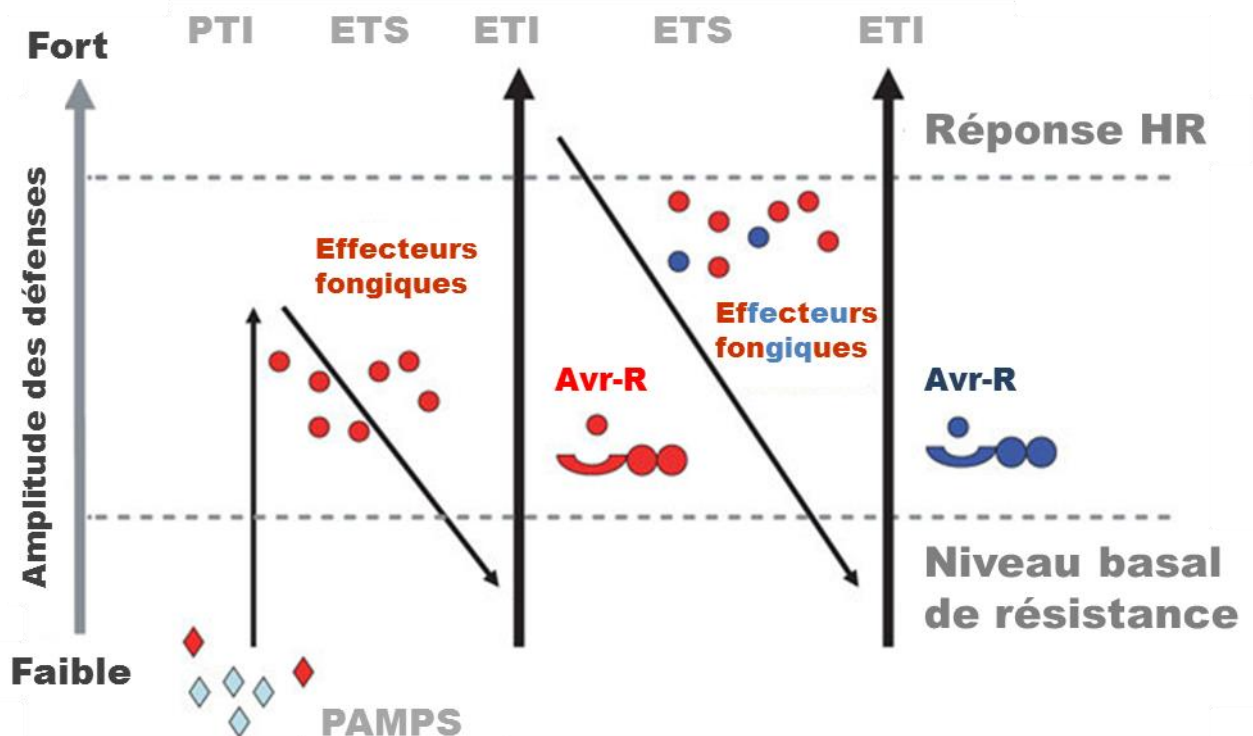


Figure I\_12 : Mécanisme moléculaire du modèle en "zig-zag" adapté de (Jones & Dangl 2006).

Construit sur la base d'un seul gène favorable, les résistances qualitatives ont été, et restent largement utilisées dans les systèmes agricoles actuels. Cependant, certains pathogènes fongiques (et oomycètes) présentent un potentiel évolutif remarquable leur permettant de contourner rapidement ces résistances. Ce potentiel est lié à des niveaux de diversité génétique accrus, qui leurs permettent de s'adapter rapidement à de nouvelles conditions. L'ensemble des forces évolutives majeures bénéfiques à l'adaptation des pathogènes fongiques est détaillé dans l' « Encadré I\_1 ».

**Encadré I\_1** : Les cinq forces évolutives pouvant contribuer à l'adaptation des pathogènes fongiques:

- **Taux de mutation** : certains pathogènes, notamment par le biais des éléments transposables présents dans leurs génomes, vont créer rapidement de nouvelles formes alléliques de gènes. La création de nouveaux allèles d'effecteurs permet notamment d'échapper aux systèmes de reconnaissance de plantes.
- **La taille des populations** : les pathogènes présentant de grandes tailles de populations, moins sensibles à la fixation d'allèles par dérive génétique, présentent une diversité et une richesse allélique plus importante.
- **Les flux de gènes** : la production de larges sources d'inoculum, dispersées sur de longues distances, permet des échanges génétiques plus importants.
- **Les cycles de reproduction** : les cycles mixtes, alternant reproduction sexuée et asexuée, sont des plus dangereux en créant de nouveaux haplotypes, rapidement dispersés en cas de sélection favorable.
- **La sélection** : de fortes pressions de sélection, notamment par la mise en place de culture monovariétale, vont favoriser l'émergence d'isolats pathogènes.

- La relation quantitative :

Bien que partielle, les résistances de type quantitatives sont supposées être plus durables, du fait du nombre important de gènes impliqués (Poland et al. 2009; Jorge et al. 2005; Fabre et al. 2012). En théorie, les résistances quantitatives sont aussi amenées à sélectionner des pathogènes plus agressifs et à une érosion des résistances au cours du temps (Gandon & Michalakis 2000; McDonald & Linde 2002). Cette théorie a été vérifiée expérimentalement chez de nombreux pathosystèmes notamment chez le couple *Zymoseptoria tritici* /blé (septoriose du blé) (Cowger & Mundt 2002), chez *Phytophthora infestans* (oomycète)/pomme de terre (mildiou de la pomme de terre) (Andrivon et al. 2007), chez *Puccinia triticina*/blé (rouille brune du blé) (Pariaud et al. 2009), chez *Melampsora larici-populina*/peuplier (la rouille du peuplier) (Dowkiw et al. 2010), chez *Venturia inaequalis*/pommier (tavelure du pommier) (Caffier et al. 2014), ou chez *Plasmopara viticola* (oomycète)/vigne (mildiou de la vigne) (Delmotte et al. 2014).

Face à cette généralisation, il apparaît comme important **de comprendre et caractériser les bases génétiques de l'agressivité** des champignons phytopathogènes, afin notamment d'en **prédire les évolutions** et d'en **limiter leurs conséquences** pour la stabilité des agrosystèmes (Lannou 2012). De telles études ont notamment été entreprises chez *Plasmopara viticola* (Delmas et al. 2016), *Melampsora larici-populina* (Pernaci et al. 2014) ou bien *Zymoseptoria tritici* (Lendenmann et al. 2014; Lendenmann et al. 2016; Stewart et al. 2016).

### **VI.2. Comment étudier les bases génétiques de l'agressivité?**

Les éléments en lien avec ce paragraphe sont présentés sous la forme d'une revue scientifique. Cette revue, en cours de construction, sera proposée ultérieurement dans *Critical Reviews in Microbiology* et traite du développement des approches de génétique quantitative chez les champignons. Bien que perfectible, elle présente les bases de la génétique quantitative, ainsi que des différentes stratégies potentielles pour l'étude de l'agressivité chez *F. graminearum*.

## Quantitative genetics in fungi: moving forward

Benoit LAURENT<sup>1</sup>, Marie-Hélène BALESSENT<sup>2</sup>, Jean CARLIER<sup>3</sup>, Pascal FREY<sup>4</sup>, Agathe MAUPETIT<sup>4</sup>, Léa PICARD<sup>3</sup>, Jessica L. SOYER<sup>2</sup>, Marie FOULONGNE-ORIOU<sup>1</sup>.

<sup>1</sup> MycSA, INRA, 33882 Villenave d'Ornon, France, <sup>2</sup> Bioger, INRA, 78000 Versailles, France, <sup>3</sup> BGPI, CIRAD, 34000 Montpellier, France, <sup>4</sup> IAM, INRA, Université de Lorraine, 54000 Nancy, France.

Address for correspondance: Marie Foulongne-Oriol; phone: +33 557 12 26 35, e-mail : [mfoulong@bordeaux.inra.fr](mailto:mfoulong@bordeaux.inra.fr)

Mailing address : INRA, UR1264 Mycologie et Sécurité des Aliments, 71 avenue Edouard Bourlaux, CS20032, 33882 Villenave d'Ornon cedex, France

Phone: +33 (0)5 57 12 26 35

Fax: +33 (0)5 57 12 25 00

E-mail: [mfoulong@bordeaux.inra.fr](mailto:mfoulong@bordeaux.inra.fr)

## **Abstract:**

In fungi, numerous traits exhibit a quantitative variation within natural populations, reflecting a complex genetic architecture. A deeper understanding of the genetic control of these traits could foster our society for many aspects. How many factors are involved in the genetic control of these traits? Where are they located on the genome? What are their relative importance and their possible interaction? All these questions are addressed through quantitative genetic approaches based on association between genotype and phenotype variation, measured in recombinant population. Two major strategies have been developed: classical quantitative trait locus (QTL) mapping and genome wide association studies (GWAS). Both strategies have received less interest in non-model fungi than in plants or animals. However, their relevancies and their uses are increasing, especially thanks to the contribution of next-generation sequencing. In this review, we will give an overview of some basics of quantitative genetics, the pros and cons of classical QTL mapping and GWAS, and according to the advantages and drawbacks that characterize fungal biology. These thoughts will be illustrated by examples and will be discussed in the light of current technical improvement and further challenges.

Keywords: Quantitative trait, genetic architecture, QTL mapping, GWA, fungal species.



## Introduction

Most of the phenotypes of living organisms show continuous and large variations within natural population. This diversity typically results from the interplay of different alleles, located at a same or at different location (*loci*) of the genomes, of which expression can be further shaped by environmental factors (Mackay et al. 2009). The basic of quantitative genetic is to look for these loci, namely quantitative trait loci (QTL), to estimate their number and their contribution for the variation of a trait, to calculate their additive effects and to test their interaction with other elements of a genome (*i.e.* epistasis). Born from mathematical models (Fisher 1918) and first experimented on model species like *Drosophila melanogaster* (Clayton et al. 1957), the current scopes of quantitative genetic are diverse, with an important contribution for plant and animal breeding (Hill et al. 2010; Xu et al. 2016), or for human disease researches (Smith & Newton-Cheh 2009). Two major strategies have been developed over time, namely classical QTL mapping (CQM) for the former and genome wide association studies (GWAS) for the second. The two approaches aim to identify statistical association between allelic polymorphism and phenotypic variation in large recombinant population (Mackay et al. 2009). In fungi, much of the quantitative genetic studies have focused yet on the model species *Saccharomyces cerevisiae*, mainly for method development and for fundamental purposes (Liti and Louis 2012; Parts 2014 for dedicated reviews). Still, the fungal kingdom is probably the most diverse and complex one, gathering ~100,000 currently known species, while probably several other millions remain to be discovered (Bass & Richards 2011; Hawksworth 2001; Tedersoo et al. 2014). In addition to their large number, fungi show numerous quantitative traits that can have negative or positive impacts on our societies. The aggressiveness/virulence of pathogens, the yield-related components of cultivated mushroom and industrial fungi, are a few examples of fungal traits of interest. In this sense, having a deeper understanding of the genetic bases of quantitative trait in fungi is of great concern for predicting for example the evolution of fungal pathogen or for speeding the breeding programs up in fungi of interest for industrial or agronomical purposes. In the recent years, the ongoing efforts made for genotyping, for phenotyping and for cultivating non-model fungi have moved the field of quantitative genetics toward a promising and appealing avenue to study those species.

In order to encourage and facilitate its uses, we propose an overview of the basics of quantitative genetics, presenting its scopes and its applications. CQM and GWAS, the two

# **Introduction Générale**

major strategies available to dissect quantitative traits will be summarized, taking in consideration their uses when working with non-model fungi, and illustrating these points with concrete examples and available tools. Finally, the recent contribution and opportunities brought by next generation sequencing, and the future challenges will be discussed.

## **What is a quantitative trait?**

Quantitative genetics focus as its name suggests on quantitative traits, namely every trait that exhibits an infinite number of values in a population. This continuous distribution supposes that several genetic factors, influenced by environmental factors, contribute to a trait (Fisher 1918). In simple terms, the phenotypic variance (Var (P)) can be broken down by

$$\mathbf{Var ( P ) = Var ( G ) + Var ( E ) + Var ( G \times E )}$$

with **var ( G )** representing the genotypic variance, **var ( E )** the environmental variance and **var ( G x E )** the genotype by environment interaction variance (Falconer & Mackay 1996). The heritability of a trait is the part of the phenotypic variance explained by the genetic variance. At its broad sense, the heritability considers the additive, dominant and epistatic (*i.e* interaction) effect of the genotypic variance. A null value of heritability means that the environment is the major player in the variation of a trait. In the other end, positive values of heritability indicate that genetic factors exist. As such, quantitative genetics propose tools to dissociate the singular genetic elements involved in the trait variation, locate them on a genome (and referring to QTL for quantitative trait locus, Figure 1) and finally estimate their additive, dominance and interaction effects (Hill et al. 2010).

## **Key parameters for the development of quantitative genetic study in fungi**

Compared to functional studies, quantitative genetics is really convenient to study polygenic traits. It is not limited by the gene number involved in a trait, nor the effect of the genes. Furthermore, it is really convenient for trait with any prior consideration and does not need mutagenesis to be handled. However, knowing the life cycle of some fungi, it is obvious that all species do not fit the frame of quantitative genetic studies. The first question that needs to be considered is whether the species can reproduce sexually or only asexually? Since both major strategies are based on segregation analyses, studies on strictly asexual fungi are highly limited. The second important parameter to consider is the phenotype. Does the trait I am interested in segregate in the population and what are the phenotyping procedure available to study that trait? For the fungi of which genotypic and phenotypic segregation occur, QCM and GWAS are powerful strategy to identify the QTL underlying the variation of a trait.

## **The different strategies: CQM, GWAS or both?**

CQM and GWAS both identify genetic factors underlying a trait on the basis on its position on the genome and follow the same principles: 1) the selection/construction of a mapping population 2) the genotyping of the progeny 3) its phenotyping and finally 4) the statistical association between the genotyping and the phenotyping values (Figure 1) (Mackay et al. 2009). The confidence brought at the position is determined by an interval of which the size will determine the possibility of identifying candidate genes (Salvi & Tuberosa 2005). Therefore, the smallest is the confidence interval, the fewest will be the gene number in the interval and the greatest the chance to identify the causal one. The size of the confidence interval ultimately depends on the marker density and the number of recombination event observed in strain genomes. As such the density of markers needs to be optimized to reveal all the recombination events that can be analysed (the historic recombination events of population for GWAS and the recombination events occurring in the cross for CQM) (Darvasi et al. 1993).

- The mapping population

CQM focus on artificial population, constructed from known progenitors that have been crossed, while GWAS focus on natural populations. Obviously, choosing one or the other method can be driven by the ability to cross strain, or the availability of a natural population to work on. For CQM more precisely, the choice of the parental strains is essential too, as it should enhance the genetic and phenotypic segregation, while staying within the intraspecies boundaries to avoid improper segregation of the chromosomes in the progeny (Cumagun 2004). Furthermore, selecting recombinant progeny can be challenging when working with homothallic species, even though example made in *F. graminearum* proved that solution exist by the induction of spontaneous nitrate utilizing mutant (Bowden & Leslie 1992) or by targeting the deletion of mating type gene in parental strains (Laurent, et al. 2016a). By definition, the allelic diversity is wider and the size of linkage disequilibrium in shorter in natural population than in experimental ones (Xu et al. 2016). Therefore, the numbers of QTL identified by GWAS are generally greater and the confidence intervals delineated at the QTL position are shorter than for CQM (Korte & Farlow 2013). However, GWAS necessitate bigger populations to reach statistical significance. Alternatively, the allelic diversity of CQM studies can be improved by multiplying population/cross number (Lendenmann et al. 2014).

The resolution of the QTL position can be improved by increasing the population size and thus the number of segregation events. CQM can be fastidious when working with fungi showing small recombination rate as observed in *Agaricus bisporus* (Sonnenberg et al. 2008). In this sense, using multigenerational mapping population, as made in plant with recombinant inbred lines for example (Darvasi et al. 1993; Xu et al. 2016) could be a judicious alternative. To our knowledge, none QTL mapping studies in fungi have been based on such population. Moreover, the phenomenon of inbreeding depression reported in some species as *Neurospora crassa* or *Agaricus bisporus* will be a limiting factor (Leslie & Raju 1985; Xu 1995).

- The genotyping

The goal of genotyping is to identify the genetic linkages that occur between the heritable elements of a genome by screening molecular markers. Because GWAS use the historical admixture that has occurred in natural population, large number of molecular markers are needed (Kruglyak 2008). For CQM, the number of molecular markers can be adapted when the recombination rate of the species is known. Nonetheless, the genotyping output tends to be not a limiting factor in current days with the arrival of high-throughput and cost effective sequencing-based genotyping methods (Davey et al. 2011 for a review). The source of variation of fungal genomes is multiple, although single nucleotide polymorphisms seems to be the most numerous in several fungal genomes (Laurent et al. 2016b; Persoons et al. 2014; Atwell et al. 2015). It doesn't mean however than only SNPs are involved in trait variation since other mutations, as CNV and or larger events can have important role for trait variation (Marian 2012; Raffaele & Kamoun 2012).

- The phenotyping

Phenotyping procedures are highly important as inaccurate measurements will impact the QTL position and the size of the confidence interval (Mackay et al. 2009). As for other organisms, the trait of interest should be appropriate to be scored on large sampling. There is no difference between CQM and GWAS for the way to measure a phenotype. Advantageously, fungal life-cycles are generally shorts and they exhibit clonal reproduction. Some are easily cultivated and stored while other can be more fastidious to study as strict biotrophic fungi. Some specific fungal characteristic can also challenge the phenotyping. This is the case for example when focusing of fruiting body formation of mushroom since the

material used for genotyping (haploid fungus) may differ on the one phenotyped (dikaryotic or diploid fungus formed using a testing strain) (Larraya et al. 2003; Foulongne-Oriol et al. 2012). In other cases, like for pathogenic fungi, it is worth noting that the phenotype can be measured on the host rather than on the fungi directly. It is also understandable that measuring interaction can be more sensitive to environment. As such, it is recommended to increase replicate number to improve the statistical power. Advantageously, the clonal reproduction of fungi allows multiplying easily the number of replicates. In some cases, highly-controlled environment can be preferred for increasing QTL detection power (Rahaman et al. 2015). The statistical analysis

QCM and GWAS differ by the statistical methods used to identify QTL. These methods are summarized in the Box 1. Natural populations exhibit highly variable allele frequency. Therefore, GWAS can be limited to detect allele with minor frequency ( $< 5\%$ ) as they will not bring enough statistical insight to be associated with a trait - even though they are involved in its variation (Gorlov et al. 2011). Furthermore, GWAS are more sensitive to false positive association, especially when strong structures exist within groups (Korte & Farlow 2013). These structure can be corrected but reducing the power of the analysis (Box 1). In this sense, CQM is generally thought to be more powerful to detect QTL. As for genotyping, there is no particular way to analyse data in fungi (Box 1) and most of the tools and software primarily developed for plant or animal (Xu et al. 2016), can be used for fungi. One could note however that with the increasing number of molecular markers produced by sequencing-based genotyping methods, pioneer software developed for CQM can appear to be outdated.

To conclude, GWAS and CQM are no better option to dissect quantitative trait. Both methods can complement each other limitations, making the best strategy to adopt the combination of the two. This has been illustrated in maize, showing that the power of detection of QTL by CQM, can be combined with the resolution of GWAS to delineate shorter interval at the QTL position (Mammadov et al. 2015). However, besides few examples in plants, the combination of both strategies has never been developed for studying fungal traits.

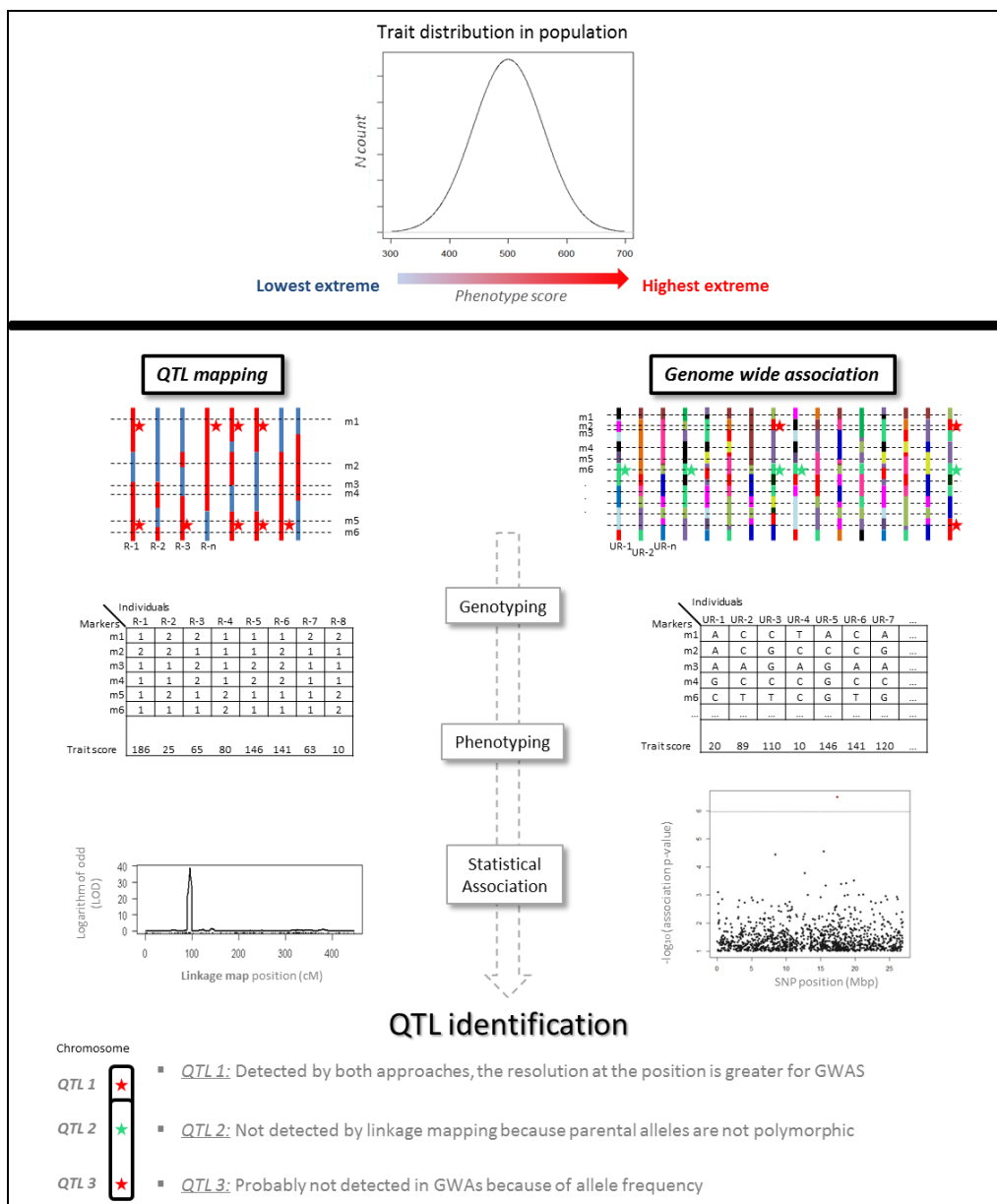


Figure 1: Illustration of CQM and GWAS, adapted from Mackay et al. 2009.

# Introduction Générale

**Table 1:** Non exhaustive list of quantitative genetic analysis in fungi.

Species	Traits	references
<b>QTL mapping</b>		
<i>Agaricus bisporus</i>	yield-related traits, cap color, disease resistance traits, bruising sensitivity	Moquet et al. 1999; Foulongne-Oriol et al. 2011, 2012a,b; Gao et al. 2015
<i>Amylostereum areolatum</i>	vegetative growth	van der Nest, 2009
<i>Aspergillus nidulans</i>	virulence	Christians et al. 2011.
<i>Cryptococcus neoformans</i>	virulence, hyphal growth	Lin et al. 2006
<i>Fusarium fujikoroii</i> species complex	growth morphology, pathogenicity	de Vos et al. 2011
<i>Fusarium graminearum</i>	pathogenicity, aggressiveness	Cumagun et al. 2003, 2004
<i>Heterobasidium annosum</i> s.l.	somatic incompatibility, intraspecific interaction, virulence, growth rate, biotic interaction	Olson, 2006; Lind et al. 2007a,b; Samils et al. 2008
<i>Lentinula edodes</i>	vegetative growth rate, fruiting body-related traits	Gong et al. 2014, 2016
<i>Neurospora crassa</i>	circadian clock related traits	Kim et al. 2007
<i>Pleurotus ostreatus</i>	vegetative growth rate, ligning-degrading enzymatic activities, yield and quality related traits	Larrraya et al. 2002, 2003; Santoyo et al. 2008
<i>Zymoseptoria tritici</i>	melanization, fungicide resistance, thermal adaptation, growth morphology	Lendenmann et al. 2014, 2015, 2016
<b>Association genetics/Genome wide association analysis</b>		
<i>Fusarium graminearum</i>	aggressiveness, deoxyrnivalenol production, azole sensitivity	Talás et al., 2012, 2016
<i>Heterobasidium annosum</i> s.s.	virulence	Dalman et al. 2013
<i>Neurospora crassa</i>	microbial communication	Palma-Guerrero et al. 2013
<i>Magnaporthe oryzae</i>	virulence	Yoshida et al. 2009



## Box 1:

**CQM:** three classes of statistical methods exist: regression (Haley and Knott, 1992), maximum-likelihood (Churchill and Doerge, 1994), and Bayesian models (Sillanpää and Arjas, 1998). The significance of an association is indicated by the LOD score, corresponding to the decimal logarithm of odd to have an association. Different methods have been proposed to choose appropriate thresholds (Churchill and Doerge, 1994; Van Ooijen, 1999; Vieland, 1998). Single-marker analysis (regression) is the simplest way to identify a QTL and identify association on the basis of difference between the mean values of the phenotypes in relation to marker genotype. This method lacks however accuracy on QTL positioning and cannot estimate its effects (Li et al. 2007). Interval mapping (maximum-likelihood or regression interval mapping) increment the genetic distances between markers, bringing a deeper estimation of QTL location. Composite interval mapping (CIM), by combining interval mapping and multiple-regression overcome these issues. Multiple QTL mapping (MQM), based on Bayesian statistics, also do and infer different situations as for example, the number of QTL, the genomic positions and the genetic effects (Yi & Shriner 2008).

**GWAS:** Mixed linear model on multiple SNPs are used to investigate association with a trait of interest (Eu-ahsunthornwattana et al. 2014 for review). To avoid spurious associations, the linkage disequilibrium is corrected with population structure, represented by the kinship matrix (K).

$$Y_i = \beta_0 + x_{ij}\beta_j + g_i + \varepsilon_i, \text{ where } g_i \sim N(0, \sigma_g^2 K) \wedge \varepsilon_i \sim N(0, \sigma_\varepsilon^2 I)$$

Where  $Y_i$  is a vector of phenotypic observation of the  $i$ th individual;  $\beta_0$  is the intercept term;  $x_{ij}\beta_j$  represents fixed effects for the  $j$ th SNP;  $g_i$  is a vector of all SNPs effects; and  $\varepsilon$  is a vector of residual effects;  $K$  is an  $(n \times n)$  matrix of relative kinship coefficients that define the degree of genetic covariance between a pair of individuals; and  $I$  is an  $(n \times n)$  identity matrix (Yu et al. 2006). Very low  $p$ -values are needed to consider an association although the significance threshold depends on gene effect sizes and sample size (Burton et al. 2007). The mixed model can control the confounding effects of a background of multiple loci of small effects, but it is not expected to handle the confounding caused by loci of moderate to large effects.

## Quantitative genetics in fungi: examples and prospects?

In this section, we propose to give an overview of the application scopes of quantitative genetics in fungi, based on concrete research examples (Table 1) and on more projective ones that could be considered in a close future.

**For breeding purposes:** mushrooms constitute a part of the human diet since earliest history, due to their nutrient properties, peculiar flavor or medicinal attributes (Valverde et al. 2015).

As a crop species, mushroom breeding programs have been developed in order to create new varieties in response to cultivation constraints, market trends and consumer demands. The methodologies used in mushroom breeding are quite similar to those developed for plant breeding (Sonnenberg et al. 2008). The genetics of various agronomical traits such as yield, earliness or cap color have been investigated by QTL mapping in *Agaricus bisporus*, *Pleurotus spp.*, and *Lentinula edodes* (Foulongne-Oriol et al. 2012; Gong et al. 2016; Im et al. 2016; Larraya et al. 2003). Technological traits related to industrial mushroom production have been also studied such as bruising sensitivity in *A. bisporus* (Gao et al. 2015) or cold storage effect in *Pleurotus ostreatus* (Larraya et al. 2003). The genetic control of the resistance to biotic stress has been explored in *A. bisporus*: QTLs for resistance to the three major diseases affecting *A. bisporus* culture, which are bacterial brown blotch (caused by *Pseudomonas tolaasii*), dry bubble (caused by *Lecanicillium fungicola*), and green mold (caused by the fungal competitor *Trichoderma aggressivum*), have been found (Foulongne-Oriol et al. 2010; Moquet et al. 1999). In the context of global warming, abiotic stress-related traits, as the adaptation to higher fruiting temperature of *A. bisporus* has also been studied (Foulongne-Oriol et al. 2014). It is noteworthy that all the quantitative genetic studies cited above are based on a classical QTL mapping strategy. So far, no genetic association study, neither with a restricted number of loci nor at whole genome level, has been developed for cultivated mushroom species. Following the advance made in plant quantitative genetics, together with the increasing affordability of sequencing technology, we may expect that GWAS will become a powerful tool for detecting natural variation underlying complex traits in mushroom crops.

Overall, the increasing powers of quantitative genetics open the field to new possibilities in mushroom breeding, as illustrated recently by the development of a linkage map in *A. subrufescens*, a cultivable mushroom with pharmaceutical properties (Foulongne-Oriol et al. 2016). Improving agronomical traits in edible ones, or enhancing the production of nutraceutical and pharmaceutical important secondary metabolites in others, are promising and feasible perspectives.

***For fungal pathogens studies:*** knowledge of the genetic architecture of qualitative and quantitative traits and their mapping on the genome is an important step in understanding the mechanisms of pathogenicity. In human and animal pathogens, quantitative genetic studies are mainly focused on virulence (in the sense of animal pathology, i.e. quantitative virulence as defined by Stewart et al. 2016). In the animal pathogen *Aspergillus nidulans*, a QTL

mapping approach allowed the identification of genes required for virulence that could be used as targets for novel treatments (Christians et al. 2011). Similarly, Lin et al. (2006) applied quantitative trait loci (QTL) mapping to an inbred population of *Cryptococcus neoformans* in order to identify genes controlling biological processes central to fungal development and pathogenicity, such as hyphal growth, melanin production, and high temperature growth. Several QTLs associated with virulence-related traits (melanin production, cell size, cell wall thickness, frequency of capsule production) and drug (fluconazole) resistance were discovered in a hybrid cross between *C. neoformans* and *C. deneoformans* (Vogan et al. 2016). In 2011, Muller et al. proposed the first GWAS study in fungi and identified 20 significant SNPs linked to *S. cerevisiae* pathogenicity, among which 12 were located within genes. More recently, the study of fungal plant pathogens is bringing much attention. Indeed, whereas resistant plant breeding is a promising and environment-friendly alternative to the use of pesticides, it becomes confronted to the adaptive potential of fungal pathogens. Much of the breeding efforts have yet focused on the use of qualitative resistance (i.e. major resistance genes) due to their immediate and qualitative efficacy and the easiness to introduce single gene resistance in new varieties. In annual crops, and even more in perennial crops, the wide inequality between the pathogen's rapid generation time and the time needed to deploy new host varieties has led to rapid breakdown of qualitative resistance genes (Xu et al. 2012). This prompted plant breeders to search for quantitative resistances, which are postulated to be polygenic, race non-specific and more durable than qualitative resistances (Brun et al. 2010; Fabre et al. 2012; Jorge et al. 2005). Nevertheless, quantitative resistance can also be challenged by the evolution of aggressiveness, understood as the quantitative component of the pathogenicity and determined by several disease-associated traits (Andrivon et al. 2007; Caffier et al. 2014; Delmas et al. 2016; Dowkiw et al. 2010; Pariaud et al. 2009). Thus it is of primary importance to assess the potential evolution of aggressiveness traits in the pathogen and the potential trade-offs between these traits (Lannou 2012). In this sense, either QTL mapping or GWAS were used to identify genomic regions involved in traits directly linked to pathogenicity in several Ascomycetes and Basidiomycetes, such as mycelial growth rate (Lind et al. 2007; Olson 2006; De Vos et al. 2011), lesion size (Dalman et al. 2013; Lind et al. 2007; Stewart et al. 2016), disease severity measured in the field (Cumagun & Miedaner 2004; Talas et al. 2016). QTL mapping and association genetics also allowed identifying genomic regions involved in virulence (in the sense of plant pathology, i.e. the qualitative component of pathogenicity) in *Pyrenophora teres* f. *teres* (Shjerve et al. 2014) and in *Magnaporthe oryzae* (Yoshida et al. 2009). Other traits, less

directly linked to pathogenicity, were also mapped, such as mycotoxins production (Cumagun & Miedaner 2004; Talas et al. 2012; Talas et al. 2016), melanization (Lendenmann et al. 2014; Lendenmann et al. 2015), fungicide resistance (Lendenmann et al. 2015; Talas et al. 2016) or temperature sensitivity (Lendenmann et al. 2016).

***For fundamental purposes:*** Among the basic biological traits, circadian clocks play a crucial role in fungal life cycle completion (Hevia et al. 2016; Montenegro-Montero et al. 2015). Those clocks are molecular circuits that allow organisms to temporally coordinate numerous biological processes in synchrony with daily environmental cycles. An original study describing QTL for circadian rhythms related traits in the fungal model *Neurospora crassa* has been proposed by Kim et al. (2007) and identified loci unlinked to any previously known clock genes. Similarly, the self and non-self-recognition is fundamental in fungi to maintain mycelium integrity and genotype stability. If numerous single genes have been described (Hall et al. 2010), the somatic incompatibility reaction is not only a strict binary character, as illustrated by the QTL mapping of intraspecific interactions related traits in *Heterobasidion annosum* (Lind et al. 2007). The fungal communication underlying recognition process has been investigated through GWAS and allowed the identification of several loci involved in germling communication in *N. crassa* (Palma-Guerrero et al. 2013). These examples of quantitative genetics in *N. crassa* illustrate very well the relevance of such kind of forward genetics approach to disentangle complex traits up to the identification of underlying molecular pathways. Indeed, in each case, it allows the uncovering of novel candidate genes that have never been described before as being involved in the trait studied (Kim et al. 2007; Palma-Guerrero et al. 2013).

***Towards new challenges:*** Since any trait related to fungal biology could be considered under the quantitative perspective, a few examples for which quantitative genetics analyses have been or could be considered are given below. For example, the ectomycorrhizal association between plants and fungi is of key importance for the stability of ecosystem. The model symbiosis between *Poplar* spp. and the basidiomycete *Laccaria bicolor* has been already investigated through QTL mapping in the plant (Labbé et al. 2011), but similar understanding from its fungal partner is also foreseeable and feasible (Labbé et al. 2008; Sanders & Croll 2010). Although often forgotten, the initial purpose of quantitative genetics was to bear evolutionary explanations (Fisher 1932). It is now able to answer critical questions about evolution of species (Rajon & Plotkin 2013). Despite the lack of concrete examples,

quantitative genetics could be employed to understand the evolution of fungal species. An environmental challenge involving fungi is the development of alternative to fossil fuels through the bioconversion processes and the transformation of plant polysaccharides in bio-alcohol. The production of fungal lignocellulolytic enzymes is of importance in the emergence of new biotechnologies for the treatment of lignocellulosic material. In *Pleurotus ostreatus*, the enzymatic activities of key enzymes dedicated to lignin degradation have been studied through QTL mapping (Santoyo et al. 2008). By this way, the authors have identified various genomic regions that do not map to genomic regions sheltering the corresponding structural genes and that may contain regulators of the enzyme activities. Perhaps not so surprisingly, quantitative genetic studies on *Penicillium spp.* are inexistent despite their industrial importance in food technology. The development of classical QTL mapping is hardly conceivable since these species are commonly classified as being “imperfect”, i.e. unable to have a sexual reproduction. However, the recent evidences of sexual reproduction and recombination, together with a great phenotypic and genetic diversity in *P. roqueforti* (Gillot et al. 2015; Ropars et al. 2014) offers opportunities for further quantitative genetic studies of technological related traits in this species using either QTL mapping or GWAS.

## **From QTL to QTN: the added-value of NGS**

The ultimate goal of QTL mapping is to reduce the confidence interval from large chromosomal segments encompassing multiple genes to a reduced number of candidate genes and eventually to detect the single quantitative trait nucleotides (QTN) explaining a trait. Whereas this remains a major challenge, it has been achieved in a few examples in plant and even recently in a phytopathogenic fungus (Lendenmann et al. 2016). The advent of new sequencing technologies combined with a constant decrease of the sequencing costs renders possible the sequencing of a high number of individuals from an experimental or a natural population. Restriction site associated DNA (RAD) sequencing (Baird et al. 2008), and other strategy using reduced representation of a genome are especially relevant (Davey et al. 2011), by optimizing the number of markers and the number of individuals genotyped (Talas & McDonald 2015; Laurent, et al. 2016a; Lendenmann et al. 2014). This strategy increases the resolution of QTLs detection and allows the identification of the nucleotide variation underlying a phenotype (QTN) (Talas et al. 2016). In the case of CQM, NGS makes important contribution to the identification of candidate gene by the sequencing of whole genome sequences, or at least from the confidence interval sequence. In the case of classical mapping

approach, the sequence of the parental genome allows identifying candidate genes affected by non-synonymous mutations (Stewart et al. 2016; Anderson et al. 2016; Christians et al. 2011). Whereas sequencing full genomes of several hundreds of fungi has been not undertaken for GWAS yet, the continuous decrease of the sequencing operating costs let us foresee such approach feasible in a close future to identify causal mutation in natural population.

## **Where are we going to?**

We have seen that the genotyping is not considered anymore as a limitation for the development of quantitative genetics in fungi. The classical marker-based genotyping approach is being replaced by genome-based methods, allowed by the next generation sequencing technologies or high-throughput genotyping platforms. We are not far from genotyping by whole genome resequencing. Therefore, beyond some biological constraints inherent to some species that definitely exclude the possibility to perform quantitative genetics, the most critical step that the fungal geneticists will have to face in the future is the assessment of the phenotype in its whole. Examples cited in this review are far from the 223 traits evaluated for a GWAS study in the yeast *S. pombe* (Jeffares et al. 2015). Accessing the phenome through reliable and repeatable high-throughput phenotyping concepts and technologies is one of the next challenges. For example, for fungal biotechnological traits of interest, automated and standardized in vitro assay in micro-cultivation process could be developed as done for yeast (Jung et al. 2015). The measurement of agronomical traits in some mushroom could be based on high resolution morphometric analyses achieved through imaging technologies (Li et al. 2014; Rahaman et al. 2015). To phenotype host-pathogen interaction, similar approaches have been developed to quantify plant disease symptoms (Mutka & Bart 2014). A recent automated image analysis has been developed to estimate aggressiveness related traits of the wheat pathogen *Zymoseptoria tritici* (Stewart & McDonald 2014). Besides, this method has been used to measure those traits on two mapping populations encompassing near 520 strains, and leading to the detection of underlying QTL (Lendenmann et al. 2016). Examples of quantitative genetics applied in fungi available in the literature are all based on phenotypic data, considered as the ultimate expression of the genotype. However, such a point of view is quite restrictive. Numerous variants associated in quantitative traits are found in non-coding regions of the genome, suggesting that they are therefore likely to be involved in gene regulation (Talas et al. 2016). Indeed, regarding the association between the allelic polymorphism and the phenotypic value omits all of the

intermediate steps which led to the observed trait variation from the DNA sequence. Intermediate molecular phenotypes such as transcript abundance, proteins representativeness or metabolites contents may also vary genetically in populations and in this, could be considered themselves as quantitative traits (Mackay et al. 2009). Loci that underlie those traits are referred to as eQTL, pQTL and mQTL respectively. Another level of variation which could be taken into account as a quantitative trait is the dynamic interplay between chromatin structure and DNA sequence factors (Johannes et al. 2008; Taudt et al. 2016). Quantifying the impact of such heritable epigenetic variation on complex traits is an emerging challenge in population genetics (Cortijo 2014) and example of epi-QTL - that refers to QTL influencing chromatin states - are still scarce (Kooke & Keurentjes 2015; Long et al. 2011). The analysis of molecular traits allowed the genetic integration of the different levels of regulation that govern the phenotype variation (Albert & Kruglyak 2015). Up to now, the reconstruction of such QTL networks are limited to model organisms (Jansen et al. 2009; Fu et al. 2009; Civelek & Lusis 2014), although possibilities of development in non-model fungi are existing and encouraging.

## **CONCLUSION**

The study of fungal quantitative traits is of ecological, industrial and societal importance. GWAS and CQM are two complementary and powerful approaches to dissect the genetic basis of these traits, of which uses in non-model fungi is rising. The recent increase number of study case, certainly boosted by the contribution of next generation sequencing, highlight the scopes of application of these methods. Whereas major advances have been made in model species as in yeast, the margin of progress in non-model fungal species is real and promising.

## **Acknowledgements**

The authors would like to thank Bernadette Rubio for constructive talks and for comments.

## **Declaration of interest**

The authors report no declaration of interest.

## References:

- Albert, F.W. & Kruglyak, L., 2015. The role of regulatory variation in complex traits and disease. *Nature Reviews Genetics*, 16(4), p.197-212.
- Anderson, C. et al., 2016. Genome analysis and avirulence gene cloning using a high-density RADseq linkage map of the flax rust fungus, *Melampsora lini*. *BMC Genomics*, 17(1), p.667.
- Andrivon, D. et al., 2007. Adaptation of *Phytophthora infestans* to Partial Resistance in Potato: Evidence from French and Moroccan Populations. <http://dx.doi.org/10.1094/PHYTO-97-3-0338>.
- Atwell, S. et al., 2015. Whole genome resequencing of *Botrytis cinerea* isolates identifies high levels of standing diversity. *Front Microbiol*, 6, p.996.
- Baird, N.A. et al., 2008. Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLoS One*, 3(10), p.e3376.
- Bass, D. & Richards, T.A., 2011. Three reasons to re-evaluate fungal diversity 'on Earth and in the ocean'. *Fungal Biology Reviews*, 25(4), p.159-164.
- Bowden, R.L. & Leslie, J.F., 1992. Nitrate-nonutilizing mutants of *Gibberella zeae* (*Fusarium graminearum*) and their use in determining vegetative compatibility. *Experimental Mycology*, 16(4), p.308-315.
- Brun, H. et al., 2010. Quantitative resistance increases the durability of qualitative resistance to *Leptosphaeria maculans* in *Brassica napus*. *The New phytologist*, 185(1), p.285-99.
- Burton, P.R. et al., 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, 447(7145), p.661-678.
- Caffier, V. et al., 2014. Erosion of quantitative host resistance in the apple×*Venturia inaequalis* pathosystem. *Infection, Genetics and Evolution*, 27, p.481-489.
- Christians, J.K. et al., 2011. Quantitative trait locus (QTL) mapping reveals a role for unstudied genes in *Aspergillus* virulence. *PLoS One*, 6(4), p.e19325.
- Civelek, M. & Lusk, A.J., 2014. Systems genetics approaches to understand complex traits. *Nature reviews. Genetics*, 15(1), p.34-48.
- Clayton, G.A., Morris, J.A. & Robertson, A., 1957. An experimental check on quantitative genetical theory I. Short-term responses to selection. *Journal of Genetics*, 55(1), p.131-151.
- Cortijo, S., 2014. Mapping the epigenetic basis of complex traits. *Science*.
- Cumagun, 2004. Genetic mapping of pathogenicity and aggressiveness of *Gibberella zeae* (*Fusarium graminearum*) toward wheat. *Genetics and Resistance*.
- Cumagun, C.J.R. & Miedaner, T., 2004. Segregation for aggressiveness and deoxynivalenol production of a population of *Gibberella zeae* causing head blight of wheat. *European Journal of Plant Pathology*, 110(8), p.789-799.
- Dalman, K. et al., 2013. A genome-wide association study identifies genomic regions for virulence in the non-model organism *Heterobasidion annosum* s.s. *PLoS One*, 8(1), p.e53525.
- Darvasi, A. et al., 1993. Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. *Genetics*, 134(3).
- Davey, J.W. et al., 2011. Genome-wide genetic marker discovery and genotyping using next-



- generation sequencing. *Nat Rev Genet*, 12(7), p.499-510.
- Delmas, C.E.L. et al., 2016. Adaptation of a plant pathogen to partial host resistance: selection for greater aggressiveness in grapevine downy mildew. *Evolutionary Applications*, 9(5), p.709-725.
- Dowkiw, A., Voisin, E. & Bastien, C., 2010. Potential of Eurasian poplar rust to overcome a major quantitative resistance factor. *Plant Pathology*, 59(3), p.523-534.
- Eu-ahsunthornwattana, J. et al., 2014. Comparison of Methods to Account for Relatedness in Genome-Wide Association Studies with Family-Based Data G. R. Abecasis, éd. *PLoS Genetics*, 10(7), p.e1004445.
- Fabre, F. et al., 2012. Durable strategies to deploy plant resistance in agricultural landscapes. *New Phytologist*, 193(4), p.1064-1075.
- Falconer, D.S. & Mackay, T., 1996. *Introduction to Quantitative Genetics*. , Ed. 4.
- Fisher, R., 1932. The bearing of genetics on theories of evolution. *Science Progress in the Twentieth Century (1919-1933)*.
- Fisher, R.A., 1918. The correlation between Relatives on the Supposition of Mendelian Inheritance.
- Foulongne-Oriol, M. et al., 2010. An expanded genetic linkage map of an intervarietal *Agaricus bisporus* var. *bisporus* × *A. bisporus* var. *burnettii* hybrid based on AFLP, SSR and CAPS markers sheds light on the recombination behaviour of the species. *Fungal Genetics and Biology*, 47(3), p.226-236.
- Foulongne-Oriol, M. et al., 2014. Deciphering the ability of *Agaricus bisporus* var. *burnettii* to produce mushrooms at high temperature (25°C). *Fungal Genetics and Biology*, 73, p.1-11.
- Foulongne-Oriol, M. et al., 2012. Quantitative trait locus mapping of yield-related components and oligogenic control of the cap color of the button mushroom, *Agaricus bisporus*. *Appl Environ Microbiol*, 78(7), p.2422-2434.
- Foulongne-Oriol, M. et al., 2016. The Genetic Linkage Map of the Medicinal Mushroom *Agaricus subrufescens* Reveals Highly Conserved Macrosynteny with the Congeneric Species *Agaricus bisporus*. *G3 & Genes/Genomes/Genetics*, 6(5), p.1217-1226.
- Fu, J. et al., 2009. System-wide molecular evidence for phenotypic buffering in *Arabidopsis*. *Nature Genetics*, 41(2), p.166-167.
- Gao, W. et al., 2015. Quantitative trait locus mapping for bruising sensitivity and cap color of *Agaricus bisporus* (button mushrooms). *Fungal Genetics and Biology*, 77, p.69-81.
- Gillot, G. et al., 2015. Insights into *Penicillium roqueforti* Morphological and Genetic Diversity H. T. Lumbsch, éd. *PLOS ONE*, 10(6), p.e0129849.
- Gong, W. et al., 2016. Genetic dissection of fruiting body-related traits using quantitative trait loci mapping in *Lentinula edodes*. *Applied Microbiology and Biotechnology*, 100(12), p.5437-5452.
- Gorlov, I. et al., 2011. Evolutionary evidence of the effect of rare variants on disease etiology. *Clinical Genetics*, 79(3), p.199-206.
- Grover, A. & Sharma, P.C., 2016. Development and use of molecular markers: past and present. *Critical Reviews in Biotechnology*, 36(2), p.290-302.
- Grünwald, N.J., McDonald, B.A. & Milgroom, M.G., 2016. Population Genomics of Fungal and Oomycete Pathogens. *Annual Review of Phytopathology*, 54(1), p.323-346.

- Hall, C. et al., 2010. Evolution and diversity of a fungal self/nonself recognition locus. *PloS one*, 5(11), p.e14055.
- Hawksworth, D.L., 2001. The magnitude of fungal diversity: the 1.5 million species estimate revisited. *Mycological Research*, 105(12), p.1422-1432.
- Hevia, M.A., Canessa, P. & Larrondo, L.F., 2016. Circadian clocks and the regulation of virulence in fungi: Getting up to speed. *Seminars in Cell & Developmental Biology*.
- Hill, W.G. et al., 2010. Understanding and using quantitative genetic variation. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 365(1537), p.73-85.
- Im, C.H. et al., 2016. Construction of a genetic linkage map and analysis of quantitative trait loci associated with the agronomically important traits of *Pleurotus eryngii*. *Fungal Genetics and Biology*, 92, p.50-64.
- Jansen, R.C. et al., 2009. Defining gene and QTL networks. *Current opinion in plant biology*, 12(2), p.241-6.
- Jeffares, D.C. et al., 2015. The genomic and phenotypic diversity of *Schizosaccharomyces pombe*. *Nature genetics*, 47(3), p.235-41.
- Johannes, F., Colot, V. & Jansen, R.C., 2008. Epigenome dynamics: a quantitative genetics perspective. *Nature Reviews Genetics*, 9(11), p.883-890.
- Jorge, V. et al., 2005. Genetic architecture of qualitative and quantitative *Melampsora larici-populina* leaf rust resistance in hybrid poplar: genetic mapping and QTL detection. *New Phytologist*, 167(1), p.113-127.
- Jung, P.P. et al., 2015. Protocols and programs for high-throughput growth and aging phenotyping in yeast. *PloS one*, 10(3), p.e0119807.
- Kim, T.-S. et al., 2007. Quantitative trait loci for the circadian clock in *Neurospora crassa*. *Genetics*, 177(4), p.2335-47.
- Kooke, R. & Keurentjes, J.J.B., 2015. Epigenetic variation contributes to environmental adaptation of *Arabidopsis thaliana*. *Plant signaling & behavior*, 10(9), p.e1057368.
- Korte, A. & Farlow, A., 2013. The advantages and limitations of trait analysis with GWAS: a review. *Plant methods*, 9, p.29.
- Kruglyak, L., 2008. The road to genome-wide association studies. *Nature Reviews Genetics*, 9(4), p.314-318.
- Labbé, J. et al., 2008. A genetic linkage map for the ectomycorrhizal fungus *Laccaria bicolor* and its alignment to the whole-genome sequence assemblies. *New Phytologist*, 180(2), p.316-328.
- Labbé, J. et al., 2011. Identification of quantitative trait loci affecting ectomycorrhizal symbiosis in an interspecific F1 poplar cross and differential expression of genes in ectomycorrhizas of the two parents: *Populus deltoides* and *Populus trichocarpa*. *Tree Genetics & Genomes*, 7(3), p.617-627.
- Lannou, C., 2012. Variation and selection of quantitative traits in plant pathogens. *Annu Rev Phytopathol*, 50, p.319-338.
- Larraya, L.M. et al., 2003. Mapping of Genomic Regions (Quantitative Trait Loci) Controlling Production and Quality in Industrial Cultures of the Edible Basidiomycete *Pleurotus ostreatus*. *Applied and Environmental Microbiology*, 69(6), p.3617-3625.
- Laurent, B., Palaiokostas, C., et al., 2016a. High resolution mapping of the recombination landscape of

- the phytopathogen *Fusarium graminearum* suggests two-speed genome evolution. *Molecular Plant Pathology*, Submitted.
- Laurent, B., Moinard, M., et al., 2016b. Landscape of genomic diversity and host adaptation in *Fusarium graminearum*. *BMC genomics*, Submitted.
- Lendenmann, M.H. et al., 2016. QTL mapping of temperature sensitivity reveals candidate genes for thermal adaptation and growth morphology in the plant pathogenic fungus *Zymoseptoria tritici*. *Heredity*, 116(4), p.384-394.
- Lendenmann, M.H. et al., 2014. Quantitative trait locus mapping of melanization in the plant pathogenic fungus *Zymoseptoria tritici*. *G3: Genes/Genomes/Genetics*, 4(12), p.2519-2533.
- Lendenmann, M.H., Croll, D. & McDonald, B.A., 2015. QTL mapping of fungicide sensitivity reveals novel genes and pleiotropy with melanization in the pathogen *Zymoseptoria tritici*. *Fungal Genetics and Biology*, 80, p.53-67.
- Leslie, J.F. & Raju, N.B., 1985. Recessive mutations from natural populations of *Neurospora crassa* that are expressed in the sexual diplophase. *Genetics*, 111(4), p.759-77.
- Li, H., Ye, G. & Wang, J., 2007. A modified algorithm for the improvement of composite interval mapping. *Genetics*, 175(1), p.361-74.
- Li, L., Zhang, Q. & Huang, D., 2014. A Review of Imaging Techniques for Plant Phenotyping. *Sensors*, 14(11), p.20078-20111.
- Lin, X. et al., 2006. Virulence attributes and hyphal growth of *C. neoformans* are quantitative traits and the MAT $\alpha$  allele enhances filamentation. *PLoS genetics*, 2(11), p.e187.
- Lind, M. et al., 2007. Identification of quantitative trait loci affecting virulence in the basidiomycete *Heterobasidion annosum* s.l. *Current Genetics*, 52(1), p.35-44.
- Liti, G. & Louis, E.J., 2012. Advances in quantitative trait analysis in yeast. *PLoS genetics*, 8(8), p.e1002912.
- Long, Y. et al., 2011. Epigenetic QTL mapping in *Brassica napus*. *Genetics*, 189(3), p.1093-102.
- Mackay, T.F., Stone, E.A. & Ayroles, J.F., 2009. The genetics of quantitative traits: challenges and prospects. *Nat Rev Genet*, 10(8), p.565-577.
- Mackay, T.F.C., Stone, E.A. & Ayroles, J.F., 2009. The genetics of quantitative traits: challenges and prospects. *Nature Reviews Genetics*, 10(8), p.565-577.
- Mammadov, J. et al., 2015. Combining powers of linkage and association mapping for precise dissection of QTL controlling resistance to gray leaf spot disease in maize (*Zea mays* L.). *BMC Genomics*, 16(1), p.916.
- Marian, A.J., 2012. Elements of « missing heritability ». *Current opinion in cardiology*, 27(3), p.197-201.
- Montenegro-Montero, A., Canessa, P. & Larrondo, L.F., 2015. Chapter Four – Around the Fungal Clock: Recent Advances in the Molecular Study of Circadian Clocks in *Neurospora* and Other Fungi. *Advances in Genetics*, 92, p.107-184.
- Moquet, F. et al., 1999. A Quantitative Trait Locus of *Agaricus bisporus* Resistance to *Pseudomonas tolaasii* Is Closely Linked to Natural Cap Color. *Fungal Genetics and Biology*, 28(1), p.34-42.
- Muller, L.A.H. et al., 2011. Genome-wide association analysis of clinical vs. nonclinical origin provides insights into *Saccharomyces cerevisiae* pathogenesis. *Molecular Ecology*, 20(19),

p.4085-4097.

- Mutka, A.M. & Bart, R.S., 2014. Image-based phenotyping of plant disease symptoms. *Frontiers in plant science*, 5, p.734.
- Olson, Å., 2006. Genetic linkage between growth rate and the intersterility genes S and P in the basidiomycete *Heterobasidion annosum* s.lat. *Mycological Research*, 110(8), p.979-984.
- Palma-Guerrero, J. et al., 2013. Genome Wide Association Identifies Novel Loci Involved in Fungal Communication J. M. Akey, éd. *PLoS Genetics*, 9(8), p.e1003669.
- Pariaud, B. et al., 2009. Aggressiveness Components and Adaptation to a Host Cultivar in Wheat Leaf Rust. *Phytopathology*, 99(7), p.869-878.
- Parts, L., 2014. Genome-wide mapping of cellular traits using yeast. *Yeast (Chichester, England)*, 31(6), p.197-205.
- Persoons, A. et al., 2014. Patterns of genomic variation in the poplar rust fungus *Melampsora larici-populina* identify pathogenesis-related factors. *Frontiers in Plant Science*, 5.
- Raffaele, S. & Kamoun, S., 2012. Genome evolution in filamentous plant pathogens: why bigger can be better. *Nature Reviews Microbiology*, 10(6), p.417.
- Rahaman, M.M. et al., 2015. Advanced phenotyping and phenotype data analysis for the study of plant growth and development. *Frontiers in plant science*, 6, p.619.
- Rajon, E. & Plotkin, J.B., 2013. The evolution of genetic architectures underlying quantitative traits. *Proc Biol Sci*, 280(1769), p.20131552.
- Ropars, J. et al., 2014. Induction of sexual reproduction and genetic diversity in the cheese fungus *Penicillium roqueforti*. *Evolutionary applications*, 7(4), p.433-41.
- Salvi, S. & Tuberosa, R., 2005. To clone or not to clone plant QTLs: present and future challenges. *Trends in Plant Science*, 10(6), p.297-304.
- Sanders, I.R. & Croll, D., 2010. Arbuscular mycorrhiza: the challenge to understand the genetics of the fungal partner. *Annual review of genetics*, 44, p.271-92.
- Santoyo, F. et al., 2008. Quantitative linkage mapping of lignin-degrading enzymatic activities in *Pleurotus ostreatus*. *Enzyme and Microbial Technology*, 43(2), p.137-143.
- Shjerve, R.A. et al., 2014. Evaluation of a *Pyrenophora teres* f. *teres* mapping population reveals multiple independent interactions with a region of barley chromosome 6H. *Fungal Genetics and Biology*, 70(0), p.104-112.
- Smith, J.G. & Newton-Cheh, C., 2009. Genome-Wide Association Study in Humans. In p. 231-258.
- Sonnenberg, A.S.M., Baars, J.J.P. & Kerrigan, R.W., 2008. Mushroom breeding: hurdles and challenges.
- Stewart, E.L. et al., 2016. *QTL mapping reveals complex genetic architecture of quantitative virulence in the wheat pathogen Zymoseptoria tritici.*, Cold Spring Harbor Labs Journals.
- Stewart, E.L. & McDonald, B.A., 2014. Measuring Quantitative Virulence in the Wheat Pathogen *Zymoseptoria tritici* Using High-Throughput Automated Image Analysis. *Phytopathology*, 104(9), p.985-992.
- Talas, F. et al., 2012. Association of single nucleotide polymorphic sites in candidate genes with aggressiveness and deoxynivalenol production in *Fusarium graminearum* causing wheat head

- blight. *Bmc Genetics*, 13(1), p.1-10.
- Talas, F. et al., 2016. Genome-wide association study identifies novel candidate genes for aggressiveness, deoxynivalenol production and azole sensitivity in natural field populations of *Fusarium graminearum*. *Molecular Plant-Microbe Interactions*.
- Talas, F. & McDonald, B., 2015. Genome-wide analysis of *Fusarium graminearum* field populations reveals hotspots of recombination. *BMC Genomics*, 16(1), p.996.
- Taudt, A., Colomé-Tatché, M. & Johannes, F., 2016. Genetic sources of population epigenomic variation. *Nature Reviews Genetics*, 17(6), p.319-332.
- Tedersoo, L. et al., 2014. Fungal biogeography. Global diversity and geography of soil fungi. *Science (New York, N.Y.)*, 346(6213), p.1256688.
- Valverde, M.E. et al., 2015. Edible Mushrooms: Improving Human Health and Promoting Quality Life. *International Journal of Microbiology*, 2015, p.1-14.
- Vogan, A.A. et al., 2016. Identification of QTLs Associated with Virulence Related Traits and Drug Resistance in *Cryptococcus neoformans*. *G3 (Bethesda, Md.)*.
- De Vos, L. et al., 2011. Genetic analysis of growth, morphology and pathogenicity in the F(1) progeny of an interspecific cross between *Fusarium circinatum* and *Fusarium subglutinans*. *Fungal Biol*, 115(9), p.902-908.
- Xu, J., 1995. Analysis of inbreeding depression in *Agaricus bisporus*. *Genetics*, 141(1), p.137-45.
- Xu, X. et al., 2012. Super-races are not likely to dominate a fungal population within a life time of a perennial crop plantation of cultivar mixtures: a simulation study. *BMC Ecology*, 12(1), p.16.
- Xu, Y. et al., 2016. Genetic mapping of quantitative trait loci in crops. *The Crop Journal*.
- Yi, N. & Shriner, D., 2008. Advances in Bayesian multiple quantitative trait loci mapping in experimental crosses. *Heredity*, 100(3), p.240-252.
- Yoshida, K. et al., 2009. Association genetics reveals three novel avirulence genes from the rice blast fungal pathogen *Magnaporthe oryzae*. *Plant Cell*, 21(5), p.1573-1591.
- Yu, J. et al., 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics*, 38(2), p.203-208.

## VI.3. *L'agressivité chez F. graminearum?*

La relation qui lie *F. graminearum* à ses hôtes est régie uniquement par sa composante quantitative. Le succès de cette relation est variable : d'importantes variations du niveau d'agressivité au sein des populations de *F. graminearum* sont observées, et ce à l'échelle même du champ (Talas et al. 2012a; Ward et al. 2008; Cumagun & Miedaner 2003). Des études faites en serre et aux champs montrent que ce caractère est soumis à contrôle génétique et est supposé polygénique (Cumagun & Miedaner 2003). Un tel contrôle génétique laisse présager une évolution de ce caractère dans les populations de *F. graminearum*. Expérimentalement, une hausse des niveaux d'agressivité comparés à ceux des générations précédentes a été démontrée (Cumagun & Miedaner 2004). En Amérique du Nord, la propagation de nouvelles populations d'isolats plus agressifs a été rapportée (Ward et al. 2008; Foroud et al. 2012). Ces données sont d'autant plus préoccupantes face à l'incertitude du devenir de ces populations de pathogènes dans un contexte de modifications climatiques. Par exemple, en Chine, une étude de modélisation prédit l'avancement de la date de floraison des blés en réponse aux modifications climatiques, rendant ces derniers plus sensibles à la fusariose des épis (Zhang et al. 2014).

Alors que ces observations soulèvent des inquiétudes, la compréhension des bases génétiques de l'agressivité de ce pathogène est, à ce jour, limitée. Son étude se heurte notamment à son caractère polygénique. En effet, à cause du nombre et de la redondance fonctionnelle des gènes impliqués dans ce caractère, les études de génétique fonctionnelle classique sont difficilement envisageables (Jia & Tang 2015). Face à ces contraintes, des approches de génétique quantitative, basées sur l'observation de phénotype dans des populations (naturelles ou expérimentales), pourraient s'avérer être particulièrement efficaces (Grünwald et al. 2016). A ce jour, trois études ont étudié les bases génétiques de l'agressivité chez *F. graminearum* (Tableau 1). Toutes ont été réalisées chez le blé.

La première étude, développée en 2004, proposait de cartographier les bases génétiques de l'agressivité d'une population expérimentale de *F. graminearum* sur une carte génétique (Jurgenson et al. 2002; Cumagun 2004). Les auteurs ont été confrontés

au manque d'information disponible à l'époque sur la phylogénie du complexe d'espèces de *F. graminearum* et ont croisé une souche de *F. graminearum* avec une souche qui s'est révélée plus tard être de l'espèce *F. asiaticum*. Malgré les différents biais de ségrégations observés sur la carte de génétiques dus au croisement interspécifique, les auteurs ont identifié un QTL majeur, responsable de 50% de la variation d'agressivité observée (Tableau I\_1). Ce QTL co-localise avec le cluster *Tri* et a été associé à la capacité à produire du NIV ou du DON, les souches produisant du DON étant plus agressives dans les conditions testées. La deuxième étude a été développée à partir de populations naturelles isolées en Allemagne et s'est focalisée sur le polymorphisme de gènes candidats (Talas et al. 2012b). Les auteurs ont attribué 26% de la variation d'agressivité à un polymorphisme génétique localisé sur le gène *MetAP1* (impliqué dans le développement fongique), et 4.4% de la variation de production en DON à un polymorphisme génétique localisé sur le gène *Tri1*, présent dans le cluster *Tri* et codant une P450 oxygénase (Tableau I\_1). En 2016, les mêmes auteurs ont développé une approche sans *a priori* de GWA et ont identifié 50 nucléotides polymorphiques associés à la variation d'agressivité (QTNs pour « quantitative trait nucleotides »). Vingt-six de ces QTN sont localisés dans des gènes (Tableau I\_1). Vingt-neuf QTN ont été associés à la variation de production en DON (Talas et al. 2016). La corrélation entre ces deux caractères n'a pas été pas testée, et aucun QTN n'a été associé à la fois à l'agressivité et à la production de mycotoxines. L'effet maximal apporté par un QTN était de 24% pour l'agressivité et 19% pour la production en mycotoxine. De manière surprenante, beaucoup de ces QTNs sont localisés en dehors de séquences codantes et les mécanismes moléculaires sous-jacents restent méconnus (Tableau I\_1, Talas et al. 2016).

**Tableau I\_1 :** Liste de gènes précédemment associés à des variations du niveau d'agressivité.

Gènes impliqués?	Fonction	% variance	Auteurs	Approche
Cluster gènes <i>Tri</i>	TCTB biosynthesis	51%	Cumagun et al. 2004	Cartographie de QTL classique
MetAP1	Uncharacterized	26%	Talas et al. 2012b	Approche gènes candidats
FGSG_06583	ATP binding	25%		
FGSG_00355	RAS GTPas activator	24%		
FGSG_00586	Metal ion binding	18%		
FGSG_01935	Transferase activity	18%		
FGSG_08910	CAS/CSE protein family	18%		
FGSG_12420	Hypothetical protein	18%		
FGSG_13008	Metal ion binding/Transcription factor activity	18%		
FGSG_12140	Metallophos 2	17%		
FGSG_12204	Sugar transporter	17%		
FGSG_12685	NOT2,3,5 domain	17%		
FGSG_03620	Hypothetical protein	16%		
FGSG_11179	Hydrolase activity	16%		
FGSG_01339	DNA binding/TATA binding	15%		
FGSG_13878	AMP binding	15%	Talas et al. 2016	GWA
FGSG_06075	Microtubule binding/Metal ion binding	15%		
FGSG_05475	Protein binding	15%		
FGSG_09541	Hydrolase	15%		
FGSG_02259	ATP adenylyltransferase II	14%		
FGSG_02352	Acyltransferase activity	14%		
FGSG_01563	Hypothetical protein	14%		
FGSG_08608	PAS2/His Kinase A/Response Regulator/GAF	14%		
FGSG_08010	Zn-cluster	13%		
FGSG_03751	L-amino acid transport	13%		
FGSG_02531	Transcriptional control (zinc binding)	12%		
FGSG_09604	Uncharacterized	9%		
FGSG_13639	Hypothetical protein	9%		

Plusieurs questions restent à ce jour sans réponse : est-ce que les acteurs génétiques déjà caractérisés sont similaires chez des souches issues de régions différentes? Quels sont les mécanismes moléculaires sous-jacents à ces variations du niveau d'agressivité ? Quel est le rôle des variations d'agressivité pour l'évolution des populations de *F. graminearum*?



# VII. L'hypothèse et les objectifs de la thèse

Durant ces travaux, nous avons proposé d'apporter de nouveaux éléments de réponses sur les bases génétiques de la pathogénicité chez *F. graminearum*, à travers l'étude de i) l'agressivité sur blé ii) la production de toxines, iii) la relation entre ces deux caractères. Pour cela, nous disposons au laboratoire d'une collection de diversité, composée majoritairement d'isolats français (Pinson-Gadais et al. 2013).

L'hypothèse qui a été testée au cours de ces travaux de thèse est que l'agressivité des souches de *F. graminearum* de cette collection est variable et que cette variabilité est sous le contrôle de facteurs génétiques. La stratégie qui a été développée pour vérifier cette hypothèse est une approche de cartographie génétique de QTL. Comme développé précédemment, cette stratégie repose sur l'étude de la ségrégation des caractères d'intérêts au sein d'une population recombinante afin de localiser sur une carte génétique, les QTL en lien avec la variation du caractère d'intérêt.

L'objectif principal de ces travaux a donc été d'étudier les relations entre les différents caractères en lien avec l'**agressivité** au sein d'une **population recombinante** de *F. graminearum* et de caractériser les bases génétiques de l'agressivité dans cette population : quelle est la part génétique impliquée dans la variation d'agressivité ? Quelles sont les régions génomiques (QTL) impliquées dans cette variation ? Quelle est leur importance, quels sont leurs effets sur le caractère ? Interagissent-ils ?

Pour atteindre cet objectif, 3 tâches majeures ont été effectuées (Figure I\_13).

1) La première tâche a consisté à caractériser la diversité génétique et phénotypique de souches issues de la collection de diversité disponible au laboratoire (Pinson-Gadais et al. 2013) afin d'identifier des souches parentales potentielles pour la construction de la population recombinante. Ces résultats sont présentés en Chapitre I.

2) La deuxième tâche a consisté à créer la population recombinante, à la génotyper et à construire une nouvelle carte génétique à partir de cette population. Ces résultats sont présentés en Chapitre II.

3) La troisième tâche a consisté à phénotyper la population recombinante pour plusieurs caractères en lien avec l'agressivité des souches, et à étudier la ségrégation des caractères dans la population recombinante, afin de détecter et localiser des QTLs sur la carte génétique pour ces caractères. Cette dernière tâche, qui nous a permis d'atteindre notre objectif principal, est présentée en Chapitre III.

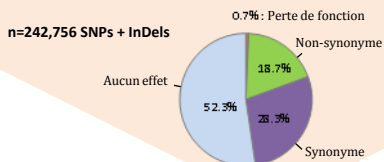
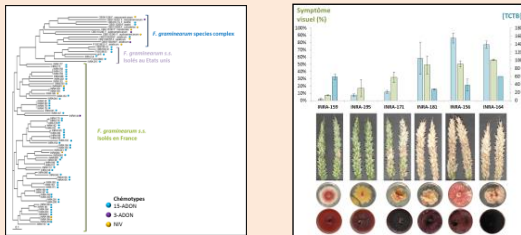
Un quatrième chapitre est proposé en fin de manuscrit. Il résume les résultats majeurs de ces travaux, et discute notamment du potentiel d'évolution et d'adaptation de ce pathogène, ainsi que des nouvelles interrogations soulevées par ces travaux.

## Hypothèse

*L'agressivité des souches de F. graminearum est variable et sous le contrôle de facteurs génétiques*

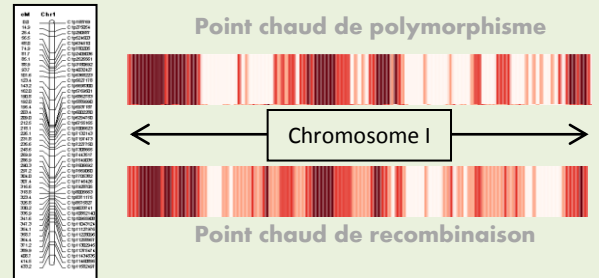
### Chapitre 1

Caractérisation de la **diversité génétique** et **phénotypique** de souches de *F. graminearum*

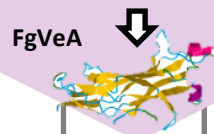
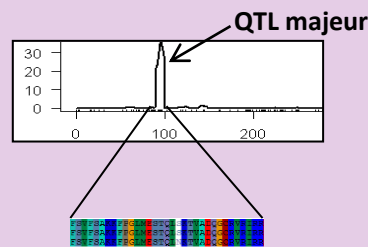


### Chapitre 2

La **recombinaison** méiotique chez *F. graminearum* : un mécanisme évolutif majeur, à la base de la **cartographie génétique**.



Détection de QTL impliqués dans la **variation de production de mycotoxines** et d'**agressivité** chez *F. graminearum*



### Chapitre 3

## Objectif principal

*Etudier les relations entre les différents caractères en lien avec l'agressivité au sein d'une population recombinante de F. graminearum et caractériser les bases génétiques de l'agressivité dans cette population*

Figure I\_13 : Organisation des travaux de thèse



**Chapitre 1 :**  
**Caractérisation de la diversité  
génétique et phénotypique de  
souches de *F. graminearum***



## **I. Etudes préliminaires de la collection de *F. graminearum***

Comme nous l'avons vu en introduction, un point crucial dans le développement d'une approche de génétique quantitative classique réside dans le choix des parents pour la création d'une population expérimentale adaptée à la problématique posée. Plusieurs paramètres critiques guident ce choix : le polymorphisme génétique, qui permettra d'optimiser la construction d'une carte génétique, le contraste phénotypique qui présume de la ségrégation du caractère d'intérêt dans la descendance et enfin, la possibilité de réaliser des croisements.

Au laboratoire, nous disposons d'une collection d'une centaine de souches de *F. graminearum* comprenant des isolats récoltés en France pour la plupart et également des souches de référence issues de collections internationales. A l'initiative du projet, ces souches étaient en partie caractérisées biochimiquement pour leur niveau de production de toxine en conditions contrôlées mais aucune information n'était disponible sur leur potentielle diversité génétique. Afin d'identifier et de sélectionner des souches pertinentes pour l'élaboration d'une approche QTL, un premier travail était donc de caractériser la diversité génétique et de compléter la caractérisation phénotypique de la collection disponible.

### ***I.1. Caractérisation génétique et phénotypique de *F. graminearum****

Ces travaux ont été réalisés par Fatima Zahra SHAILIA dans le cadre d'un stage de Master 2 réalisé en 2012, et encadré par Marie FOULONGNE-ORIOU. Les procédures expérimentales, ainsi que le rapport de ce stage présentant ces travaux sont disponibles en Annexe I.

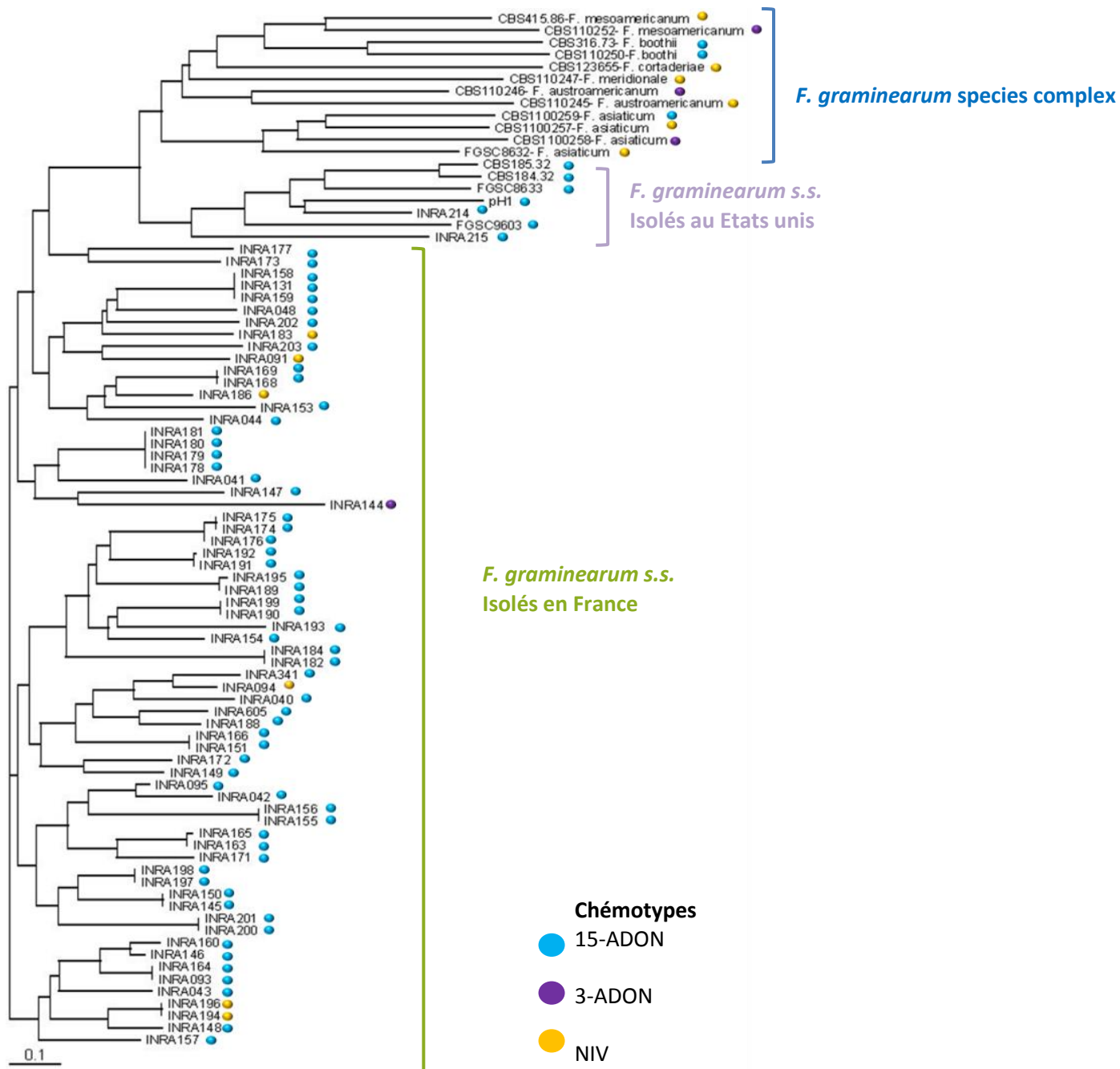
Au total 85 souches (66 isolats français, 7 souches de référence, 12 souches appartenant au complexe d'espèces) ont été génotypées à l'aide de 20 marqueurs de types microsatellites (ou SSR pour l'acronyme anglais Simple Sequence Repeats ») (Figure CI\_1). L'analyse des 66 isolats français a permis de révéler un niveau important de

diversité avec 60 haplotypes différents identifiés (Figure 1). Le niveau de polymorphisme entre ces haplotypes varie de 13 % à 80 %. (Shailia et al. 2012).

La majorité des souches de cette collection a été caractérisée comme appartenant aux chémotypes DON/15-ADON, confirmant la prédominance de ce chémotype en France (Figure CI\_1) (Boutigny et al. 2014). En parallèle, 66 isolats de *F. graminearum* ont été caractérisés pour leur production de mycotoxines en condition *in vitro* sur grains de blé. Les niveaux de production révélés sont variables entre souches, suggérant que le niveau de diversité génétique observé précédemment s'accompagne d'un niveau important de diversité phénotypique. Cette étude préliminaire a confirmé l'intérêt de la collection de souches de *F. graminearum* disponible au laboratoire en tant que ressources génétiques et phénotypiques pour la mise en place d'une approche de génétique quantitative.

Les travaux développés dans la suite du manuscrit font partie intégrante de mon travail de thèse.

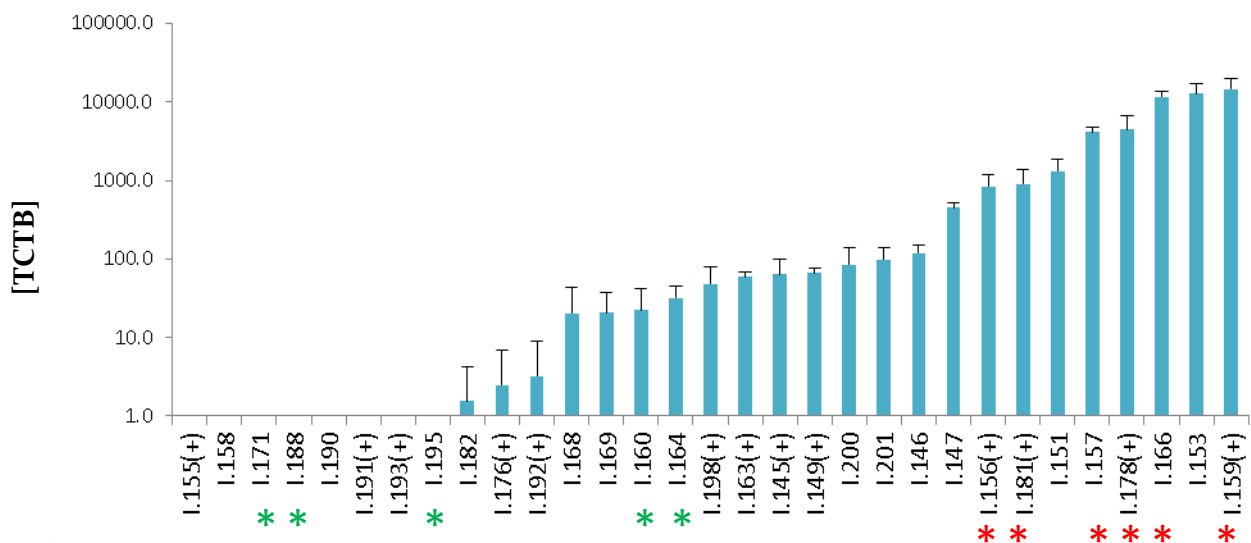




**Figure CI\_1:** Arbre construit par Neighbor-Joining basé sur les distances euclidiennes calculées à partir de 20 marqueurs microsatellites génotypés sur 66 souches de *F. graminearum* isolées en France et 19 issues d'autres collections internationales (Pinson-Gadais et al. 2013).

## I.2. Vers la sélection de six souches de *F. graminearum*

Afin de réduire progressivement le nombre de souches pour la sélection de génotypes parentaux, 31 souches de la collection ont également été phénotypées en milieu synthétique minimum liquide (milieu MS, favorisant la production de mycotoxines, Figure CI\_2). Toutes ces souches appartiennent aux chémotypes DON/15-ADON. Cette démarche présente un double intérêt. La source de carbone disponible pour le champignon affectant sa capacité à produire des toxines (Jiao et al. 2008), une caractérisation de la production de toxine élargie à d'autres milieux nutritifs apparaît pertinente pour mieux rendre compte du caractère. D'autre part, la culture sur grain ne pouvant être envisagée que pour un petit nombre de souches, il était indispensable d'avoir une méthode de caractérisation phénotypique adaptée aux grands échantillons comme le requiert l'analyse d'une descendance de plusieurs dizaines d'individus.



**Figure CI\_2 :** Concentration en TCTB ( $\mu\text{g}$  de toxines par g de biomasse fongique) pour les différentes souches INRA cultivées en milieu synthétique contenant du glucose. '+' signifie que la souche produit plus de 50ppm de TCTB totaux en culture sur grains. Les souches considérées et sélectionnées comme très toxigènes sont marquées par une étoile rouge. Les souches considérées et sélectionnées comme peu ou pas toxigènes sont marquées par une étoile verte.

Des différences de niveau de production de mycotoxines en fonction du milieu de culture ont été observées. Par exemple, la souche INRA-155 produit beaucoup de toxine sur milieu à base de grains de blé mais très peu en culture liquide. Les souches ont été cultivées en milieu MS contenant du glucose ou du saccharose comme source carbonée. D'une manière générale la production de TCTB est augmentée significativement d'un facteur 15 en présence de saccharose comparée aux milieux de culture contenant du glucose (non présenté). Ce résultat est en accord avec les travaux de Jiao et al. (2008) qui ont démontré que, chez des souches DON/3-ADON cultivées en milieu liquide, la production de TCTB est induite par le saccharose et les sucres contenant des  $\alpha$ -glucosides. De plus, ce résultat est cohérent avec les mécanismes d'infection de ce pathogène puisque il a été montré que les trichothécènes sont produits majoritairement lors de la propagation dans l'épi de blé, tissus riches en saccharoses et fructo-oligosaccharides (Jiao et al. 2008). Le milieu liquide synthétique supplémenté en glucose est celui le plus utilisé au laboratoire. Nos résultats montrent que l'utilisation du saccharose peut être intéressante car ce sucre exacerbe la production de toxine chez *F. graminearum* et le milieu synthétique semble se rapprocher des conditions naturelles. Son utilisation permet aussi d'employer des méthodes de détection et de quantification de toxines moins sensibles. De fait, le milieu MS contenant du saccharose a été utilisé pour la réalisation des expérimentations conduites pendant la thèse.

L'ensemble de ces données a servi à la sélection de souches dites 'faibles productrices' et de souches dites 'fortes productrices'. Dans la mesure du possible, nous avons tenté de choisir des parents producteurs et faiblement producteurs quel que soit le milieu de culture (Figure CI\_2). Les 11 souches présélectionnées ont ensuite été testées pour leur potentiel à produire des périthèces dans nos conditions de laboratoires, prérequis indispensable pour la réalisation de croisement (Tableau CI\_1). Le comportement des souches pour ce caractère s'est avéré très contrasté, avec des souches ne produisant pas ou très peu de périthèces (INRA-159) et des souches produisant des périthèces en grand nombres (INRA-171). Les différentes données de phénotypes ont ensuite été confrontées aux données de génotypage obtenues

précédemment. Seules les souches montrant un niveau de polymorphisme supérieur à 40% avec au moins un autre partenaire potentiel ont été sélectionnées (Tableau CI\_1).

**Tableau CI\_1** : Polymorphisme estimé sur 20 marqueurs SSR génotypés, « +++ » signifie que la souche produit des périthèces abondamment, « - » signifie qu'elle n'en produit pas dans nos conditions de culture. Rouge : souches productrices ; Vert : souches non à faiblement productrices.

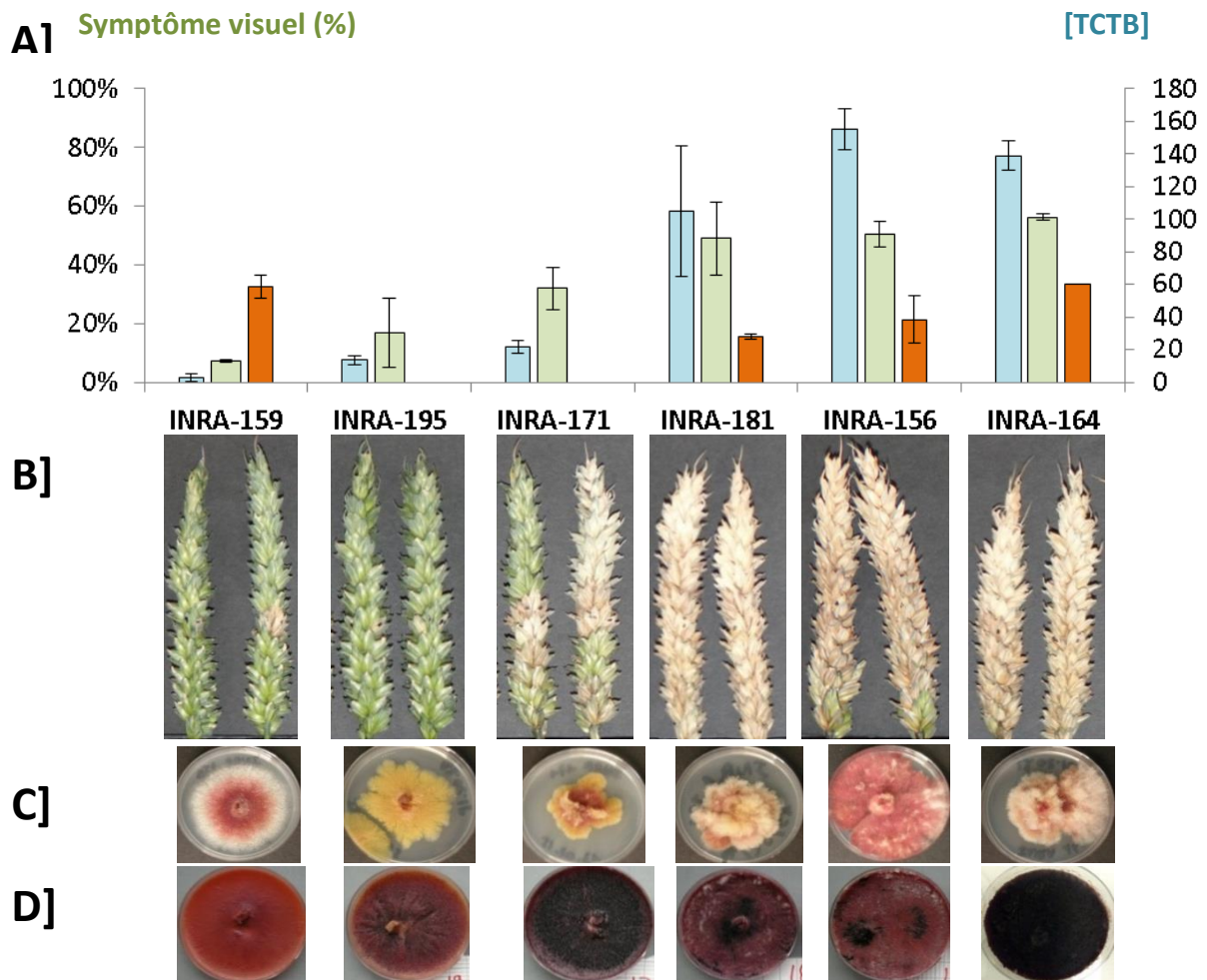
Souche INRA et potentiel autoreproductif	I.171 (+++)	I.188 (-)	I.195 (+++)	I.160 (-)	I.164 (-)
I.156 (+++)	0.50	0.73	0.70	0.56	0.65
I.181 (++)	0.45	0.53	0.50	0.38	0.30
I.157 (+)	0.50	0.40	0.55	0.31	0.35
I.178 (+)	0.30	0.53	0.50	0.38	0.30
I.166 (++)	0.40	0.40	0.45	0.38	0.40
I.159 (-)	0.40	0.40	0.55	0.56	0.55

Finalement, 6 souches ont été gardées pour la suite de l'analyse, 3 fortement toxigènes (INRA-156, INRA-159 et INRA-181) et 3 faiblement toxigènes (INRA-164, INRA-171 et INRA-195).

### 1.3. Du modèle *in vitro* vers le modèle *in planta*

Dans le cadre de la collaboration avec l'équipe « Génétique, Diversité, Ecophysiologie des Céréales » (GDEC) du centre INRA de Clermont-Ferrand, ces six souches ont été testées pour leur pathogénicité sur blé. La variété Roysac décrite comme plutôt sensible à la fusariose a été utilisée (Figure CI\_3A et CI\_3B). Bien qu'ayant un potentiel toxigène fort dans nos conditions *in vitro*, La souche INRA-159 s'est avérée ne pas être agressive sur blé. Dans nos conditions d'inoculation, le mycélium de cette souche ne se propage pas en dehors de l'épi inoculé. La souche INRA-164, classée précédemment parmi les modérément toxigènes *in vitro*, s'est avérée être hautement agressive et hautement toxigène sur blé *in planta*. De plus, cette souche présente un fort potentiel de production de toxine en milieu liquide contenant du saccharose. Des expérimentations complémentaires ont révélées des problèmes de traçabilité entre les différentes versions de la souche de travail utilisée.

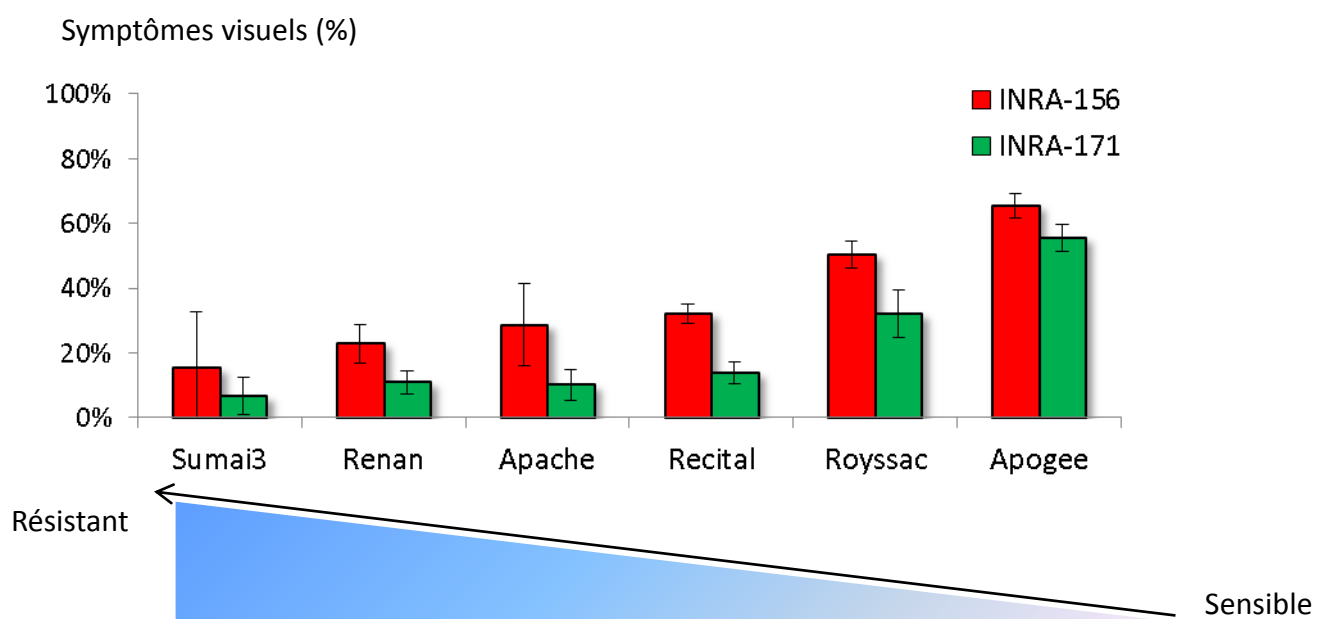
Notre échantillonnage est constitué en fait de trois souches hautement agressives sur blé (INRA-156 ; INRA-164 ; INRA-181) et trois souches faiblement ou non-agressives sur blé (INRA-159 ; INRA-171 ; INRA-195). Les concentrations en TCTB mesurées dans les épis infectés pour ces 6 souches sont cohérentes avec les symptômes visuels (Figure CI\_3A). Les morphologies de ces souches sont très variées (Figure CI\_3C et CI\_3D)



**Figure CI\_3 :** A] En vert : notation des symptômes visuels (en % d'épillet infecté par épi) à 22 jours après inoculation sur les plants de blé (Roysac). En bleu, les concentrations en TCTB mesurées à partir des épis inoculés, en µg de toxine par gramme de blé sec. En orange, les concentrations en TCTB mesurées à partir de cultures liquides *in vitro*, en mg de TCTB par g de biomasse fongique sèche. B] Photos prises à partir des épis inoculés pour chaque souche. C] Photos des différentes souches cultivées sur milieu solide PDA (potatoes dextrose agar, croissance végétative). D] Photos des différentes souches cultivées sur milieu Carrot agar, induisant la formation des périthèces.

Nous avons par la suite testé l'interaction hôte-isolat sur une gamme plus large de variétés de blé, afin de déterminer notamment la variété à utiliser pour de futurs essais. Pour cela, 6 variétés présentant des niveaux de résistances variables pour la fusariose

ont été inoculées par deux souches de cet échantillonnage (INRA-156 et INRA-171) présentant des niveaux contrastés d'agressivité sur la variété Roysac (Figure CI\_4). Alors que les niveaux moyens d'agressivité varient en fonction des résistances inhérentes à chaque variété, la souche INRA-156 semble être toujours plus agressive que la souche INRA-171. La différence d'agressivité qui différencie ces deux souches semble donc être respectée et ces dernières ne montrent pas de spécificité d'hôte. Parmi les différentes variétés, la variété Roysac nous a semblé la plus adaptée pour la suite de l'analyse car présentant une sensibilité modérée à la fusariose et mettant bien en valeur les différences d'agressivités observées entre souches.



**Figure CI\_4 :** Graphique représentant les niveaux d'agressivités de la souche INRA-156 (hautement agressive) et INRA-171 (modérément agressive) observé chez 6 variétés de blé, contrastées pour leurs niveaux de résistance à la fusariose (échelle en bleue).

Suite à cette caractérisation assez détaillée du phénotype de ces six souches de *F. graminearum*, nous avons décidé de caractériser leur diversité génétique, non plus à l'échelle de quelques marqueurs microsatellites, mais à l'échelle de leur génome. Le but d'une telle approche est de :

- caractériser de manière la plus complète possible la diversité génétique de ses souches.

- d'aider le choix des souches à utiliser pour la réalisation d'une population recombinante.
- de guider les étapes de génotypages pour la construction d'une carte génétique.

Ces travaux sont présentés dans le paragraphe « 2. Diversité génomique de *F. graminearum* et rôle pour l'adaptation à l'hôte ».

## **I.4. Matériels et méthodes :**

*Sont décrits ici les matériels et méthodes utilisés dans les paragraphes I.2 et I.3*

**Matériels biologiques :** 31 souches de *F. graminearum*, isolées à partir de blés et de maïs infectés issues la collection du laboratoire ont été utilisées (Pinson-Gadais et al. 2013). Les souches sont dorénavant stockés au Centre International de Ressources Microbiennes Champignon Filamenteux à Marseille (<https://www6.inra.fr/cirm/>) et disponibles à la demande.

**Test de reproduction sexuée :** La production des périthèces a été mise en place au laboratoire en se basant sur le protocole publié par Cavinder et al., (2012). Les souches ont été inoculées sur milieu solide Carrot Agar (carottes broyées 400 g/L ; agar 20 g/L) et incubées sous lumière blanche permanente, à température constante (22°C). Une fois que la masse mycélienne a recouvert le milieu, la partie aérienne du mycélium est raclée à l'aide d'une anse stérile, et 1 mL de TWEEN® 60 concentré à 2,5% est ajouté à la surface du milieu, induisant un stress et amenant à la formation de périthèces.

**Production d'inoculum et mise en culture :** Les différentes souches, maintenues à 4°C sur milieu PDA (Potato Dextrose Agar 39 g/l), sont mises en culture sur milieu PDA et incubées à l'obscurité à 25°C pendant 5 à 9 jours. Afin de s'affranchir de l'effet « âge du mycélium », l'inoculum primaire est constitué de conidies. Pour normaliser la production d'inoculum pour un grand nombre de souches, le protocole suivant a été mis au point : pour chaque souche, 4 implants de culture mycélienne de 2.5 cm de diamètre sont prélevés et incubés dans un flacon contenant 12 ml de milieu liquide CMC (Carboxymethyl cellulose 15 g/L, extrait de levure 1 g/L, MgSO<sub>4</sub>.7H<sub>2</sub>O 0,5 g/L, NH<sub>4</sub>NO<sub>3</sub> 1 g/L, KH<sub>2</sub>PO<sub>4</sub> 1 g/L) puis incubés pendant 2 jours à 25°C sous agitation

constante (170rpm-180 rpm, Infors™ Multitron Standard). Les spores ainsi produites sont récupérées après filtration sur maille de 100 µm et concentrées après centrifugation à 11 570 g pendant 10 minutes dans 2 ml de milieu. La densité optique (DO) à 500 nm de chaque solution est corrélée avec le nombre de spores. Pour chaque expérimentation, une courbe étalon est construite à partir d'échantillons internes dont le nombre de spores est calculé à partir de comptage sur cellule de Thoma.

**Production de TCTB en condition *in vitro* :** Pour chaque souche, les spores sont inoculées dans 1,2 ml de milieu synthétique liquide à une concentration finale d'environ  $21 \times 10^3$  spores/ml (Glucose 20g/L ou Saccharose 20g/L,  $\text{KH}_2\text{PO}_4$  0,5g/L,  $\text{K}_2\text{HPO}_4$  0,6g/L,  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$  0,017g/L,  $(\text{NH}_4)_2\text{SO}_4$  1g/L, éléments traces de Vogel, 0,1ml/L). A partir d'une même solution de CMC, 3 réplicats sont réalisés pour la culture en milieu synthétique. La culture est réalisée en plaques de 24 puits, les plaques sont parafilmées et maintenues à l'obscurité à 25°C pendant 10 jours avant de doser les toxines dans le milieu.

**Test de pathogénicité :** Une solution conidiale de 100 spores ajustée dans 10µL de milieu liquide CMC a été inoculée dans les cavités florales médianes d'épis de blé (deux paires opposées par épi). Les plantes ont été cultivées à température et humidité optimales et ont été inoculées à l'anthèse. Trois plantes par souches ont été utilisées à raison de 3 à 4 épis inoculés par plante. Les symptômes visuels correspondent au nombre d'épillets infectés par nombre d'épillets total par épis, comme précédemment proposé (Talas et al. 2012b). La notation a été réalisée 22 jours après inoculation.

**Dosage des TCTB : *In vitro* :** Après 10 jours d'incubation, les cultures mycéliennes sont récupérées dans des tubes de 2 mL et centrifugées à 8370 g à 4°C. Le culot contenant la masse mycélienne est séché à 80°C pendant 3 à 4 jours afin de peser la biomasse sèche. Le surnageant contenant les toxines sécrétées par le champignon est filtré sur maille de 0,22 µm avant passage direct en chromatographie. La concentration en toxines accumulées dans le milieu extracellulaire a été déterminée après analyse en HPLC-MS-MS (QTrap 2000). Le protocole utilisé est celui décrit par Picot et al. (2012) en modifiant le solvant A en méthanol : eau (90 : 10, v : v) et en ajustant le volume d'injection à 20 µl. L'interprétation des spectrogrammes est réalisée via le logiciel



Analyst® 1.4 (AB Sciex). Les concentrations sont mesurées en ng/mL et sont ramenées en µg de toxines par g de biomasse sèche de champignon. ***In planta*** : Un gramme de poudre d'épis préalablement séché a été dilué dans 8mL d'un mélange composé d'acétonitrile et d'eau (86 : 16, volume : volume), puis vortexé pendant une minute. Cinq mL de surnageant ont été prélevés et évaporés. Les résidus ont ensuite été dissouts dans 200µL d'un mélange composé de méthanol et d'eau (50 : 50, volume/volume) et filtré à 0.22 µm (Phenomenex, Torrance, USA). La concentration en toxines a été déterminée de la même manière que décrit précédemment. La concentration en TCTB est mesurée en µg par gramme de poudre de blé sec.

## **II. Diversité génomique de *F. graminearum* et rôle pour l'adaptation à l'hôte**

Ce sous-chapitre a fait l'objet d'un article scientifique, soumis à BMC genomics.

### **Landscape of genomic diversity and host adaptation in *Fusarium graminearum***

Benoit Laurent<sup>1</sup>, Magalie Moinard<sup>1</sup>, Cathy Spataro<sup>1</sup>, Nadia Ponts<sup>1</sup>, Christian Barreau<sup>1</sup>, Marie Foulongne-Oriol<sup>1\*</sup>

<sup>1</sup> MycSA, INRA, 33882, Villenave d'Ornon, France

\*Corresponding author

Mailing address: INRA, UR1264 Mycologie et Sécurité des Aliments, bâtiment Qualis, 71 avenue Edouard Bourleaux, CS 20032, F-33882 Villenave d'Ornon cedex, France

Phone: +33 (0)5 57 12 26 35

Fax: +33 (0)5 57 12 25 00

E-mail: [mfoulong@bordeaux.inra.fr](mailto:mfoulong@bordeaux.inra.fr)

## **Abstract**

**Background:** *Fusarium graminearum* is one of the main causal agents of the *Fusarium* Head Blight, a worldwide disease affecting cereal cultures, whose presence can lead to contaminated grains with chemically stable and harmful mycotoxins. Resistant cultivars and fungicides are frequently used to control this pathogen, and several observations suggest an adaptation of *F. graminearum* that raises concerns regarding the future of current plant disease management strategies. To understand the genetic basis as well as the extent of its adaptive potential, we investigated the diversity of genomic landscapes for six French isolates of *F. graminearum*, at single-nucleotide resolution using whole-genome re-sequencing.

**Results:** A total of 242,756 high confidence genetic variants were detected when compared to the reference genome, among which 96 % are single nucleotides polymorphisms. One third of these variants were observed in all isolates. Seventy-seven percent of the total polymorphism is located in 32 % of the total length of the genome, comprising telomeric/subtelomeric regions as well as discrete interstitial sections, delineating clear variant-enriched genomic regions - 7.5 times in average. About 80% of all the *F. graminearum* protein-coding genes were found polymorphic. Biological functions are not equally affected: genes potentially involved in host adaptation are preferentially located within polymorphic islands and show greater diversification rate than genes fulfilling basal functions. We further identified 29 putative effector genes enriched with non-synonymous effect mutation.

**Conclusions:** Our results highlight a remarkable level of polymorphism in the genome of *F. graminearum* distributed in a specific pattern. Indeed, the landscape of genomic diversity follows a bi-partite organization of the genome according to polymorphism and biological functions. We measured, for the first time, the level of sequence diversity for the entire gene repertoire of *F. graminearum* and revealed that the majority are polymorphic. Those assumed to play a role in host-pathogen interaction are discussed, in the light of the subsequent consequences for host adaptation. The annotated genetic variants discovered for this major pathogen are valuable resources for further genetic and genomic studies.

**Keywords:** Fungal pathogen, *Fusarium* Head Blight, Whole genome re-sequencing, Genome-wide polymorphism, Single Nucleotides Polymorphism, Host-Pathogen interaction, Evolution, Two-speed genome

## **Background**

The ascomycete *Fusarium graminearum* (teleomorphe *Gibberella zeae*) is an hemibiotrophic pathogen commonly described as one of the main causal agent of the *Fusarium* Head blight (FHB), a devastating disease affecting small grains cereals worldwide [1]. In addition to the defect on annual yield, major concerns arise from contamination of grains by stable and harmful fungal metabolites so-called mycotoxins which are present in feed and food constitute a real threat for consumers and livestock [2]. Molecules belonging to the type B family of trichothecenes (TCTB) are probably the most concerning due to their frequent occurrence and demonstrated toxic effects [3]. The genes acting in TCTB production, named *Tri* genes, are clustered for the majority and expressed after plant penetration with an implication in pathogenicity [4,5]. Despite the wide array of trichothecenes potentially produced by *F. graminearum isolates*, the spectrum of production observed in individual strains is more limited, defining chemotypes [6,7]. To date, three chemotypes of TCTB-producing isolates have been described according to their ability to produce deoxynivalenol along with 15-acetyldeoxynivalenol (DON/15-ADON), deoxynivalenol and 3-acetyldeoxynivalenol (DON/3-ADON), and nivalenol and acetylated form (NIV). These chemotypes are associated with quantitative difference in pathogenicity; the strains producing DON instead of NIV are, for example, more aggressive against wheat [8]. In some cases, levels of TCTB have also been found to be correlated with the visual symptoms on the spike [9,10]. Nevertheless, other factors were identified in *F. graminearum* with the detection of 50 quantitative trait nucleotides linked to aggressiveness variation [11].

Cultivars resistant against FHB and mycotoxin accumulation as well as fungicides are frequently used to control this pathogen [12]. However, there is now evidence that *F. graminearum* is adapting to such strategies, as demonstrated by the emergence of fungicide-resistant strains [13,14] and the rapid shift towards more aggressive isolates in some part of the world [15]. Cultural management practices must therefore keep up with the “arm race”, which requires a detailed knowledge of the fungus adaptive potential with a special focus on the evolution of pathogenicity-related traits.

Grounds for *F. graminearum* adaptation are certainly provided for by intensive gene flow and large amounts of genetic diversity between and within field populations [16–24]. In *F. graminearum* specifically, these elements are further supported by particular biological features that favor the emergence of genetic diversity, namely a mixed reproduction system

based on clonality, selfing and outcrossing [16,24,25] as well as both local and long range dispersal of the different spores produced [26–30]. Such combination is indeed particularly efficient to create new haplotypes of which the favorable ones will rapidly spread [31]. The molecular mechanisms underlying the emergence of more aggressive isolates of *F. graminearum* remains remain sparsely documented.

Deep sequencing technologies have been successfully used to investigate genome-wide polymorphism in various fungi, eventually leading to the proposition of candidate genes implicated in phenotype variations [32–40]. In the case of *F. graminearum*, an annotated genome of reference is available, based on the sequencings of a North-American isolate [41–43]. The latest version consists of 38 Mb distributed in four scaffolds assigned to the four expected chromosomes and has been predicted to contain 14,160 nuclear protein coding genes [41]. Although the function of the majority of these genes remains unknown, specific efforts of manually curated genome-mining coupled to proteomics and transcriptomics studies revealed a large arsenal of potential effectors, including potential secreted proteins or secondary metabolites other than the currently known mycotoxins [41,44–48]. The state of the diversity of these particular genes between isolates and its implication for host-adaptation remain however largely unknown.

In France, *F. graminearum* strains belonging to the DON/15-ADON chemotypes, are the most frequently encountered *Fusarium* species on infected wheat [20]. In the present study, we considered the genomic diversity of *F. graminearum* strains isolated from various locations in France. The first objective of our analysis is therefore to quantify the overall genetic diversity of French isolates compared to the reference genome. We also aim to evaluate the contribution of this diversity for phenotypic diversity by a systematic variant annotation and an estimation of the encoding-effects for variants located within genes; with a special attention on genes potentially implicated, or previously suggested to be implicated for host-pathogen interaction. These results will be discussed in the light of pathogen adaptation.

## Results

### SNPs and InDels discovery

Our strategy of genome re-sequencing applied to six *F. graminearum* strains generated a total of 125 million of read pairs of 100 base pairs (bp) in length, corresponding to 37.0 - 44.7 million raw reads per genome (Additional file 1). Quality trimming and filtering of reads resulted in 35.5 – 42.9 million paired-end reads per genome with an average read length of 91 bp. Between 88.4 % and 94.8 % of these reads were aligned correctly on the reference genome – a total genome coverage of 98.8 % (considering all reads produced, 99 % for mitochondrial genomes) and sequencing depths ranging from 79.5 X to 93.2 X depending on the considered isolate (Additional file 1). Only 13 protein coding genes of the 14,160 described in the reference nuclear genome were not covered by read in any of the isolate genomes presented herein (Additional file 2). These genes were discarded for downstream analysis.

The locations of genetic variations were investigated (Table 1). Variants were called on the basis of a variation compared to the sequence of the reference genome (RRES v4.0). Variant calling was fine-tuned to detect preferentially short size variants, *i.e.*, Single Nucleotide Polymorphisms (SNPs) and short Insertions or Deletions (InDels), and obtained a final dataset of 242,756 highly-confident variants, all strains considered, consisting of 234,151 SNPs (96 %) and 8,605 InDels (Table 1, Additional file 3). Regarding the insertion and deletion events, 52 % and 50 % of them, respectively, concerned single nucleotide positions. The largest insertion is 25 nucleotide-long and the largest deletion is 36 nucleotide-long, with mean lengths for both events being 2.8 bp and -2.7 bp respectively (Additional file 4: Figure S1). The number of variants per strain ranges from 143,283 to 146,849 for SNPs, and from 4,844 to 5,073 for InDelS (Table 1). Among them, 82,882 variants (34.1 %) are common between all six French isolates. For simplicity purposes, this particular subset of variants will be referred to as the “common block of diversity”. Beside this baseline of diversity, each isolate differs from the other five French isolates by 67,157 genetic variations in average (65,157 – 69,040; Table 1). Pairwise comparison of isolates shows that INRA-156, with an average of 69,165 variants with each other French isolates, has the most polymorphic genome whereas the genomes of INRA-164 and INRA-181 are the least different with 35,153 variants identified (Table 2). Among the complete set of variable loci identified in this analysis, 1,235

(0.5 %) presented different alleles between French alleles, all different that the reference one (*i.e.* multi-allelic variants).

## Genomic distribution of variants

Variant average genome-wide density reached 6.6 variants per kilobase (kb) considering the all genomes, ranging from 3.9 to 4.0 variants per kb per individual genome (Table 1). The distribution of the variants is not uniform between and within chromosome. At the inter-chromosomal level, Chromosome II, with 5.4 – 5.6 variants per kb per genome always exhibits the greatest variant density (Figure 1). The number of variants detected in the mitochondrial genomes dropped considerably (less than 0.3 variant per kb) compared to nuclear genomes, all variants being localized outside of annotated genic sequences (Additional file 3, Table S3). At the intra-chromosomal level, the contribution of chromosome segments to the overall polymorphism is not linear (Figure 2A).

**Table 1:** Variant calling statistics, considering strain-specific reads and considering total reads produced.

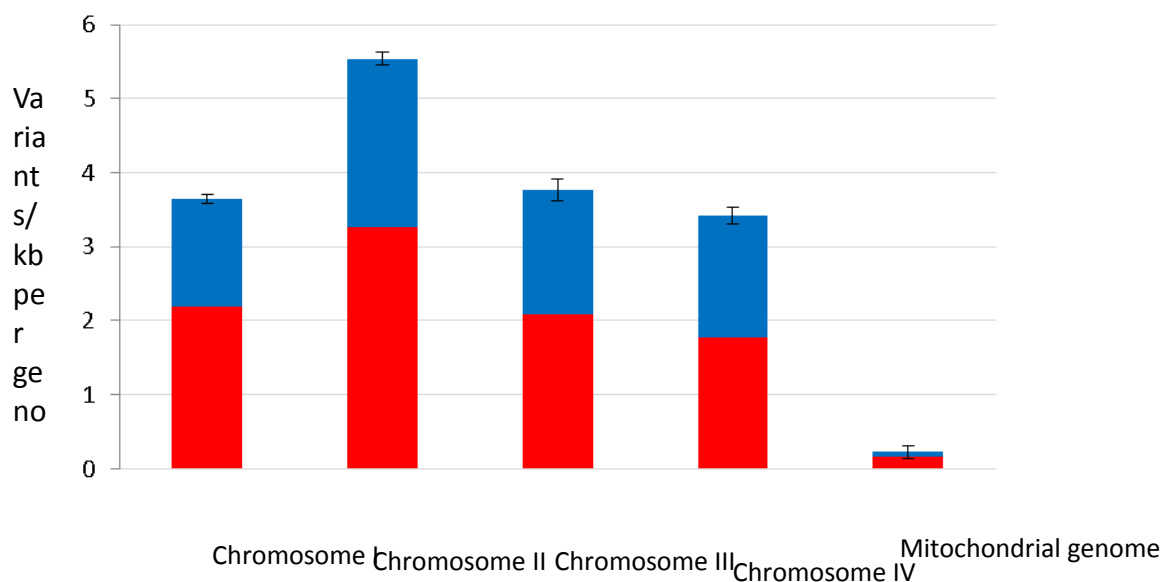
Isolate	INRA-156	INRA-159	INRA-164	INRA-171	INRA-181	INRA-195	Total <sup>a</sup>
% of reference genome callable <sup>b</sup>	97.7%	97.7%	97.6%	97.6%	97.5%	97.7%	98.3%
Number of SNPs	144,679	146,849	144,802	143,283	145,071	145,840	234,151
Number of InDels	4,929	5,073	4,953	4,844	4,938	4,970	8,605
Total number of variants	149,608	151,922	149,755	148,127	150,009	150,810	242,756
Mean depth of sequencing at variant position (X)	72.3	83.5	78.0	79.5	79.6	80.3	78.9
Mean variant density (variants per kb)	3.9	4.0	3.9	3.9	3.9	4.0	6.4
Number of exonic variants	70,850	71,488	70,259	69,130	70,659	71,676	111,975
Number of intronic variants	10,779	10,821	10,509	10,320	10,682	10,786	17,095
Number of non-genic variants	67,979	69,613	68,987	68,677	68,668	68,348	113,686
Number of variants with French genomes	66,726	69,040	66,873	65,245	67,127	67,928	159,874

<sup>a</sup> : considering all reads produced by whole genome sequencing of the six isolates

<sup>b</sup> : excluding the end of the chromosome IV from 7,953,943 bp onwards, corresponding to repeated RNA encoding sequence (see Methods)

**Table 2:** Genome-wide comparison of variants between pairs of isolates. Upper diagonal considers number of variants by pair, lower diagonal considers the part of the overall diversity (242,756 variants) in percent explained by this pair.

Strains	PH-1	INRA-156	INRA-159	INRA-164	INRA-171	INRA-181	INRA-195
PH-1	-	149,608	151,922	149,755	148,127	150,009	150,809
INRA-156	61.6%	-	70,856	66,763	65,709	63,758	78,739
INRA-159	62.6%	29.2%	-	48,501	55,771	62,330	71,769
INRA-164	61.7%	27.5%	20.0%	-	48,453	35,153	68,191
INRA-171	61.0%	27.1%	23.0%	20.0%	-	36,719	66,139
INRA-181	61.8%	26.3%	25.7%	14.5%	15.1%	-	65,433
INRA-195	62.1%	32.4%	29.6%	28.1%	27.2%	27.0%	-

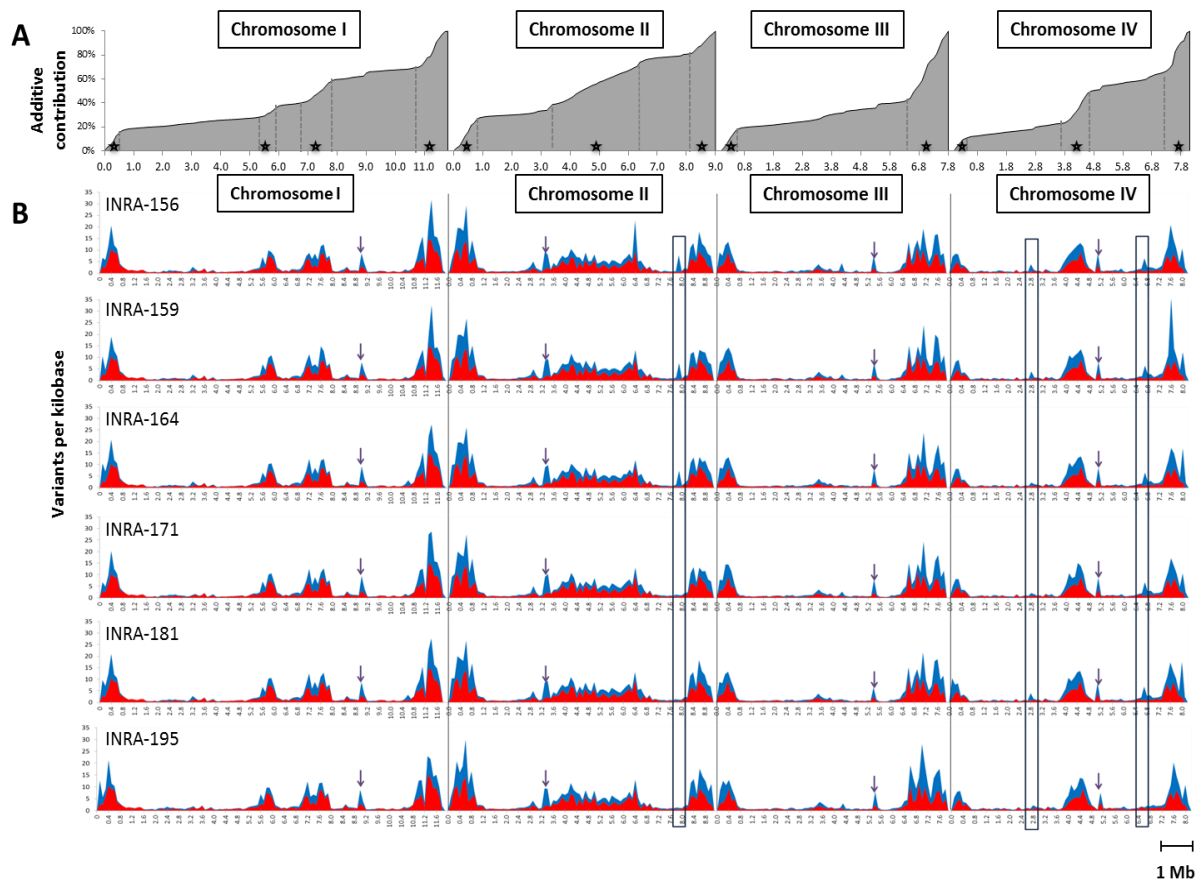


**Figure 1:** Average variant density by strain for the four chromosomes and the mitochondrial genome. Variant density is represented in variants/kb. The density of variants belonging to the common block of diversity (observed in all French isolates) is in red; the density of variant belonging to the diversity observed between French isolates is in blue.

Telomeric/subtelomeric ends and discrete interspersed interstitial regions participate actively to the total polymorphism. Polymorphic islands are distinguished easily (Figure 2A, delimited by dot lines and dark stars; accounted for when longer than 200 kb and showing at least a two-fold increase in variant density compared to the genome-wide median density). Such regions present in average a 7.5-fold increase of variant density compared to others (16.0 variants/kb vs. 2.1 variants/kb). The additive length of these regions represents 31.5 % of total



nuclear genome length while containing 76.7 % of the total polymorphism (Additional file 5: Table S4). The presence of polymorphic islands at both chromosome ends are a common feature between chromosomes, whereas the number and size of interstitial polymorphic regions differ: for example, chromosome I exhibits two distinct variant-rich regions, chromosome II has a long continuous variant-rich region spreading over one third of total chromosomal size, chromosome IV displays a single ~1Mb-long variant-rich region, and chromosome III has none (Figure 2A, 2B). The predicted positions of centromeres [41] also appear to collocate with variant-rich regions (Figure 2B), whereas too short in length to be accounted for polymorphic islands. Variant density is not uniform within polymorphic islands either (Figure 2B). General variant density profiles are conserved between genomes (Figure 2B); and between the common block of diversity and the diversity recorded between French isolates (Figure 2B). This tendency does not exclude occasional differences observed between strains (examples delimited by black rectangles, Figure 2B). For instance, the region ranging from 7.8 Mb to 8 Mb on chromosome II is rich in variants in the genomes of INRA-156, INRA-159 and INRA-164 but not in those of the other three strains.

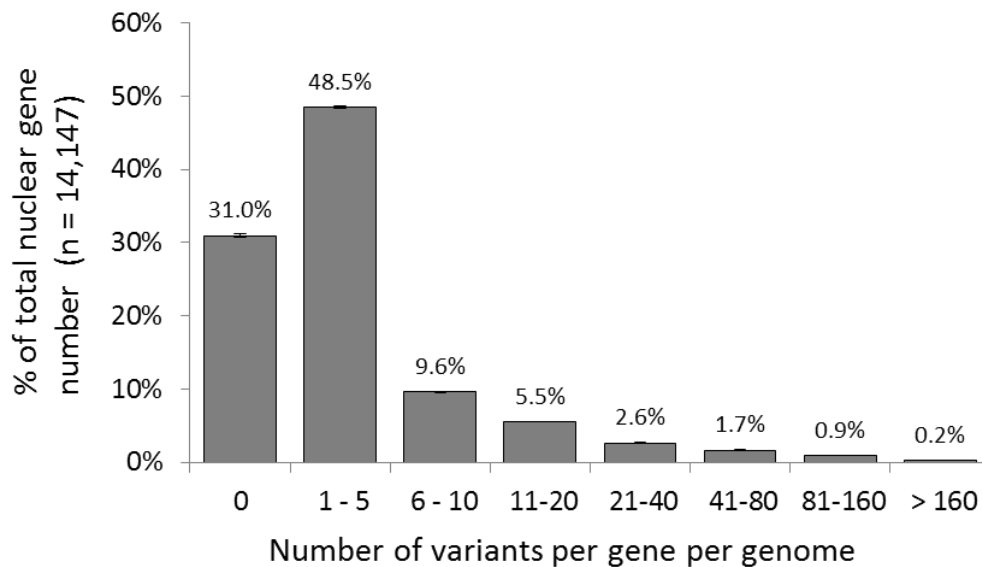


**Figure 2:** Profiles of variant distribution by chromosome. Density profiles were computed for non-overlapping 100kb-long sliding windows along the four chromosomes of *F. graminearum*. A] Cumulative variant density profiles, all polymorphism considered. Star-containing intervals delineated by dotted lines indicate polymorphic islands. B] Variant density profiles along the four chromosomes of *F. graminearum* for each strain. The density of variants belonging to the common block of diversity (observed in all French isolates) is in red; the density of variant belonging to the diversity observed with other French isolates is in blue. Black rectangles highlight selected differences between isolates. The arrows indicate the positions of centromeres.

## Functional annotation of variants

All strains considered, 129,070 variants are found within genic (introns and exons) sequences and 113,686 variants are found elsewhere in the genome (Table 1). Although significant due to the large number of genes, variant density observed within genic sequences does not appear to be greatly reduced compared to the variant density of other sequences (1.05-fold;  $p$ -value < 0.001). Intronic variants (total: 17,095; per genome: 10,320 – 10,821) are overrepresented by 5.3-fold ( $p$ -value < 0.001) whereas exonic variants (total: 111,975; per genome: 69,130 – 71,676) are slightly underrepresented by 0.9-fold ( $p$ -value < 0.001). Considering all protein-coding nuclear genes ( $n=14,147$  excluding not covered genes), 80 % present at least one mutation in at least one isolate - 69 % of genes in average when strains are considered individually (Figure 3). Median number of variants per gene per genome is 1, whereas the

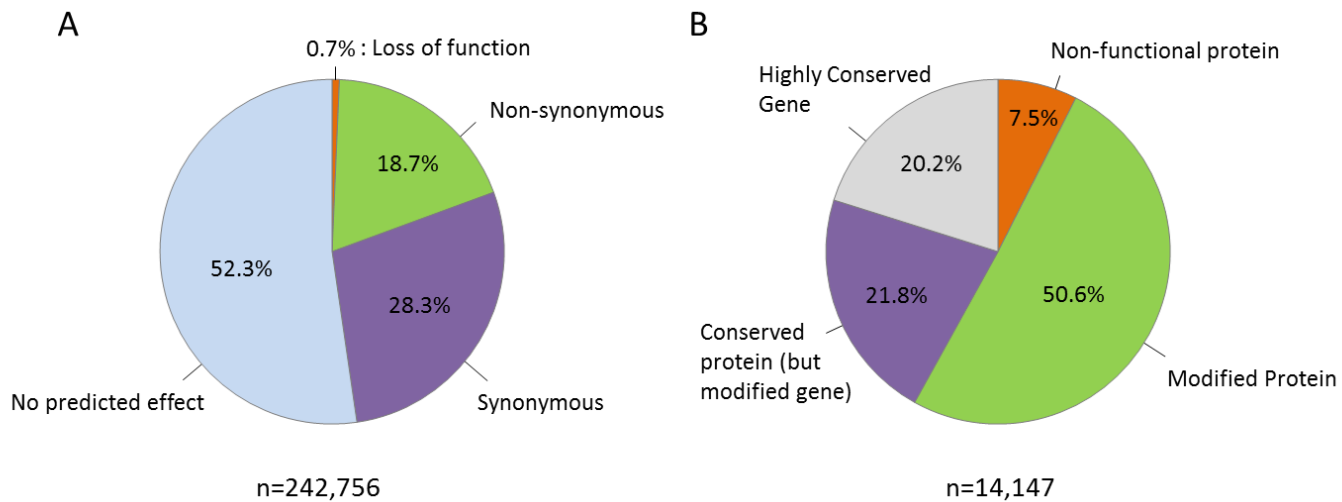
distribution of variant number per gene is skewed due to extreme variant content exhibited by a small percent of genes (Figure 3).



**Figure 3:** Distribution of average variant content per gene per genome. Values are expressed in percent of total nuclear protein encoding gene number (n=14,147). Bars are mean values for the count of variant considered and error bars the standing deviations per genome.

In order to identify biological functions possibly more affected than others by variants, we estimated the consequences of genic variants in all strains considered (including introns and exons; Figure 4A and Additional file 3: Table S3). A little more than half of the variants (52.3 %) are predicted to not change protein sequences because they are located in intergenic and intronic regions, outside of splicing sites. Another 28.3 % have synonymous effects (*i.e.*, a codon exchange leading to no change in amino acid), 0.7 % of total variants have a predicted loss-of-function effect (LoF, in our case the introduction of a frameshift, a stop codon, the loss of the codon start or a critical mutation within the splicing-site), 18.7 % have a non-synonymous effect (*i.e.*, a codon exchange leading to a change in amino acid). Genes can also be organized according to their content in variants and their predicted effects (Figure 4B and Additional file 6: Table S5). Four categories can be defined: the “non-functional” category consists of the 1,057 genes (7.5 % of the protein-coding genes) that contain at least variants predicted to lead to a loss of function in at least one isolate; the “Modified Protein” and “Conserved Protein” categories includes 7,164 genes (50.6 % of the protein-coding genes) with non-synonymous variant(s) and 3,085 genes (21.8 % of the protein-coding genes) with synonymous variant(s) respectively; finally the “Highly Conserved Gene” category

(Additional file 6: Table S5) includes genes with no variant identified in any of the isolates (n = 2,841, 20.1 % of the protein-coding genes).



**Figure 4:** Variant effect prediction and subsequent gene classification. A) Classification of variants according to their predicted effects (n=242,756). Orange: variants leading to a loss of function (LoF) of the proteins; Green: variants with non-synonymous effects (including intronic and exonic variants); Purple: variants with no predicted effect; Blue: variants located outside of genic sequences. B) Classification of genes according to the type of variant (predicted effect) they contain. Orange: genes containing at least variant(s) leading to a loss of function (LoF) of the proteins; Green: genes containing at least variants with non-synonymous effects (including intronic and exonic variants, and containing no LoF variant); Purple: genes containing only variants with no predicted effect; Blue: genes of which no variants have been detected.

### Biological functions that can be affected by genetic variants

We investigated the putative functions of the genes belonging to the different categories described above. A gene ontology (GO) term enrichment approach was used to discover top functions represented in gene lists belonging to each category. Results are summarized in Table 3. Category “Non-functional” is significantly enriched in genes implicated in chitin catabolism; category “Modified Protein” is enriched in genes involved in the regulation of transcription, in oxidation and reduction processes and in the regulation of primary metabolic process; category “Conserved Protein” is enriched in genes acting in signalization and communication, translation, protein transport and several process involved for example in carbohydrate metabolism ; finally, the “Highly Conserved Gene” category is enriched in genes involved in more universal cellular process, such as cytoplasmic transport including Golgi vesicle transport, protein folding and macromolecule assemblies, translation, as well as several biosynthetic and catabolic processes (Table 3).

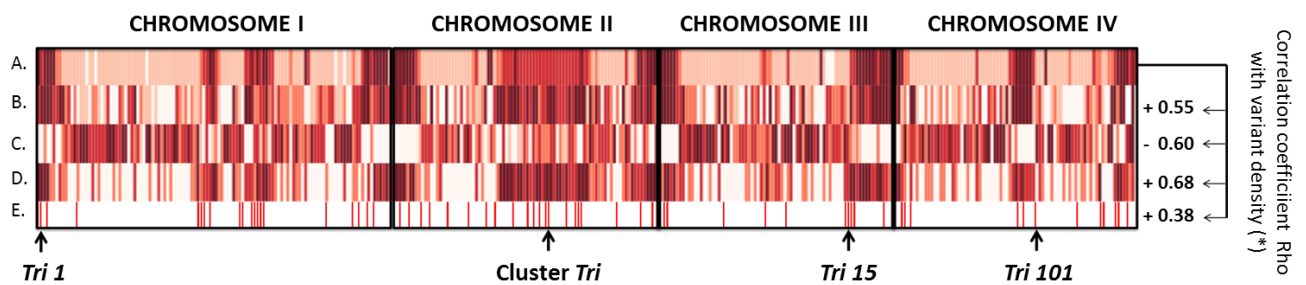
**Table 3:** Significant (p-value < 0.01) gene ontology enrichment of the categories built from their variant contents and downstream coding-effect.

Genes Lists	GO ID	GO Term	Genes in the GO list	Theoretical gene number	Observed gene number	Fold enrichment
"Non-functional"	GO:0006030	chitin metabolic process	21	1	5	4.9
"Modified Protein"	GO:0006355	regulation of transcription, DNA-dependent	458	239	294	1.2
	GO:0055114	oxidation-reduction process	763	398	478	1.2
	GO:0060255	regulation of macromolecule metabolic process	482	250	300	1.2
	GO:0080090	regulation of primary metabolic process	497	258	307	1.2
"Conserved protein"	GO:0044262	cellular carbohydrate metabolic process	30	8	17	2.2
	GO:0007264	small GTPase mediated signal transduction	43	11	22	2.0
	GO:0015031	protein transport	121	31	59	1.9
	GO:0044723	single-organism carbohydrate metabolic process	82	21	36	1.7
	GO:0072521	purine-containing compound metabolic process	90	23	38	1.6
	GO:0006412	translation	181	47	76	1.6
	GO:0007154	cell communication	140	36	56	1.5
	GO:0044267	cellular protein metabolic process	509	132	179	1.4
"Highly conserved genes"	GO:0006888	ER to Golgi vesicle-mediated transport	8	1	6	4.5
	GO:0048193	Golgi vesicle transport	15	3	10	4.0
	GO:0016482	cytoplasmic transport	26	4	12	2.8
	GO:0034622	cellular macromolecular complex assembly	25	4	11	2.6
	GO:0006457	protein folding	40	7	16	2.4
	GO:0022607	cellular component assembly	35	6	14	2.4
	GO:0044283	small molecule biosynthetic process	132	22	46	2.1
	GO:1901136	carbohydrate derivative catabolic process	53	9	18	2.0
	GO:0046394	carboxylic acid biosynthetic process	110	18	37	2.0
	GO:0006996	organelle organization	66	11	22	2.0
	GO:0006412	translation	181	30	59	1.9
	GO:0008652	cellular amino acid biosynthetic process	86	14	27	1.9
	GO:1901565	organonitrogen compound catabolic process	75	13	23	1.8
	GO:0071840	cellular component organization or biogenesis	134	22	39	1.7
	GO:1901566	organonitrogen compound biosynthetic process	190	32	54	1.7
	GO:0044267	cellular protein metabolic process	509	85	128	1.5
	GO:0044281	small molecule metabolic process	463	77	110	1.4
	GO:1901564	organonitrogen compound metabolic process	400	67	95	1.4
	GO:0044249	cellular biosynthetic process	788	132	174	1.3
	GO:0019538	protein metabolic process	680	114	141	1.2

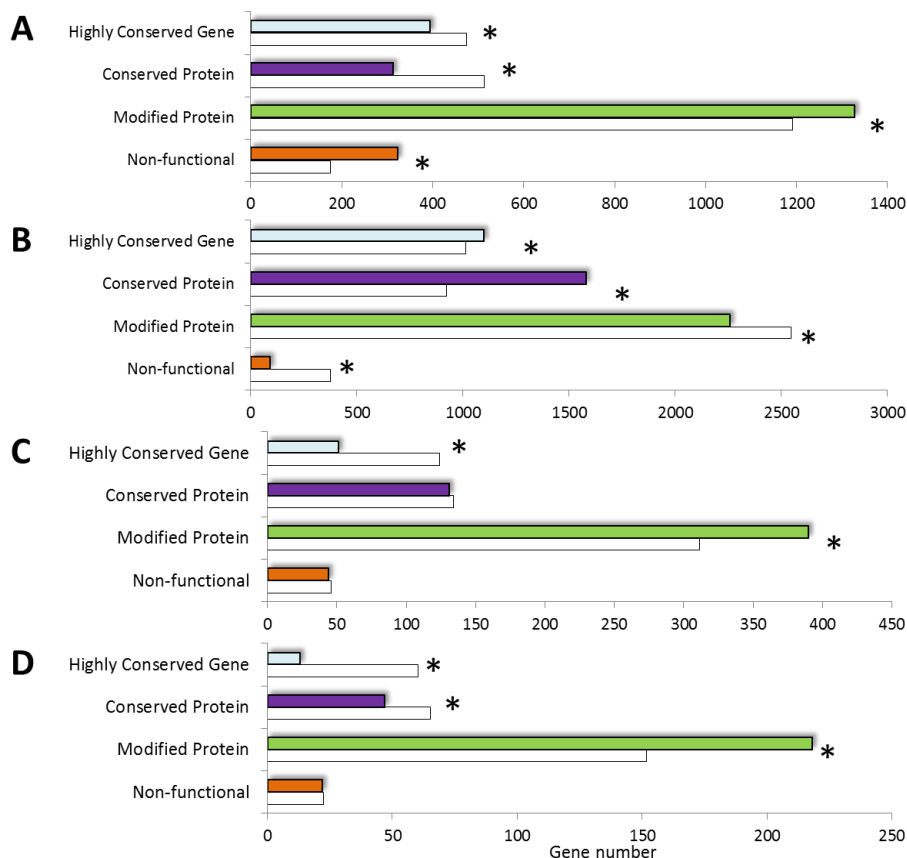
GO term enrichment analyses are however prone to ontology mapping-related biases [49]. Forty five percent of the totality of nuclear protein-coding genes of *F. graminearum* lack GO term annotation [41]. Therefore, we developed a second approach that consist in using *F. graminearum*-specific gene lists compiled from transcriptomic experiments and genome-mining efforts and available from the literature: transcriptomic data from *in planta* experiment, genes coding for putative secreted proteins, genes belonging to predicted secondary metabolite clusters [41,48,50]. The first list derives from *in planta* transcriptomic experiments that identified genes showing unique host-specificity of expression (17 % of total nuclear gene number,  $n = 2,353$ ) by contrast with genes showing constitutive expression (36% of total nuclear gene number  $n = 5,029$ ) suggested to correspond to basal and universal mechanism of host infection ([50], Additional file 6: Table S5). We observed a positive correlation between locations of polymorphisms and location of host-specific genes (Spearman rank order  $Rho = 0.55$ , Figure 5 lane B). Host-specific genes are found overrepresented in the categories “Non-functional” and “Modified Protein” and underrepresented in the categories “Conserved Protein” and “Highly Conserved Gene” (Figure 6A). This observation suggests than non-synonymous mutations tend to be accumulated into these genes. Indeed, loss-of-function and non-synonymous variants are particularly found within these genes with a 2.1-fold and 1.8-fold enrichment, respectively (Additional file 7). Conversely, the locations of genes expressed constitutively in all *in planta* conditions is negatively correlated to the locations of variants ( $Rho = - 0.60$ , Figure 5 lane C). These genes are overrepresented in the categories “Highly Conserved Gene” and “Conserved Protein”, and underrepresented in the categories “Modified Protein” and “Non-functional” (Figure 6B). Similarly, these genes contain less loss-of-function and other non-synonymous variants (5.6 times and 2.5 times respectively; Additional file 7).

The second list consists of genes with typical motifs suggesting that they code for secreted proteins that could therefore be potential effectors ( $n = 616$ ; 126 have been shown to be expressed in a host-specific manner). The spatial distribution of these genes positively correlates with the genome-wide distribution of polymorphisms ( $Rho = 0.68$ , Figure 5 lane D). These secreted protein-encoding genes are found overrepresented in the category “Modified Protein” and underrepresented in the category “Highly Conserved Gene” by 1.25 and 0.41-fold respectively (Figure 6C). These genes are further enriched in non-synonymous mutations (other than loss-of-function) by 1.38 fold (Additional file 7).

Finally, we investigated genes predicted to be implicated in the biosynthesis of secondary metabolites and (mostly) organized in clusters on the genome (n = 300). The genomic distribution of these genes is significantly correlated with polymorphism ( $Rho = 0.38$ , Figure 5 lane E). They are significantly overrepresented in the category “Modified Protein” and significantly underrepresented in the categories “Highly Conserved Gene” and “Conserved Protein” (Figure 6D). These genes are indeed enriched in non-synonymous variants, but show in the other hand a reduction of LoF mutations (Additional file 7). Remarkable secondary metabolites are the type B trichothecenes (TCTB), including the deoxynivalenol (DON), reported to be involved in pathogenicity [51]. We examined the polymorphisms affecting *Tri* genes (n = 15) involved in the biosynthesis of TCTB (12 of them are clustered on chromosome II as indicated on Figure 5; Additional file 8: Table S6). An overall of 252 variants have been identified within the genic sequences and the intergenic sequences of *Tri* genes (located in the upstream and downstream sequences for the non-clustered *Tri* genes; Additional file 8: Table S7). Among these variants, 131 belong to the common block of diversity (observed in all six genomes analyzed herein). Only four of the rest of the variants are predicted to bring non-synonymous effects other than loss-of-function. All of them are located within the coding sequence of *Tri15* and affect the strains INRA-159, INRA-164, INRA-171 and INRA-181 (Additional file 8, Table S7). *Tri15* gene encodes a putative transcription factor and does not seem to be implicated in TCTB production [5].



**Figure 5:** Heatmap representation of variant and gene counts per 100 kb-long non-overlapping windows. Spearman rank order correlation coefficients were computed between variant and gene counts. The star \* indicates that all correlations are significant at the threshold  $p = 0.01$ . A. Genetic variants ( $n = 242,756$ ). B. Host-specific genes ( $n = 2,353$ ) [50]. C. *In planta*-constitutive genes ( $n = 5,029$ ) [50]. D. Secreted protein-encoding genes ( $n = 616$ ) [41]. E. Secondary metabolite-encoding gene clusters ( $n = 67$ ) [48]. The positions of the *Tri* cluster and the not-clustered *Tri* genes *Tri1*, *Tri15* and *Tri101* are indicated by arrows.



**Figure 6:** Selected *F. graminearum*-specific gene content of each category of predicted variant effect. For each category, actual gene counts (colored bars) are compared to the theoretical counts expected under hypothesis of random distribution of variants (white). The star \* means Chi-squared test was significant ( $p$ -value  $< 0.001$ ). A. Host-specific genes ( $n = 2,353$ ) [50]. B. *In planta*-constitutive genes ( $n = 5,029$ ) [50]. C. Secreted protein-encoding genes ( $n = 616$ ) [41]. D. Clustered secondary metabolite-encoding gene ( $n = 300$ ).



## Genes showing an excess of non-synonymous effect mutations

In order to identify genes accumulating non-synonymous effect mutations, we consider the total polymorphism detected in this analysis and extracted 797 genes that accumulated either or both non-synonymous (NS) and LoF mutations (NS + LoF > total number of mutation, minimum total number of mutation = 4; Additional file 6: Table S5). The large majority of them (64 %) is located within polymorphic islands (Additional file 6: Table S5). Twenty-nine of them have been previously shown to be both expressed *in planta* and predicted to be secreted (Table 4). Fifteen have been shown to be expressed in a host-specific manner and only one has been shown to be expressed constitutively in all *planta* conditions tested (Table 4). Remarkably, all of them have no known function according to reference genome annotation [41], with the exception of FGRRES\_04689 that code for a rhamnogalacturonase A, involved in cell wall polysaccharide degradation. Seven of them contain LoF variants (FGRRES\_16333, FGRRES\_03521, FGRRES\_12210, FGRRES\_04646\_M, FGRRES\_13876, FGRRES\_07699, and FGRRES\_09118). For FGRRES\_04646\_M, the mutation is present in every French isolates tested. This gene is unlikely to be an essential effector during infection of wheat as several strains of this sample have been shown to be highly aggressive (Table 5; Additional file 3: Table S3). In the other hand, the gene FGRRES\_07699 is predicted to be non-functional in the highly aggressive strain INRA-156 only; the gene FGRRES\_12210 is predicted to be non-functional in the less aggressive strain INRA-195 only. These genes represent interesting effectors that could have escaped from the host defense for the first case or implicated in aggressiveness reduction for the second case. The knowledge on the diversity of these genes might help further investigations.

**Table 4:** Putative effectors showing an excess of non-synonymous effect mutations.

Gene ID	Chrom	Gene start (bp)	Gene end (bp)	Gene description	LoF <sup>a</sup>	NS <sup>b</sup>	Total <sup>c</sup>	<i>In Planta</i> Expression (Harris et al. 2016)
FGRRES_11675	I	356,230	357,118	Uncharacterized protein	0	7	10	Wheat
FGRRES_01778	I	5,860,579	5,861,567	Uncharacterized protein	0	5	9	Wheat
FGRRES_02228	I	7,225,618	7,227,797	Uncharacterized protein	0	10	19	Wheat Barley
FGRRES_02269	I	7,357,559	7,358,332	Uncharacterized protein	0	4	7	Wheat Barley
FGRRES_13692	I	9,626,040	9,628,066	Uncharacterized protein	0	4	7	Wheat Barley
FGRRES_07993	II	110,904	113,251	Uncharacterized protein	0	5	7	Wheat Barley
FGRRES_17022	II	1,652,257	1,656,378	Uncharacterized protein	0	5	6	Wheat
FGRRES_16333	II	4,194,219	4,196,939	Uncharacterized protein	3	41	80	Barley
FGRRES_03274	II	4,695,334	4,698,042	Uncharacterized protein	0	12	22	Wheat Barley
FGRRES_03521	II	5,366,512	5,367,123	Uncharacterized protein	1	5	7	Wheat
FGRRES_03612	II	5,604,284	5,605,254	Uncharacterized protein	0	4	7	Wheat Barley

FGRRES_12405_M	II	5,622,275	5,622,943	Uncharacterized protein	0	8	9	Wheat Barley
FGRRES_03944	II	6,465,510	6,466,808	Uncharacterized protein	0	5	8	Wheat Barley
FGRRES_03972	II	6,548,953	6,550,914	Uncharacterized protein	0	7	10	Barley
FGRRES_04429	II	7,989,077	7,992,064	Uncharacterized protein	0	4	7	Wheat
FGRRES_12210	II	8,620,515	8,622,358	Uncharacterized protein	1	10	18	Wheat
FGRRES_04646_M	II	8,655,498	8,656,180	Uncharacterized protein	2	12	22	Wheat
FGRRES_04689	II	8,765,660	8,767,148	Rhamnolacturonase A	0	7	13	Wheat Barley
FGRRES_05719	III	3,177,333	3,180,794	Uncharacterized protein	0	4	6	Wheat Barley
FGRRES_05847	III	3,549,004	3,550,475	Uncharacterized protein	0	3	5	Wheat Barley Maize
FGRRES_12835	III	3,658,250	3,659,015	Uncharacterized protein	0	4	7	Wheat
FGRRES_16623	III	4,751,122	4,755,561	Uncharacterized protein	0	6	11	Wheat
FGRRES_13876	III	6,370,523	6,371,374	Uncharacterized protein	3	18	40	Wheat
FGRRES_17469	III	6,711,891	6,714,195	Uncharacterized protein	0	7	12	Barley
FGRRES_11379	III	7,412,583	7,413,959	Uncharacterized protein	0	12	23	Wheat
FGRRES_06610	IV	582,670	585,011	Uncharacterized protein	0	5	9	Wheat Barley
FGRRES_07686_M	IV	4,196,885	4,197,838	Uncharacterized protein	0	8	14	Wheat Barley
FGRRES_07699	IV	4,231,666	4,232,190	Uncharacterized protein	1	5	10	Wheat Barley
FGRRES_09118	IV	7,533,260	7,536,888	Uncharacterized protein	3	10	21	Wheat

<sup>a</sup>: variant with loss of function effect

<sup>b</sup>: variant with other non-synonymous effect

<sup>c</sup>: total variant numbers include variant detected within the 100 base pairs located in upstream and downstream of the genic sequence.

**Table 5:** Phenotypic information and geographical origin for the six strains

Strains	TCTB production <i>in vitro</i>	Aggressiveness on wheat	TCTB production in wheat	Location in France/Administrative division
<b>INRA-156</b>	++	+++	+++	Center / Cher (18)
<b>INRA-159</b>	++++	0	0	Center / Cher (18)
<b>INRA-164</b>	++++	++++	++++	North / Seine Maritime (76)
<b>INRA-171</b>	0	++	++	South-West / Gers (32)
<b>INRA-181</b>	++	+++	+++	North / Eure (27)
<b>INRA-195</b>	0	+	+	North-East / Meuse (55)

## **Discussion**

The presented work examines the level of variation that can be observed between the genomes of different *F. graminearum* isolates at the sequence level. In addition to describing genome-wide polymorphisms, this analysis proposes, for the first time, to quantify the downstream effects of the observed variants, particularly exonic as well as intronic variants that can lead to important consequences on the translation product [52]. Here, we applied a whole-genome reference-based DNA re-sequencing strategy rather than *de novo* assembly previously described as more sensitive to sequencing errors [53]. Using a re-sequencing method, the accuracy of variant calling greatly depends on the quality of the read alignments on the reference genome and the depth of read coverage per base. Filters must be applied to differentiate true variants from sequencing errors while keeping the false negative rate low. There is no « one-size-fit-all » situation and settings must be adjusted according to the type of genetic variant investigated [53]. For *F. graminearum*, stringent filtration is further possible (and recommended) as this fungus is haploid and one allele is expected at the position. A preliminary test-run indicated that SNPs and short InDels are mostly found in the genome of *F. graminearum* with very few variants of larger sizes (data not shown). Consequently, the filters applied to the analysis presented here were set as optimal for SNPs and short InDels detection for enhanced variant discovery.

### ***F. graminearum* genome-wide polymorphism is consistent with its pathogenic lifestyle**

The comparison of six genomes of French isolates with the reference genome of the PH-1 strain of *F. graminearum* [41] produced a highly confident set of 242,756 distinct variants total. Each of the six genomes presented an average of 150,039 variations when compared to the PH-1 reference genome, and ~ 67,157 variations when compared to each other. This number is much higher than the 10,495 SNPs identified in the first published investigation of the genome-wide polymorphism between another North-American isolate and the PH-1 strain [42]. The much lower number of variants then observed is certainly linked to a very low fold-coverage, 0.4X, being insufficient for exhaustive and confident variant calling [42]. More recently, Walkowiak et al. [54] assembled two genomes of *F. graminearum* representative of the two chemotype-based populations observed in Canada, DON/3-ADON and DON/15-ADON. They reported 147,555 and 103,774 SNPs with the reference genome respectively, as well as 148,978 SNPs between the two Canadian isolates, approximating the level of polymorphism reported herein. Altogether, an estimate of ~150,000 SNPs seem to be a typical variant content expected to be observed between genomes of geographically distal isolates or

belonging to different populations. Along the same line, the reduced level of polymorphism observed between French isolates may suggest that these strains are likely to belong to the same population.

On a broader scale, the genome-wide level of polymorphism observed between isolates of *F. graminearum* (~4 SNP/ Kb) is consistent with levels of polymorphism exhibited by other pathogenic fungi, as reported in the causal agent of poplar rust *Melampsora larici-populina* (~ 2 SNP/Kb; [37]), the causal agent of the wheat stripe rust *Puccinia striiformis* (~ 5 SNPs per Kb;[55]) or in the human pathogen *Coccidioides immitis* (~ 5 SNP/Kb);[56]). These levels are lower than the one revealed between isolates of *Botrytis cinerea* that could reach 10 variants per kb [38]. Authors suggested that such genome-wide diversity is linked to the ability of this pathogen to infect a remarkably broad range of hosts. In comparison, the human genome exhibits between 1.2 and 1.5 variants per kb [57]. The higher values observed in fungal pathogen genomes may be a consequence from their parasitic lifestyle that pushes their need to evolution up [58]. According to the criteria given by McDonald and Linde [23] to estimate the potential of evolution of fungal pathogen, *F. graminearum* can be considered as a high-adaptive potential pathogen. Paradoxically, the various analyses aiming at identifying parts of the genome that are under selection for diversification failed to identify genomic regions under strong selection [11,59]. In the present analysis, we use total genome information and identify 797 candidate genes accumulating missense and nonsense mutations. The functions of these genes are unknown for the majority but their potential implication in pathogenicity and adaptation certainly calls for in-depth investigations.

## **The multi-scaled location of polymorphisms in the genome**

The genome-wide average value of variant density is not a metric sufficient enough to fully comprehend the patterns of polymorphism in *F. graminearum*; indeed it does not describe the remarkable discrete variations that we observed at both inter- and intra- chromosomal dimensions. For example, chromosome II is more polymorphic than the other chromosomes. This higher speed of diversification of this individual chromosome has been previously reported in genome-wide comparisons of close species of the *Fusarium* genus [59]. The authors hypothesized that the chromosome II of *F. graminearum* could play a preponderant role for host infection and adaptation [59,60]. Our analysis is in line with such a hypothesis. Distribution of polymorphisms is also highly uneven within each chromosome. The single-base resolution of our analysis enabled the accurate definition of polymorphism islands along chromosomes. Polymorphisms are preferentially located at the ends of chromosomes. Such

phenomenon is common in fungal pathogens and more generally in eukaryotic genomes [58]. The interstitial polymorphic islands are, for their part, more original. These regions have been previously investigated and are described as telomeric-like and subtelomeric-like regions that originate from ancestral chromosomes fusion events [42,60]. Remarkably, the chromosomal landscapes of polymorphism reported herein follow striking similitude with the lower-coverage analysis reported in 2007 [42]. Considering these patterns of polymorphism highly conserved between isolates, we may hypothesize that genome architecture plays a predominant role in shaping the polymorphism landscape, instead of evolution forces. Several mechanisms have been previously proposed as driving genome structure of fungal pathogens [61], among which the action of meiotic recombination may play an important role through the preferential shuffling of particular chromosome regions [62]. An inconsistency of recombination rate has been already reported along *F. graminearum* genome, and the increases of recombination activity seem to collocate with variant rich regions [42]. These same regions were further showed to be enriched in specific epigenetic mark [63], of which implication for meiotic recombination through chromatin remodeling has never been tested. Thus, the weight of individual contributions and inter-connections between the different proposed elements of regulation remains however unclear [61]. A comprehensive investigation of these different phenomena could shed light on the events driving the organization of *F. graminearum* genome and its evolution.

Finally, we observed that polymorphism rates are highly variable at gene level, with introns being more polymorphic than exons. Such situation could be the result of selective pressures since exonic variants can more directly affect protein function, as such, can be rapidly unselected [64]. Nevertheless, this does not mean that variants located in introns, and typically classified as synonymous mutations, have no contribution in protein polymorphism. Indeed, the demonstration has been done that mutations located in introns can have important effect, notably by altering the splicing process [52,65]. Accordingly, the presented work takes all variants into consideration during the annotation process [66], identifying 1,647 variants with predicted loss of function effect and 45,196 variants with other predicted non-synonymous effects.

## **Further evidence of a two-speed genome organization in *F. graminearum***

Our data reveal a remarkable positive correlation between specific biological functions and polymorphism along the genome. For example, polymorphic islands are enriched in genes with roles in biotic and abiotic adaptation, these genes exhibiting a greater level of polymorphism than genes with basal and vital functions. This result is especially true for genes coding for secreted proteins or belonging to secondary metabolite biosynthesis clusters; both categories of genes that have been suggested to play preponderant roles during pathogenesis [41,44]. As a whole, plant-specifically expressed genes, which translate host-specific mechanisms of infection [50], are overrepresented in these hotspots of diversity and are more diverse than other genes. In the line of our above hypothesis, we propose that this correlation arise from the preferential location of certain biological functions according to the organization of the polymorphism in the genome rather than a result of the historical and ongoing diversifying selection acting directly on these genes. Such genomic organization could argue in favor of contrasted abilities of evolution of *F. graminearum* gene repertoire - with genes implicated in basal process being placed in conserved compartments and genes with a bigger need of evolution being placed in highly diversifying chromosomal segments.

With the convergence of knowledge about fungal genomes, such “two-speed” organization seems to be a frequent feature in filamentous pathogens [58,61,67–70]. Nevertheless, in several pathogenic species, a faster speed of evolution has been attributed to gene sparse, repeat-rich compartment, as in *Leptosphaeria maculans*, or dedicated dispensable chromosomes, as observed in *Zymoseptoria tritici* [61,67]. Compared to closely related species, as in *Fusarium oxysporum*, no dispensable chromosome has been ever observed in *F. graminearum*. Moreover, its genome is quite compacted and relatively poor in repeated sequences [41,42]. This decreased number of repeated sequences compared to some other *Fusarium* species has been previously attributed to the action of the repeat induced mutation system (RIP) that introduces point mutations within repeated sequence and therefore protect the genome from the mobility of transposons [42]. Therefore, this system may have participated to the organization of the polymorphism in the genome of *F. graminearum* [42,61]. Altogether, *F. graminearum* is an original and interesting model for the investigation of genome architecture in evolution as well as improve our understanding about the theory of two-speed genome evolution of fungal pathogens.

## **The added-value of genomic data to identify genes involved in pathogenicity.**

The level of polymorphism affecting gene sequence is remarkable. Indeed per genome, up to 69 % of the genes are polymorphic compared to the reference sequence. Such polymorphism should be taken into account for genome editing approaches, often designed from the consensus sequence of the reference genome.

Quantitative variation of mycotoxin production and aggressiveness has been previously observed within field populations of *F. graminearum* [71]. Although variable according to environment, heritability of these traits has been demonstrated and suggests the action of multiple genetic factors [11]. To associate genetic variation with phenotype changes remains however challenging. Polymorphism of genes involved on trichothecene of type B production, as the biosynthetic *Tri* gene cluster, has been previously suggested to be responsible to aggressiveness variation in *F. graminearum* [8]. Likewise, a genome-wide association survey associated variants located in 27 different genes to variations in aggressiveness, all of them unlinked to mycotoxin production [11,72]. The isolates used in the present survey presented contrasted level of aggressiveness in wheat, correlated to contrasted level of DON production (Table 5). These phenotypic characteristics were further shown to be stable, as repeatable and not dependent to the wheat variety inoculated (data not shown). Although these phenotypes may be related to genetic variation, none of the variants described previously were retrieved in French isolates; and may indicate that the genetic bases of aggressiveness and TCTB production may be more complex. As a preliminary investigation, we separated our sample into two discrete groups according to aggressiveness and DON production and consider the distribution of variants. Therefore, forty eighty four genes were conserved in highly toxinogenic and aggressive isolates, and accumulating non-synonymous mutation(s) in less toxinogenic and less aggressive isolates (data not shown). Genes coding for vesicle trafficking were found to be significantly impacted by mutation in the less toxinogenic and less aggressive isolates. This observation is consistent with the vesicle-mediated secretion of TCTB in *F. graminearum* [73], the regulation of secondary metabolism and cellular compartmentalization of biosynthesis pathways being tightly linked in fungi [74].

## **Conclusion**

Whole-genome sequencing of six *F. graminearum* isolates revealed a remarkable number of polymorphisms, with an overall of 242,756 highly confident variants. Polymorphisms are preferentially found clustered in the genome and may play a role in the diversification of the gene repertoire implicated in host-pathogen interaction. We further hypothesize that fungal biological functions are organized in such a way that they take full advantage of the evolving dichotomy proposed by the intrinsic architecture of this pathogen's chromosomes. The molecular control of intrinsic chromosome features remains however to be investigated. Our observations further emphasize the high-adaptive potential of this pathogen and defend the use of more integrative pest management. As a whole, this detailed description of the genetic and functional diversity of these genomes is a milestone on the path to dissect the genetic bases of important history-life traits of *F. graminearum*.

## **Methods**

### **Fungal isolates**

Six strains of *Fusarium graminearum sensu stricto* were isolated from wheat plants cultivated in several French regions between 2001 and 2002 ([75], Table 5). These strains exhibit various trends of pathogenicity and quantitative profiles of DON/15-ADON production that are representative of the genotypic and phenotypic diversity observed within a larger French collection of isolates ([75], Table 5).

### **Extraction of genomic DNA and sequencing**

Genomic DNA was extracted from ~50 mg of lyophilized mycelium previously grown for five days on potato dextrose agar (39 g/l, Difco). Mycelia were lysed in 600 µL of a buffer containing 100 mM Tris-HCl (pH 9.0), 10 mM EDTA, 1 % sarkosyl, and proteinase K 200 µg/mL for 2 h at 65 °C. After centrifugation (10 min at 10,000 g), the supernatant was extracted successively with 1 volume of phenol, 1 volume of phenol: chloroform (50:50) and finally 1 volume of chloroform. Nucleic acids were precipitated with 0.1 volume of cold sodium acetate (pH 5.5, 3M) and two volumes of isopropanol and agitated twice. Solution was then centrifuged 10 min at 10,000 g and supernatant were eluted. DNA precipitate was washed twice with 1 mL cold 70% ethanol for 5 minutes. After centrifugation, DNA was air dried 5 minutes using SpeedVac. DNA was dissolved in 100 µL nuclease-free water.



Preparation of the libraries and sequencing was performed at the Montpellier GenomiX sequencing platform (France [76]). Briefly, quantities of genomic DNA were measured using a Qubit® Fluorometer (Life Technologies) and DNA integrity was verified by electrophoresis on Bioanalyzer (Agilent). DNA libraries were prepared from one µg of DNA per strain using TruSeq DNA sample preparation kit (Illumina) following the manufacturer's instructions for 6-plexed samples (library size 350 bp +/-50 bp). Sequencing was performed on one lane of Illumina HiSeq 2000 generating 100 bp-long paired-ends reads. Post-run read quality was verified using FastQC ([77], v0.11.2).

## **SNP & Indel discovery and analysis**

Reads were cleaned up using PRINSEQ v0.17.1 [78]. Briefly, duplicated reads were removed and the 9th first 5' nucleotides were systematically trimmed due to skewed base composition introduced by sequencing preparation. Reads with an overall mean Phred-scaled value less than 20 were discarded. Remaining reads were further 3' trimmed for quality (Phred scale threshold of 20); high quality paired-end reads with length greater than 20 nucleotides were aligned on the genome version RRes V4.0 ([41], [EMBL-EBI accessions HG970331, HG970332, HG970333, HG970334, and HG970335]) using BWA (v0.7.8) and BWA-MEM with standard parameters and a seed size of 15 nucleotides [79]. Invalid paired with aberrant insertion size or unpaired alignments were filtered out using Samtools v 0.1.19 [80]. Valid alignments were annotated using Picard tools (v1.88, [81]) for further processing with the GATK suite (v2.4,[82]). Identification of genomic regions that could not be mapped or called for variants considering our sequencing set up was conducted using the CallableLoci module of GATK. Alignments on the last 1.38 Mb of the end of chromosome IV systematically showed poor quality scores and were removed from the analysis. This result arises from the large number of rRNA-encoding repetitive sequences units [41] which lengths approximate the length of the sequenced reads resulting in non-unique alignments that are excluded from the analysis. This specific region, however, does not contain known protein-coding genes.

The six BAM files were simultaneously used to discover SNPs and InDels using GATK and the Unified Genotyper walker in haploid mode [82]. Depth of read coverage per 100 kb-long non-overlapping bins was calculated using the bamCoverage tool of the Bamtools suite (version 1.5.9.1.0). Positions of InDels were recalibrated using the RealignerTargetCreator and the IndelRealigner modules of GATK with standard parameters. A variant was called when observed on at least 80% of the reads aligned at the position with a read coverage more

than 5 reads, and a confidence threshold greater than 250 (SelectVariants module of GATK, v2.4). Other parameters were standards. Variant densities for 100 kb-long non-overlapping windows and other statistics on the resulting Variant Call Format (VCF) files were obtained using Vcftools (v0.1.12a, [83]). Regions longer than 200 kb showing at least a two-fold increase in variant density compared to the genome-wide median density were considered as polymorphic islands.

## **Functional variant annotation**

Genetic variants from specific genomic compartments (*i.e.*, exons, introns, and intergenic regions) were annotated and their effect predicted using snpEff [66] and a custom database constructed from the reference genome (RRes v4.0) and the GTF file associated. Briefly, the software classifies variants found within genes (exons and introns) according to their downstream effect at the protein level (under the standard genetic code). A full list of mutation event classification is available [84]. Using this classification, four categories of genes were distinguished: i) the category “Non-functional”, *i.e.*, genes containing variant(s) predicted to lead to a loss of the function of the protein; ii) the category “Modified Protein”, *i.e.*, genes containing only variant(s) predicted to bring one or few amino acid composition changes; iii) the category “Conserved Protein”, *i.e.*, genes containing only variant(s) which do(es) not change amino composition of protein; iv) the category “Highly Conserved Gene”, *i.e.*, genes with no variant detected. Then, we calculated the ratio between the number of mutations with non-synonymous effect (including non-synonymous exonic and intronic variants as well as loss-of-function variants) and the total number of variants within genic sequences (including 100 bp downstream and upstream). Genes were accepted to show an excess of mutation with non-synonymous effect if they contained at least 4 mutations and a ratio of non-synonymous effect mutation on total number of mutation greater than 0.5 (corresponding to a two-fold increase of the genome-wide median number). PROVEAN software [85] was finally used to estimate the functional importance of missense mutations found in genes coding for secreted proteins (as described in King et al., 2015 [41]), genes expressed in wheat, barley and/or maize (Harris et al., 2016 [50]), and genes predicted to belong to secondary metabolite pathways (retrieved from Sieber et al., 2014 [48]).

## **Statistical analyses**

Annotation enrichment analysis was conducted using the Gene Ontology Enrichment tools proposed online by the EuPathDB project [43,86] using Biological ontology and InterPro predictions. Enrichments were accepted for *p*-value lower or equal to 0.01. Chi-squared test

was used to compare the observed distribution of the number of genes affected by mutations in the genome with the theoretical distribution of the number genes affected by mutations under the hypothesis of a random distribution of variants in the genome. Similarly, variant enrichment in the polymorphic islands or in genes was tested using the Chi-squared test. Over and underrepresentation were accepted for  $p$ -value  $< 0.001$ . Gene-enrichment in variant-rich region was determined by calculating the density of both variants and genes in non-overlapping 100 kb windows and testing each window for Spearman rank order correlation at the significance threshold of  $p$ -value  $< 0.001$ .

## **Availability of data and materials**

Strains are available from the International Center for Microbial Resources (CIRM,[87]), with codes 1974, 1977, 1981, 1988, 1999 and 2013 for strains INRA-156, INRA-159, INRA-164, INRA-171, INRA-181 and INRA-195 respectively.

The entire set of raw reads that was generated in this project has been deposited in fastq format at the GenBank Sequence Read Archive under the accession number SRP064374.

The description and annotation of variant supporting the conclusions of this article is accessible in Additional File 1: Table S2.

## **Declarations**

The authors declare that they have no competing interests.

## Additional figure and table legends:

**Additional file 1: Table S1:** Metrics and statistics of read sequencing and alignment on the reference genome (version RRES V4.0).

Isolates	INRA-156	INRA-159	INRA-164	INRA-171	INRA-181	INRA-195	Total*
<b>Number of raw reads</b>	36,994,372	44,672,194	40,304,694	42,545,510	42,058,676	43,721,972	250,297,418
<b>Number of cleaned reads</b>	35,517,410	42,924,658	38,958,456	38,669,436	40,253,380	39,241,388	235,564,728
<b>GC %</b>	47	46	47	45	47	46	46
<b>Number of reads aligned on the reference</b>	34,407,939	40,523,004	37,891,608	37,497,652	39,202,629	36,797,230	226,320,062
<b>Number of properly paired and aligned reads (% of cleaned read)</b>	33,236,718 (93.6%)	37,927,124 (88.4%)	35,504,178 (91.1%)	36,667,030 (94.8%)	38,030,258 (94.5%)	36,057,402 (91.9%)	217,422,710 (92.4%)
<b>Sequencing depth (X)</b>	79.5	93.2	87.9	82.9	89.7	84.5	517.6
<b>Genome coverage (in %)</b>	98.1	98.2	98.0	98.0	97.9	98.3	98.8

\* considering all reads produced by whole genome sequencing of the six isolates

**Additional file 2: Table S2:** Genes described the reference genome (RRES v4.0) not covered by reads in any isolate genomes.

Gene ID (RRES v4.0)	Gene ID (FGDB v3.1)	Chrom	Gene start (bp)	Gene end (bp)	Gene description
FGRRES_11584	FGSG_11584	1	26,109	26,441	Uncharacterized protein
FGRRES_15050	FGSG_15050	1	449,564	449,815	-
FGRRES_20068	FGSG_20068	1	7,039,832	7,040,080	-
FGRRES_13817	FGSG_13817	1	11,726,770	11,727,350	Uncharacterized protein
FGRRES_13818	FGSG_13818	1	11,728,645	11,729,228	Uncharacterized protein
FGRRES_20107	FGSG_20107	2	23,530	23,856	-
FGRRES_15458	FGSG_15458	2	98,299	98,593	-
FGRRES_20171	FGSG_20171	2	5,016,684	5,016,917	-
FGRRES_04705	FGSG_04705	3	56,134	56,618	Uncharacterized protein
FGRRES_15259	FGSG_15259	3	347,938	348,183	Uncharacterized protein
FGRRES_20329	FGSG_20329	4	15,034	15,517	-
FGRRES_20408	FGSG_20408	4	7,963,922	7,964,143	-
FGRRES_15535	FGSG_15535	4	7,969,380	7,969,592	-

**Additional file 3: Table S3:** Information and annotations for all genetic variants detected in this analysis.

*# Here a description of the different columns:*

column A: Chromosome

column B: Position on the reference genome

column C: Nature of the reference allele

column D: Nature of the alternative allele. In the case of multiallelic alternative allele (*F. graminearum* is haploid but alternative allele can be different depending of the isolate) the allele are written following this code: "first allele,seconde allele".

Colone E: Allelic option at this position in INRA-156 strain. "0": reference allele; "1": alternative allele. "Filtered": alternative allele was called in primary call but filtered due to bad quality, "NA" read deapth or alignment quality didn't pass the filters and allelic otpion could not be determined.

column F: Allelic option at this position in INRA-159 strain. "0": reference allele; "1": alternative allele. "2": alternative allele correspond to the second choice proposed. "Filtered": alternative allele was called in primary call but filtered due to bad quality, "NA" read deapth or alignment quality didn't pass the filters and allelic otpion could not be determined.

column G: Similar for INRA-164

column H: Similar for INRA-171

column I: Similar for INRA-181

column J: Similar for INRA-195

column K: Type of variants

column L: Classification proposed by SnpEFF (see SnpEff manual for more details: [http://snpeff.sourceforge.net/SnpEff\\_manual.html](http://snpeff.sourceforge.net/SnpEff_manual.html))

column M: Name of the genes surrounding or affected by the genetic variation

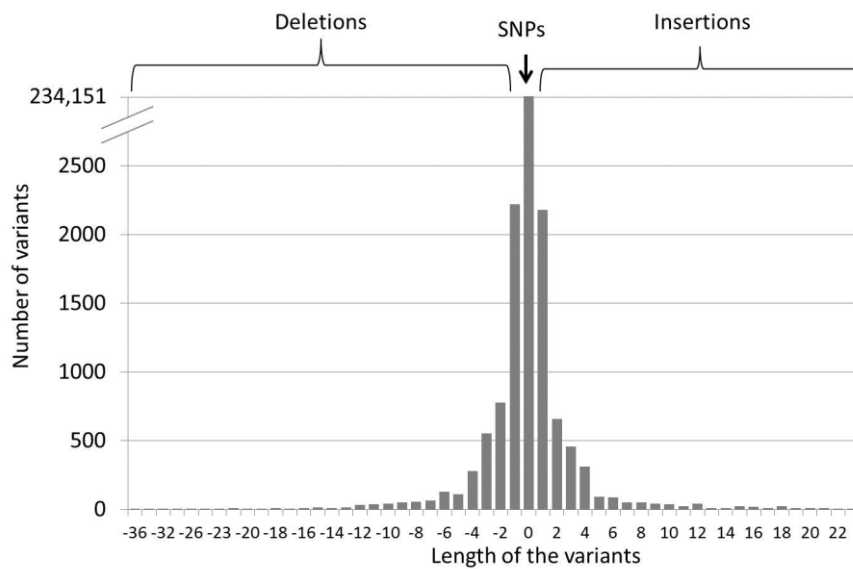
column N: Compartment affected

column O: Nucleotidic change brought by the variation: in the case of intergenic variant, position in the genome is given, in the case of genic variant, position in the gene sequence is given

column P: In the case of exonic variant, the corresponding amino acid after reading the new codon, in addition to it position in the protein sequence.

column Q: allele distribution among French isolates. "Common block of diversity" refer to the genetic variant identified with the reference but common between French isolates.

**Additional file 4: Figure S1:** Distribution of the length of the variants identified in this analysis. Polymorphisms of length equal to zero are SNPs.



**Additional file 5: Table S4** Details of the polymorphic islands delineated during the analysis.

Chromosomes	Start (Mb)	End (Mb)	Number of variants	Length (Mb)	Mean variants /kb
Chrom-I	0	0.6	11,290	0.6	16.1
Chrom-I	5.5	6	5,372	0.5	10.7
Chrom-I	6.9	7.9	11,934	1	11.9
Chrom-I	10.8	11.7	18,950	0.9	21.1
Chrom-II	0	0.9	21,879	0.9	24.3
Chrom-II	3.5	6.6	29,345	3.1	9.5
Chrom-II	8.2	9	14,508	0.8	18.1
Chrom-III	0	0.6	9,251	0.6	15.4
Chrom-III	6.3	7.8	29,355	1.5	19.6
Chrom-IV	0	0.3	4,968	0.3	12.4
Chrom-IV	3.8	4.7	12,730	0.9	14.1
Chrom-IV	7.2	8.1	16,528	0.9	18.4
<b>Overall</b>			186,110	12.0	16.0
% compared to genome total number			76.7%	31.5%	

**Additional file 6: Table S5:** Gene annotation according to the polymorphism detected in the analysis.

*# Here a description of the different columns:*

**Gene ID (RRES v4.0):** gene name according to the reference genome annotation RRES V4.0

**Gene start (bp):** the 5' position in the genome

**Gene end (bp):** the 3' position in the genome

**Length: length of the coding sequence**

**Chromosome**

**Transcript stable ID**

**Gene name**

**Gene description**

**Gene ID (FGSB v3.1):** correspondance of gene name according to theFGDB v3.1 reference genome

**Loss of function variants:** number of variants of which encoding effect is predicted to lead to a loss of function

**Non-synonymous effect (Introns+Exons):** number of variant located in genic sequence and that is predicted to modify amino-acid composition of the protein

**Synonymous variants exons:** number of variant located within exonic sequence and that is predicted to do not modify amino-acid composition of the protein

**Synonymous variants others:** other variant located within genic sequence and that is predicted to do not modify amino-acid composition of the protein

**Variants located within 100 bp upstream and downstream sequences (UD):** number of variants located the 100 bp downstream and upstream of the genic sequence

**Total Including UD sequences:** total number of variants (Loss of Functions + Synonymous exons + Synonymous others + UD sequences).

**Ratio Diversification:** calculated as follow: (number of loss of function variants + number of non-synonymous)/ total variant number

**Accumulation of non-synonymous effect variants: Yes:** total variant number > 4 and ratio of diversification > 0.5

**Category:** category of which the gene has been classified, see Experimental Procedure

**Polymorphic Islands:** gene is located within polymorphic island determined during the analysis (Additional file 5: Table S4)

**InPlanta Expression:** profile of expression according to Harris et al., 2016. "-" Gene was not showed to be expressed in planta conditions, "Plant name" gene was shown to be expressed in this planta condition (crop tested: barley, wheat and maize).

**SecretionPrediction:** gene has been previously predicted to encode a secreted protein (Brown et al. 2012, revised by King et al. 2015)

**SM clustering prediction:** genes has been previously predicted to belong to secondary metabolite cluster (Sieber et al. 2014)

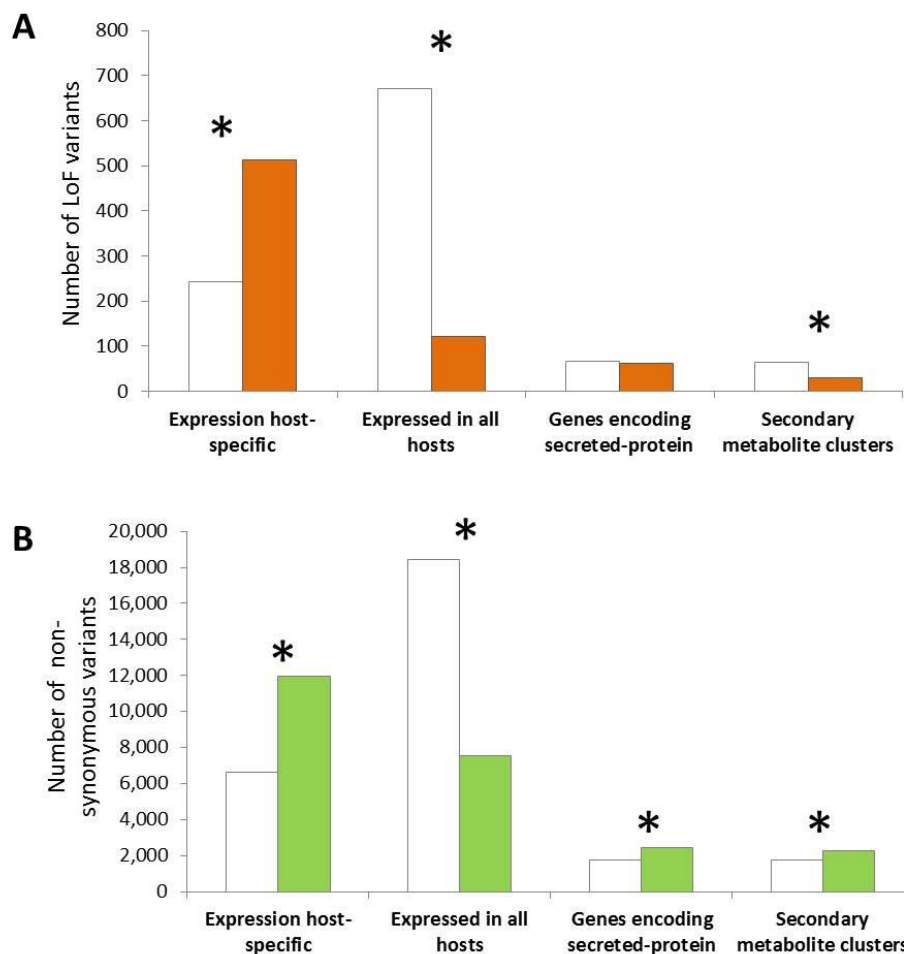
**PHI-base ID:** gene has a hit in the Pathogen Host Interaction database, that regroup functional studies of genes whose role in host-pathogen interaction has been investigated

**Pathogenic phenotype:** in the case of gene with a hit in PHI-base: what phenotype was obtain after deletion, replacement or disruption of the gene

**Plant host:** in the case of gene with a hit in PHI-base: which host has been tested

**Gene Ontology Evidences:** Has the gene been attributed a gene ontology code?

**Additional file 7: Figure S2:** Variant enrichment of genes with candidate functions. Representation of the observed variant number (in color) compared to the theoretical number expected under hypothesis of random distribution of variants (white). A. Variants predicted to lead to a loss of function of the protein. B. Variants predicted to have non-synonymous effects on the protein. “\*” means Chi-squared test was significant ( $p$ -value < 0.001).





**Additional file 8: Table S6:** Summary of known *Tri* genes, implicated in trichothecene of type B (TCTB) production. **Table S7: Not shown in this manuscript:** Annotation of *Tri* genes and surrounding sequences according to the polymorphism detected during this analysis.

Stable_ID	Name	FGDB_ID	Chromosome	Start	Stop	Sens	Translated protein
FGRRES_00071	Tri1	FGSG_00071	1	291,505	293,259	(+)	cytochrome P450 monooxygenase
FGRRES_03532	Tri8	FGSG_03532	2	5,392,461	5,393,798	(-)	trichothecene 3-O-esterase
FGRRES_03533	Tri7	FGSG_03533	2	5,395,009	5,395,867	(+)	related to TRI7 - trichothecene biosynthesis gene cluster
FGRRES_03534_M	Tri3	FGSG_03534	2	5,396,677	5,398,444	(+)	trichothecene 15-O-acetyltransferase
FGRRES_03535	Tri4	FGSG_03535	2	5,399,398	5,401,139	(-)	trichodiene oxygenase [cytochrome P450]
FGRRES_16251	Tri6	FGSG_16251	2	5,402,308	5,402,964		trichothecene biosynthesis positive transcription factor
FGRRES_03537	Tri5	FGSG_03537	2	5,405,852	5,407,038	(+)	trichodiene synthase [sesquiterpene cyclase]
FGRRES_03538	Tri10	FGSG_03538	2	5,407,706	5,409,058	(+)	regulatory protein
FGRRES_03539	Tri9	FGSG_03539	2	5,409,654	5,409,785	(-)	hypothetical protein
FGRRES_03540	Tri11	FGSG_03540	2	5,410,953	5,412,693	(-)	isotrichodermin C-15 hydroxylase
FGRRES_03541	Tri12	FGSG_03541	2	5,413,251	5,415,227	(-)	trichothecene efflux pump
FGRRES_03542_M	Tri13	FGSG_03542	2	5,416,141	5,416,628	(-)	probable cytochrome P450
FGRRES_03543	Tri14	FGSG_03543	2	5,418,308	5,419,477	(+)	putative trichothecene biosynthesis gene
FGRRES_11025	Tri15	FGSG_11025	3	6,422,848	6,423,822	(-)	putative C2H2 zinc finger transcription factor
FGRRES_07896	Tri101	FGSG_07896	4	4,805,149	4,806,504	(-)	trichothecene 3-O-acetyltransferase

## Competing interests

The authors declare that they have no competing interests.

## Authors' contribution

BL carried out the bioinformatical and statistical analyses and wrote the manuscript. BL, MM and CS conducted the wet lab experiments. NP assisted to the data mining and manuscript writing. CB contributed to conceive and design the study. MFO was responsible for overall research management. All authors have read and approved the final manuscript.

## Funding

This study was supported by a grant from the research department Plant Health and Environment of the French National Institute of Agricultural Research. BL received a PhD fellowship from the French Research Ministry.

## Acknowledgements

We would like to acknowledge the Genotoul bioinformatics platform Toulouse Midi-Pyrénées and Sigenae group for providing help along the bioinformatics analysis and for providing computing and storage resources (Anon s.d.). We are grateful to Vessela Atanasova-Pénichon, Pauline Lassere-Zuber and Richard Blanc for their help in strain characterizations.

## References:

1. Trail F. For blighted waves of grain: *Fusarium graminearum* in the postgenomics era. *Plant Physiol.* 2009;149:103–10.
2. Waskiewicz A, Golinski P. Mycotoxins in foods, feeds and their components. *Krmiva.* 2013;55:35–45.
3. Morgavi DP, Riley RT. *Fusarium* and their toxins: Mycology, occurrence, toxicity, control and economic impact. *Anim. Feed Sci. Technol.* 2007;137:199–200.
4. Brown NA, Bass C, Baldwin TK, Chen H, Massot F, Carion PW, et al. Characterisation of the *Fusarium graminearum*-wheat floral interaction. *J Pathog.* 2011/01/01. 2011;2011:626345.
5. Kimura M, Tokai T, Takahashi-Ando N, Ohsato S, Fujimura M. Molecular and genetic studies of *Fusarium* trichothecene biosynthesis: pathways, genes, and evolution. *Biosci. Biotechnol. Biochem.* 2007;71:2105–23.
6. Alexander NJ, McCormick SP, Waalwijk C, van der Lee T, Proctor RH. The genetic basis for 3-ADON and 15-ADON trichothecene chemotypes in *Fusarium*. *Fungal Genet Biol.* 2011;48:485–95.
7. Lee T, Han YK, Kim KH, Yun SH, Lee YW. Tri13 and Tri7 determine deoxynivalenol- and nivalenol-producing chemotypes of *Gibberella zeae*. *Appl Env. Microbiol.* 2002/04/27. 2002;68:2148–54.
8. Cumagun. Genetic mapping of pathogenicity and aggressiveness of *Gibberella zeae* (*Fusarium graminearum*) toward wheat. *Genet. Resist.* 2004;
9. Jin F, Bai G, Zhang D, Dong Y, Ma L, Bockus W, et al. *Fusarium*-damaged kernels and deoxynivalenol in *Fusarium*-infected U.S. winter wheat. *Phytopathology.* 2014;104:472–8.
10. Paul PA, Lipps PE, Madden L V. Relationship between visual estimates of *Fusarium* head blight intensity and deoxynivalenol accumulation in harvested wheat grain: a meta-analysis. *Phytopathology.* 2005;95:1225–36.
11. Talas F, Kalih R, Miedaner T, McDonald BA. Genome-wide association study identifies novel candidate genes for aggressiveness, deoxynivalenol production and azole sensitivity in natural field populations of *Fusarium graminearum*. *Mol. Plant-Microbe Interact. Scientific Societies;* 2016;
12. Terzi V, Tumino G, Stanca AM, Morcia C. Reducing the incidence of cereal head infection and mycotoxins in small grain cereal species. *J. Cereal Sci.* 2014;59:284–93.
13. Chen Y, Zhou M-G. Characterization of *Fusarium graminearum* isolates resistant to both carbendazim and a new fungicide JS399-19. *Phytopathology. The American Phytopathological Society ;* 2009;99:441–6.
14. Talas F, McDonald BA. Significant variation in sensitivity to a DMI fungicide in field populations of *Fusarium graminearum*. *Plant Pathol.* 2015;64:664–70.

15. Ward TJ, Clear RM, Rooney AP, O'Donnell K, Gaba D, Patrick S, et al. An adaptive evolutionary shift in *Fusarium* head blight pathogen populations is driving the rapid spread of more toxigenic *Fusarium graminearum* in North America. *Fungal Genet. Biol.* 2008;45:473–84.
16. Zeller KA, Bowden RL, Leslie JF. Population differentiation and recombination in wheat scab populations of *Gibberella zeae* from the United States. *Mol. Ecol.* 2004;13:563–71.
17. Akinsanmi OA, Backhouse D, Simpfendorfer S, Chakraborty S. Genetic diversity of Australian *Fusarium graminearum* and *F. pseudograminearum*. *Plant Pathol.* Blackwell Publishing Ltd; 2006;55:494–504.
18. Gale LR, Ward TJ, Balmas V, Kistler HC. Population Subdivision of *Fusarium graminearum* sensu stricto in the upper midwestern United States. *Phytopathology.* 2008/10/24. 2007;97:1434–9.
19. Leslie JF, Anderson LL, Bowden RL, Lee Y-W. Inter- and intra-specific genetic variation in *Fusarium*. *Int. J. Food Microbiol.* 2007;119:25–32.
20. Boutigny A-L, Ward T, Ballois N, Iancu G, Ioos R. Diversity of the *Fusarium graminearum* species complex on French cereals. *Eur. J. Plant Pathol.* Springer Netherlands; 2014;138:133–48.
21. Liang JM, Xayamongkhon H, Broz K, Dong Y, McCormick SP, Abramova S, et al. Temporal dynamics and population genetic structure of *Fusarium graminearum* in the upper Midwestern United States. *Fungal Genet Biol.* 2014/10/15. 2014;73:83–92.
22. Van der Lee T, Zhang H, van Diepeningen A, Waalwijk C. Biogeography of *Fusarium graminearum* species complex and chemotypes: a review. *Food Addit Contam Part A Chem Anal Control Expo Risk Assess.* 2014/12/23. 2015;32:453–60.
23. McDonald B, Linde C. The population genetics of plant pathogens and breeding strategies for durable resistance. *Euphytica.* Kluwer Academic Publishers; 2002;124:163–80.
24. Talas F, McDonald B. Genome-wide analysis of *Fusarium graminearum* field populations reveals hotspots of recombination. *BMC Genomics.* 2015;16:996.
25. Yun SH, Arie T, Kaneko I, Yoder OC, Turgeon BG. Molecular organization of mating type loci in heterothallic, homothallic, and asexual *Gibberella/Fusarium* species. *Fungal Genet Biol.* 2000;31:7–20.
26. Keller M, Bergstrom G, Shields E. The aerobiology of *Fusarium graminearum*. *Aerobiologia (Bologna).* Springer Netherlands; 2014;30:123–36.
27. Maldonado-Ramirez SL, Schmale Iii DG, Shields EJ, Bergstrom GC. The relative abundance of viable spores of *Gibberella zeae* in the planetary boundary layer suggests the role of long-distance transport in regional epidemics of *Fusarium* head blight. *Agric. For. Meteorol.* 2005;132:20–7.
28. Prussin AJ, Li Q, Malla R, Ross SD, Schmale DG. Monitoring the long-distance transport of *Fusarium graminearum* from field-scale sources of inoculum. *Plant Dis. Scientific Societies;* 2013;98:504–11.
29. Prussin AJ, Szanyi NA, Welling PI, Ross SD, Schmale DG. Estimating the production and release of ascospores from a field-scale source of *Fusarium graminearum* inoculum. *Plant Dis. Scientific Societies;* 2013;98:497–503.
30. Schmale Iii DG, Leslie JF, Zeller KA, Saleh AA, Shields EJ, Bergstrom GC. Genetic structure of atmospheric populations of *Gibberella zeae*. *Phytopathology. Scientific Societies;* 2006;96:1021–6.
31. Stukenbrock EH, McDonald BA. The origins of plant pathogens in agro-ecosystems. *Annu Rev Phytopathol.* 2008/08/06. 2008;46:75–100.
32. Pombert J-F, Xu J, Smith DR, Heiman D, Young S, Cuomo CA, et al. Complete genome sequences from

- three genetically distinct strains reveal high intraspecies genetic diversity in the microsporidian *Encephalitozoon cuniculi*. Eukaryot. Cell. 2013;12:503–11.
33. Ojeda DI, Dhillon B, Tsui CKM, Hamelin RC. Single-nucleotide polymorphism discovery in *Leptographium longiclavatum*, a mountain pine beetle-associated symbiotic fungus, using whole-genome resequencing. Mol. Ecol. Resour. 2014;14:401–10.
34. Zander M, Patel DA, Van de Wouw A, Lai KT, Lorenc MT, Campbell E, et al. Identifying genetic diversity of avirulence genes in *Leptosphaeria maculans* using whole genome sequencing. Funct Integr Genomics. 2013;13:294–308.
35. Abdolrasouli A, Rhodes J, Beale MA, Hagen F, Rogers TR, Chowdhary A, et al. Genomic context of azole resistance mutations in *Aspergillus fumigatus* determined using whole-genome sequencing. MBio. 1752 N St., N.W., Washington, DC: American Society of Microbiology; 2015;6:e00536-15.
36. Hane JK, Anderson JP, Williams AH, Spersneider J, Singh KB. Genome sequencing and comparative genomics of the broad host-range pathogen *Rhizoctonia solani* AG8. PLoS Genet. 2014/05/09. 2014;10:e1004281.
37. Persoons A, Morin E, Delaruelle C, Payen T, Halkett F, Frey P, et al. Patterns of genomic variation in the poplar rust fungus *Melampsora larici-populina* identify pathogenesis-related factors. Front. Plant Sci. 2014;5.
38. Atwell S, Corwin JA, Soltis NE, Subedy A, Denby KJ, Kliebenstein DJ. Whole genome resequencing of *Botrytis cinerea* isolates identifies high levels of standing diversity. Front Microbiol. 2015/10/07. 2015;6:996.
39. McDonald MC, Williams AH, Milgate A, Pattemore JA, Solomon PS, Hane JK. Next-generation resequencing as a tool for rapid bioinformatic screening of presence and absence of genes and accessory chromosomes across isolates of *Zymoseptoria tritici*. Fungal Genet. Biol. 2015;79:71–5.
40. Fedorova ND, Khaldi N, Joardar VS, Maiti R, Amedeo P, Anderson MJ, et al. Genomic islands in the pathogenic filamentous fungus *Aspergillus fumigatus*. PLoS Genet. Public Library of Science; 2008;4:e1000046.
41. King R, Urban M, Hammond-Kosack M, Hassani-Pak K, Hammond-Kosack K. The completed genome sequence of the pathogenic ascomycete fungus *Fusarium graminearum*. BMC Genomics. 2015;16:544.
42. Cuomo CA, Guldener U, Xu JR, Trail F, Turgeon BG, Di Pietro A, et al. The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. Science (80-. ). 2007;317:1400–2.
43. Wong P, Walter M, Lee W, Mannhaupt G, Münsterkötter M, Mewes H-W, et al. FGDB: revisiting the genome annotation of the plant pathogen *Fusarium graminearum*. Nucleic Acids Res. Oxford University Press; 2011;39:D637–9.
44. Brown NA, Antoniw J, Hammond-Kosack KE. The predicted secretome of the plant pathogenic fungus *Fusarium graminearum*: a refined comparative analysis. PLoS One. 2012;7:e33731.
45. Rampitsch C, Day J, Subramaniam R, Walkowiak S. Comparative secretome analysis of *Fusarium graminearum* and two of its non-pathogenic mutants upon deoxynivalenol induction in vitro. Proteomics. 2013;13:1913–21.
46. Lowe RGT, McCorkelle O, Bleackley M, Collins C, Faou P, Mathivanan S, et al. Extracellular peptidases of the cereal pathogen *Fusarium graminearum*. Front. Plant Sci. 2015;6.
47. Phalip V, Delalande F, Carapito C, Goubet F, Hatsch D, Leize-Wagner E, et al. Diversity of the exoproteome of *Fusarium graminearum* grown on plant cell wall. Curr. Genet. 2005;48:366–79.

48. Sieber CMK, Lee W, Wong P, Münsterkötter M, Mewes H-W, Schmeitzl C, et al. The *Fusarium graminearum* genome reveals more secondary metabolite gene clusters and hints of horizontal gene transfer. PLoS One. Public Library of Science; 2014;9:e110311.
49. Timmons JA, Szkop KJ, Gallagher IJ, Keller P, Vollaard N, Gustafsson T, et al. Multiple sources of bias confound functional enrichment analysis of global -omics data. Genome Biol. 2015 161. BioMed Central; 2015;16:46–59.
50. Harris LJ, Balcerzak M, Johnston A, Schneiderman D, Ouellet T. Host-preferential *Fusarium graminearum* gene expression during infection of wheat, barley, and maize. Fungal Biol. 2016;120:111–23.
51. Bai GH, Desjardins AE, Plattner RD. Deoxynivalenol-nonproducing *Fusarium graminearum* causes initial infection, but does not cause disease spread in wheat spikes. Mycopathologia. 2002/05/10. 2002;153:91–8.
52. Pagani F, Baralle FE. Opinion: Genomic variants in exons and introns: identifying the splicing spoilers. Nat. Rev. Genet. Nature Publishing Group; 2004;5:389–96.
53. Sims D, Sudbery I, Illott NE, Heger A, Ponting CP. Sequencing depth and coverage: key considerations in genomic analyses. Nat Rev Genet. 2014/01/18. 2014;15:121–32.
54. Walkowiak S, Bonner CT, Wang L, Blackwell B, Rowland O, Subramaniam R. Intraspecies interaction of *Fusarium graminearum* contributes to reduced toxin production and virulence. Mol Plant Microbe Interact. 2015/07/01. 2015;28:1256–67.
55. Cantu D, Segovia V, MacLean D, Bayles R, Chen X, Kamoun S, et al. Genome analyses of the wheat yellow (stripe) rust pathogen *Puccinia striiformis* f. sp. *tritici* reveal polymorphic and haustorial expressed secreted proteins as candidate effectors. BMC Genomics. BioMed Central; 2013;14:270.
56. Neafsey DE, Barker BM, Sharpton TJ, Stajich JE, Park DJ, Whiston E, et al. Population genomic sequencing of *Coccidioides* fungi reveals recent hybridization and transposon control. Genome Res. Cold Spring Harbor Laboratory Press; 2010;20:938–46.
57. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, et al. A global reference for human genetic variation. Nature. Nature Research; 2015;526:68–74.
58. Gladieux P, Ropars J, Badouin H, Branca A, Aguilera G, de Vienne DM, et al. Fungal evolutionary genomics provides insight into the mechanisms of adaptive divergence in eukaryotes. Mol. Ecol. 2014;23:753–73.
59. Sperschneider J, Gardiner DM, Thatcher LF, Lyons R, Singh KB, Manners JM, et al. Genome-wide analysis in three *Fusarium* pathogens identifies rapidly evolving chromosomes and genes associated with pathogenicity. Genome Biol. Evol. 2015;
60. Ma L-J, van der Does HC, Borkovich KA, Coleman JJ, Daboussi M-J, Di Pietro A, et al. Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. Nature. Macmillan Publishers Limited. All rights reserved; 2010;464:367–73.
61. Raffaele S, Kamoun S. Genome evolution in filamentous plant pathogens: why bigger can be better. Nat. Rev. Microbiol. Nature Publishing Group; 2012;10:417.
62. Croll D, Lendenmann MH, Stewart E, McDonald BA. The impact of recombination hotspots on genome evolution of a fungal plant pathogen. Genetics. 2015;201:1213–28.
63. Connolly LR, Smith KM, Freitag M. The *Fusarium graminearum* histone H3 K27 methyltransferase KMT6 regulates development and expression of secondary metabolite gene clusters. PLoS Genet. 2013/11/10. 2013;9:e1003916.

64. McVicker G, Gordon D, Davis C, Green P. Widespread genomic signatures of natural selection in hominid evolution. *PLoS Genet. Public Library of Science*; 2009;5:e1000471.
65. Supek F, Miñana B, Valcárcel J, Gabaldón T, Lehner B. Synonymous mutations frequently act as driver mutations in human cancers. *Cell*. 2014;156:1324–35.
66. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*. 2012/06/26. 2012;6:80–92.
67. Dong S, Raffaele S, Kamoun S. The two-speed genomes of filamentous pathogens: waltz with plants. *Curr. Opin. Genet. Dev.* 2015;35:57–65.
68. Raffaele S, Farrer RA, Cano LM, Studholme DJ, MacLean D, Thines M, et al. Genome evolution following host jumps in the Irish potato famine pathogen lineage. *Science. American Association for the Advancement of Science*; 2010;330:1540–3.
69. Croll D, McDonald BA. The accessory genome as a cradle for adaptive evolution in pathogens. *PLoS Pathog. Public Library of Science*; 2012;8:e1002608.
70. Galazka JM, Freitag M. Variability of chromosome structure in pathogenic fungi—of “ends and odds.” *Curr. Opin. Microbiol.* 2014;20:19–26.
71. Talas F, Kalih R, Miedaner T. Within-field variation of *Fusarium graminearum* isolates for aggressiveness and deoxynivalenol production in wheat head blight. *Phytopathology*. 2012;102:128–34.
72. Talas F, Wurschum T, Reif JC, Parzies HK, Miedaner T. Association of single nucleotide polymorphic sites in candidate genes with aggressiveness and deoxynivalenol production in *Fusarium graminearum* causing wheat head blight. *BMC Genet.* 2012;13:14.
73. Menke J, Weber J, Broz K, Kistler HC. Cellular development associated with induced mycotoxin synthesis in the filamentous fungus *Fusarium graminearum*. *PLoS One. Public Library of Science*; 2013;8:e63077.
74. Kistler HC, Broz K. Cellular compartmentalization of secondary metabolism. *Front Microbiol. Frontiers Media S.A.*; 2015;6:68.
75. Pinson-Gadais L, Foulongne-Oriol M, Ponts N, Barreau C, Richard-Forget F. The French *Fusarium* Collection: a living resource for mycotoxin research. 12th Eur. *Fusarium Semin.* 2013;
76. <http://www.mgx.cnrs.fr>. <http://www.mgx.cnrs.fr>.
77. Andrews S. FastQC: a quality control tool for high throughput sequence data. Available online at : <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>. 2010;
78. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics*. 2011;
79. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. 2009;25:1754–60.
80. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009/06/10. 2009;25:2078–9.
81. <http://picard.sourceforge.net>.
82. DePristo MA, Banks E, Poplin R, Garimella K V, Maguire JR, Hartl C, et al. Additional data:A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.*; 2011;43:491–8.
83. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and

VCFtools. Bioinformatics. 2011/06/10. 2011;27:2156–8.

84. [http://snpeff.sourceforge.net/SnpEff\\_manual.html](http://snpeff.sourceforge.net/SnpEff_manual.html).

85. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the Functional Effect of Amino Acid Substitutions and Indels. PLoS One. Public Library of Science; 2012;7:e46688.

86. Stajich JE, Harris T, Brunk BP, Brestelli J, Fischer S, Harb OS, et al. FungiDB: an integrated functional genomics database for fungi. Nucleic Acids Res. Oxford University Press; 2012;40:D675–81.

87. <http://www6.inra.fr/cirm>.

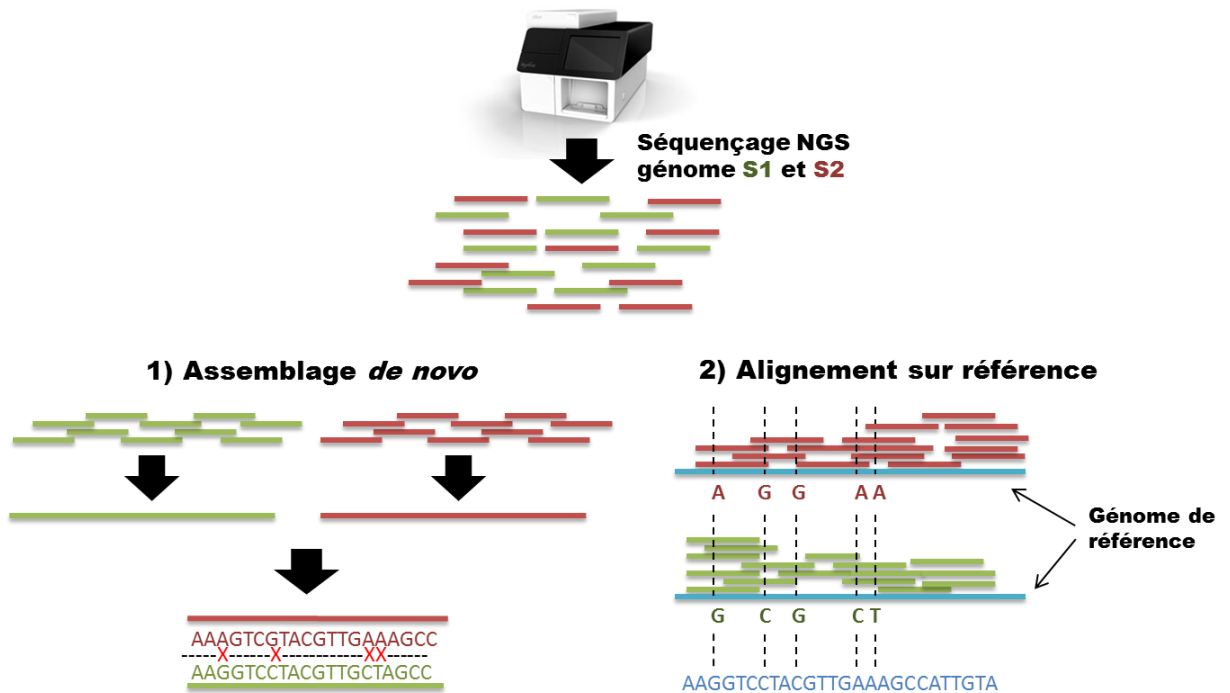
88. <http://bioinfo.genotoul.fr/>.

### **III. Le séquençage des génomes : stratégies et enjeux.**

L'accès à l'information des génomes révolutionne les sciences biologiques contemporaines. Il permet de raccorder de nombreuses disciplines et aspects variés de la biologie : de la génétique fonctionnelle à la génétique des populations, de la structure des chromosomes à l'évolution des espèces. Cette information permet notamment de caractériser de manière complète la diversité génétique des individus d'une même espèce ; cette diversité génétique permet aux espèces d'évoluer dans leurs environnements malgré des contraintes biotiques et abiotiques changeantes.

Deux stratégies majeures existent pour caractériser le polymorphisme d'un génome à partir des petites séquences (*i.e.* reads) produites par les séquenceurs de nouvelles générations (Figure CI\_5) (Sims et al. 2014). La première approche consiste à assembler ces reads pour reconstruire la séquence complète des chromosomes. Les séquences chromosomiques sont ensuite comparées. La deuxième approche consiste à aligner et comparer les reads directement sur une séquence génomique de référence, si celui-ci est disponible. *F. graminearum* dispose d'un génome de référence, mais comme discuté en introduction, plusieurs versions sont disponibles.





**Figure CI\_5 :** les deux principales stratégies pour identifier des variations génétiques à partir des séquences produites par les technologies NGS (de type Illumina sur cet exemple).

Plusieurs points peuvent justifier des préférences d'utilisation entre ces deux stratégies : l'alignement sur un génome de référence est moins performant pour l'identification des variants génétiques de grandes tailles (transfert horizontal, variation du nombre de copies, grands InDels, modifications structurales...). En effet, la taille de ces modifications génétiques diminue la qualité d'alignement des reads sur le génome et écarte souvent les variants associés de l'analyse (Sims et al. 2014). A l'inverse, l'alignement des reads sur le génome de référence apporte la notion de profondeur d'alignement – non disponible dans l'autre cas – permettant de discriminer facilement les erreurs de séquençage de vraies variations génétiques (Sims et al. 2014). De fait, ce type d'analyse est plus fiable pour l'analyse de variants de petites tailles (SNPs, petit InDels) et moins sensible aux erreurs de séquençage.

Un des objectifs de ces travaux était de caractériser le polymorphisme global du génome de nos souches, pour optimiser ultérieurement le processus de génotypage. Nos attentes étaient donc de 1) caractériser le maximum de variants génétiques et 2) d'assurer un seuil de confiance élevé à ces variants. En cohérence avec nos objectifs, et

puisque les variants de petites tailles composent généralement la majorité (en nombre) des modifications génétiques d'un génome (Sims et al. 2014), l'alignement sur le génome de référence nous a semblé être la stratégie la plus adaptée. Le génome de référence initialement utilisé fut celui proposé par le Broad Institute. Cette référence est aujourd'hui difficilement accessible à la communauté depuis la fermeture de la base de données dédiée à *F. graminearum* du Broad Institute. Ce génome a été révisé durant la thèse par King et al. (2015). Ce point illustre de l'un des risques de cette stratégie: les résultats dépendent fortement de la version du génome de référence, et l'analyse peut être remise en question si la version du génome de référence évolue. Plus globalement, les différentes versions de la référence du génome de *F. graminearum* peuvent poser des problèmes pour la comparaison/ l'interprétation/ la réutilisation de certains résultats. Des initiatives, visant à proposer une méthodologie harmonieuse et synergique de travail au sein de la communauté scientifique ont été récemment évoquées (Corby Kistler, *Fusarium* workshop, 2015 Fungal Genetics Conference).

Dans notre cas, l'analyse a été reconduite en utilisant cette nouvelle version (King et al. 2015). Les scripts développés lors de la première analyse ont pu être adaptés et réutilisés. Une synthèse des scripts écrits pour l'utilisation des différents logiciels bio-informatiques durant ces travaux de thèse est disponible en Annexe II.

Plusieurs logiciels sont disponibles afin de détecter des variations génétiques à partir de l'alignement de reads sur un génome de référence. Le choix du logiciel n'est pas anodin : le nombre et la nature des variants génétiques d'un même échantillon, analysés par des logiciels différents, peuvent varier de manières significatives (Liu et al. 2013). Dans notre cas, le logiciel GATK (Genome analysis toolkit) a été préféré. Ce logiciel présente l'avantage de pouvoir analyser des génomes haploïdes et d'intégrer l'information des différents génomes simultanément. Ainsi, le logiciel réévalue la présence d'un variant à une position en prenant en compte l'existence de ce variant à la même position chez d'autres souches (Liu et al. 2013; McKenna et al. 2010). D'autres avantages comme la gratuité d'utilisation, la facilité d'apprentissage, ainsi que l'interactivité et la richesse des forums associés, ont également orienté notre choix.

## IV. Conclusion du chapitre 1

L'objectif de ce travail était de caractériser la diversité génétique et phénotypique de souches issues de la collection du laboratoire. Cette caractérisation était motivée par l'identification de parents potentiels pour la création d'une population recombinante. Six souches ont été présélectionnées comme présentant un niveau intéressant de diversité phénotypique. Ces six souches ont été caractérisées pour leurs aptitudes à provoquer des symptômes sur blé, leurs potentiels toxigènes sur blé et *in vitro*. Ces travaux ont aussi permis d'adapter et de conduire en routine des techniques de cultures pour la production de périthèces par *F. graminearum* au laboratoire, étape indispensable pour croiser les souches entre elles. Leur aptitude à produire des périthèces en condition de laboratoire a ainsi été testée. La diversité génomique de ces six souches a ensuite été caractérisée, révélant un nombre remarquable de variations génétiques entre les souches. Ce deuxième point est encourageant pour la deuxième tâche de la thèse, qui est de construire la première carte génétique à haute résolution de *F. graminearum*. En effet, ce nombre important de variants génétiques laisse penser que l'identification de marqueurs moléculaires ne devrait pas être un facteur limitant. Cette étude a de plus fournis des informations plus globales sur l'organisation du polymorphisme dans le génome de *F. graminearum*. L'organisation des fonctions biologiques en fonction du polymorphisme suggère un rôle pour l'adaptation du pathogène dans son environnement.



**Chapitre 2 :**  
**La recombinaison méiotique chez**  
***F. graminearum* : un mécanisme**  
**évolutif majeur, à la base de la**  
**cartographie génétique.**

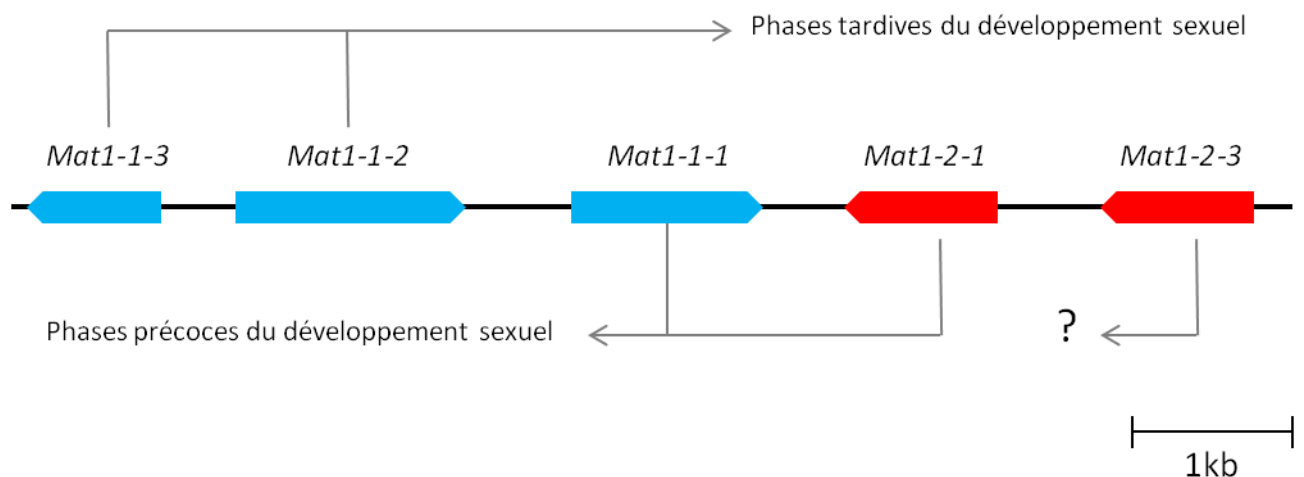


# I. Construction de la population recombinante

*F. graminearum* est une espèce homothallique. Les organes fructifères peuvent-être initiés par la fusion d'hyphes issus du même individu (« autofécondation », Yun et al. 2000). En effet, les différents idiomorphes des gènes *Mat* sont localisés dans le même noyau et permettent une auto-compatibilité (Yun et al. 2000). Ce mode de reproduction complique la possibilité de faire des croisements au laboratoire.

## I.1. *Le contrôle de la reproduction sexuée chez F. graminearum*

Le système moléculaire sous-jacent à l'homothallisme est contrôlé chez *F. graminearum* par les gènes de type *Mat*<sub>1-1</sub> (*Mat*<sub>1-1-1</sub>, *Mat*<sub>1-1-2</sub> et *Mat*<sub>1-1-3</sub>) et les gènes de type *Mat*<sub>1-2</sub> (*Mat*<sub>1-2-1</sub> et *Mat*<sub>1-2-3</sub>), codant les idiomorphes sexuels « compatibles » (Figure CII\_1). Une même cellule dispose des différents idiomorphes sexuels, dont les gènes sont groupés en cluster sur le même chromosome (Lee et al. 2003; Kim et al. 2012). Tous les gènes *Mat* semblent être exprimés spécifiquement durant les phases sexuées (Kim et al. 2012). Cependant, le rôle de *Mat*<sub>1-2-3</sub> reste inconnu puisque le mutant pour ce gène est capable de se reproduire par autofécondation (Kim et al. 2012). Les gènes *Mat*<sub>1-1-2</sub> et *Mat*<sub>1-1-3</sub> semblent eux, agir lors de la maturation des périthèces : les souches mutantes pour ces gènes forment des structures sexuées mais sont incapables de former les ascospores (Kim et al. 2012). La souche mutante pour le gène *Mat*<sub>1-2-1</sub> ne présente aucune structure sexuée (Kim et al. 2012) alors que la souche mutante pour le gène *Mat*<sub>1-1-1</sub> présente des périthèces anormalement petits. Ces gènes semblent donc être impliqués plus tôt, comme par exemple lors de la reconnaissance nucléaire, la formation d'hyphes ascogènes ou la fusion nucléaire (Kim et al. 2012).



**Figure CII\_1 :** base génétique de l'homothallisme chez *F. graminearum* et rôle supposé des différents gènes *Mat*. Les gènes de type *Mat 1-1* sont en bleu et les gènes de type *Mat 1-2* sont en rouge (d'après Kim et al. 2012).

La réalisation de croisement contrôlé de souches de *F. graminearum* en condition de laboratoire n'est pas simple (Bowden & Leslie 1999; Lee et al. 2003). Ce phénomène contraste pourtant les observations faites au champ où la diminution rapide des déséquilibres de liaisons observée au sein des population suggèrent des fréquences importantes de croisements (Zeller et al. 2004; Talas & McDonald 2015). *F. graminearum* n'est d'ailleurs pas une exception puisque un niveau fréquent de croisements a été supposé chez d'autres espèces fongiques homothalliques comme *Sclerotinia sclerotiorum*, un autre pathogène de plantes (Attanayake et al. 2014).

Chez *F. graminearum*, peu d'informations sont disponibles sur la fréquence des événements de croisements. Est-ce que les individus se croisent à une fréquence trop faible pour être perçue au laboratoire mais suffisante pour être un moteur de diversité dans la nature *via* une forte production d'inoculum et des cycles de reproduction courts ? Est-ce que des facteurs externes sont perçus par le champignon et favorisent le croisement avec un partenaire ? Ces éléments restent sans réponse mais leur étude pourrait fournir de nouvelles connaissances sur les modes de reproduction fongique ainsi que de nouvelles pistes pour le contrôle de la maladie au champ.

Pour pallier la difficulté rencontrée pour croiser les souches de *F. graminearum*, Bowden et Leslie (1999) ont proposé de croiser des souches avec des mutants



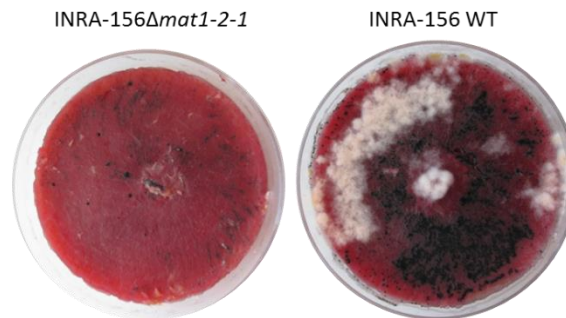
présentant des déficiences d'utilisation du nitrate. Chaque parent présente un phénotype « muté » caractéristique. Les individus présentant un phénotype « sauvage », possible par complémentation des deux allèles mutantes, sont issus d'une fécondation croisée (Bowden & Leslie 1999). Avec cette technique, les souches recombinantes sont identifiées à une fréquence de ~35% (Bowden & Leslie 1999). Cette stratégie expérimentale a été utilisée pour la construction des deux cartes génétiques précédemment proposées et impliquant des souches de *F. graminearum* (Gale et al. 2005; Jurgenson et al. 2002). Du fait de la sélection systématique de souches présentant le phénotype recombinant, cette technique introduit des biais de ségrégation dans la descendance.

Depuis, de nouvelles techniques ont été proposées pour contourner l'homothallisme et « forcer » le croisement entre souche (Lee et al. 2003). Cette stratégie repose sur la délétion d'un des idiomorphes des gènes *Mat*, permettant de rapprocher la structuration moléculaire des souches de *F. graminearum* de celle d'individus hétérothalliques (Lee et al. 2003). Si cette stratégie semble pertinente pour l'obtention de populations recombinantes, elle n'a jamais été employée pour la construction de carte génétique.

### **I.2. Construction de souches hétérothalliques**

Durant ces travaux, des souches ont été génétiquement modifiées pour les gènes du mating type (Lee et al. 2003; Kim et al. 2012). Compte tenu de la structure du locus *Mat* chez *F. graminearum*, plusieurs gènes peuvent être ciblés pour être délétés. Nous avons préféré la construction de souches  $\Delta Mat_{1-2-1}$ , qui présentent l'avantage de ne produire aucun périthèce si elles ne sont pas croisées avec une souche présentant l'allèle *Mat*  $1-2-1$  fonctionnel (Kim et al. 2012). Trois des six souches présentées en Chapitre I ont été transformées avec succès : INRA-156, INRA-171 et INRA-195. Ces souches sont dénommées par la suite INRA-156 $\Delta Mat$ , INRA-171 $\Delta Mat$  et INRA-195 $\Delta Mat$  respectivement. Dans nos conditions, les souches  $\Delta Mat$  générées produisent tout de même des périthèces mais immatures et réduits en taille (Photo CII\_1). Ce dernier point facilite donc la sélection des souches recombinantes : la formation de périthèces matures sur les cultures en milieu solide de la souche  $\Delta Mat$ , et précédemment

fécondée par une suspension conidiale de souche sauvage, résulte en théorie d'une fécondation croisée (Cavinder et al. 2012). Les souches  $\Delta Mat$  se comportent donc comme une souche hétérothallique.



**Photo CII\_1 :** A gauche, INRA-156 $\Delta mat$  et à droite INRA-156 WT, cultivés dans les mêmes conditions sur milieu Carot Agar induisant le développement de périthèces. Notons que la souche INRA-156 $\Delta mat$  est capable d'initier la formation de périthèces qui restent immatures (petites structures noires), alors que chez la souche INRA-156 sauvage les périthèces sont plus gros, matures et fonctionnels.

### 1.3. *Les tentatives de croisement*

Le séquençage des génomes des six souches de la collection démontre que la diversité génétique n'est pas un facteur limitant dans notre cas. Les différentes souches utilisées présentent au minimum 35 153 variations génétiques (~1 variant génétique/kb). Le choix des parents potentiels a donc été guidé par les données de phénotypage de production de TCTB *in vitro* (Tableau CII\_1).

**Tableau CII\_ 1 :** les différentes combinaisons testées pour la construction de la population recombinante.

	INRA-156	INRA-159	INRA-164	INRA-171	INRA-181	INRA-195
INRA-156 $\Delta Mat$			X	X		X
INRA-195 $\Delta Mat$	X	X			X	

Après plusieurs tentatives, seul le croisement réalisé entre la souche INRA-156 $\Delta$ Mat et la souche INRA-171 a permis l'isolement d'une population recombinante. Environ 150 souches ont été isolées à partir de ce croisement. Plus tardivement dans la thèse, le croisement entre la souche INRA-156 $\Delta$ Mat et INRA-195 a été retenté et a abouti à l'obtention d'une nouvelle population recombinante. Pour ce croisement toutefois, les périthèces n'atteignaient pas la maturité. Pour obtenir la population recombinante, la solution d'ascospores recombinantes a été obtenue après avoir broyé manuellement le périthèce dans une solution d'eau stérile.

Enfin, ces résultats démontrent que l'utilisation de souche  $\Delta$ Mat ne garantit pas l'obtention systématique d'une descendance recombinante, et que le contrôle de la reproduction sexuée ne se limite probablement pas qu'au simple gène *Mat*. Des systèmes d'incompatibilité/compatibilité végétative, contrôlant la fusion d'hyphes entre individus différents, ont déjà été observés pour *F. graminearum* (Bowden & Leslie 1992; Ramirez et al. 2006; Gilbert et al. 2002). Ces travaux décrivent d'une grande variabilité de groupes de compatibilités végétatives. Par exemple, Ramirez et al. (2006) ont montré que pour 70 souches de *F. graminearum* isolées en Argentine, 55 groupes de compatibilité végétative différents pouvaient être distingués. Ces mécanismes de compatibilité végétative pourraient donc aussi jouer un rôle important dans la réussite ou non des croisements.

### **I.4. Matériels et méthodes**

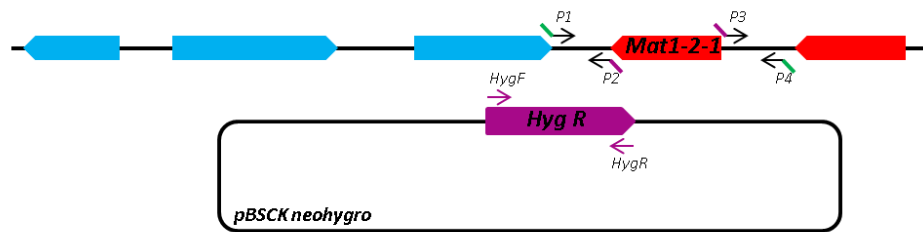
*Sont décrits ici le matériel et méthode utilisés dans les paragraphes I.2 et I.3*

**Construction des souches  $\Delta$ Mat :** la copie du gène *Mat*<sub>1-2-1</sub> (FGRRES\_08893) a été remplacée par la cassette de résistance à l'hygromycine en suivant la méthode du « split marker » (Goswami 2012). La stratégie expérimentale est présentée en Figure CII\_2. Premièrement, les régions externes correspondant à 576 pb en amont du codon Start ont été amplifiées en utilisant la KAPA2G Robuste<sup>®</sup> DNA polymérase (Kapa Biosystems) (95°C 3min, 35 x [95°C 30s, 64°C 15s, 72°C 30-60s /kb], 72°C 45s) en utilisant le couple d'amorces P1/P2 (Tableau CII\_2) à partir d'ADNg de la souche INRA-812 ainsi qu'à 988 pb en aval du codon stop en utilisant la GoTaq<sup>®</sup> DNA polymérase

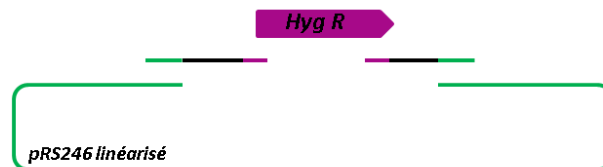
(Promega) et les couples d'amorces P3/P4 (95°C 5min, 35 x [95°C 30s, 64°C 30s, 72°C 1min /kb], 72°C - 2 min). La cassette de résistance à l'hygromycine (2.1 kb) contenant le gène *hph* sous le contrôle du promoteur CPC<sub>1</sub> de *Neurospora crassa* et le terminateur *trpC* d'*Aspergillus nidulans* a été amplifiée en utilisant le couple d'amorces HygF/HygR à partir du plasmide pBCSK neohygro, (Tableau CII\_2) avec la KAPA HiFi DNA polymérase (KapaBiosystems) (95°C 5min, 35 x [98°C 20s, 64°C 15s, 72°C 60s /kb], 72°C 5min). Pour chaque enzyme, les mélanges réactionnels utilisés sont ceux recommandés par les fournisseurs. Les produits PCR ont été purifiés en utilisant le kit Wizard PCR clean-up System® (Promega) en suivant les instructions du fournisseur. Ces trois produits de PCR ont été assemblés par transformation chez la levure (souche FY1679) par recombinaison homologue des régions chevauchantes en suivant le protocole décrit par Colot *et al.*, 2006 en utilisant le plasmide répliatif linéarisé pRS426. Les fragments 5'UTR-Hygro5' et Hygro3'-3'UTR, dont les séquences se chevauchent sur 466pb, ont été amplifiés à partir de l'ADN des levures transformées en utilisant la KAPA Hifi DNA® polymérase (Kapa Biosystems) dans les mêmes conditions que décrit précédemment sauf pour la température d'hybridation qui était de 65°C (Tableau CII\_2). Ces deux amplicons ont été utilisés pour la transformation chimique (PEG et Ca<sup>2+</sup>) de protoplastes de *F. graminearum* en suivant le protocole décrit par (Merhej *et al.* 2011b).

L'intégration du gène de résistance à l'hygromycine au locus du gène *Mat1-2-1* a été validée par PCR, séquençage, et Southern blot (résultats non présentés).

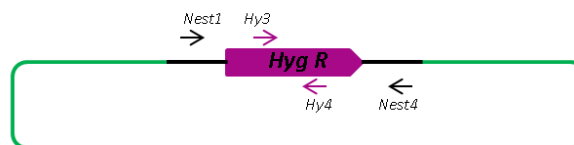
## 1. Amplification chez *F. graminearum* et chez le plasmide



## 2. Recombinaison homologue chez la levure



## 3. Amplification chez la levure



## 4. Recombinaison homologue chez *F. graminearum*



Figure CII\_2 : Stratégie expérimentale pour la délétion du gène *Mat 1-2-1*

Tableau CII\_2 : Les amorces utilisées pour la délétion du gène *Mat 1-2-1*.

Nom de l'amorce	Sequence 5'-3'	Tm (°C)
P1	GTAACGCCAGGGTTTTCCAGTCACGACGTCGTTCCAGAACATGCAGTC	75.0
P2	GCGTTTCGGGTTTACCTCTTCTAGTGGCCGATAATCTCCTCGAC	71.0
P3	GGACACCGGAGACAGGGCGGCCCTGCTGGCGTTCTGAGAGTT	79.0
P4	GCGGATAACAATTTACACAGGAAACAGCCAAGCGCAGACTCAGGTACA	73.0
Nest1	GACTTGGGCAAAGACACGTA	62.7
Nest4	GGCATAGAGTCGTCCCAGAA	64.2
HY3	GGATGCCTCCGCTCGAAGTA	68.6
HY4	CGTTGCAAGACCTGCCTGAA	68.7
HygF	CTAGAAGAGGTAACCCGAAACG	64.0
HygR	GGCCGCCCTGTCTCCGGTGTCCCT	64.0

**Croisement et isolement de la descendance** : les procédures de croisement des souches sont décrites dans le prochain paragraphe (suite). Chaque souche recombinante a été isolée 2 fois successivement à partir d'une spore en cours de germination.

### ***1.5. Une carte génétique : un outil unique pour des applications multiples***

Outre l'utilisation pour la cartographie de QTL, la construction d'une carte génétique permet aussi, une fois alignée sur le génome, de localiser les événements de « crossovers » sur les chromosomes (Foulongne-Oriol 2012). Ce genre d'étude apporte des informations importantes sur l'aptitude de recombinaison d'une espèce. Il a été montré chez de nombreux organismes que le taux de recombinaison était variable le long des chromosomes et entre chromosomes et pouvaient avoir des répercussions importantes sur l'organisation du polymorphisme sur le génome (Spencer et al. 2006). Dans le chapitre précédent, nous avons mis en lumière une organisation spécifique, et très conservée, du polymorphisme sur le génome. Ce phénomène suggère des mécanismes de régulation plutôt qu'un simple hasard. Nous avons donc profité de la création d'une carte génétique pour étudier la recombinaison méiotique chez *F. graminearum*, son organisation dans le génome vis-à-vis d'autres attributs génomiques, et plus particulièrement vis-à-vis du polymorphisme.

# II. La cartographie à haute résolution des profils de recombinaison chez le phytopathogène *F. graminearum* apporte de nouvelles preuves sur l'organisation à deux-vitesses de son génome

Ce sous-chapitre a fait l'objet d'un article scientifique accepté pour publication après corrections mineures dans *Molecular Plant Pathology*. Il est présenté ci-dessous tel qu'il a été soumis initialement.

## High resolution mapping of the recombination landscape of the phytopathogen *Fusarium graminearum* suggests two-speed genome evolution

Benoit Laurent<sup>1</sup>, Christos Palaiokostas<sup>2</sup>, Cathy Spataro<sup>1</sup>, Magalie Moinard<sup>1</sup>, Enric Zehraoui<sup>1</sup>, Ross D. Houston<sup>2</sup> and Marie Foulongne-Oriol<sup>1\*</sup>

<sup>1</sup> MycSA, INRA, Université de Bordeaux, 33882 Villenave d'Ornon, France

<sup>2</sup> The Roslin Institute, University of Edinburgh, Midlothian EH25 9RG, Scotland, UK

\*Corresponding author

Mailing address: INRA, UR1264 Mycologie et Sécurité des Aliments, bâtiment Qualis, 71 avenue Edouard Bourlaux, CS 20032 F-33882 Villenave d'Ornon cedex, France

Email : mfoulong@bordeaux.inra.fr

**Keywords:** *Gibberella zeae*, dense genetic linkage map, Genotyping-by-Sequencing, RAD-seq, meiosis, crossover rate

### Summary

Recombination is a major evolutionary force, increasing genetic diversity and permitting efficient coevolution of fungal pathogen(s) with their host(s). The ascomycete *Fusarium graminearum* is a devastating pathogen of cereal crops, and can contaminate food and feed with harmful mycotoxins. Previous studies suggest a high-adaptive potential of this pathogen, illustrated by an increase of pathogenicity and resistance to fungicides. In this study, we aim to provide a detailed picture of the crossovers events occurring during meiosis and discuss the role of recombination may play for pathogen evolution. An experimental recombinant population (n=88) was created and genotyped using 1,306 polymorphic markers obtained from Restriction site Associated DNA sequencing (RAD-seq) and aligned to the reference genome. The construction of a high-density linkage map, anchoring 99% of the total length of the reference genome, allowed the identification of 1,451 putative crossovers, positioned at a median resolution of 24 kilobases. The majority of crossovers (87.2 %) occurred in a relatively small portion of the genome (30 %). All chromosomes demonstrated recombination-active sections which had a near 20-fold higher crossover rate than non-active recombinant ones. Recombination rate had a strong positive correlation with nucleotide diversity, and recombinant active regions were enriched for genes with a putative role in host-pathogen interaction, as well as putative diversifying genes. Our results are in line with preliminary analysis observed in other *F. graminearum* strains and suggest a conserved 'two-speed' recombination landscape.



### **Introduction**

*Fusarium graminearum sensu stricto* is one of the main causal agent of *Fusarium* Head Blight (FHB), a disease of cereal crops that constitutes a limiting factor for global production (Trail, 2009). The most damaging aspect of FHB infection is contamination of grains with stable and harmful mycotoxins such as deoxynivalenol (DON), whose presence in food and feed are widely regulated in the world (Waskiewicz and Golinski, 2013). Therefore, research targeting improved understanding of this pathogen and its control are a critical component of food security.

Despite its homothallic reproduction, field isolates of *F. graminearum* have been shown to outcross at high frequency (Chen and Zhou, 2009; Liang *et al.*, 2014; Talas and McDonald, 2015). Fungal pathogens that exhibit such features can evolve rapidly, thereby endangering the efficiency and the durability of control strategies (McDonald and Linde, 2002; Pariaud *et al.*, 2009). Concerning *F. graminearum*, a shift towards more aggressive endogenic populations has been reported (Ward *et al.*, 2008), as well as the emergence of isolates resistant to fungicide (Chen and Zhou, 2009; Talas and McDonald, 2015). Furthermore, experimental studies have demonstrated evolution of more aggressive strains following sexual outcrossing (Cumagun and Miedaner, 2004), illustrating the role of sexual reproduction and recombination for evolution. A better characterization of the recombination landscape of *F. graminearum* is therefore needed to understand pathogen evolution, and to further adapt FHB management and control strategies.

*F. graminearum* is haploid for most of its life cycle, but meiosis can occur in the fruiting body (perithecium) formed after fusion of two haploid hyphae (Trail, 2009). Meiosis is a highly-conserved mechanism in eukaryotic genomes, of which double-strand breaks are formed between homologous chromosomes, ensuring their proper transmission and leading to either crossover events or non-crossover events (Mezard *et al.*, 2015). Both events can have direct mutagenic effects due to gene conversion (Mezard *et al.*, 2015). However, only crossovers will break parental haplotypes, limiting hitchhiking of large chromosomal segments during selection (Charlesworth and Campos, 2014). Crossover rate has been shown to be highly variable along eukaryotic genomes (Bensasson, 2011; Bhakta *et al.*, 2015; Brachet *et al.*, 2012; Croll *et al.*, 2015; Mezard, 2006; Petes, 2001; Sonnenberg *et al.*, 2016; Tsai *et al.*, 2016; Yelina *et al.*, 2015). Furthermore, recent insight from several fungal plant pathogen genomes revealed a bipartite distribution of recombination activity, diversity, and

gene function, often referred to as the “two-speed” genome dogma (Croll and McDonald, 2012; Dong *et al.*, 2015; Raffaele *et al.*, 2010).

In *F. graminearum*, the first draft of the genome assembly revealed a relatively condensed genome (36 Mb), of which 13,718 genes were annotated using automated approach (Cuomo *et al.*, 2007; Wong *et al.*, 2011) while more specific analysis suggested a significant number of effectors as secreted proteins or secondary metabolites (Brown *et al.*, 2012; Sieber *et al.*, 2014). Alignment of the reference genome with a linkage map suggested an increase of recombination rate within telomeric / subtelomeric regions, in addition to interstitial regions (Cuomo *et al.*, 2007; Gale *et al.*, 2005). However, due to the fact that both the linkage map and the genome assembly were uncompleted, the interpretation of these results is limited. Strikingly, the genes supposed to be implicated in host-pathogen interactions were enriched in recombinant regions (Cuomo *et al.*, 2007), as well as an increase in polymorphism frequency (Cuomo *et al.*, 2007). Despite the recent revision and improvement of the reference genome assembly (38 Mb, 14,164 protein-coding genes, King *et al.*, 2015), the linkage maps currently available lack the required resolution to provide better insight into the patterns of recombination across the genome. Additionally, questions arise also concerning the organization of polymorphism in different genetic backgrounds, and the role of recombination in the distribution of polymorphisms. Recently, sequencing-based genotyping strategies have revolutionized the process of obtaining genome-wide marker data (Davey *et al.*, 2011). Of these techniques, restriction site associated DNA (RAD) sequencing has been widely employed for generation of high density linkage maps for several species (e.g. Baird *et al.*, 2008; Davey *et al.*, 2013; Gonen *et al.*, 2014; Lendenmann *et al.*, 2014; Palaiokostas *et al.*, 2013).

In the current study, the primary aim was to assess the meiotic recombination landscape in *F. graminearum*, and to understand the role recombination may play for this pathogen for evolution. To achieve this, we aimed to develop an accurate, high resolution crossover landscape of *F. graminearum* using a linkage map constructed using RAD sequencing data from a progeny of 88 strains, and to align this map with the latest version of the reference genome. Altogether, our observations provide strong evidences of “two-speed” genome architecture in *F. graminearum*, and this will be discussed in the light of *F. graminearum* evolution.

### Results

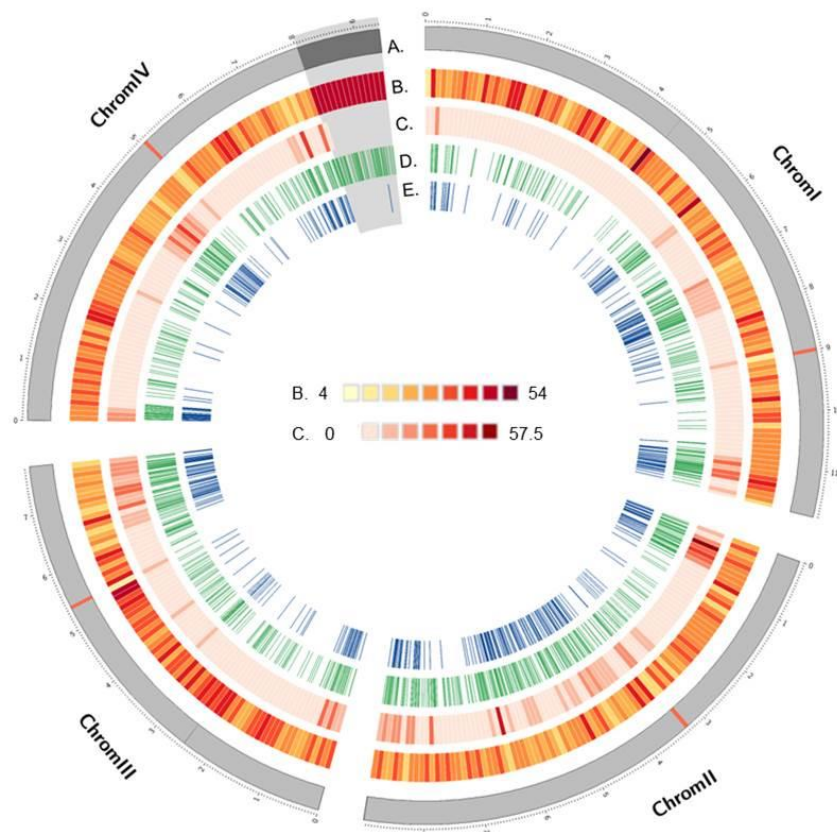
#### Strains and genotyping

The parental strains used for this analysis, namely INRA-156 and INRA-171, were isolated from wheat in France. We resequenced their genomes to help with downstream analysis, and to enable an accurate reference-based RAD sequencing approach to genotyping. In total, 63,486 single nucleotide polymorphisms (SNPs) were discovered. Using parental genomes, we simulated *Pst*I enzyme digestion of the INRA-156 and INRA-171 genomes, which predicted 12,610 and 12,634 fragments with a mean size of 3,016 base pairs (bp) and 3,010 bp respectively. Cutting site distribution for *Pst*I was predicted to be random (KS test,  $p$ -value > 0.01) avoiding introducing subsequent bias (Figure 1, track B), suggesting that the use of *Pst*I was appropriate.

Digestion of genomic DNA from the 88 recombinant progeny and from the two parental strains using the *Pst*I enzyme was followed by restriction site associated DNA (RAD-seq) library preparation and Illumina paired-end sequencing. A total of 401,726,418 paired end reads of 125 bp were produced, resulting in a total of 50.2 Gbp of sequence data. Filtering and demultiplexing resulted in a variable number of reads assigned per strain (ranging from 308,147 – 10,818,740), the majority of which were successfully aligned with the reference genome (97%). A catalogue of sequenced loci was constructed, recovering 17% of the total length of the reference genome. Strain specific sequences matching to the catalogue resulted in the identification of 1,866 polymorphic loci of which 1,306 passed the quality control requirements (Figure 1, track D). Segregation profiles of 31 randomly chosen RAD loci SNPs were matched to corresponding profiles obtained from cleaved amplified polymorphic sequences (CAPS) assays. Only 15 mismatches were observed out of the 2,714 combinations with known genotypes, giving an overall correspondence of 99.4% (data not shown).

The dataset of markers used for downstream analysis was composed of 1,306 RAD markers, 21 markers genotyped during the validation of recombinant strain isolation, together with three additional markers named KSNP100, KSNP101 and KSNP102 (Supplemental file 1). In order to identify possible bias introduced during genotyping and prior to linkage map construction, the distribution of markers according to the reference genome position was estimated. Median physical distance between markers was 12.9 kilobases (kb) ranging from 11.6 for Chrom II to 15.1 for Chrom IV (Table 1). Distribution of markers along

chromosomes was not random (Figure 1 track D,  $p$ -value  $< 0.01$ ), with maximal distance between pairs of markers of 498.8kb (chromosome I at the position 4.0Mb to 4.5Mb). Marker density was highest in subtelomeric and interstitial regions within chromosomes (Figure 1, track D), and correlated with the density of SNPs detected between parental genomes (Figure 1, track C,  $Rho=0.76$ ,  $p$ -value $<2.2E-16$ ). An exception to this pattern was observed at the end of chromosome IV (~1.3 megabases), where highly repetitive rRNA-encoding DNA content resulted in difficulty in read alignment and subsequent variant calling. Due to the repetitive nature of the DNA in this region, there were very few informative polymorphic markers, despite an apparent abundance of *Pst*I restriction sites (Figure 1). For the remainder of the genome, the pattern of polymorphic marker distribution along the chromosomes is likely to reflect polymorphism rate rather than restriction site frequency, due to the random distribution of *Pst*I sites.



**Figure 1:** Circos plot of the distribution of several genomic attributes and genotyping features along the four chromosomes of *F. graminearum*: A. Representation (in Mb) of the four chromosomes of *F. graminearum*, with the positions of the predicted centromeres in red. B) Predicted *Pst*I cutting sites based on parental sequences. C) Single nucleotide polymorphism frequency between whole genome sequences of parental strains. D) Frequency of RADseq polymorphic marker sites. E) Position of markers with non-redundant segregation profiles.

### **Segregation analysis, linkage map construction and alignment to the reference genome**

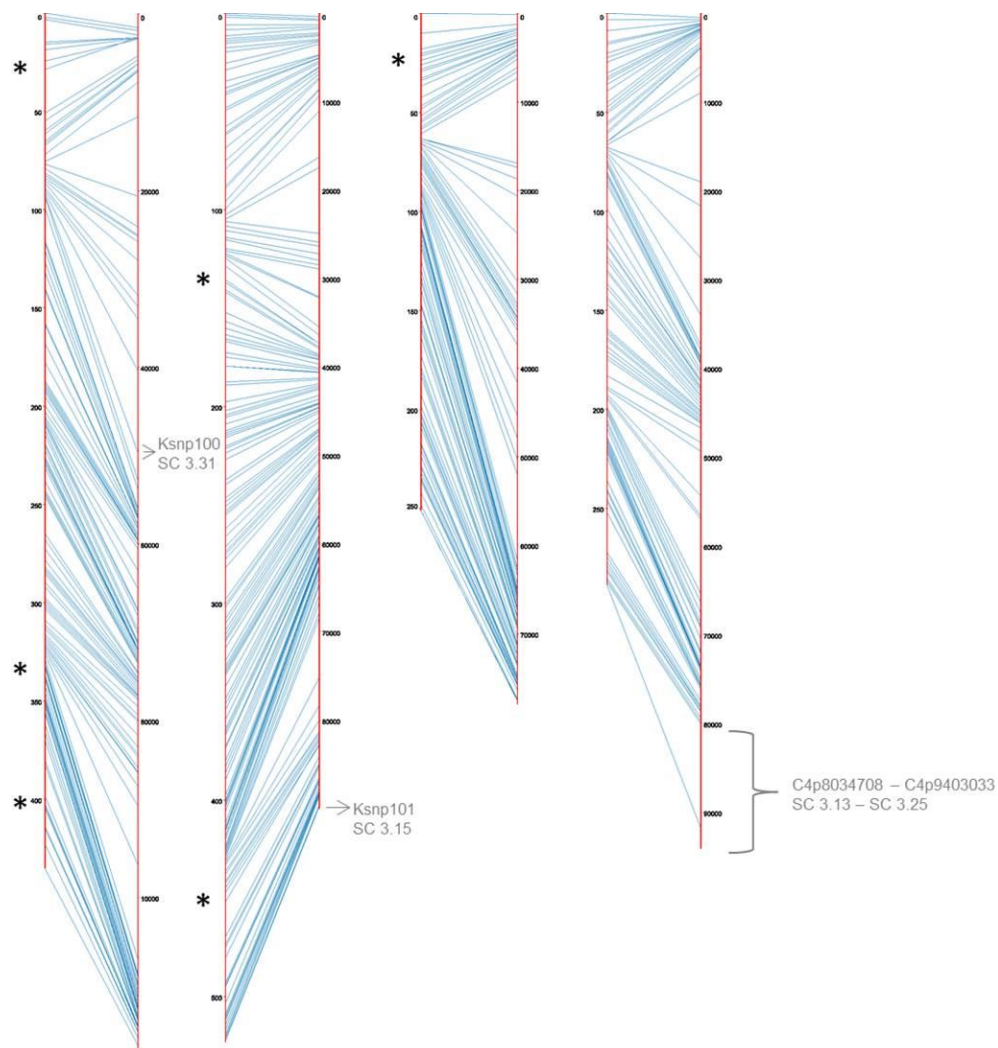
Only five markers, sparsely located on the genome were found to exhibit segregation distortion ( $p$ -value  $< 0.01$ , Supplemental file 2 Table S1). Regarding the segregation pattern of these markers, we assumed that the observed bias of segregation could be reasonably attributed to missing data or genotyping errors. In contrast to the segregation of chromosomal markers, segregation analysis of mitochondrial markers and KSNP102 marker (Supplemental file 1) revealed uniparental inheritance from INRA-156 $\Delta$ *mat* strain of mitochondria and of HG970330 sequence respectively.

Only markers with unique segregation profiles were used to construct the framework linkage map ( $n=483$ ; Figure 1, Table 1). Four linkage groups, named “LG-1”, “LG2”, “LG-3” and “LG-4” were constructed (LOD threshold = 6), corresponding respectively to chromosomes I, II, III and IV described in the reference genome assembly (version 4.0). Markers were then ordered and genetic distances were calculated using Kosambi function (Table 1). Alignment of the linkage map to the reference genome revealed a remarkable collinearity (Figure 2). Only thirteen pairs of markers exhibited inverted order out of a possible 479 pairs of successive markers (2.7%). Manual investigation of these inverted pairs revealed that the order of seven could be inverted and resulted from uncertainty due to the map resolution at these positions. The remaining six pairs of markers remained inverted, and may reflect errors in the linkage map or reference genome assembly order (Figure 2, black stars). The final map length was 1,497.1 cM with an average genetic distance between markers of 3.1 cM (Table 1).

Marker orders and distances given by this linkage map were then used to add the 847 co-segregating markers (those without unique segregation profiles), anchoring a total of 99.0% of the markers to the reference genome (Table 1). The markers KSNP100 and KSNP101, designed to align the supercontig 3.31 and supercontig 3.15 of the reference genome FGDG v3.1 have been successfully anchored to the chromosome I and II respectively (Figure 2), and this is in agreement with location proposed in RRes V4.0 assembly (Supplemental file 2 Table S2). Similarly, the RADseq markers C4p8034708 to C4p9403033 aligned as expected the different supercontigs that have been merged in the latest assembly and composing the end of the actual chromosome IV (Figure 2, Supplemental file 2 Table S2).

**Table 1:** General features of genotyping procedure and linkage map construction

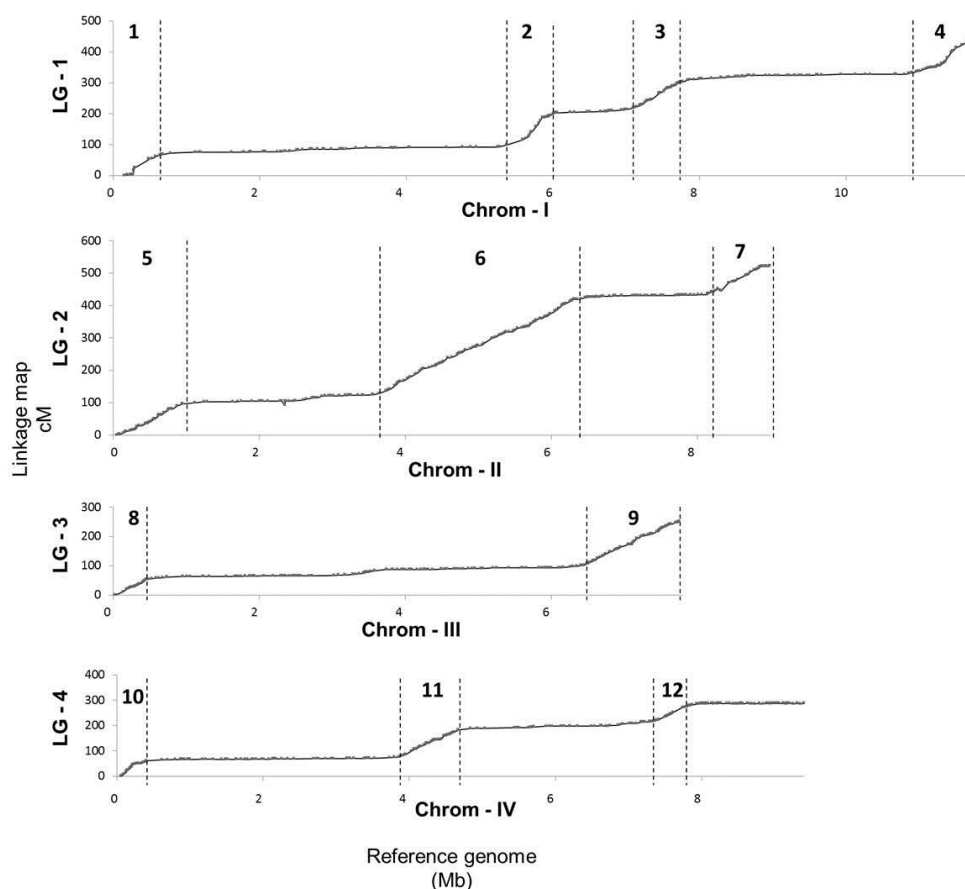
Linkage group / Chromosome	Marker number	Average marker number per 100kb	Median distance between markers (kb)	Max distance between markers (kb)	Marker used for linkage map construction*	Average spacing (cM)	Linkage group size (cM)	Chromosome size (Mbp)	Recombination rate (cM/Mb)	% anchored	Average number of CO per strain
LG-1 / Chr. I	342	2.9	11.7	499	130	3.4	435	11.76	37.0	98.0%	4.5
LG-2 / Chr. II	387	4.3	11.6	351.7	164	3.2	522.9	9.00	58.1	99.4%	5.5
LG-3 / Chr. III	278	3.6	12.8	301.1	92	2.8	250.8	7.79	32.2	99.2%	2.8
LG-4 / Chr. IV	323	3.4	15.1	398.7	97	3	288.4	9.41	30.7	99.4%	3.1
Overall	1330	3.5	12.9	499	483	3.1	1497.1	37.96	39.4	99.0%	15.9



**Figure 2:** Alignment of the linkage map (left side) to the reference genome RRES v4.0 (right side). The dark stars represent the six inversions remaining in the final linkage map. The location of the sequence placed by King *et al.* in the assembly and anchored by the linkage map is indicated in grey.

## Recombination landscape

By using the segregation information of the 1,330 markers ordered on the linkage map, 1,451 putative crossovers were mapped, within intervals of a median length of 24 kb. The average recombination rate across the entire genome was high at 39.4 cM/Mb (+/- 12.1 cM/Mb, Table 1). Each strain exhibited an average of 4.4 crossovers per chromosome, giving an average of one crossover every 2.4 Mb (Table 1). However, recombination rate was variable between chromosomes and ranged between 30.7 cM/Mb for chromosome IV up to 58.1 cM/Mb for chromosome II (Table 1). According to the limited number of chromosomes, no correlation could be tested between recombination rates and chromosome lengths, and no tentative pattern was evident. A high level of variation in recombination rate within chromosomes was typically observed, as seen in Figure 3 that represents the alignment of the linkage map (y-axis) with the physical map (x-axis). Thus, recombination-active sections (and corresponding to slope in the curve) and recombination deserts (and corresponding to flat portion of the curve) were defined.



**Figure 3:** Scatterplot showing the linkage map position (y-axis) and physical position (x-axis). The dotted lines delimitate recombination-active sections (numbered with arabic numerals) from recombination-desert sections.

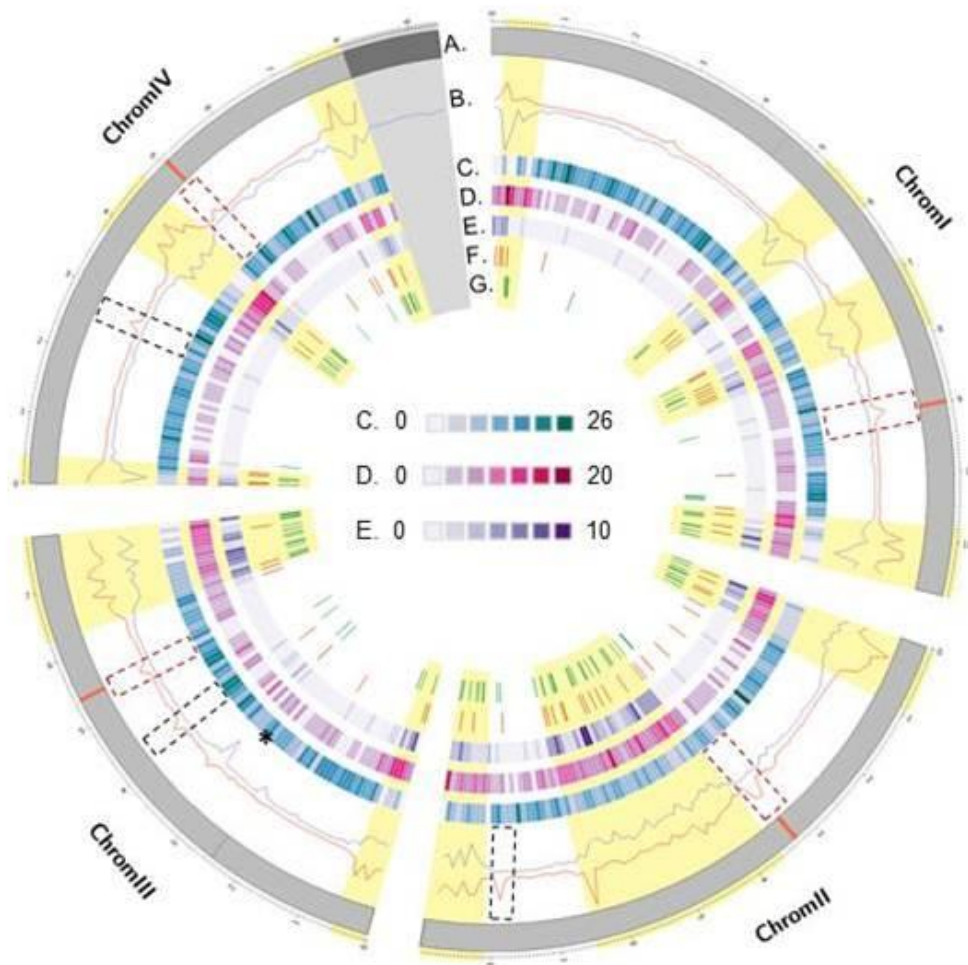
Overall, 12 chromosomal segments were identified to be recombinant-active [at least a two-fold increase of genome-wide average of recombination rate ( $> 78.8$  cM/Mb) and a size larger than 0.5Mb (Supplemental file 2 Table 3)]. Recombination hotspots, defined as a high crossover (CO) density ( $> 4$  CO) observed within a short genomic interval ( $< 20$  kb) were also detected and exhibited at least a ten-fold increase of recombination rate compared to genome-wide average (Supplemental file 2 Table S4). However, isolated recombination hotspots did not typically result in recombinant-active sections as illustrated by the hotspot located on chromosome III at 3.4 Mb, not associated with any recombinant-section (Figure 4, dark star, recombinant-active section in yellow). In total, recombinant-active sections represented 30.2% of the total length of the genome, and 87.2% of the total linkage map length. Chromosome I contained two distal recombinant-active sections, and two proximal and central sections, each flanked by a recombination desert of 1 Mb in size. Chromosome II contained three recombinant active sections, with the intermediate one alone encompassing more than one half of recombination activity recorded in the entire chromosome. Chromosome III contained two recombinant active regions spanning more than half of the linkage group. Chromosome IV contained three recombinant-active sections, but in contrast to other chromosomes, one subtelomeric region contained a non-recombinant chromosomal segment, corresponding to rRNA rich encoding region (Figure 3, Figure 4 track A in dark grey). Overall, average recombination rate was drastically different between non-active and active sections and varied overall by a  $\sim 20$  fold factor (Figure 5A). Recombination rate in the defined recombinant-active sections ranged from 104.6 – 125.9 cM / Mb, while recombinant rate of recombination ‘deserts’ ranged from 6.4 to 9.2 cM/Mb (Figure 5A). The total physical length of recombinant-active sections per chromosome was positively correlated with overall chromosome recombination rate (Figure 5B).

### **Functional and sequence enrichment analysis of highly recombinant regions**

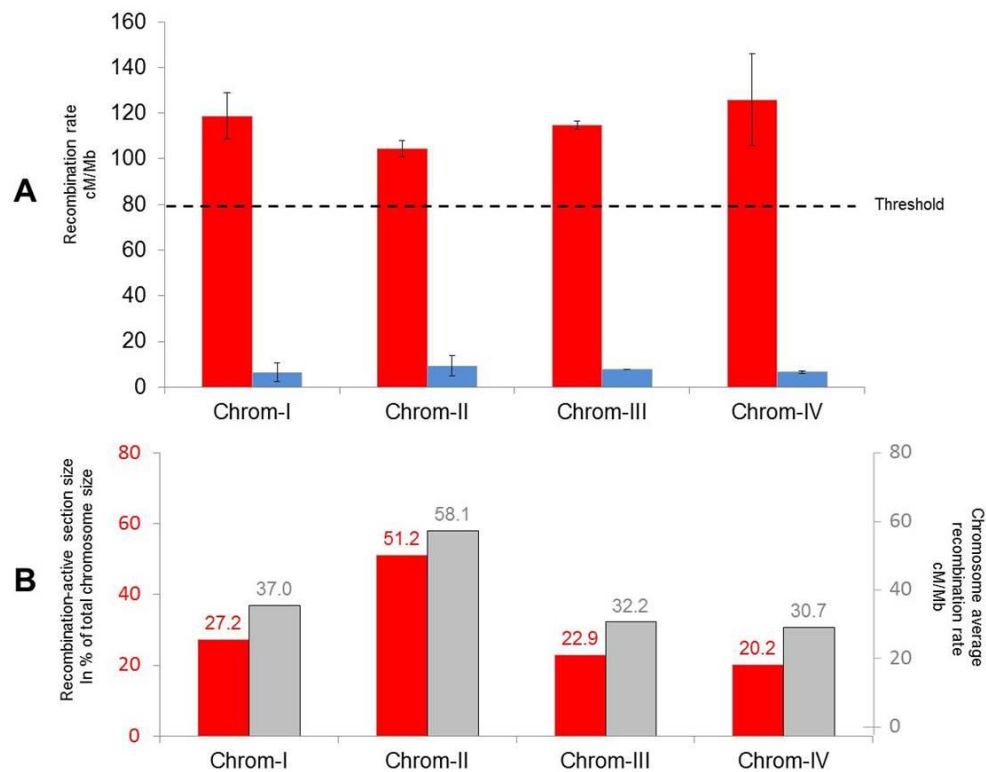
Overall, crossover density was positively correlated with polymorphism rate (measured in SNPs/kb) along the genome (Figure 4, track B,  $\text{Rho}=0.67$   $p\text{-value} < 2.20\text{E-}16$ ). However, several distinct regions deviate from this pattern, with four of them being located within the predicted centromeres of the chromosomes (Figure 4, red squares on track A and track B). However, three others regions, dispersed throughout the genome, showed similar patterns (Figure 4 track B, dashed rectangles). Crossover density was also positively correlated with gene density along the genome ( $\text{Rho}=0.50$ ,  $p\text{-value} < 2.20\text{E-}16$ ), and negatively with GC content ( $\text{Rho}= -0.56$ ,  $p\text{-value} < 2.20\text{E-}16$ ). No correlation was observed between non-



synonymous / synonymous ratio and crossover count, as calculated in 100 kb bins. However, the majority of genes showing an excess of mutation with non-synonymous effects identified in this analysis (92 of the 121 genes identified), and supposed to represent genes under diversifying selection, were located into recombinant active sections (Figure 4 track G). The 12 recombinant-active regions contained a total of 4,947 genes, representing 35% of the total number of protein encoding genes annotated in the reference genome (Supplemental file 3A). Gene ontology analysis of these genes revealed several enriched categories (Supplemental file 3B). Highest significant enrichment was recorded for amino-acid and transmembrane transport, oxidation reduction and carbohydrate metabolic processes, as well as the regulation of several cellular processes as transcription or nitrogen compound metabolism (1.97-fold - 1.27-fold, Supplemental file 3B). At the other end of the spectrum, recombinant-desert sections were enriched for gene categories arguably associated with basal mechanism, like localization and protein transport and metabolism or translation (Supplemental file 3B). Nonetheless, it is worth noting that 50.4% percent and 42.8 % percent of the protein coding genes located in active and desert section respectively has no predicted GO code. Following the general gene ontology approach described above, a more specific enrichment analysis was performed using specific datasets of *F. graminearum* retrieved from the literature. For example, genes predicted to encode for secreted proteins, and suggested to be putative effectors, were enriched by 2 fold in recombinant active regions ( $p$ -value  $< 0.001$ , Figure 4 track E and Supplemental file 3C), as were genes predicted to act in secondary metabolite biosynthesis (Figure 4 track F, 1.4 fold,  $p$ -value  $< 0.001$ ), or those previously reported to show host-specificity of expression (2.0 fold,  $p$ -value  $< 0.001$ , Figure 4 track D and Supplemental file 3C). Motif enrichment analysis of genes located in recombinant-active sections (including 500 bp upstream and downstream sequences) revealed a significant overrepresentation of motifs similar to C2H2 zing finger factors of *S. cerevisiae* ( $p$ -value  $< 1E-14$ ), as well as motifs similar to High Mobility Group of *Mus musculus* (Supplemental file 2 table S5,  $p$ -value  $< 1E-13$ ). In contrast, genes expressed constitutively in *all planta* conditions, and previously hypothesized to represent more conserved cellular functions, were under-represented by 1.5 fold in recombinant active regions, and over-represented in non-recombinant section ( $p$ -value  $< 0.001$ , Figure 4 track C and Supplemental file 3C).



**Figure 4:** Circos plot of the distribution of several genomic sequence features along the four chromosomes of *F. graminearum*. A. Representation of the four chromosomes of *F. graminearum* (in Mb), the dashed red rectangles delimitate the positions of the predicted centromeres. B. In red the SNP density (SNPs/kb) is calculated in windows of 100 kb bins along parental genomes. In blue, the number of CO calculated in separated windows of 100 kb bins across the progeny. The dashed red rectangular boxes locate typical patterns observed at centromere positions; dashed dark boxes locate centromere-like patterns. Recombinant-active sections are highlighted in yellow. C. Protein coding genes expressed constitutively in *all planta* conditions (Harris *et al.*, 2016). D. Protein coding genes expressed in host-specific conditions (Harris *et al.*, 2016). For C and D, gene density was calculated in 100 kb bin windows E. Location of genes predicted to code for secreted proteins (Brown *et al.*, 2012; King *et al.*, 2015). F. Location of predicted secondary metabolite clusters (Sieber *et al.*, 2014). G. Location of genes showing evidence of diversification.



**Figure 5:** A. Recombination rate distribution according to chromosomal section. Recombinant-active sections are given in red, and recombinant-desert sections in blue. The dashed line refers to the recombination rate threshold to be considered as recombinant-active section (2-fold increase of genome-wide recombination rate). B. The percentage of total size of chromosome allocated to recombinant-active sections is given in red, and the chromosome average recombination rate (cM/Mb) in grey.

## Discussion:

In the current study, the recombination activity of *F. graminearum* species, its pattern across the genome, and its potential role in the evolution of this devastating pathogen was investigated. The first high density linkage map of the species was created using RAD sequencing. The linkage map is almost fully integrated with the reference genome sequence (99%). It also improves the resolution of previously published genetic map of *Fusarium graminearum* by six fold (Gale *et al.*, 2005; Lee *et al.*, 2008), allowing an in depth characterization of the recombination landscape of the species.

### **Crossover mapping reveals chromosome-specific landscape of recombination.**

Recombination rate and its distribution is a key component of genome biology, and can vary tremendously across organisms. In sexually reproducing fungi, recombination rate was reported to vary from ~11cM / Mb in the edible mushroom *Agaricus bisporus* var. *bisporus* (Sonnenberg *et al.*, 2016), up to ~600 cM / Mb in *Saccharomyces cerevisiae* (Mancera *et al.*, 2008). The recombination rate estimates in this analysis for *F. graminearum* (~39 cM / Mb) is substantially lower overall than *S. cerevisiae*, and other pathogenic fungi like the causal agent of the wheat Septoria Tritici Blotch, *Zymoseptoria tritici* (130 cM / Mb, Lendenmann *et al.*, 2014). However, this recombination rate is consistent with a previously reported estimate for *F. graminearum* (~34 cM / Mb, Gale *et al.*, 2005), and may be higher due to the increased power to detect crossover using high resolution genomic tools (Sonnenberg *et al.*, 2016). Overall it suggests that *F. graminearum* does not have a particularly high genome-wide recombination rate in comparison to other pathogenic fungi. However, this feature could be offset by other phenomena, like a large production of inoculum as observed in *F. graminearum* (Osborne and Stein, 2007). Furthermore, genome-wide average recombination rates do not reflect the variation observed along chromosomes. Indeed, in many eukaryotic species, an increase in recombination rate is observed in subtelomeric regions, and a decrease near centromeres (Bhakta *et al.*, 2015; Croll and McDonald, 2012; Cuomo *et al.*, 2007; Jensen-Seaman *et al.*, 2004; Limborg *et al.*, 2016; Mancera *et al.*, 2008; Sonnenberg *et al.*, 2016; Tsai *et al.*, 2016). As already highlighted in a previous study (Cuomo *et al.*, 2007; Gale *et al.*, 2005), *F. graminearum* is no exception to the rule. However, additional regions of the genome with high recombination rate have also been described in interstitial regions corresponding to previously suggested ancestral chromosomal fusions sites (Cuomo *et al.*, 2007; Ma *et al.*, 2010). The recombination landscape revealed in the present study is strikingly consistent with previous reports (Cuomo *et al.*, 2007; Gale *et al.*, 2005), albeit at a higher resolution. It suggests a tight control of recombination activity across strains of *F. graminearum*, and the downstream analysis discussed below provides some insight into potential explanations of this phenomenon.

The high density linkage map created in the current study was also used to investigate the accuracy of the latest version of the *F. graminearum* reference genome assembly, which is based solely on short read alignment (King *et al.*, 2015). The supercontigs not positioned in the genome reference FGDB v3.1 (Cuomo *et al.*, 2007; Wong *et al.*, 2011) that were assembled in version RRES v4.0, were consistent with our linkage map assignments. The predicted positions of centromeres proposed by King *et al.* (2015) were also consistent with the typical recombination pattern of centromeres identified in the current study. Centromeric regions typically have high DNA polymorphism, whereas recombination is suppressed due to physical constraints (Bensasson *et al.*, 2008; Henikoff *et al.*, 2001). Interestingly, other genomic positions demonstrate similar characteristics of centromeres, and could correspond to ancestral centromeres following chromosome fusions or neocentromere positions.

Finally, the last remaining unassigned sequence contig (HG970330) in RRES v4.0 was found to show cytoplasmic inheritance, in line with the hypothesis of phage DNA given by King *et al.* (2015). It is however surprising to find such DNA sequence conserved in distal strains. That is not to dismiss that the four genes encoded on this sequence (King *et al.*, 2015) may have important cellular function(s).

### **Characterization of recombinant-active genome regions**

Overall, the recombination landscape of *F. graminearum* and its close relative *F. pseudograminearum* (Gardiner *et al.*, 2016) is rather unique amongst fungal species studied to date. There are clearly defined peripheral and central genomic regions where the recombination activity differs by a ~20-fold factor in comparison to recombinant-desert regions. The fact that chromosomes contain specific recombinant-active regions in interstitial DNA supports the theory of distinct chromosome fusion events (Ma *et al.*, 2010), that have retained their recombination characteristics, possibly related to conserved molecular mechanisms.

Tremendous advances have been made in model species to understand the molecular mechanisms controlling meiotic recombination, including the important role of epigenetics (Brachet *et al.*, 2012; Galazka and Freitag, 2014; Mezard *et al.*, 2015). It is now commonly accepted that recombination and chromatin accessibility are linked, with crossover-rich regions demonstrating chromatin accessibility and crossover-poor regions being more restricted. However, a paradox remains in *F. graminearum* as the epigenetic mark associated with "recombinophobic" chromatin state in most organisms is found enriched in recombination rich regions (Connolly *et al.*, 2013).

Beyond chromatin accessibility, recombination rate variation can also be directly attributable to double strand break variation along the genome, or to preferential repair of double strand break in crossover rather than in non-crossover (NCO) events. In yeast, the balance between CO and NCO formation following double-strand break repair is under genetic control, and some of the underlying factors have been already identified (Girard *et al.*, 2015; Jain *et al.*, 2016; Lorenz *et al.*, 2012). Knowledge about such regulation in *F. graminearum* is lacking. However, it is worth noting that yeast effectors, for example the antagonist of crossover formation FANCM-like protein (Lorenz *et al.*, 2012), seems to be present in *F. graminearum* genome (FGRRES\_17603, unpublished data) and could merit additional functional studies to investigate its role in recombination control in this species.

### **Recombinant-active sections are linked to several crossover hotspots**

Relatively small genomic regions (1 to 5 kb) with high crossover rate have been identified in a wide range of organisms (mammals, plants, yeast), and are commonly referred to as “hotspots” (Mancera *et al.*, 2008; Mezard, 2006; Mezard *et al.*, 2015; Paigen and Petkov, 2010). Hotspots are often located in promoters of genes, raising the hypothesis that they are associated with chromatin accessibility (Brachet *et al.*, 2012; Comeron *et al.*, 2012; Goodstadt, 2011). High recombination levels can be defined in two ways; tightly defined smaller regions corresponding to the precise length of crossover hotspots, and a broader scale genomic region corresponding to recombinant-active sections (Comeron *et al.*, 2012; Duret and Arndt, 2008; Myers *et al.*, 2005), as analyzed in the current study. Enrichment of motifs similar to yeast C2H2 zinc finger, or to High Mobility Group motifs in genes located in recombinant active regions is interesting because these motifs have previously been implicated in hotspot formation (Bergeron *et al.*, 2005; Baudat, 2010; Goodstadt, 2011; Panday *et al.*, 2016) and could be directly correlated to hotspot presence in *F. graminearum*. Furthermore, the existence of narrow hotspots has been shown in the current study, and the frequency of these may underpin the larger recombinant active regions. In *Z. tritici*, it has been demonstrated that hotspots are not always consistent between crosses (Croll *et al.*, 2015). Although the data in the current study does not exclude the possibility that the hotspots are transient in *F. graminearum*, the strong conservation of recombination landscape profiles between distant genetic backgrounds suggests that their locations are conserved, as observed in human chromosomes (Myers *et al.*, 2005).

### **Potential role of recombination landscape in pathogen evolution**

Sexual recombination has been experimentally demonstrated to play a great role for the evolution of species by increasing genetic diversity and allowing selection of specific favorable alleles and haplotypes (Goddard *et al.*, 2005). As observed in our analysis and reported in previous study (Cuomo *et al.*, 2007), the landscape of recombination and polymorphism rate are strikingly similar across the genome, suggesting that meiotic recombination is one of, if not, the major drivers of genetic diversity in this pathogen. Whereas correlation between recombination rate and nucleotide diversity is a common feature of eukaryote genomes, the implication and importance of evolutionary forces versus mutagenic effect of crossovers are still debated (Charlesworth and Campos, 2014; Manzano-Winkler *et al.*, 2013; Noor, 2008; Roselius *et al.*, 2005; Spencer *et al.*, 2006). Several elements suggest that, in *F. graminearum*, recombination events lead to polymorphism, both from an evolutionary and a molecular perspective. The high linkage disequilibrium imposed by the suppression of recombination, coupled to the presence of genes under positive and purifying selection, are likely to reduce nucleotide diversity over time via selective sweeps (Smith and Haigh, 1974). This hypothesis is strengthened by findings that revealed that constitutively-expressed genes in plants (Harris *et al.*, 2016), which may be related to more basal and conserved biological function, are overrepresented in these regions. At the

other end of the scale, the overrepresentation of genes showing host-specific expression in recombinant-active regions, corresponding possibly to host-specific effectors such as secreted proteins or genes implicated in secondary metabolite production (Brown *et al.*, 2012; Harris *et al.*, 2016; King *et al.*, 2015; Sieber *et al.*, 2014), may be related to increased nucleotide diversity due to diversifying selection (Sperschneider *et al.*, 2015; current study), made efficient by the frequent break up and reshuffle of haplotypes. Secondly, due to frequent outcrossing (Talas and McDonald, 2015), recombination activity may increase diversity by mutation events associated with double strand break and repair occurring more frequently in these regions.

Overall, the distribution and characterization of the recombination landscape of *F. graminearum* is in line with the emerging concept of a bipartite architecture of genome evolution in pathogenic fungi, also referred as “two speed” genome. Under this concept, genes that are critical to host-pathogen interaction cluster in genomic regions associated with high recombination, to facilitate a more rapid evolutionary response (Croll *et al.*, 2015; Dong *et al.*, 2015). However, the recombination landscape of *F. graminearum* differs from other fungal pathogens in several respects: it implies unordinary chromosome architecture while the epigenetic marks associated to recombination activity contradict common observations. The uniqueness of the *F. graminearum* recombination landscape and genome organization makes this pathogen a relevant model to improve knowledge of fungal genome organization and its role in evolution. Overall, these results shed light on the high potential of adaptation inherent to the recombination landscape of this pathogen, with implications for current strategies of control and effective FHB management.

In conclusion, a high density linkage map of *F. graminearum* species was constructed using RAD genotyping by sequencing. This high-resolution map facilitated a detailed identification of crossover location and the recombination landscape of these species, giving new insights into its genome architecture and its potential role in pathogen evolution. Genomic regions could broadly be split into two categories according to recombination rate, DNA diversity, and location of genes potentially related to pathogenicity. The results support the hypothesis of a two-speed genome organization, although more efforts are needed to understand the molecular mechanisms underlying this pattern. As such, the role of the epigenetic landscape of *F. graminearum* genome in regulating recombination patterns may be of future interest. The high density linkage map proposed herein may also be applied in QTL studies to detect genomic regions associated with important life-history traits of this pathogen.

### **Experimental procedures**

#### **Progeny construction and culture conditions:**

The INRA-156 and INRA-171 strains were isolated from wheat in 2001 and 2002 in different fields of Center and South-West of France respectively. Strains were characterized as *Fusarium graminearum sensu stricto* and are both pathogenic and toxinogenic on wheat. Eighty-eight (88) recombinant strains were produced from a cross between polymorphic INRA-156 $\Delta$ mat strain and INRA-171 strain. Recombinant strains were further validated using 21 targeted marker assays (see “Targeted marker genotyping” section and Supplemental file 1 & 5). Potato dextrose agar (PDA, Difco-France) was used throughout vegetative growth. INRA-156 $\Delta$ mat was obtained from INRA-156 strain after deletion of the *Mat 1-2-1* gene (Supplemental file 4). The crossing procedure was adapted from Lee *et al.*, 2003, except for incubation that has been conducted under continuous white light. 3ml to 4ml of Aerial mycelium of INRA-156 $\Delta$ mat was removed with a sterile stick, and 1ml of INRA-171 conidial suspension diluted in 2.5% Tween® 60 was distributed on the INRA-156 $\Delta$ mat plate and gently spread with a sterile stick. The plate was incubated at 25°C under continuous white light, until perithecia reached maturity (Cavinder *et al.*, 2012). Visible cirrhi were collected in sterile water, spread and incubated on PDA plates.

#### **Genomic resources:**

The *F. graminearum* genome version RRes V4.0 (King *et al.*, 2015) used in the study is available at EMBL-EBI (HG970331, HG970332, HG970333, HG970334, HG970335). Genes predicted to code for the secreted proteins of *F. graminearum* were retrieved from King *et al.*, (2015). Genes previously described as expressed in wheat, barley and/or maize were retrieved from Harris *et al.*, (2016) and the genes predicted to belong to secondary metabolite pathway clusters were retrieved from Sieber *et al.*, (2014).

#### **Genomic DNA extraction:**

Genomic DNA was extracted from lyophilized mycelium previously grown on potato dextrose agar (39 g/l, Difco). Mycelia were lysed in a buffer containing 100 mM Tris-HCl (pH 9.0), 10 mM EDTA, 1% sarkosyl and 200  $\mu$ g/mL proteinase K for 2 h at 65 °C. After centrifugation (10 min at 10,000 g), the supernatant was extracted successively with phenol, phenol: chloroform (50:50) and finally chloroform. Nucleic acids were precipitated with cold ammonium acetate (3M) and isopropanol, washed in 70% ethanol and dissolved in 100  $\mu$ L nuclease free water.



### **Whole genome sequencing of parental strains and analysis:**

Whole nuclear DNA from the parental isolates were sequenced by the MGX platform in Montpellier, France (<http://www.mgx.cnrs.fr/>) using the Illumina sequencing technology (HiSeq™ Sequencing Systems). The reads were filtered and trimmed using Prinseq software (Schmieder and Edwards, 2011). They were then aligned to the reference genome RRes V4.0 (King *et al.*, 2015) using BWA (v 0.7.8) and the BWA-MEM algorithm with standard parameters, and a seed size of 15 nucleotides (Li and Durbin, 2009). SNP calling was performed using GATK (v 2.4) and the Unified Genotyper walker in haploid mode (DePristo *et al.*, 2011). Parental consensus sequences were constructed using VcfToFasta tool, and by correcting the reference genome with variants identified in each parental genome. *In silico* digestion of the genomes by *PstI* enzyme was conducted using EMBOSS and the “Restrict” tool. The 121 genes showing an excess of mutations with non-synonymous effects, and supposed to represent putative genes under diversification, were identified as exhibiting more than 75% of non-synonymous mutations compared to total number of mutation into genic sequence (upper 3<sup>rd</sup> quartile of genome wide distribution) and at least 4 genic mutations (median mutation number per gene x 2).

### **RAD-seq library preparation and sequencing**

Library and sequencing was prepared and conducted by the MGX platform team in Montpellier, France (<http://www.mgx.cnrs.fr/>). The libraries were prepared according to Baird *et al.* protocol (Baird *et al.*, 2008) using the *PstI* restriction enzyme. The main modification was the use of Ampure XP beads for the different purification steps. A detailed version of RAD-seq procedure is available in Supporting Materials. Sequence data are retrievable from GenBank database under SRA accession SRP083578.

### **RAD-seq SNP discovery and genotyping analysis**

The quality of the 125-nt long paired-end reads was analyzed using FastQC v0.11.2. Reads were demultiplexed according to barcode sequences using ‘process\_radtags’ program from Stacks software pipeline v1.32 with the following parameters: *PstI* enzyme cutting site recognition, removal of reads with at least one uncalled base and/or an average quality score below the phred score of 20. For each strain, reads were then mapped on the reference genome RRes V4.0 (King *et al.*, 2015) using Bowtie2 v0.12.9 using default parameters. SAM files were then sorted and transformed to BAM files using Samtools v1.1 (Li *et al.*, 2009). The stacks software (Catchen *et al.*, 2011) was used all along to identify SNP and InDel markers. The ref\_map.pl program was used to i) compare both parental BAM files and the 93 BAM files from the progeny in order to build RAD loci and call SNPs, ii) create a catalog of all loci and iii) match each sample against the catalog. The minimum depth of coverage to call a stack was 3, one mismatch was allowed between loci when building the catalog. Markers were

retained if they were bi-allelic, contained up to 10 SNPs compared to the consensus sequences, and at least 80% of the progeny were genotyped.

### **Targeted marker genotyping**

Single Nucleotide Polymorphism (SNP) derived markers genotyped using either Cleaved Amplified Polymorphic Sequences (CAPS) or Kompetitive Allele Specific PCR (KASP) techniques have been designed from the whole genome polymorphism information between the two parental sequences. Microsatellite (Simple Sequence Repeat SSR) markers have been retrieved from literature (Brygoo and Gautier, 2007; Giraud *et al.*, 2002; Naef *et al.*, 2006; Naef and Défago, 2006; Vogelgsang *et al.*, 2009). Ten SSR and 11 CAPS markers were used to confirm the recombinant progeny. Three KASP markers were designed to align supercontig 3.31, supercontig 3.15 and supercontig 3.12 respectively. These sequences corresponded to supercontigs that were not anchored during initial assembly (Cuomo *et al.*, 2007). The first two supercontigs have been in the version RRes V4.0 of the assembly (King *et al.* 2015). Supercontig 3.12 remains unanchored to any chromosomes and has been renamed HG970330. All details about primers and experiment conditions are available in Supplemental File 5.

### **Genetic map construction:**

The linkage map was constructed using R/qtl (Broman *et al.*, 2003). Firstly, markers exhibiting redundant segregation patterns in the population were set aside from the mapping procedure. Linkage groups were formed using a maximum recombination fraction of 0.35 and a decimal logarithm of odd (LOD) score threshold of 6. Markers were ordered and genetic distance was calculated using the Kosambi function. Groups of co-segregating markers were repositioned manually at marker position (the one used for linkage map construction) and following reference genome order.

### **Alignment of genetic and physical maps and analysis of recombination**

The linkage map was aligned on the reference genome (RRES v4.0). For RAD-tags containing several genetic variants, only the position of the first variant was considered. The ArkMap software was used to construct the illustration showing the alignment of the linkage map to the reference genome (<http://www.bioinformatics.roslin.ed.ac.uk/arkmap/>). The recombination rate of overall chromosome, or recombinant-active and recombinant desert sections, were calculated by dividing linkage group size by chromosome or section size. Detection of crossover was performed manually using inheritance of markers along the chromosome according to the linkage map order. The resolution of crossover position was obtained by calculating the median size spanning two markers adjacent to the corresponding crossover. Recombinant-active sections were defined when they contained a two-fold increase in recombination rate compared to the genome-wide average, over a region larger than 0.5 Mb. Recombination hotspots were defined as the genomic loci delimited by markers of which more

than 4 crossovers were recorded and presenting an estimated recombination rate greater than a ten-fold increase of genome-wide recombination rate ( $> 394\text{cM/Mb}$ ).

### **Statistical analysis:**

The Poisson distribution was used to test the distribution pattern of predicted *Pst*I cutting site and markers site in the genome. Marker distribution analysis was carried out as described in Bhakta *et al.*, (2015) using an interval of 100 kb. Significant differences between observed and expected frequencies of the number of markers per interval were detected through Kolmogorov Smirnov tests. Enrichment of functional categories was calculated using the Gene Ontology Enrichment tools proposed online by the EuPathDB project (<http://fungidb.org>) using Biological ontology and InterPro predictions. Enrichments were defined for  $p$ -value  $< 0.001$ . A Chi-squared test was used to compare the observed distribution of genes located in recombinant active regions compared to the theoretical distribution under the hypothesis of a random distribution in the genome. Over and underrepresentation were defined for  $p$ -value  $< 0.001$ . Gene density, GC content and observed crossovers were calculated using different windows of 100 kb bins. Correlations were calculated using Spearman rank order test correlation and accepted for  $p$ -value  $< 0.001$ .

### **Acknowledgements**

Authors would like to acknowledge the genotoul bioinformatic platform Toulouse Midi-Pyrénées and Sigenae group, especially Marie-Stéphane Trotard, for providing help along the bioinformatic analysis and for providing computing and storage resources (<http://bioinfo.genotoul.fr/>). The authors would like to sincerely thank Donald Gardiner to share his interest in linkage mapping development in *Fusarium* species and Nadia Ponts for thoughtful discussions. BL received a PhD fellowship from the French Research Ministry. Collaboration between the French National Institute for Agricultural Research (INRA) and The Roslin Institute was made possible thanks to a grant delivered by the Agreenium Institute and the Plant Health and Environment INRA division. RDH and CP were supported by the Bioscience and Biotechnology Research Council (BBSRC) funds BB/J004235/1 and BB/J004324/1.

### **Reference**

- Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A. and Johnson, E.A.** (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* **3**, e3376.
- Bensasson, D.** (2011) Evidence for a high mutation rate at rapidly evolving yeast centromeres. *Bmc Evol. Biol.* **11**.
- Bensasson, D., Zarowiecki, M., Burt, A. and Koufopanou, V.** (2008) Rapid evolution of yeast centromeres in the absence of drive. *Genetics* **178**, 2161–2167.

- Bergeron, S., Madathiparambil, T. and Swanson, P.C.** (2005) Both high mobility group (HMG)-boxes and the acidic tail of HMGB1 regulate recombination-activating gene (RAG)-mediated recombination signal synapsis and cleavage *in vitro*. *J. Biol. Chem.* **280**, 31314–31324.
- Bhakta, M.S., Jones, V.A. and Vallejos, C.E.** (2015) Punctuated distribution of recombination hotspots and demarcation of pericentromeric regions in *Phaseolus vulgaris* L. *PLoS One* **10**, e0116822.
- Brachet, E., Sommermeyer, V. and Borde, V.** (2012) Interplay between modifications of chromatin and meiotic recombination hotspots. *Biol Cell* **104**, 51–69.
- Broman, K.W., Wu, H., Sen, S. and Churchill, G.A.** (2003) R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**, 889–890.
- Brown, N.A., Antoniw, J. and Hammond-Kosack, K.E.** (2012) The predicted secretome of the plant pathogenic fungus *Fusarium graminearum*: a refined comparative analysis. *PLoS One* **7**, e33731.
- Brygoo, Y. and Gautier, A.** (2007) Molecular polymorphism of *Fusarium* strains isolated from wheat and corn ears in France Collects 2003 and 2004. *Progrès Perspect. la Rech. sur les mycotoxines Fusarium dans les céréales*.
- Catchen, J.M., Amores, A., Hohenlohe, P., Cresko, W. and Postlethwait, J.H.** (2011) Stacks: building and genotyping loci *de novo* from short-read sequences. *G3 Genes, Genomes, Genet.* **1**, 171–182.
- Cavinder, B., Sikhakolli, U., Fellows, K.M. and Trail, F.** (2012) Sexual development and ascospore discharge in *Fusarium graminearum*. *J Vis Exp*.
- Charlesworth, B. and Campos, J.L.** (2014) The relations between recombination rate and patterns of molecular variation and evolution in *Drosophila*. <http://dx.doi.org/10.1146/annurev-genet-120213-092525>.
- Chen, Y. and Zhou, M.-G.** (2009) Sexual recombination of carbendazim resistance in *Fusarium graminearum* under field conditions. *Pest Manag. Sci.* **65**, 398–403.
- Comeron, J.M., Ratnappan, R., Bailin, S., et al.** (2012) The many landscapes of recombination in *Drosophila melanogaster* Petrov, D.A., ed. *PLoS Genet.* **8**, e1002905.
- Croll, D., Lendenmann, M.H., Stewart, E. and McDonald, B.A.** (2015) The impact of recombination hotspots on genome evolution of a fungal plant pathogen. *Genetics* **201**, 1213–1228.
- Croll, D. and McDonald, B.A.** (2012) The accessory genome as a cradle for adaptive evolution in pathogens. *PLoS Pathog* **8**, e1002608.
- Cumagun, C.J.R. and Miedaner, T.** (2004) Segregation for aggressiveness and deoxynivalenol production of a population of *Gibberella zeae* causing head blight of wheat. *Eur. J. Plant Pathol.* **110**, 789–799.
- Cuomo, C.A., Guldener, U., Xu, J.R., et al.** (2007) The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. *Science* (80-. ). **317**, 1400–1402.

- Davey, J.W., Cezard, T., Fuentes-Utrilla, P., Eland, C., Gharbi, K. and Blaxter, M.L.** (2013) Special features of RAD sequencing data: implications for genotyping. *Mol. Ecol.* **22**, 3151–3164.
- Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M. and Blaxter, M.L.** (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat Rev Genet* **12**, 499–510.
- DePristo, M.A., Banks, E., Poplin, R., et al.** (2011) Additional data: A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* **43**, 491–498.
- Dong, S., Raffaele, S. and Kamoun, S.** (2015) The two-speed genomes of filamentous pathogens: waltz with plants. *Curr. Opin. Genet. Dev.* **35**, 57–65.
- Duret, L. and Arndt, P.F.** (2008) The impact of recombination on nucleotide substitutions in the human genome Przeworski, M., ed. *PLoS Genet.* **4**, e1000071.
- F. Baudat, J.B.C.G.A.F.-A.C.O.M.P.G.C.B. de M.** (2010) PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science* **327**, 836.
- Galazka, J.M. and Freitag, M.** (2014) Variability of chromosome structure in pathogenic fungi--of “ends and odds.” *Curr Opin Microbiol* **20**, 19–26.
- Gale, L.R., Bryant, J.D., Calvo, S., et al.** (2005) Chromosome complement of the fungal plant pathogen *Fusarium graminearum* based on genetic and physical mapping and cytological observations. *Genetics* **171**, 985–1001.
- Girard, C., Chelysheva, L., Choinard, S., et al.** (2015) AAA-ATPase FIDGETIN-LIKE 1 and helicase FANCM antagonize meiotic crossovers by distinct mechanisms. *PLoS Genet.* **11**, e1005369.
- Giraud, T., Fournier, E., Vautrin, D., Solignac, M., Vercken, E., Bakan, B. and Brygoo, Y.** (2002) Isolation of eight polymorphic microsatellite loci, using an enrichment protocol, in the phytopathogenic fungus *Fusarium culmorum*. *Mol. Ecol. Notes* **2**, 121–123.
- Goddard, M.R., Godfray, H.C.J. and Burt, A.** (2005) Sex increases the efficacy of natural selection in experimental yeast populations. *Nature* **434**, 636–640.
- Gonen, S., Lowe, N.R., Cezard, T., Gharbi, K., Bishop, S.C. and Houston, R.D.** (2014) Linkage maps of the Atlantic salmon (*Salmo salar*) genome derived from RAD sequencing. *BMC Genomics* **15**.
- Harris, L.J., Balcerzak, M., Johnston, A., Schneiderman, D. and Ouellet, T.** (2016) Host-preferential *Fusarium graminearum* gene expression during infection of wheat, barley, and maize. *Fungal Biol.* **120**, 111–123.
- Henikoff, S., Ahmad, K. and Malik, H.S.** (2001) The centromere paradox: stable inheritance with rapidly evolving DNA. *Science (80-. )*. **293**, 1098–1102.
- Jain, S., Sugawara, N., Mehta, A., Ryu, T. and Haber, J.E.** (2016) Sgs1 and Mph1 Helicases Enforce the recombination execution checkpoint during DNA double-strand break repair in *Saccharomyces cerevisiae*. *Genetics* **203**, 667–75.

- Jensen-Seaman, M.I., Furey, T.S., Payseur, B.A., Lu, Y., Roskin, K.M., Chen, C.-F., Thomas, M.A., Haussler, D. and Jacob, H.J.** (2004) Comparative recombination rates in the rat, mouse, and human genomes. *Genome Res.* **14**, 528–538.
- King, R., Urban, M., Hammond-Kosack, M., Hassani-Pak, K. and Hammond-Kosack, K.** (2015) The completed genome sequence of the pathogenic ascomycete fungus *Fusarium graminearum*. *BMC Genomics* **16**, 544.
- L Goodstadt, C.P.P.** (2011) Is the control of recombination conserved among diverse eukaryotes? *Heredity (Edinb)*. **106**, 710.
- Lee, J., Jurgenson, J.E., Leslie, J.F. and Bowden, R.L.** (2008) Alignment of genetic and physical maps of *Gibberella zeae*. *Appl Env. Microbiol* **74**, 2349–2359.
- Lendenmann, M.H., Croll, D., Stewart, E.L. and McDonald, B.A.** (2014) Quantitative trait locus mapping of melanization in the plant pathogenic fungus *Zymoseptoria tritici*. *G3 Genes/Genomes/Genetics* **4**, 2519–2533.
- Li, H. and Durbin, R.** (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760.
- Liang, J.M., Xayamongkhon, H., Broz, K., Dong, Y., McCormick, S.P., Abramova, S., Ward, T.J., Ma, Z.H. and Kistler, H.C.** (2014) Temporal dynamics and population genetic structure of *Fusarium graminearum* in the upper Midwestern United States. *Fungal Genet Biol* **73**, 83–92.
- Limborg, M.T., McKinney, G.J., Seeb, L.W. and Seeb, J.E.** (2016) Recombination patterns reveal information about centromere location on linkage maps. *Mol. Ecol. Resour.* **16**, 655–661.
- Lorenz, A., Osman, F., Sun, W., et al.** (2012) The fission yeast FANCM ortholog directs non-crossover recombination during meiosis. *Science (80-. )*. **336**, e299–e299.
- Ma, L.-J., Does, H.C. van der, Borkovich, K.A., et al.** (2010) Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* **464**, 367–373.
- Mancera, E., Bourgon, R., Brozzi, A., Huber, W. and Steinmetz, L.M.** (2008) High-resolution mapping of meiotic crossovers and non-crossovers in yeast. *Nature* **454**, 479–485.
- Manzano-Winkler, B., McGaugh, S.E., Noor, M.A.F., et al.** (2013) How hot are drosophila hotspots? Examining recombination rate variation and associations with nucleotide diversity, divergence, and maternal age in *Drosophila pseudoobscura* Palsson, A., ed. *PLoS One* **8**, e71582.
- McDonald, B. and Linde, C.** (2002) The population genetics of plant pathogens and breeding strategies for durable resistance. *Euphytica* **124**, 163–180.
- Mezard, C.** (2006) Meiotic recombination hotspots in plants. *Biochem. Soc. Trans.* **34**, 531–534.
- Mezard, C., Jahns, M.T. and Grelon, M.** (2015) Where to cross? New insights into the location of meiotic crossovers. *Trends Genet* **31**, 393–401.
- Myers, S., Bottolo, L., Freeman, C., McVean, G. and Donnelly, P.** (2005) A fine-scale map of recombination rates and hotspots across the human genome. *Science* **310**, 321–4.
- Naef, A. and Défago, G.** (2006) Population structure of plant-pathogenic *Fusarium* species in

- overwintered stalk residues from Bt-transformed and non-transformed maize crops. *Eur. J. Plant Pathol.* **116**, 129–143.
- Naef, A., Senatore, M. and Défago, G.** (2006) A microsatellite based method for quantification of fungi in decomposing plant material elucidates the role of *Fusarium graminearum* DON production in the saprophytic competition with *Trichoderma atroviride* in maize tissue microcosms. *FEMS Microbiol. Ecol.* **55**, 211–20.
- Noor, M.A.F.** (2008) Connecting recombination, nucleotide diversity and species divergence in *Drosophila*. *Fly (Austin)*. **2**, 255–6.
- Osborne, L.E. and Stein, J.M.** (2007) Epidemiology of *Fusarium* head blight on small-grain cereals. *Int. J. Food Microbiol.* **119**, 103–108.
- Paigen, K. and Petkov, P.** (2010) Mammalian recombination hot spots: properties, control and evolution. *Nat. Rev. Genet.* **11**, 221–233.
- Palaiokostas, C., Bekaert, M., Khan, M.G.Q., Taggart, J.B., Gharbi, K., McAndrew, B.J. and Penman, D.J.** (2013) Mapping and validation of the major sex-determining region in Nile Tilapia (*Oreochromis niloticus* L.) using RAD sequencing. *PLoS One* **8**.
- Panday, A., Grove, A., Hergeth, S., et al.** (2016) The high mobility group protein HMO1 functions as a linker histone in yeast. *Epigenetics Chromatin* **9**, 13.
- Pariaud, B., Ravigné, V., Halkett, F., Goyeau, H., Carlier, J. and Lannou, C.** (2009) Aggressiveness and its role in the adaptation of plant pathogens. *Plant Pathol.* **58**, 409–424.
- Petes, T.D.** (2001) Meiotic recombination hot spots and cold spots. *Nat Rev Genet* **2**, 360–369.
- Raffaele, S., Farrer, R.A., Cano, L.M., et al.** (2010) Genome evolution following host jumps in the Irish potato famine pathogen lineage. *Science* **330**, 1540–3.
- Roselius, K., Stephan, W. and Städler, T.** (2005) The relationship of nucleotide polymorphism, recombination rate and selection in wild tomato species. *Genetics* **171**, 753–763.
- Schmieder, R. and Edwards, R.** (2011) Quality control and preprocessing of metagenomic datasets. *Bioinformatics*.
- Sieber, C.M.K., Lee, W., Wong, P., et al.** (2014) The *Fusarium graminearum* genome reveals more secondary metabolite gene clusters and hints of horizontal gene transfer. *PLoS One* **9**, e110311.
- Smith, J.M. and Haigh, J.** (1974) The hitch-hiking effect of a favourable gene. *Genet. Res.* **23**, 23.
- Sonnenberg, A.S.M., Gao, W., Lavrijssen, B., et al.** (2016) A detailed analysis of the recombination landscape of the button mushroom *Agaricus bisporus* var. *bisporus*. *Fungal Genet. Biol.* **93**, 35–45.
- Spencer, C.C.A., Deloukas, P., Hunt, S., Mullikin, J., Myers, S., Silverman, B., Donnelly, P., Bentley, D. and McVean, G.** (2006) The influence of recombination on human genetic diversity. *PLoS Genet.* **2**, e148.
- Sperschneider, J., Gardiner, D.M., Thatcher, L.F., Lyons, R., Singh, K.B., Manners, J.M. and Taylor, J.M.** (2015) Genome-wide analysis in three *Fusarium* pathogens identifies rapidly

evolving chromosomes and genes associated with pathogenicity. *Genome Biol. Evol.*

- Talas, F. and McDonald, B.** (2015) Genome-wide analysis of *Fusarium graminearum* field populations reveals hotspots of recombination. *BMC Genomics* **16**, 996.
- Talas, F. and McDonald, B.A.** (2015) Significant variation in sensitivity to a DMI fungicide in field populations of *Fusarium graminearum*. *Plant Pathol.* **64**, 664–670.
- Trail, F.** (2009) For blighted waves of grain: *Fusarium graminearum* in the postgenomics era. *Plant Physiol* **149**, 103–110.
- Tsai, H.Y., Robledo, D., Lowe, N.R., Bekaert, M., Taggart, J.B., Bron, J.E. and Houston, R.D.** (2016) Construction and annotation of a high density SNP linkage map of the atlantic salmon (*Salmo salar*) genome. *G3&#58; Genes/Genomes/Genetics* **6**, 2173–2179.
- Vogelgsang, S., Widmer, F., Jenny, E. and Enkerli, J.** (2009) Characterisation of novel *Fusarium graminearum* microsatellite markers in different *Fusarium* species from various countries. *Eur. J. Plant Pathol.* **123**, 477–482.
- Ward, T.J., Clear, R.M., Rooney, A.P., et al.** (2008) An adaptive evolutionary shift in *Fusarium* head blight pathogen populations is driving the rapid spread of more toxigenic *Fusarium graminearum* in North America. *Fungal Genet. Biol.* **45**, 473–484.
- Waskiewicz, A. and Golinski, P.** (2013) Mycotoxins in foods, feeds and their components. *Krmiva* **55**, 35–45.
- Wong, P., Walter, M., Lee, W., Mannhaupt, G., Münsterkötter, M., Mewes, H.-W., Adam, G. and Güldener, U.** (2011) FGDB: revisiting the genome annotation of the plant pathogen *Fusarium graminearum*. *Nucleic Acids Res.* **39**, D637–D639.
- Yelina, N., Diaz, P., Lambing, C. and Henderson, I.R.** (2015) Epigenetic control of meiotic recombination in plants. *Sci. China-Life Sci.* **58**, 223–231.



## Supporting information

**Supplemental file 1:** Dataset of markers. A) Lab markers B) RAD markers.

**Supplemental file 2: Table S1:** List of markers exhibiting non-Mendelian segregation. **Table S2:** Overview of the supercontigs anchored during RRes4.0 assembly proposal, with information concerning the marker used for anchoring these supercontigs. **Table S3:** List of the chromosomal segments classified as recombinant-active. **Table S4:** List of the hypothetical crossover hotspots identified in this analysis. **Table S5:** Motifs significantly overrepresented in genes located in recombinant active genes (gene sequence in addition to 500 bp downstream and upstream sequence).

**Table S1:** List of markers exhibiting non-Mendelian segregation.

Marker ID	Missing genotype	INRA-156 genotype	INRA-171 genotype	<i>p</i> -value*
C3p3172164	3	55	30	0.00669531
C3p7699843	14	52	22	0.00048769
C4p4545092	13	49	26	0.00791179
C4p4545047	12	50	26	0.00590539
C4p7999747	17	23	48	0.00300766

\*Probability associated with chis-square tests performed on the observed segregation pattern to fit the 1:1 expected Mendelian ratio

**Table S2:** Overview of the supercontigs anchored during RRes4.0 assembly proposal, with information concerning the marker used for anchoring these supercontigs.

Supercontig number FGDB v3.1	Chromosome	Position in the reference genome RRES v4.0	Corresponding marker	Anchored
3.12	Unanchored	0 - 15,604	KSNP102	No
3.13	Mitochondria	5,891 - 17,410	- (*)	-
3.15	II	8,995,262 - 8,997,558	KSNP101	Yes
3.26	III	2,466,775 - 2,469,779	- (*)	-
3.31	I	4,541,855 - 4,543,915	KSNP100	Yes
3.13-3.25	IV	8,012,341 - 9,397,177	C4p8034708- C4p9403033	Yes

\* No polymorphism was detected between parental sequences

**Table S3:** List of the chromosomal segments classified as recombinant-active.

Band ID	Chromosome	Start (Mb)	End (Mb)	Length (Mb)	Start (cM)	End (cM)	Length (cM)	cM/Mb
1	Chrom-I	0.2	0.8	0.6	2.1	72.2	70.1	117
2	Chrom-I	5.2	6	0.8	93.2	199.3	106.1	133
3	Chrom-I	7	7.9	0.9	214.1	311.1	97.0	108
4	Chrom-I	10.8	11.7	0.9	328.1	435.0	106.9	119
5	Chrom-II	0.04	0.95	0.91	0.0	96.1	96.1	106
6	Chrom-II	3.6	6.4	2.8	124.3	420.7	296.4	106
7	Chrom-II	8.1	9	0.9	433.3	522.9	89.6	100
8	Chrom-III	0	0.48	0.48	0.0	54.2	54.2	113
9	Chrom-III	6.4	7.7	1.3	100.8	250.8	150.0	115
10	Chrom-IV	0	0.4	0.4	0.0	61.3	61.3	153
11	Chrom-IV	3.8	4.7	0.9	75.3	183.0	107.7	120
12	Chrom-IV	7.3	7.9	0.6	214.7	285.0	70.3	117

**Table S4:** List of the hypothetical crossover hotspots identified in this analysis.

Chromosome	Start	End	Length	Number of CO	Recombination rate (cM/Mb)
III	3,361,859	3,373,238	11,379	11	1,099
I	11,358,151	11,375,421	17,270	8	526
I	5,803,420	5,822,699	19,279	7	413
I	282,208	290,586	8,378	6	814
IV	200,160	210,848	10,688	6	638
I	5,667,112	5,677,518	10,406	5	546
I	7,514,533	7,524,890	10,357	5	549
II	312,763	319,384	6,621	5	858
II	4,506,134	4,519,339	13,205	5	430

**Table S5:** Motifs significantly overrepresented in genes located in recombinant active genes (gene sequence in addition to 500 bp downstream and upstream sequence).

Motifs identified	<i>p</i> -value	% background	% targets	Best match of known motifs
	1 <sup>E</sup> -14	11.7%	16.1%	<i>S. cerevisiae</i> C2H2 zinc finger factors Score: 89%
	1 <sup>E</sup> -13	37.8%	43.9%	<i>M. Musculus</i> High Mobility Group box (HMG) Score: 78%

**Supplemental file 3:** Recombination and gene functions. See A], B] and C]

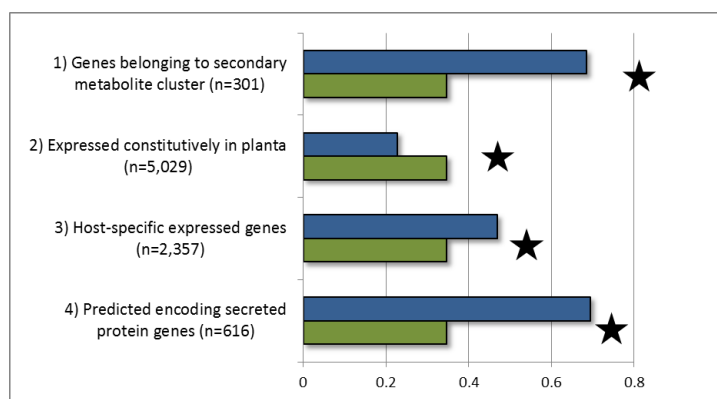
**Example A]:** All protein-coding genes described in the reference genome (RRES v4.0) with features.

RRES v4.0	Gene start (bp)	Gene end (bp)	Chromosome	FGDB v3.1	Gene description	Recombination
FGRRES_16072_M	6170	11000	I	FGSG_16072	Uncharacterized protein	Desert
FGRRES_11606	11394	12168	I	FGSG_11606	Uncharacterized protein	Desert
FGRRES_20000	19069	19668	I	FGSG_20000	-	Desert
FGRRES_15662	21463	22193	I	FGSG_15662	-	Desert
FGRRES_16069	23519	25108	I	FGSG_16069	Cytochrome P450 55A2	Desert
FGRRES_11584	26109	26441	I	FGSG_11584	Uncharacterized protein	Desert
FGRRES_16068	27171	30578	I	FGSG_16068	Uncharacterized protein	Desert
FGRRES_11582	31131	31487	I	FGSG_11582	Uncharacterized protein	Desert
FGRRES_11581	32088	34196	I	FGSG_11581	Uncharacterized protein	Desert
FGRRES_11580	36525	37114	I	FGSG_11580	Uncharacterized protein	Desert
FGRRES_16067	37732	39108	I	FGSG_16067	Uncharacterized protein	Desert
FGRRES_11641	42300	42965	I	FGSG_11641	Uncharacterized protein	Desert
FGRRES_11605	46799	48110	I	FGSG_11605	Uncharacterized protein	Desert
FGRRES_11642	51001	51726	I	FGSG_11642	Uncharacterized protein	Desert
FGRRES_11643	52395	52854	I	FGSG_11643	Uncharacterized protein	Desert
FGRRES_11600	53937	55904	I	FGSG_11600	Uncharacterized protein	Desert
FGRRES_11601	57837	58687	I	FGSG_11601	Uncharacterized protein	Desert
FGRRES_00001	59208	59741	I	FGSG_00001	Uncharacterized protein	Desert
FGRRES_00002	60405	62603	I	FGSG_00002	Uncharacterized protein	Desert
FGRRES_11644	62745	64667	I	FGSG_11644	Uncharacterized protein	Desert
FGRRES_00003	67014	70396	I	FGSG_00003	Uncharacterized protein	Desert
FGRRES_20001	70886	71196	I	FGSG_20001	Uncharacterized protein	Desert
FGRRES_15677_M	70921	71208	I	FGSG_15677	-	Desert
FGRRES_00004	71902	72697	I	FGSG_00004	Uncharacterized protein	Desert
FGRRES_00005	74163	76352	I	FGSG_00005	Uncharacterized protein	Desert

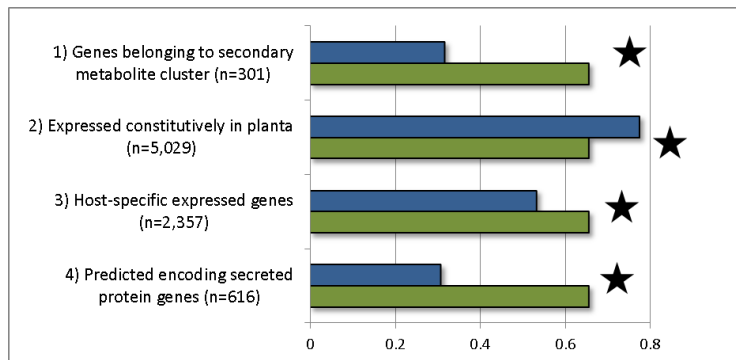
## B] Gene ontology enrichment of recombinant-active and desert sections.

Recombination	GO ID	GO Term	Total gene number with this GO term	Genes in the section with this term	Percent of total gene number	Fold enrichment	p-value
active sections	GO:0003333	amino acid transmembrane transport	51	33	64.7	1.97	2.14E-03
	GO:0055085	transmembrane transport	607	303	49.9	1.52	4.62E-10
	GO:0055114	oxidation-reduction process	763	354	46.4	1.41	9.56E-09
	GO:0005975	carbohydrate metabolic process	324	147	45.4	1.38	6.15E-04
	GO:0006355	regulation of transcription, DNA-dependent	458	199	43.4	1.32	4.30E-04
	GO:0051171	regulation of nitrogen compound metabolic process	479	200	41.8	1.27	1.94E-03
	GO:0051179	localization	874	352	40.3	1.23	3.16E-04
deserts sections	GO:0051649	establishment of localization in cell	119	114	95.8	1.44	3.24E-03
	GO:0015031	protein transport	121	112	92.6	1.39	7.02E-03
	GO:0006412	translation	181	164	90.6	1.36	2.34E-03
	GO:0051716	cellular response to stimulus	200	172	86	1.29	7.54E-03
	GO:0044267	cellular protein metabolic process	509	439	86.2	1.29	2.18E-05
	GO:0006793	phosphorus metabolic process	363	292	80.4	1.21	7.84E-03
	GO:0044260	cellular macromolecule metabolic process	1016	803	79	1.19	4.62E-05
	GO:0019538	protein metabolic process	680	533	78.4	1.18	1.70E-03
	GO:0046483	heterocycle metabolic process	754	577	76.5	1.15	4.34E-03
	GO:0044249	cellular biosynthetic process	788	603	76.5	1.15	3.55E-03
	GO:0034641	cellular nitrogen compound metabolic process	760	584	76.8	1.15	3.27E-03
	GO:0006139	nucleobase-containing cpnd metabolic process	684	521	76.2	1.14	8.37E-03
	GO:1901360	organic cyclic compound metabolic process	786	597	76	1.14	5.62E-03
	GO:0006807	nitrogen compound metabolic process	927	702	75.7	1.14	3.19E-03

## C] Representation of specific gene lists within recombinant-active and desert sections



Observed distribution of genes belonging to specific gene lists in recombinant-section (in blue) compared to the theoretical distribution under hypothesis of random distribution of genes in the genome (in green). Representation are in % compared to the total number of genes in the list (n= XXX)



Observed distribution of genes belonging to specific gene lists in recombinant-desert sections (in blue) compared to the theoretical distribution under hypothesis of random distribution of genes in the genome (in green). Representation are in % compared to the total number of genes in the list (n= XXX)

★ : Chi-squared test is significant (p-value < 0.001)

#### Supplemental file 4: Supporting procedures Restriction Site Associated DNA sequencing:

Library was prepared according to Baird et al. protocol (Baird et al. 2008). The main modification was the use of Ampure XP beads for the different purification steps. Briefly, each individual DNA was quantified using microfluorimetry. 500ng to 1µg of DNA was digested with 100 units of *Pst*1\_HF enzyme (NEB, #R3140T). Digested DNA were purified using Ampure XP beads (Beckman Coulter, #A63881) and each DNA was ligated to a different indexed adaptor P1 using concentrated T4 DNA ligase (NEB, # M0202M) at 4°C overnight. Ligation was heat inactivated and purified. Ligated DNA was quantified using microfluorimetry and equally pooled to obtain a 2 µg solution of indexed DNA. The DNA pool was sonicated using a Bioruptor (Diagenode) for 10 cycles (30 sec duty on, 30 sec duty off, high energy mode) and sonication was verified on a high sensitivity DNA chip on a Bioanalyzer (Agilent). Sonicated DNA was run on a 1.5 % agarose, 1X TAE gel for 1h30 hour. A band including DNA fragments from 300 to 500bp was cut using a scalpel blade. DNA was extracted from the agarose using a MinElute kit (QiaGen, #28604). DNA was then repaired using the End-It DNA-end repair kit (Epicentre, # ER81050), purified and adenylated using 10 mM dATP and klenow exo minus (Epicentre, # KL11101K). Once purified, adenylated DNA was ligated to P2 adaptor using concentrated DNA ligase with incubation at 4°C overnight. Ligated material was purified twice to eliminate unligated adaptors and quantified using microfluorimetry. 50ng of material was used in a 12 cycles PCR reaction. Once purified, the library was verified on a high sensitivity chip on a Bioanalyzer and quantified by qPCR (Kapabiosystems kit, #KK4854) on a Lightcycler 480 (Roche). Library was clustered and sequenced on a single lane of Hiseq2500, paired end 2\*125nt mode according to the manufacturer's instructions.

## Supplemental file 5: Targeted markers development

Marker name	Type	Position (Rres 4.0)	Chromosome	Reference*	Forward primer	Reverse primer	Enzyme	Experimental protocol**
F1	SSR	1739541	III	Giraud et al. 2002	GACAAGCAAGCGATAGGAAA	CTTGATAGCACGGACCGACG	-	A
F3	SSR	2137064	IV	Giraud et al. 2002	CATATTCAACCGACCCACAA	TTGAATGATAAGGGCCAGCG	-	A
FG12	SSR	4764723	I	Brygoo et al. 2007	AGATGCCGATGATGGATATG	GTACAAGTGCTTCACGGAAA	-	A
FusSSR12	SSR	6612800	II	Vogelgsang et al. 2009	ATTAGAGAGTGGCGAGAAAAGAG	CAGACATGTTCAAGGGCAATTG	-	A
FusSSR16	SSR	1739572	III	Vogelgsang et al. 2009	CTTTCTCCAACCGAATTGGG	AGTACGGACCGACGACATAACTC	-	A
FusSSR17	SSR	2295474	III	Vogelgsang et al. 2009	AATCCGGTATCCGTACAACCC	AAGAACTACTTCGCTGCCGTG	-	A
FusSSR18	SSR	3989941	III	Vogelgsang et al. 2009	GAATACTCAAGTCCCGTAGCCG	TGCAGAGACCCATAAACGACG	-	A
FusSSR24	SSR	2348254	II	Vogelgsang et al. 2009	CGGACATCTTCACTGGTCC	GCCGCGAATACAGAGTGCTC	-	A
HM2668	SNP (CAPS)	8583643	I	this study	AGCGCAATGTCGTATTTC	GCGCGGTCAAAAGACTTACT	<i>Rsa</i> I	B
HM2724	SNP (CAPS)	8736973	I	this study	GTGTGAGTGCAGGCTCGATA	CAAATTCGCGTCTTCGACA	<i>Mse</i> I	B
HM3204	SNP (CAPS)	4506134	II	this study	CTTATACCCTCGCCCAACAA	CTGTCCGACATCGAGCAATA	<i>Rsa</i> I	B
HM346	SNP (CAPS)	1115938	I	this study	TTCTGGCCCTCAAGAACAA	CACCCCTCTCGTAAGCCATC	<i>Nde</i> II	B
HM3669	SNP (CAPS)	5753381	II	this study	CGTGTGGCTCAGCTCTATG	ATTCTGCGGACGCTAGTGG	<i>Ssi</i> I	B
HM4568	SNP (CAPS)	8422398	II	this study	CGCACAGTCCATTCAATC	CCAAAGCCGAGGTATATTC	<i>Hinf</i> I	B
HM4911	SNP (CAPS)	619077	III	this study	GAAGCCGTCAACCCACTTTA	CACATGCCATTCTGTCATC	<i>Sal</i> I	B
HM53	SNP (CAPS)	231606	I	this study	CAAGGACAGAGAGCCATTCC	TTCTGTATCCCTCTCGTC	<i>Hae</i> III	B
HM5503	SNP (CAPS)	2479119	III	this study	CATGAACCCAGGATGATG	ACCCTCTTGCAAGTGTCCAAC	<i>Hae</i> III	B
HM5903	SNP (CAPS)	3373238	III	this study	GCCAAGCTTACCAGGAG	GTTCCGTAGACACCGCTGT	<i>Nde</i> II	B
HM6616	SNP (CAPS)	607437	IV	this study	TTTGTTCGCTCAAGTCCA	GACGTTGGAGTCTTCTTC	<i>Alu</i> I	B
KSNP100	SNP (KASP)	4542047	I	this study	GGAGATACCTAGGTAGTTAGGCTGAGCG/TB	ACTTGGTATTTTGGAGGGGACGATG	-	C
KSNP101	SNP (KASP)	8987681	II	this study	GATGACTGTCTACCAACCTCGAG/A	CGGAAGAAATTCGCTGCTCTAAA	-	C
KSNP102	SNP (KASP)	2240 SC_3.12 / HG970330		this study	TTCAAGGCGGCTACGTATACAGATAAT/A	GGCCCTATTTAGTGGTGTGGTGTGC	-	C
MsFg103	SSR	2295479	III	Naef et al. 2006	GGTATCCGTACAACCCGATG	TTCTTTGATTGGACCGAGG	-	A
MsFg30	SSR	4704506	II	Naef & Defago 2006	TATGACTGCAGTTTGCTCC	AGTCTCTAGTGCAGCGAGG	-	A

\*Brygoo & Gautier, 2007, Polymorphisme moléculaire des souches de *Fusarium* isolées sur épis de blé et de maïs en France - Récoltes 2003 et 2004. Conférence "Progrès et perspectives de la recherche sur les mycotoxines de *Fusarium* dans les céréales". Arcachon, France.

Giraud et al. 2002, Isolation of eight polymorphic microsatellite loci, using an enrichment protocol, in the phytopathogenic fungus *Fusarium culmorum* Mol. Ecol. Notes 2:121-123

Naef et al. 2006, A microsatellite based method for quantification of fungi in decomposing plant material elucidates the role of *Fusarium graminearum* DON production in the saprophytic competition with *Trichoderma atroviride* in maize tissue microcosms. FEMS Microbiol Ecol 55:211-220

Naef & Defago, 2006, Population structure of plant-pathogenic *Fusarium* species in overwintered stalk residues from Bt-transformed and non-transformed maize crops. E.J.P.P 116:129-143

Vogelgsang et al. 2009, Characterisation of novel *Fusarium graminearum* microsatellite markers in different *Fusarium* species from various countries. E.J.P.P. 123:477-482

A: SSR primers were fluorescently labeled (HEX, NED, FAM) and analyzed in multiplex co-amplification. SSRs amplification were performed using Qiagen Multiplex PCR 2X. Amplicons were subjected to electrophoresis in an automated DNA analyzer ABI3130.

B: PCR amplification were performed Taq DNA polymerase under standard reaction conditions. PCR products were digested with the appropriate nuclease enzyme following the Enzyme manufacturer instructions. Digested samples were analyzed after 2% agarose gel electrophoresis.

C: The KASP genotyping assay was performed using the high Rox master mix according to manufacturer's recommendations (LGC Genomics). Amplifications were conducted on a Roche LightCycler 480.

## **III. Apports technologiques et scientifiques**

### **III.1. *Les mitochondries sont conservées***

Dans le premier chapitre, un taux de polymorphisme significativement et drastiquement réduit a été observé dans la mitochondrie comparé au reste du génome. Cette observation est cohérente avec l'étude de recombinaison qui laisse supposer une hérédité uni-parentale et « maternelle » de cette organelle. Notre travail est le premier à renseigner sur la possible hérédité des mitochondries chez *F. graminearum*. Notons toutefois que pour construire la population recombinante, un des deux parents a « supporté » le croisement (la souche « receveuse », correspondant au mutant) et a été « fertilisée » par la deuxième souche (la souche « donneuse »). Il est donc possible que ce type de croisement, peut-être non-représentatif de la situation naturelle, induise une proportion déséquilibrée des cytoplasmes parentaux (Basse 2010). De nouvelles investigations pour tester l'hypothèse d'une hérédité « maternelle » chez *F. graminearum* sont nécessaires.

### **III.2. *Validation de l'utilisation de souches $\Delta$ Mat***

Ces travaux sont, à notre connaissance, les premiers à proposer la délétion de gènes sexuels pour la construction de carte génétique chez une espèce de champignon homothalique. Il n'est pas possible de contrôler le fait que la délétion d'un idiomorphe sexuel ne perturbe pas la recombinaison méiotique chez cet organisme. Cependant l'absence de biais de ségrégation, la très bonne colinéarité de la carte génétique avec la séquence physique et la similitude des profils de recombinaison avec les expériences précédentes, sont de bons arguments pour valider l'utilisation de cette technique chez *F. graminearum*, et encourager son utilisation potentielle chez d'autres espèces fongiques homothaliques.

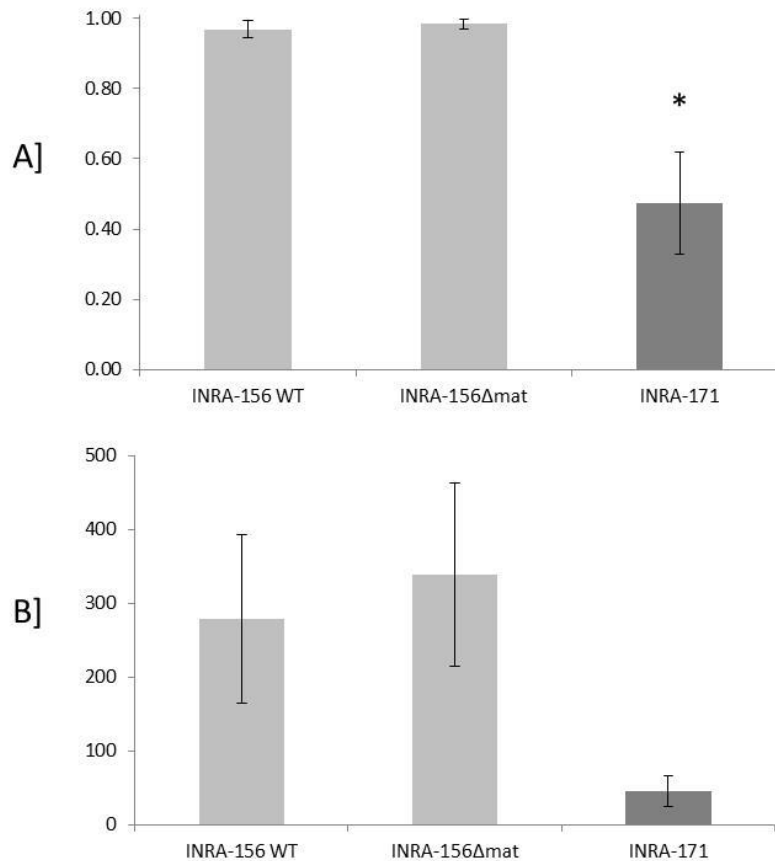
Elle apporte deux intérêts majeurs par rapport aux stratégies précédentes :

1) il est très facile d'isoler des souches recombinantes, avec un taux de récupération proche de 100%.

2) elle ne génère aucun biais de ségrégation dans la descendance.

Alors que l'utilisation de souches génétiquement modifiées pour l'étude de la recombinaison n'apparaît pas comme problématique, elle n'est cependant pas sans risque pour le reste de l'analyse. Les mutants pour ce gène présentent des modifications transcriptionnelles importantes (Kim et al. 2015). Il était donc nécessaire de vérifier que la délétion du gène *Mat 1-2-1* dans la souche INRA-156 n'entraînait pas de modifications majeures des phénotypes qui nous intéressent. Aucune différence significative n'a été observée pour l'agressivité ou pour la production de trichothécènes de type B (Figure 1) entre la souche INRA-156 sauvage et de la souche INRA-156 $\Delta$ Mat. Elles montrent cependant des différences avec la souche INRA-171, l'autre souche parentale utilisée pour la construction de la population recombinante (significatif pour le ratio d'agressivité uniquement, Figure 1). Ce résultat valide donc le choix parental pour la stratégie de détection de QTL.





**Figure CII\_3 :** Comparaisons phénotypiques entre la souche INRA-156 sauvage, la souche INRA-156 $\Delta$ mat délétée pour le gène *Mat 1-2-1* et la souche INRA-171. A] Ratio d'agressivité calculé par le nombre d'épillets infectés divisé par le nombre d'épillets totaux de l'épi. B] Concentration en trichothécènes de type B en  $\mu\text{g/g}$  d'épis secs. \*\* *p*-value associée au t-test d'indépendance inférieure à 0.05.

### III.3. *Les améliorations apportées sur la carte génétique*

L'apport majeur de la carte proposée durant ces travaux réside dans sa complétude puisqu'aucune des cartes précédentes ne propose autant de groupe de liaisons génétiques que de chromosomes. Elle présente de plus le plus grand nombre de marqueurs (Tableau CII\_3). L'utilisation du RAD-seq pour le génotypage explique très certainement les améliorations apportées par rapport à la carte génétique proposée précédemment (Gale et al. 2005). Bien que construite à partir d'un croisement interspécifique, *F. graminearum* x *F. asiaticum*, la comparaison peut aussi être faite avec la carte initialement proposée par Jurgenson et al. (2002), révisée et alignée ultérieurement sur le génome de référence par Lee et al. (2008, Tableau CII\_3). Il est

intéressant de noter que le nombre de marqueurs génotypés pour la construction de la carte développée par Jurgenson et al. (2008) est proche de celui de notre analyse mais n'a pourtant pas permis de compléter la carte. Dans leur cas, les réarrangements chromosomiques dus au croisement « interspécifique », et les biais de ségrégations produits durant la sélection de la population recombinante expliquent très certainement ce résultat.

**Tableau CII\_3** : Comparaison avec les cartes génétiques précédemment construites.

		Jurgenson et al. 2002	Lee et al. 2008	Gale et al. 2005	Carte développée durant nos travaux
<b>Nombre de marqueurs génotypés</b>		1048	115	235	1330
<b>Nombre de loci recombinants</b>		468	115	235	483
<b>Nombre de groupes de liaisons</b>		9	6	9	4
<b>Dimension de la carte (cM)</b>		1286	1140	1234	1497
<b>Génome de référence utilisé</b>		Non aligné	FGDB (36Mb)		RRES v4.0 (38Mb)
<b>Efficacité d'alignement</b>		ND	99.20%	99.80%	99.00%
<b>Taux de recombinaison (cM/Mb)</b>	Chrom-I	ND	28.1	31.9	37
	Chrom-II	ND	42.3	45.3	58.1
	Chrom-III	ND	30.9	30.3	32.2
	Chrom-IV	ND	25.6	29.2	30.7

### III.4. *L'apport du génotypage à haut-débit*

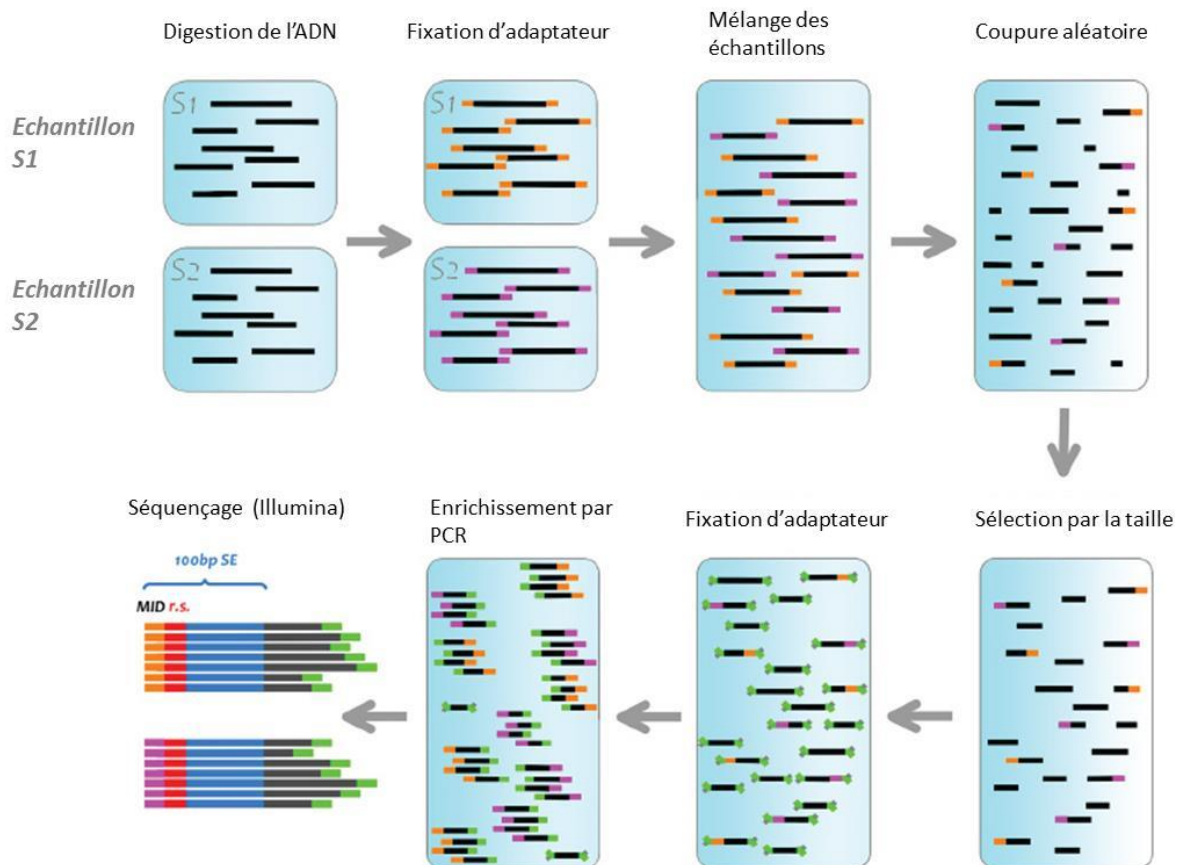
Plusieurs méthodologies de génotypage à haut débit basées sur l'utilisation de séquenceur de nouvelle génération ont été récemment développées, tel que le génotypage par séquençage à faible couverture, le génotypage par séquençage associé aux sites de coupures (RAD-seq) ou le génotypage par séquençage de bibliothèques à représentation réduite (RRL) (Davey et al., 2011 pour revue).

### III.4.1. Le principe du RAD-seq

Dans notre cas, nous avons préféré l'utilisation du RAD-seq, basé sur le séquençage de cibles génomiques plutôt que de génomes entiers. Cette stratégie en réduisant la complexité des génomes, permet d'augmenter la capacité d'individus génotypés par réaction de séquençage et donc de diminuer les prix du génotypage et/ou augmenter le nombre de marqueurs génotypés (Baxter et al., 2011; Davey et al., 2011). Lors du RAD-seq, les cibles génomiques correspondent à des fragments de restriction issus de la digestion du génome par une/des enzyme(s) de restriction (Baxter et al., 2011; Davey et al., 2011). La préparation des banques qui précède le séquençage d'ADN se passe en différentes étapes (Figure CII\_4). Ces étapes peuvent être résumées par :

- 1) la digestion du génome (avec l'enzyme *PstI* dans notre cas).
- 2) la liaison d'adaptateur souche-spécifique au niveau des sites de coupures (adaptateur P1).
- 3) le mélange de l'ADN de toutes les souches.
- 4) le découpage aléatoire des fragments d'ADN par sonication et la sélection de fragments sur des critères de tailles (comprises entre 300 et 500 pb dans notre cas).
- 5) la liaison d'un deuxième adaptateur à l'extrémité non liée.
- 6) la sélection des fragments contenant les deux adaptateurs et le séquençage de petits fragments à partir du premier adaptateur (125 nucléotides dans notre cas), et aussi à partir du deuxième adaptateur lorsque le séquençage est réalisé sur les «extrémités paires» (adaptateur P2, 125 nucléotides là aussi).

Après séquençage, les séquences sont ensuite réassignées à chaque souche par la lecture de l'adaptateur P1. Dans notre cas, 93 souches issues du croisement, ainsi que les deux souches parentales, ont été multiplexées et séquencées sur une ligne de séquenceur HiSeq™ 2500 d'Illumina®.



**Figure CII\_4** : La stratégie de génotypage par séquençage (illustré et proposé par Integrated DNA technologies®)

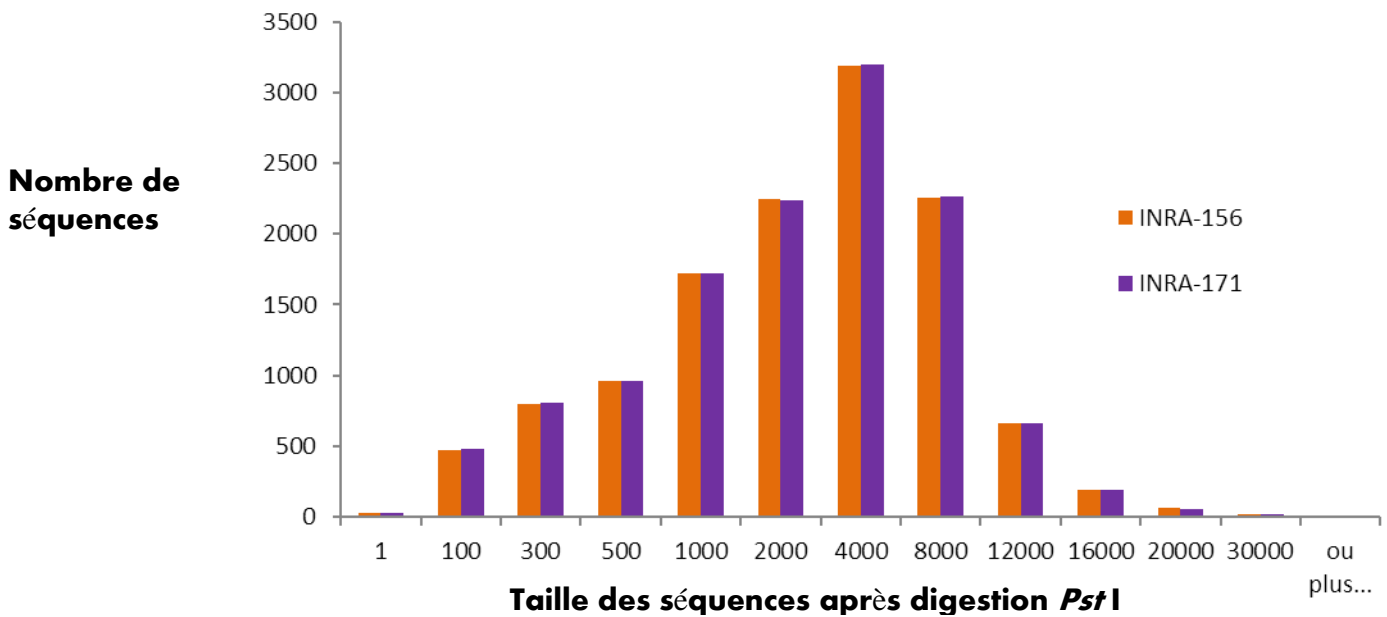
Comparée aux approches de génotypage plus classiques, l'analyse des données de RAD-seq est plus friande en calcul informatique. Cette technologie est récente et fut proposée pour la première fois en 2011. Elle a été élaborée à l'université d'Edimbourg par Simon Baxter, dans l'équipe de recherche dirigée par Mark L. Blaxter (Baxter et al., 2011). Les chercheurs du Roslin Institute (Edimbourg, Ecosse), dont Ross D. Houston, ont participé activement à son utilisation et à son développement chez des espèces à intérêt agronomiques (Tsai et al. 2016; Gonen et al. 2014; Houston et al. 2012). L'acquisition des expertises en lien à l'utilisation et l'interprétation des données de RAD-seq, non disponibles au laboratoire, a été menée directement au sein de l'équipe de Ross D. Houston dans le cadre d'un échange financé par le département « Santé des Plantes et Environnement » de l'INRA et le consortium Agreenium.

### III.4.2. Quelques facteurs importants à considérer lors du génotypage par RAD-seq

Afin d'adapter au mieux la technologie, l'expérimentation de RAD-seq fut préalablement simulée *in silico* en utilisant les données des génomes parentaux. Cela a permis notamment de prédire de manière grossière le nombre de marqueurs potentiel et de valider le choix de l'enzyme de restriction utilisée durant la préparation des banques d'ADN (Figure CII\_5 et Figure CII\_6). La comparaison entre le nombre prédit de marqueurs et le nombre observé nous a permis d'avoir un regard critique sur cette approche et d'identifier quelques facteurs limitant/importants à prendre en considération.

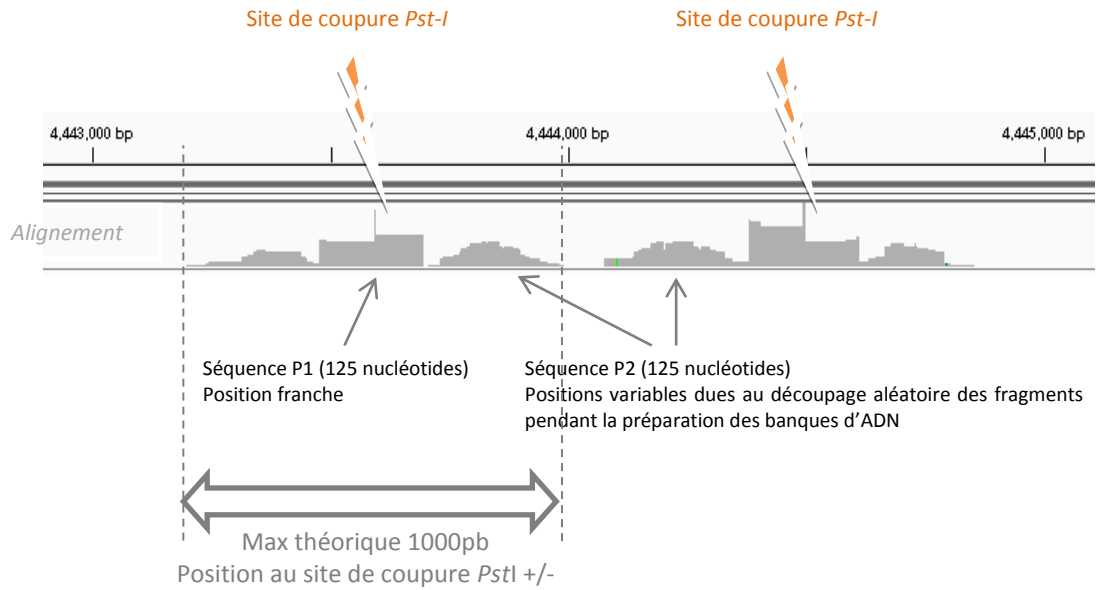
#### ➤ La prédiction du nombre de marqueurs

Pour être sélectionnés, les fragments après sonication doivent avoir une taille comprise entre 300 et 500 pb après coupure aléatoire. Sur des fragments de restriction trop courts (inférieur à 1 000 pb) la sonication risque de générer de nouveaux fragments présentant des tailles inférieures à 300pb qui ne seront donc pas sélectionnés. A l'inverse, des fragments de restrictions trop grands ne seront pas non plus sélectionnés, même après coupure aléatoire. La digestion *in silico* des génomes des souches parentales par l'enzyme de restriction *PstI* génère des fragments de tailles variables, dont la majorité a une taille comprise entre 1 000 pb et 8 000 pb (Figure CII\_5).

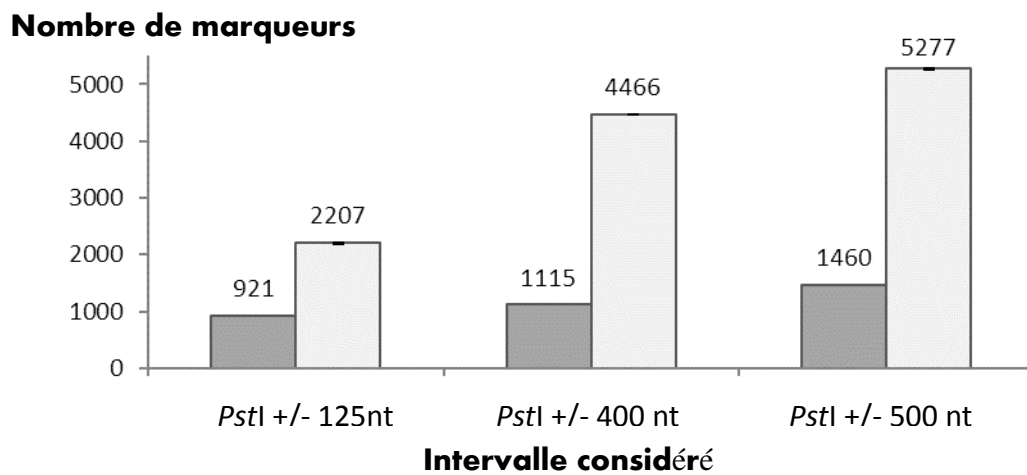


**Figure CII\_5 :** Distribution de la taille des fragments après la digestion simulée des génomes de la souche INRA-156 et INRA-171 avec l'enzyme de digestion *Pst*I

Après séquençage, on observe typiquement un ensemble d'alignement franc et uniforme de part et d'autre du site de coupure de RAD-seq et des alignements formant une courbe de Gauss pour la deuxième extrémité (Figure CII\_6). Cette courbe de Gauss reflète la coupure aléatoire des fragments (Davey et al. 2013). En théorie, une séquence de 1 000 pb est séquencée par site de coupure dans nos conditions et correspondant à la taille des fragments sélectionnés (500pb en amont et en aval du site de restriction). Ne connaissant pas la taille des fragments réellement séquencés, le nombre de marqueurs que nous aurions pu identifier après digestion a été réalisé en prenant en compte des tailles de fragments variables (125pb en amont et en aval du site de coupure, 400 pb en amont et en aval du site de coupure, et 500 pb en amont et en aval du site de coupure). Suivant ces prédictions, entre 2,200 et 5,300 marqueurs auraient dû être identifiés (Figure CII\_7).



**Figure CII\_6 :** Exemple de profils retrouvés après alignement des séquences issues du RAD-seq de la population recombinante sur le génome de référence (RRES v4.o).

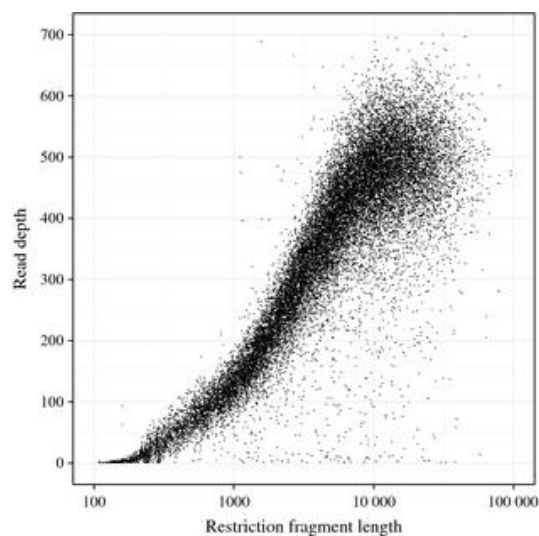


**Figure CII\_7 :** En gris foncé, le nombre de marqueurs génotypés observé dans l'intervalle entourant les sites de coupures *Pst*I prédits. En blanc, le nombre théorique de marqueurs, calculé à partir du polymorphisme des génomes parentaux.

### ➤ La comparaison entre l'attendu et l'observé

A terme, l'analyse a permis l'identification de 1 461 marqueurs polymorphiques dont 1,306 ont été gardés pour la suite de l'analyse après filtration sur la qualité. De manière intéressante, 1 460 des 1 461 marqueurs ont été identifiés dans les intervalles entre 500 pb en amont et en aval des sites de coupures de *Pst*I. En considérant donc que les marqueurs identifiés sont localisés dans cet intervalle, nous aurions pu identifier 5 277 marqueurs. Quels sont donc les facteurs qui pourraient expliquer cette réduction du nombre de marqueurs entre l'attendu et l'observé ?

- L'efficacité de l'enzyme de restriction n'est pas totale et donc le nombre de sites de coupures a été surévalué : ceci ne semble pas être le cas puisque la majorité des sites de coupures prédits (99%) sont associés à des alignements obtenus après séquençage.
- La profondeur de séquençage à la position n'est pas suffisante pour identifier les marqueurs. La taille des fragments de restriction a des conséquences importantes sur la profondeur de séquençage au *locus* (Figure CII\_8). De fait, il semblerait qu'un biais soit introduit lors de l'étape de sonication. En effet, son efficacité varie en fonction de la taille des fragments de restriction, amenant à des biais sur les profondeurs de séquençage (Davey et al. 2013). Ce biais est difficilement quantifiable et prédictible.



**Figure CII\_8 :** Relation entre la taille des fragments de restriction et la profondeur de séquençage observée à partir d'une expérience de RAD-seq utilisant *Pst*I chez *C. elegans*. Retrouvés à partir de Davey et al., 2013.



- Des sites des coupures polymorphes entre les génomes parentaux : 204 fragments de restrictions sont potentiellement différents entre les génomes parentaux.
- La variabilité de l'efficacité de séquençage entre souches qui pourrait réduire les chances de génotyper un marqueur faute d'information sur l'ensemble de la descendance. Le nombre de séquences assignées pour chaque souche varie de 308,147 à 10,818,740. Les origines de cette variabilité sont certainement multiples (lors de la préparation des banques ou lors du séquençage). Ce biais peut être là aussi difficilement prédit.

Globalement plusieurs paramètres peuvent expliquer cette diminution entre l'attendu et le réel, et l'accumulation de tous ces paramètres explique certainement la différence observée dans notre cas.

L'analyse des données de RAD-seq et l'appel des marqueurs a été réalisé en utilisant le logiciel Stacks (Catchen et al. 2013). Ce logiciel prend en charge toute l'analyse (hormis l'alignement des séquences sur le génome de référence) et fournit un set de marqueurs robustes et de bonnes qualités. Une critique à son utilisation est toutefois son opacité puisque de nombreuses opérations ne sont pas « visibles » ni modifiables par l'opérateur. De fait, une meilleure visibilité des opérations aurait permis de corriger ou d'adapter l'appel des marqueurs afin d'optimiser le nombre de marqueurs identifiés. L'identification de marqueurs par des procédés de détection de variants, tels que développés lors du premier chapitre, pourrait être une alternative intéressante pour augmenter le nombre de marqueurs génotypés si nécessaire.

### III.4.3. Les implications pour la suite de l'analyse

Dans notre cas, un nombre de marqueurs plus important aurait permis notamment d'identifier avec une meilleure précision la position des crossovers sur le génome. Dans la perspective de la stratégie de détection de QTL, le nombre de marqueurs obtenus durant notre analyse ne constitue pas un facteur limitant, puisque beaucoup de marqueurs apportaient la même information dans la descendance (marqueurs co-ségréant). Le facteur limitant pour la résolution de la carte génétique est, dans ce cas, le nombre de souches génotypées. Ce dernier étant relativement faible ( $< 100$ ), la fréquence de recombinaison entre marqueurs physiquement proches est limitée. Nous disposons au laboratoire d'environ 150 individus issus de ce croisement. La construction de la carte a été réalisée à partir de 88 d'entre eux. Le séquençage d'un plus grand nombre de souches pourrait permettre d'apporter une meilleure résolution à notre analyse.

Les marqueurs co-ségréant ne sont pas utiles pour la construction de la carte génétique, et peuvent même parfois en alourdir ou en perturber le processus. Cependant, en considérant que leur ordre génétique respecte leur ordre physique décrit dans le génome de référence, l'alignement de ces marqueurs sur la référence permet d'affiner la position d'un QTL. En prenant en compte ce dernier point, notre carte génétique, avec une distance médiane entre marqueur de 13 kb, apporte une résolution très prometteuse pour la détection de QTL.

### IV. Matériels et méthodes

*Sont décrits ici le matériel et méthode utilisé dans le paragraphe III.*

- Phénotypage *in planta* et dosage des mycotoxines dans l'épi

Ces procédures expérimentales ont été développées dans le Chapitre I.I.4.

- Prédiction du nombre de marqueurs

Les génomes parentaux ont été reconstruits avec l'outil FastaAlternateReferenceMaker de GATK (v 2.4) (McKenna et al. 2010) en utilisant la diversité génétique des souches parentales détectée dans le premier chapitre (SNPs uniquement). Ces génomes parentaux ont ensuite été digérés *in silico* avec l'outil Restrict de la suite d'outils bio-informatiques en ligne Emboss (Rice et al. 2000). Les intervalles aux sites de restriction (+/- 125 nt; +/- 400 nt; +/- 500 nt) ont été délimités manuellement sous Excel. L'outil intersectBed (v2.13.3) de la suite BedTools (Quinlan & Hall 2010) a ensuite été utilisé pour identifier les fragments de restriction contenant des variants génétiques (et correspondant à de potentiels marqueurs). L'outil CallableLoci de GATK (v 2.4) a été utilisé pour déterminer les régions du génome de référence couvertes par des séquences issues du RADseq. La comparaison entre sites de coupure prédits et régions couvertes par des séquences a été réalisée en utilisant l'outil intersectBed (v2.13.3) de la suite BedTools (Quinlan & Hall 2010).

### **V. Conclusion du chapitre 2**

Les objectifs de ce chapitre étaient de construire le matériel biologique (la population recombinante) et analytique (la carte génétique) pour mener à bien la détection de QTL.

En conclusion, ces différents points ont été réalisés avec succès. Une originalité de ces travaux réside dans les améliorations techniques proposées, à la fois par la délétion de gène sexuel pour éviter les limitations due à l'homothallisme, mais aussi par l'apport de technologie de génotypage à haut-débit pour la construction d'une carte génétique de qualité. Dans la continuité de nos découvertes sur l'organisation génomique du polymorphisme chez *F. graminearum*, cette carte génétique nous a permis d'étayer un peu plus les connaissances sur les mécanismes régulateurs de la dynamique des génomes de *F. graminearum* et des champignons, par une étude approfondie de la recombinaison méiotique. Nos résultats montrent une forte corrélation entre les profils de polymorphismes et les profils de recombinaison à l'échelle des chromosomes, et suggèrent un rôle important de l'architecture chromosomique de cette espèce pour son potentiel d'adaptation.

La construction de ces outils nous a permis d'aborder notre objectif principal qui visait à étudier les relations entre les différents caractères en lien avec l'agressivité au sein d'une population recombinante de *F. graminearum* et de caractériser les bases génétiques de l'agressivité dans cette population.

**Chapitre 3 :**  
**Elucidation des bases génétiques de**  
**la variation de production de**  
**mycotoxines et d'agressivité**  
**chez *F. graminearum***



# **I. Une approche combinée de détection de QTL et de génomique comparative identifie une mutation sur le gène *FgVeA* avec un effet majeur sur l'agressivité de *F. graminearum***

Les résultats en lien à ce sous-chapitre sont rédigés sous la forme d'un article scientifique qui sera soumis ultérieurement à *Molecular Plant-Microbe Interaction*.

## **Combined QTL detection and comparative genomic analysis identify a mutation in *FgVeA* gene with major effects on aggressiveness in *Fusarium graminearum***

B. Laurent<sup>1</sup>, M. Moinard<sup>1</sup>, C. Spataro<sup>1</sup>, F. Touza-Soage<sup>1</sup>, S. Chéreau<sup>1</sup>, E. Zehraoui<sup>1</sup>, R. Blanc<sup>2</sup>, P. Lasserre-Züber<sup>2</sup>, N. Ponts<sup>1</sup>, C. Barreau<sup>1</sup>, M. Foulongne-Oriol<sup>1\*</sup>

<sup>1</sup> MycSA, INRA, 33882, Villenave d'Ornon, France

<sup>2</sup> GDEC, INRA, 63100 Clermont-Ferrand, France

\*Corresponding author.

Mailing address: INRA, UR1264 Mycologie et Sécurité des Aliments, bâtiment Qualis, 71 avenue Edouard Bourleaux, CS 20032, F-33882 Villenave d'Ornon cedex, France

Phone: +33 (0)5 57 12 26 35

Fax: +33 (0)5 57 12 25 00

E-mail: mfoulong@bordeaux.inra.fr

Keywords: *Gibberella zeae*, fungal pathogen, *Fusarium* Head Blight, adaptive potential, secondary metabolism, Velvet complex, *FgVe1*, Mycotoxin, Deoxynivalenol.

**Abstract:** *Fusarium graminearum* is the most frequent causal agent of the *Fusarium Head Blight*, a cereal disease spread throughout the world, and leading to grain contamination by stable and harmful mycotoxins. Resistant cultivars are frequently used to control this pathogen in the fields. However, several studies suggest that *F. graminearum* is doted of a high adaptive potential, supported among other reasons by a remarkable variation of aggressiveness between isolates. In this study, we aim to characterize the genetic basis of aggressiveness variation in an experimental population. Visual symptoms on inoculated wheat spikes, mycotoxin accumulation, conidiation, fungal spreading and biomass production were measured, either or both *in vitro* and *in planta* conditions. Most of the traits followed a bimodal distribution across the progeny and were positively correlated. Fungal biomass production *in vitro* was negatively correlated to other traits. One major pleiotropic quantitative trait locus was detected on the chromosome I and explained up to 90% of the variance for visual symptom or 91% for mycotoxin in wheat spike. The confidence interval delineated at this position was aligned on the reference genome, spanning 1.2 megabases and containing 428 genes. The analysis of the polymorphism of parental sequences in this region, along with the polymorphism of previously characterized isolates, has allowed us to reduce this number to four candidate genes. The gene *FgVeA*, previously shown to be implicated in pathogenicity and secondary metabolism in several fungi including *F. graminearum*, was of the highest priority. Functional experiments further validate that a non-synonymous mutation within the Velvet domain of *FgVeA* has large implications for overall aggressiveness variation. The present study emphasizes the relevancy of QTL mapping strategy to study pathogenic fungi and is the first to associate an allelic polymorphism affecting the master regulator VeA protein in fungi. In conclusion, this analysis gives new information about the regulation of aggressiveness variation in *F. graminearum* that we supposed to be mediated, in this scenario, by contrasted ability of mycotoxin production, along with important developmental modification like hyphal spreading.



### **Introduction**

The *Fusarium head blight* (FHB) or *Fusarium ear blight* is a devastating disease affecting cereals, and is considered to be a limiting factor for global production (Stack, 2000). In addition to yield losses due to fungal infection, grains are often associated with mycotoxins. In order to limit the economic and public health concerns associated with the development of FHB in the fields, deployment of resistant cereal varieties is a frequent and efficient strategy of control (Buerstmayr et Lemmens, 2015; Ferrigo *et al.*, 2016; Terzi *et al.*, 2014). No major gene of resistance has been discovered in cereals. The resistance of FHB is based on the action of several quantitative loci conferring partial resistance. Over one hundred Quantitative Trait Loci (QTL) have been identified in wheat, and less than 30 in barley. Some of these have been already introduced in agronomical varieties of wheat and barley (Bai et Shaner, 2004; Buerstmayr *et al.*, 2009; Lv *et al.*, 2014). Because the molecular basis of plant qualitative resistance relies on the action of single genes, quantitative resistances, which are based on several genes, are supposed to be more durable (Zhang *et al.*, 2013). Nevertheless, quantitative resistance are also facing the adaptive potential of several fungal pathogens that erode their effect overtime (Andrivon *et al.*, 2007; Caffier *et al.*, 2014; Caffier *et al.*, 2016; Delmas *et al.*, 2016; Dowkiw *et al.*, 2010). This erosion is often associated to quantitative variations of pathogenicity within populations of pathogens, referred to as aggressiveness (Pariaud *et al.*, 2009; Van der Plank, 1963).

The ascomycete and hemibiotrophic *Fusarium graminearum* (*sensu stricto*, teleomorpe *Gibberella zeae*) is the most frequently encountered agent of the *Fusarium* Head blight in the world (Van der Lee *et al.*, 2015; Trail, 2009). *F. graminearum* reproduces through both clonal reproduction and sexual outcrosses, and spores can be dispersed from short to long-range scales (Osborne et Stein, 2007). These life-cycle features create frequent new haplotypes that can widely and rapidly spread under favorable conditions (McDonald and Linde, 2002). Therefore, populations of this pathogen show important levels of diversity and gene flow within populations (Akinsanmi *et al.*, 2006; Gale *et al.*, 2007; Keller *et al.*, 2014; Prussin *et al.*, 2013; Schmale Iii *et al.*, 2006; Talas *et al.*, 2012a). This diversity is accompanied by important variations of aggressiveness between and within field isolates (Cumagun and Miedaner, 2003; Gilbert *et al.*, 2002; Talas *et al.*, 2012a). In some cases, a rapid spread of more aggressive isolates has already been reported in the fields (Ward *et al.*, 2008). Therefore, a detail description of the genetic basis of aggressiveness, and more

generally about any history-life traits that could be involved in pathogen evolution, has to be conducted in order to anticipate pathogen evolution and strengthen our efforts to keep up with this “arms race” (Anderson *et al.*, 2010; Lannou, 2012).

In addition to the important health issues caused by their presence in foodstuff and feedstuff, mycotoxins seem to play an important role in the pathogenicity of *Fusarium* species. This has been especially well demonstrated for deoxynivalenol (DON) in wheat, a molecule belonging to the trichothecene of type B (TCTB) family, which can be produced by *F. graminearum* (Bai *et al.*, 2002). DON is considered as an aggressiveness factor since mutated strains for the biosynthesis of DON are still pathogenic but cannot spread in other spikelets than the one inoculated: (Bai *et al.*, 2002). DON is not the only TCTB produced by *F. graminearum*. Classically, strains are able to produce 15-acetyldeoxynivalenol or 3-acetyldeoxynivalenol in addition to DON. Some strains, however, are not able to produce DON and acetylated forms, but produce NIV and acetylated forms instead (Alexander *et al.*, 2011; Lee *et al.*, 2002; Varga *et al.*, 2015). The role of TCTBs other than DON for pathogenicity is not clearly understood (Maier *et al.*, 2006) even though strains that produce NIV instead of DON are suspected to be less aggressive in wheat (Maier *et al.*, 2006). Recently, 50 single nucleotide polymorphisms (SNPs) sparsely located on the genomes of 119 German field isolates, of which 26 genes were affected, have been associated with aggressiveness variation (Talas *et al.*, 2016). Although the molecular mechanisms involved in this polymorphism have not been characterized yet, this work suggests that other genes than the one implicated in TCTB biosynthesis are probably involved in aggressiveness variation. On a broader scale, genome-mining efforts from *F. graminearum* reference genome revealed a large repertoire of genes that could be involved in pathogenicity (Brown *et al.*, 2012; Cuomo *et al.*, 2007; King *et al.*, 2015; Sieber *et al.*, 2014). Additional studies revealed that the genetic diversity observed within population actually reflects a large source of variations along the genome and affects the genes potentially involved in pathogenicity (Laurent, Moinard, *et al.*, 2016; Walkowiak *et al.*, 2015). What is the consequence of this polymorphism on aggressiveness variation? During a preliminary study, we noticed that field isolates showed various abilities of DON and 15-ADON production *in vitro* or in wheat, and provoked symptoms of different intensities on spikes (Laurent, Moinard, *et al.*, 2016; Pinson-Gadais *et al.*, 2013). The different genetic factors previously described to be implicated in aggressiveness variation (genes implicated in DON biosynthesis and the ones described by Talas *et al.*, 2016) were not retrieved in those isolates (Laurent, Moinard, *et al.*, 2016),

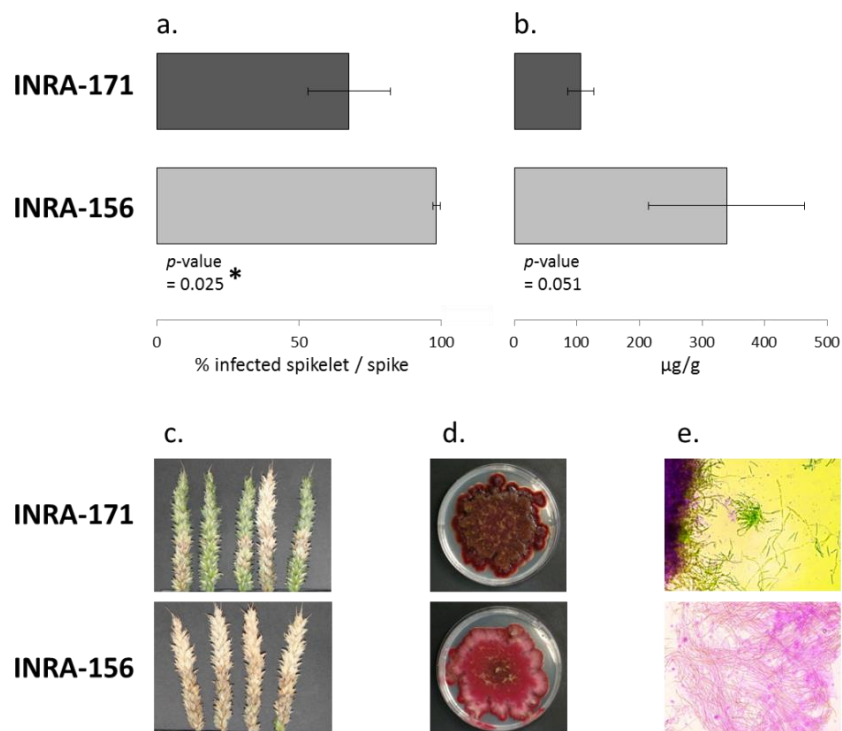
implying that other genetic factors could exist. In the present study, we sought to test this hypothesis by developing a quantitative genetic approach, and characterize the genetic basis of this variation using an experimental population constructed from two of these strains. This strategy aims to identify the genetic factors involved in trait variation, their location on the genome, and their effect (Laurent, Balesdent, *et al.*, 2016).

## Results

### Phenotyping and distribution across the offspring

A recombinant population was constructed from the cross of INRA-156( $\Delta Mat$ ) and INRA-171 strain. The strain INRA-156 $\Delta Mat$  was constructed after the deletion of the *Mat1-2-1* gene, in order to bypass homothallism restriction (Laurent, Palaiokostas, *et al.*, 2016). For clarity purposes, the strain INRA-156 $\Delta Mat$  will be named INRA-156. Both parental strains were able to infect wheat but INRA-156 induces more severe visual symptoms than INRA-171. Whereas both strains produce DON and 15-ADON, the amount of production was greater for INRA-156 (Figure 1). Only the difference observed for visual symptoms was significant (*t*-test, *p*-value < 0.01). Other contrasted phenotypes were observed on the basis of morphological aspects when grown *in vitro* condition (Figure 1). Thus, INRA-156 exhibited bright and aerial mycelia while INRA-171 showed thicker and darker mycelia with rare aerial mycelia. The typical morphologies differentiating the parental strains were easily distinguished within the progeny strains and followed a Mendelian inheritance (Chi-square test, *p*-value < 0.01). Morphological differences were also observed under microscope: INRA-156 showed long and unbranched hyphae, and INRA-171 showed hyper-branched and dense hyphae (Figure 1). In total, seven pathogenicity-related traits were phenotyped on the progeny (n=94), four of which were assayed *in vitro* conditions and correspond to fungal biomass, radial growing, conidia production (asexual spores) and TCTB production. Three phenotypes were assayed on wheat, and correspond to visual symptom measurement, TCTB and ergosterol concentration in spikes. Two experiments were conducted per trait and were significantly correlated except for ergosterol measurements (Table 1). The highest correlations between the two experiments were observed for TCTB production, assayed *in vitro* (Spearman rank order correlation  $Rho = 0.80$ ) and *in planta* ( $Rho = 0.80$ ), and for visual symptoms ( $Rho = 0.74$ , Table 1). The values of TCTB production *in vivo* measured for the second experiment was largely greater than the ones observed for the first experiment (Table 1). Symptoms could affect from 20% to 100% of the spike and TCTB concentrations (mainly DON and 15-ADON in smaller amounts) in spike could vary from 0 to 2,253  $\mu\text{g}$  per g of dried matter (Table 1). The distribution of the traits was normal for biomass production, ergosterol concentration (E1) and conidia production (*p*-value > 0.01), but was not for the other traits (Figure 2). The distribution of the other traits suggests a bimodal-like distribution (Figure 2, Supplemental file 1). For most of the traits, transgressive phenotypes were

observed (Table 2). For example, the TCTB concentration of wheat spike inoculated by A.146 strain showed a near 7-fold increase compared to TCTB concentration of wheat spike inoculated by INRA-156 strain and corresponding to the “toxinogeneous” parental strain.

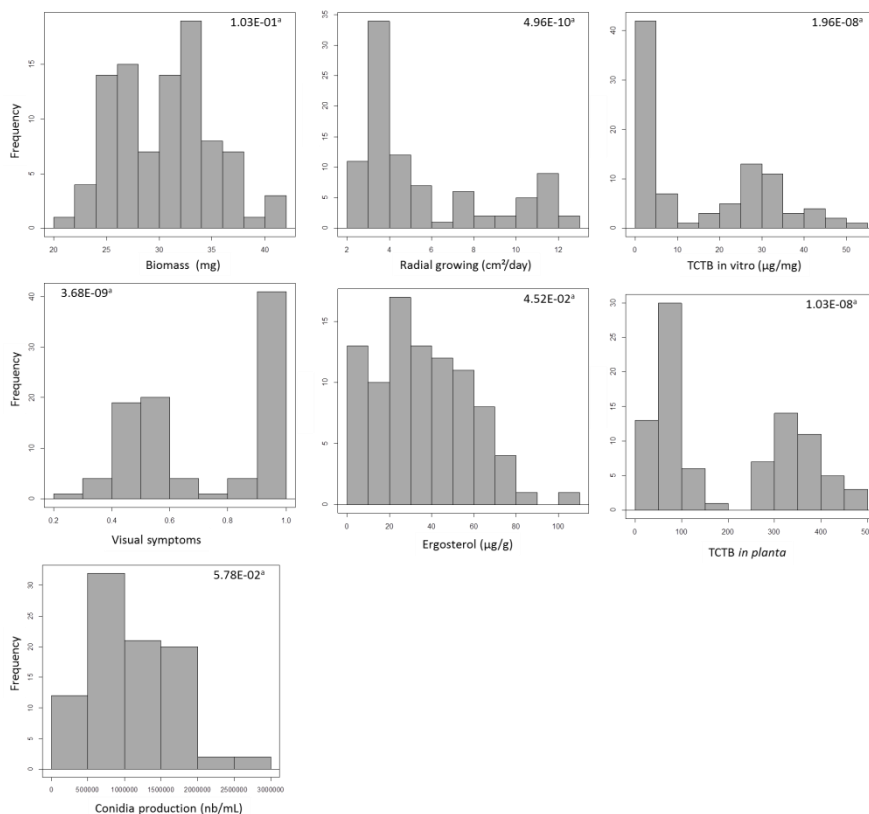


**Figure 1:** Phenotypic characteristics of parental strains, *p*-value are indicative of independent *t*-test, dark star indicates significant *p*-value. a. visual symptoms measured at 22 days post inoculation (dpi). b. Trichothecene of type B quantification in µg of mycotoxin per gram of dried spike. c. Pictures of Royssac variety wheat flowers inoculated with INRA-171 or INRA-156 strains. d. Pictures of the two parental strains grown of PDA medium. e. Microscopic view of the vegetative mycelia of the two parental strains.

**Table 1:** Minimal, maximal, median and mean values within the progeny and for the parents and Spearman rank order correlation between the experiments.

Traits	E1				E2				Correl. between exp.	Parental values	
	Min	Max	Median	Mean	Min	Max	Median	Mean		INRA-156	INRA-171
Biomass (mg)	20.5	41.2	31.3	30.5	23.7	37.1	31	31	0.28 *	24.9	32
Radial growing (cm <sup>2</sup> /day)	2.4	12.4	5.6	4	0.9	10.3	3.7	4.4	0.39 *	4.7	2.8
Conidia (spores/mL)	6.00E+04	2.60E+06	1.00E+06	1.10E+06	2.20E+05	2.90E+06	1.30E+06	1.30E+06	0.39 *	9.60E+04	5.90E+04
TCTB <i>in vitro</i> (µg/mg)	0.05	54.2	8.01	15.4	0.62	92.1	8	31.9	0.80 *	33.3	1
Visual symptom (%)	29	100	63.5	72.6	20	100	57	65.8	0.74 *	98.4	67.6
TCTB <i>in planta</i> (µg/g)	27.56	493.8	111.5	194.7	0	2,253	186.9	329.8	0.64 *	339.2	105.9
Ergosterol (µg/g)	0	107.1	34.1	35.7	0	240.7	51.1	70.3	0.2	63	40.9

**Figure 2:** Distribution of the different phenotypes assayed in the offspring.



## Correlation between traits

A significant correlation was revealed between all traits in one or both experiments (Table 2). The highest level of correlation was revealed between visual symptoms and TCTB production (measured *in planta*,  $Rho = 0.78$  for E1 and  $Rho = 0.73$  for E2). TCTB production measured in spike was well correlated with TCTB production from *in vitro* cultures ( $Rho = 0.72$  for E1 and  $Rho = 0.66$  for E2), suggesting that toxinogeneous potentials are respected between *in vitro* and *in planta* conditions. Visual symptom was moderately correlated with ergosterol concentration ( $Rho = 0.46$  for E1 and  $Rho = 0.61$  for E2) and radial spread of strains measured *in vitro* ( $Rho = 0.62$  for E1 and  $Rho = 0.50$  for E2). Fungal biomass production measured from *in vitro* cultures was negatively correlated with other developmental-related traits, namely radial growing ( $Rho = -0.66$ ) and ergosterol concentration ( $Rho = -0.37$ ). Correlation of conidia production with other traits was more contrasted, even though this trait seems to be correlated with TCTB production measured *in planta* and *in vitro* ( $Rho = 0.54$  and  $0.51$  respectively, significant for E1) and with visual symptoms on spike ( $Rho = 0.52$ , significant for E1).

**Table 2:** Spearman rank order correlation values between the different phenotypes, for the different trials. The values for the first trial are in the upper right of the diagonal of the table and the values for the second trial are in the lower left diagonal of the table. "\*" indicates significant values ( $p$ -values < 0.05)

	Biomass	Radial growing	TCTB <i>in vitro</i>	Visual symptom	TCTB <i>in planta</i>	Ergosterol	Conidia
Biomass		-0.52*	-0.78*	-0.66*	-0.58*	-0.37*	-0.43*
Radial growing	-0.11		0.70*	0.62*	0.61*	0.38*	0.35*
TCTB <i>in vitro</i>	-0.44*	0.70*		0.76*	0.72*	0.44*	0.51*
Visual symptom	-0.18	0.50*	0.69*		0.78*	0.46*	0.52*
TCTB <i>in planta</i>	-0.24*	0.45*	0.66*	0.73*		0.40*	0.54*
Ergosterol	-0.02	0.28*	0.43*	0.61*	0.47*		0.18
Conidia	0.11	0.21	0.00	0.24*	0.12	0.24*	

## Analysis of variance and heritability estimate:

ANOVA from the combined data of the two experiments revealed highly significant genotype effects for all the traits (Table 3). The models also revealed significant experiment effect and interactions by genotype effect (G x E, Table 3). The high and significant Spearman rank order correlations revealed between experiments for TCTB productions and

visual symptoms (Supplemental file 1, Table S1) suggested that the experiment effect is more likely due to changes in trait magnitude between the two experiments rather than to inconsistent behavior of genotypes. Visual symptoms and TCTB production *in vitro* show the best heritability (92.5% and 79.4% respectively, Table 3). The lowest heritabilities were observed for ergosterol concentration (23.0%) and biomass production *in vitro* (31.9%, Table 3). Overall, this data suggests that the genotype effect is the predominant source of variation for the values of the traits. Therefore, we focus on the identification of the genetic factors underlying the phenotypic variation through QTL detection. However, given the significant experiment effect, we treated QTL detection by separate experiments for all the traits.

**Table 3:** Analyses of variance and heritability estimates ( $h^2$ ) for the aggressiveness-related traits measured *in planta* and *in vitro* from combined experiment data (two experiments per trait).

Traits	Visual symptoms (%)		TCTB <i>in planta</i> ( $\mu\text{g/g}$ )		Ergosterol ( $\mu\text{g/g}$ )		Biomass (mg)		TCTB <i>in vitro</i> ( $\mu\text{g/mg}$ )		Radial growing ( $\text{cm}^2/\text{day}$ )	
	<i>df</i>	MS <sup>a</sup>	<i>df</i>	MS <sup>a</sup>	<i>df</i>	MS <sup>a</sup>	<i>df</i>	MS <sup>a</sup>	<i>df</i>	MS <sup>a</sup>	<i>df</i>	MS <sup>a</sup>
G	93	0.29**	93	280123**	93	5045**	93	52.4**	93	3250**	93	28.5**
E	1	0.42**	1	1403745**	1	107119**	1	34.3**	1	36751**	1	230**
G x E	92	0.02*	88	139951*	88	3837**	87	31.1*	86	668**	84	10.9**
Residuals	187	0.01	183	84148	183	1011	364	2.8	362	35	339	1.2
<b><math>h^2</math></b>	<b>92.50%</b>		<b>49.30%</b>		<b>23.00%</b>		<b>31.90%</b>		<b>79.40%</b>		<b>61.70%</b>	

a: mean square

\*\* $p$ -value < 0.001

\* $p$ -value < 0.01



### **QTL mapping**

QTL detection for each trait is showed in Table 4. One QTL was detected per trait and per experiment, at the exception of biomass measurement *in vitro* and conidia production *in vitro* for which no significant QTL were detected for E2. The QTL identified for TCTB production *in planta* for E2 was slightly under the estimated threshold. However, the LOD score threshold calculated for this trait was higher than for the others, and we suggest this phenomenon to be caused by the skewed and extreme distribution of the values for this trait. The QTL identified always has a major effect on the trait values (Table 4). Thus, 89.4% (88.3% – 90.5%) of aggressiveness variation is explained by the single and stable QTL identified. Similarly, QTL for TCTB production *in vitro* and *in planta* condition explain 82.0% (86.5 – 77.4) and 91.5% (E1 only) of the variation respectively. The QTL effect for ergosterol quantification and radial growing was more moderate (28.8% – 43.3% and 23.4% – 73.2% respectively, Table 4). The parental allele increasing the trait values (i.e. favorable allele) was brought by INRA-156 strain for most of the traits, except for the biomass measurement *in vitro*, of which the favorable allele was brought by INRA-171 strain (Table 4). The confidence intervals calculated at each QTL position were all located within the same region - within the 89.8 cM to 115.0 cM positions of the linkage group I. The approximate location of the QTL for the quantification of TCTB *in vitro*, the quantification of TCTB production *in planta* (E1), the visual symptoms, the ergosterol measurement *in planta* (E2) and the conidia production *in vitro* (E1) was more precise. For all these traits, the location of QTL was estimated at 92.1cM, and shared the same confidence interval (90.9 cM – 93.2 cM). The morphological aspect was also mapped as a qualitative trait at the position 92.1cM (Table 4). According to the overrepresentation of this position for the QTL of the trait of major interests - namely visual symptoms and TCTB production - we suggested focusing on this region only in order to identify the one or the many candidate genes responsible of the variation of these traits.

**Table 4:** QTL associated with phenotype variation under experiment condition and mapping of the morphological Mendelian trait.

Trait	Exp	QTL number	Chrom	QTL position (cM)	Confidence interval (cM)	Flanking position 5' <sup>a</sup>	Flanking position 3' <sup>a</sup>	LOD score	Threshold <sup>b</sup>	QTL effect (%)	Additive Effect	Favorable allele <sup>c</sup>
Biomass (mg)	E1	1	I	90.9	89.8 - 92.0	3,446,465	4,952,659	16	6.1	54.2	6.6	B
	E2	0	I	-	-	-	-	-	5.5	-	-	-
TCTB <i>in vitro</i> (µg/mg)	E1	1	I	92.1	90.9 - 93.2	4,032,446	5,267,806	39	4.9	86.5	28.5	A
	E2	1	I	92.1	90.9 - 93.2	4,032,446	5,267,806	30	5.1	77.4	58.5	A
Radial Growing (cm <sup>2</sup> /day)	E1	1	I	93.2	92.0 - 93.2	4,952,659	5,267,806	10.8	5.4	43.3	4	A
	E2	1	I	98.7	93.2 - 115.0	5,267,806	5,583,270	7.27	5.76	28.8	2.3	A
Conidia production (spores/mL)	E1	1	I	92.1	90.9 - 93.2	4,032,446	5,267,806	6.11	5.41	30.7	607,776	A
	E2	0	-	-	-	-	-	-	5.6	-	-	-
Visual symptom (%)	E1	1	I	92.1	90.9 - 93.2	4,032,446	5,267,806	41.6	4.8	88.3	45.9	A
	E2	1	I	92.1	90.9 - 93.2	4,032,446	5,267,806	45.7	5.2	90.5	57.9	A
TCTB <i>in planta</i> (µg/g)	E1	1	I	92.1	90.9 - 93.2	4,032,446	5,267,806	48	5.05	91.5	290.3	A
	E2	1	I	93.2	92.0 - 98.7	4,952,659	5,368,315	7.6	9.1	31.5	488	A
Ergosterol (µg/g)	E1	1	I	93.2	92.0 - 98.7	4,952,659	5,368,315	6.17	5.2	26.4	23.4	A
	E2	1	I	92.1	90.9 - 93.2	4,032,446	5,267,806	7.06	5.1	32.9	73.2	A
Morphological aspect	-	-	I	92.1	90.9 - 93.2	4,032,446	5,267,806	25.5	4.27	-	-	-

<sup>a</sup> corresponding to the position of the reference genome (RRES v4.0) of the closest marker from one and the other side of the confidence interval

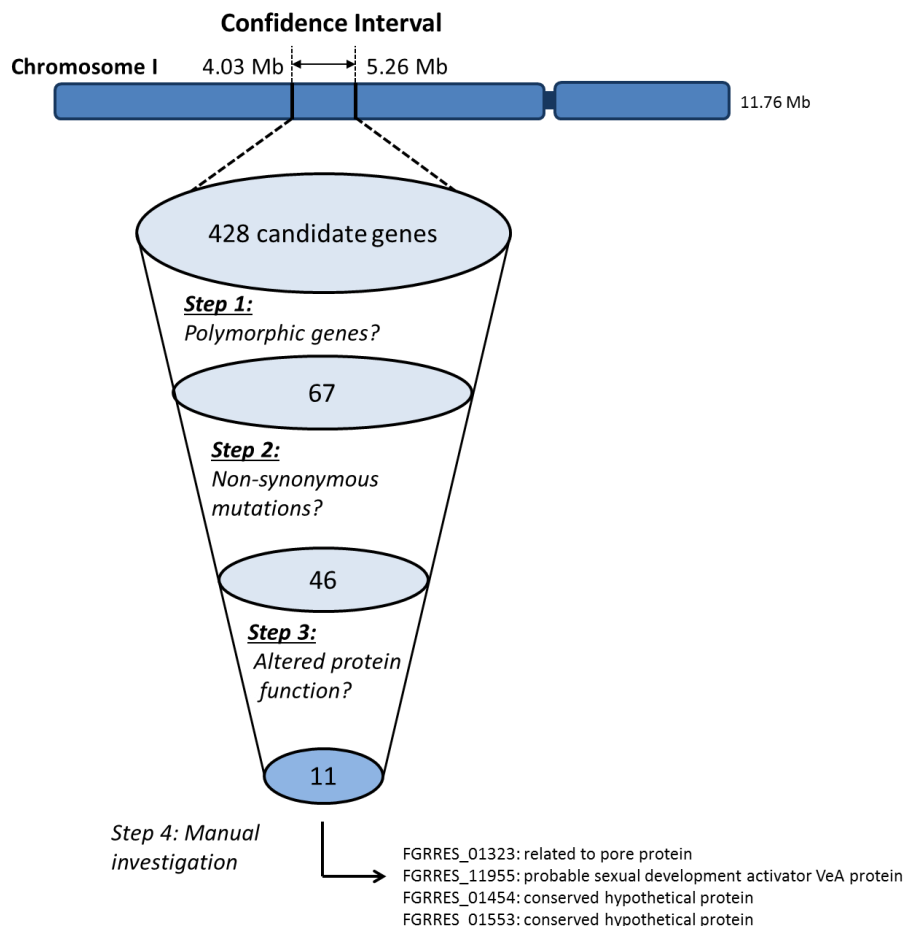
<sup>b</sup> Calculated using 500 permutations

<sup>c</sup> "A" corresponds to INRA-156 $\Delta$ mat strain and "B" to INRA-171 strain, favorable allele increase trait values

## *In silico* identification of the causal mutation(s)

The closest markers flanking the confidence intervals highlighted above (90.9 cM – 93.2 cM, Table 4) were aligned on the reference, delineating a genomic region from the 4,032,446 bp position to the 5,267,806 bp position of the chromosome I of the reference genome (RRES v4.0, Table 4). This region contains 428 putative protein coding genes (Supplemental file 2, Table S1). In order to reduce the number of genes, we used an analytic process (Figure 3) based on targeted genomic data mining from whole genome parental polymorphism annotation (Laurent, Moinard, *et al.*, 2016). Among the 428 genes, 67 showed polymorphic allelic versions between the two parents (Figure 3, Step 1; Supplemental file 2, Table S2). We hypothesized that the causal mutation was more likely to bring a non-synonymous effect or a deleterious effect on the protein translated. Forty-six genes followed this criterion (Figure 3, Step 2; Supplemental file 2, Table S2). We then considered these genes and evaluated the effect of the non-synonymous effect on the function of the protein (Figure 3, Step 3;

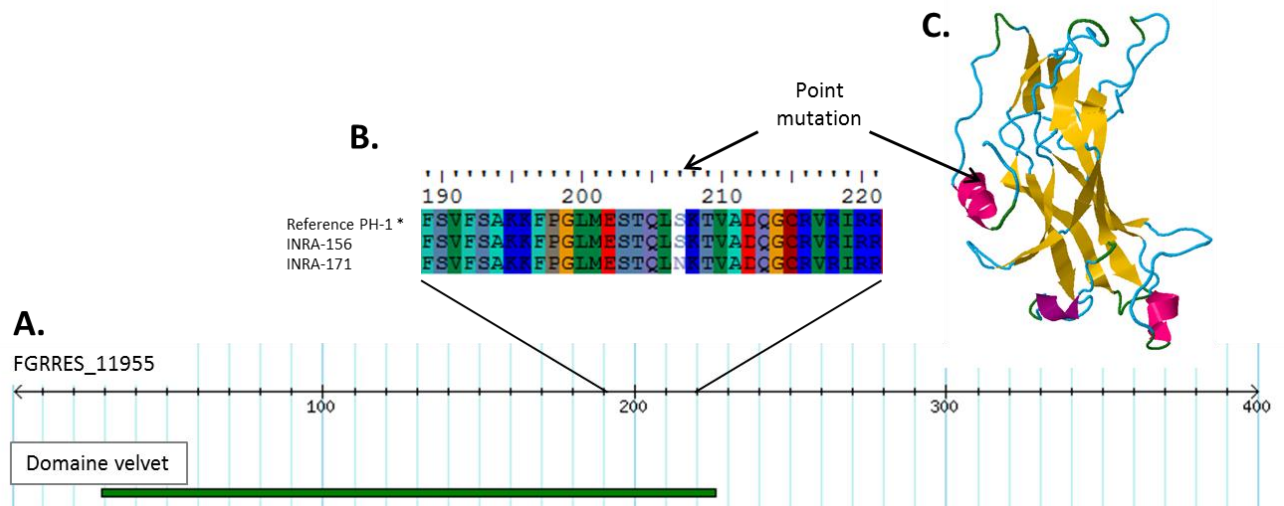
Supplemental file 2, Table S2). Only 11 proteins were predicted to have probable altered activity due to the presence of the mutation. Manual investigation was then conducted on the remaining candidates based on the full gene sequence of parental strains, and the gene sequences from four other *F. graminearum* strains, previously characterized for visual symptoms and TCTB production (Figure 3, Step 4; Supplemental file 2, Table S2) (Laurent, Moinard, *et al.*, 2016).



**Figure 3:** Analytical process of the *in silico* analysis of the genomic region of interest

Thus, the genes showing additional deleterious mutations, shared between parental genomes, or affected in a strain showing different phenotypes than the parental one, were discarded. Finally, four genes, named FGRRES\_01323, FGRRES\_11955, FGRRES\_01454, and FGRRES\_01553 were retained as potential candidates. All have been previously described to be expressed in wheat (Harris *et al.*, 2016) but only FGRRES\_01323 and FGRRES\_11955 have a predicted biological function (Figure 3, Supplemental file 2, Table S2). FGRRES\_11955, the orthologue of the *VeA* gene of *A. nidulans*, is predicted to encode

for a major transcription factor and to be involved in secondary metabolism regulation and pathogenicity of several fungal pathogens, including *F. graminearum* (Jiang *et al.*, 2011; Merhej *et al.*, 2012). The mutation in *FgVeA* has been observed in INRA-171 strain and affects the conserved Velvet domain by changing the serine located at position 207 into asparagine (Figure 4). This candidate gene shows the highest priority.

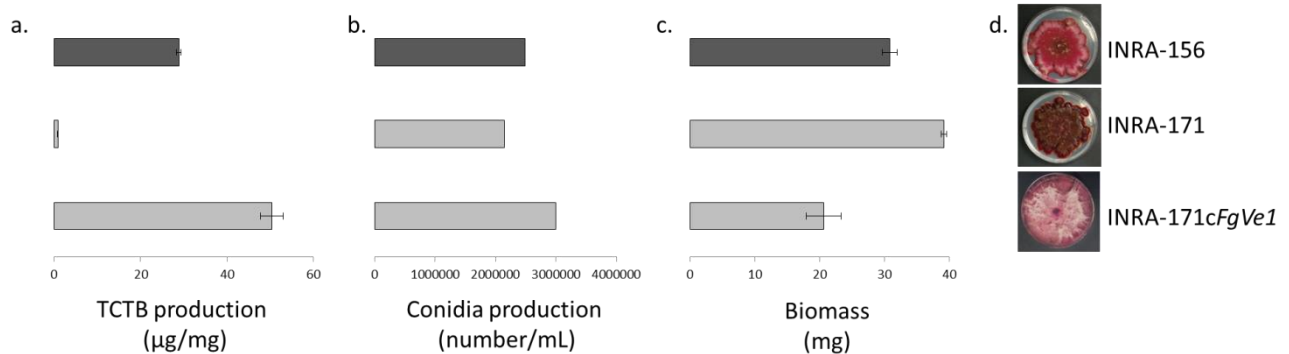


**Figure 4:** Identification of the causal mutation. A. Representation of the *FgVeA* (FGRRES\_11955) protein sequence (in amino-acids number) and location and length of the predicted “Velvet domain” in green. B. Alignment of the protein sequence at the mutation position for the reference genome sequence (\* PH-1 strain), for the INRA-156 and the INRA-171 strain. C. Three-dimensional prediction of the *FgVeA* protein and location of the point mutation (dark arrow) on the alpha-loop domain.

## Functional validation of *FgVeA*

In order to test whether or not the mutation affecting the gene *FgVeA* in the strain INRA-171 was responsible for the observed phenotypes, we replaced the INRA-156 allelic version of this gene in the INRA-171 strain. The INRA-171 strain and the transformed strain, named INRA-171*cFgVeA*, differed by one SNPs. INRA-171*cFgVeA* exhibited lighter and more aerial mycelia than the wild-type (Figure 5). The morphological aspect of INRA-171*cFgVeA* was thus more similar to the INRA-156 strain. Fungal biomass of the mutant was reduced compared to the wild type strain and was consistent with the INRA-156 phenotype. Furthermore, this strain showed a significant increase of TCTB production compared to

INRA-171 strain. The amount of TCTB production in INRA-171cFgVeA is comparable to the one observed in the INRA-156 strain (tested *in vitro* condition, Figure 5).



**Figure 5:** Phenotyping data of INRA-156 strain, INRA-171 strain and INRA-171cFgVeA strain, of which the *FgVeA* allele has been replaced by the allelic version of INRA-156 strain. a. TCTB production measured from *in vitro* cultures. b. Conidia production from *in vitro* cultures. c. Fungal biomass production from *in vitro* culture. d. Pictures of strains grown on PDA medium.

## Discussion

### Towards a better understanding of aggressiveness

In *F. graminearum*, the notation of visual symptoms on spike is the most commonly used phenotype to characterize aggressiveness variation (Cumagun and Miedaner, 2004; Talas *et al.*, 2012b; Talas *et al.*, 2016). Quantification of TCTB is also relevant because of the health issues due to its presence in food and feed, but also for its role for pathogenicity. During our analysis, we revealed a remarkable correlation between DON concentration and visual symptoms ( $Rho=0.8$ ). This result is in agreement with previous reports (Buerstmayr et Lemmens, 2015; Paul *et al.*, 2005). Indeed, DON has been shown to be produced in response to wheat defense, avoiding the thickening of the cell wall and allowing the spread of the fungus (Bai *et al.*, 2002; Brown *et al.*, 2010; Jansen *et al.*, 2005; Walter *et al.*, 2010). Therefore, it makes sense that an increase of DON production will allow a faster and more efficient spread of the fungus within the spike. TCTB production measured *in planta* shows

good consistency with TCTB production measured *in vitro* culture. This result suggests that the difference in aggressiveness is presumably a consequence of different intrinsic abilities of DON production, independent of the perception of plant defense. We also show that visual symptoms were positively correlated with the ability of the fungus to spread *in vitro* culture and its ability to produce conidia *in vitro*. Therefore, in addition to the production of DON that counteracts the establishment of physical barriers, some strains are probably able to grow faster within the spike and cause more severe symptoms. Similarly, conidia production can have important effects on disease severity: an increase in conidia concentration during inoculation has been previously shown to provoke more spike damages and higher DON concentration (Stein *et al.*, 2009). In the present study, conidia solutions were normalized prior to inoculation and could not have generated more severe symptoms. However, one could expect that a large conidia production could have magnified the effect of aggressiveness in the fields, leading to better propagation of this haplotype through larger inoculum production (Osborne et Stein, 2007). On the other hand, visual symptoms, DON concentration and fungal spreading were negatively correlated with fungal biomass production measured *in vitro*. First, it could explain the moderate correlation observed between ergosterol concentration in spike and visual symptoms: ergosterol concentration of strains showing a rapid spread in the spike could be confounded with the ergosterol concentration of strains showing a more limited growth but accumulating more biomass. The biological significance of this negative correlation is interesting. Biomass production could be linked to important cellular functions, such as reserve accumulation or cell wall thickening. Such a result could reflect a trade-off between a rapid colonization of plant-tissue and another strategy of parasitism adopted by some *F. graminearum* strains. A deeper characterization of this experimental population for this trait during other periods of its life cycle could shed light on this interesting trade-off. As a whole, this study highlights the importance of phenotyping numerous traits when focusing on aggressiveness variation (Lannou, 2012), as the sole observation of visual symptoms and DON concentration are maybe not representative of the full epidemic potential of this pathogen. The different heritability revealed for each trait further suggests that the overall aggressiveness variation is largely influenced by genetic factors, as previously suggested (Cumagan et Miedaner, 2003, Talas *et al.*, 2016).

### **Aggressiveness variation is due to a major QTL with pleiotropic effects**

For each trait, one major effect QTL was detected on chromosome I, and this result could be expected according to the bimodal-like distribution observed for the majority of these traits in the progeny (Kang, 2002). The presence of this major QTL does not exclude the existence of additional QTL with minor effects that could not be detected in our experiment. Increasing the population size would possibly reveal additional QTL of minor effects at other positions (Mackay *et al.*, 2009). The detected QTL always explained large effects for the trait values. In some cases, as for visual symptoms and TCTB production, the QTL could explain up to 91% of the variation. In 2012, Talas *et al.* associated SNPs in *MetAP1* gene explaining 26% of the genotypic variance for aggressiveness; and one SNP in *Tri1* gene explaining 4.4% of the genotypic variance for DON production. In 2016, 50 and 29 quantitative trait nucleotides were identified to be implicated in aggressiveness and DON production variation respectively (Talas *et al.*, 2016). Maximum effect of QTN was 24% for aggressiveness and 19% for DON production. Correlation between these two traits was however not tested and no QTN seemed to be pleiotropic for these two traits (Talas *et al.*, 2016). Before that, a study identified one major QTL for aggressiveness using an interspecific cross, explaining 50% of the variation observed along the progeny (Cumagun, 2004). The authors proposed that this QTL correspond to a gene belonging to the *Tri* gene cluster involved in TCTB biosynthesis, and suggested the production in either DON or NIV to be associated with aggressiveness variation. Such a scenario could not be considered in our case, as strains were only able to produce DON and 15-ADON. Actually, the positions of the QTL identified in the present study differ from all previously characterized (QTL or QTN), and the effects associated with the different traits were remarkably more important than previously reported. Moreover the QTL detected for all traits were always located on a same chromosomal region. Such a result can be explained in theory by tightly linked QTL or by a single QTL with pleiotropic effects (Studer et Doebley, 2011). We hypothesized that these QTL were more likely to be single ones. In reality, pleiotropic QTL are not rare in nature (Mackay *et al.*, 2009), and could be explained by the tightly interconnected and complex network of gene regulation, as observed in *F. graminearum* (Guo *et al.*, 2016).

### **From one QTL to the identification of *FgVeA*, the added-value of next-generation sequencing for candidate gene analysis**

QTL detection is a powerful approach to localizing the genomic region involved in quantitative trait variation, as illustrated herein for aggressiveness variation in *F. graminearum*. Next-generation sequencing is further increasing the power of mapping strategies, by facilitating high-throughput genotyping and leading to high-resolution linkage maps. It also opens great perspectives to identify candidate genes by giving access to the full parental sequences and study the implication of polymorphism of the genes (Grünwald *et al.*, 2016). In 2011, candidate genes involved in the virulence of *Aspergillus nidulans* on animal models (understood as ‘aggressiveness’ for plant pathologist) were identified on the basis of QTL location and the presence of non-synonymous mutations within genes (Christians *et al.*, 2011). A similar approach was developed to identify candidate genes implicated in *Zymoseptoria tritici* aggressiveness on wheat (Stewart *et al.*, 2016). These examples illustrate the relevancy of such an approach to investigate the genetic basis of aggressiveness in fungi. Nevertheless, and in contrast to animal and plant species (Salvi and Tuberosa, 2005; Zhu and Zhao, 2007), the functional validation of a candidate gene underlying a QTL in non-model fungi has not yet been conducted.

The QTL detected in this analysis was located in a genomic region previously reported to be non-recombinant (Cuomo *et al.*, 2007; Laurent, Palaiokostas, *et al.*, 2016). In consequence, the confidence interval of the QTL was large (~1.2Mb) and the number of genes (428) was too important to identify candidate genes on the sole basis of the QTL position on the genome. The novelty of our approach consisted of feeding the QTL location with information from parental sequences and sequences from other *F. graminearum* isolates previously characterized (Laurent, Moinard, *et al.*, 2016). In doing so, we were able to identify the gene *FgVeA* as the highest-priority candidate. The involvement of this gene for aggressiveness was further validated through more classical functional studies and strongly suggests that *FgVeA* was the gene underlying the detected QTL. This work is the first to our knowledge to report the validation of a QTL in fungi. Recently, two avirulence genes of the pathogen *Melampsora lini* rust fungi have been cloned using a RAD-seq-based mapping strategy (Anderson *et al.*, 2016). For these genes, the qualitative variation on the plant resistance certainly helped their functional validation. Similarly, the validation of *FgVeA* was likely helped by its strong and pleiotropic effects on the phenotypes.



### **Implication of velvet gene polymorphism for aggressiveness and TCTB potential**

VeA protein is a conserved major regulator protein in several fungi, involved in morphogenesis and secondary metabolism (see Calvo *et al.*, 2016 or Bayram *et al.*, 2008 for a review). The gene *VeA* was first discovered in *A. nidulans* (KAFER, 1962) and the mode of action of the VeA protein has been well defined in this species. Therefore, in this species, VeA has been shown to interact with VelB. In the dark, this heterodimer is transported through the action of KapA protein and interacts with LaeA protein (a conserved methyltransferase), forming the so-called velvet complex. The velvet complex is supposed to coordinate fungal development and secondary metabolism through the regulation of histone methylation mediated by LaeA. The heterodimer VeA-VelB is also active in the nucleus and is supposed to favor sexual fruiting body formation in the dark. VelB is able to interact with VosA, another velvet protein, inhibiting asexual development in the dark. In the light, the heterodimer VeA-VelB is located in the cytoplasm, leading to the induction of asexual reproduction and the reduction of fruiting body production.

In *F. graminearum*, *FgVeA* has been previously shown to affect fungal development, asexual reproduction, lipid metabolism and secondary metabolism, including trichothecenes biosynthesis (Jiang *et al.*, 2011; Merhej *et al.*, 2012). Therefore, mutant strains deleted for this gene were affected for pigmentation, showed a reduced hydrophobicity and aerial mycelium production, and an important decrease of DON production and pathogenicity (Jiang *et al.*, 2011; Merhej *et al.*, 2012). The results for conidia production were found to be inconsistent between these two analyses, making the implication of *FgVeA* in asexuality unclear (Jiang *et al.*, 2011; Merhej *et al.*, 2012). Furthermore, the protein *FgVeA* has been shown to interact with *FgLaeA* (Kim *et al.*, 2013) but no interaction with *FgVelB* has yet been demonstrated (Jiang *et al.*, 2012). The role of the protein *FgVeA* for sexual reproduction has never been tested. In our study, we identified a never-described allele of *FgVeA*, consisting of the same protein with one serine being replaced by an asparagine. The causal mutation does not introduce a large effect on the translated protein, as frameshift variant or codon stop gain do, and the mutation has been predicted to have no obvious effect on the three-dimensional structure of the protein itself. However, it affects a predicted alpha-loop, conserved within the Velvet domain characterizing all the Velvet proteins, and previously suggested to bind DNA (Ahmed *et al.*, 2013). The mutated strain presented, at first, similitudes to  $\Delta FgVeA$  strain phenotypes - a reduction of aerial mycelia, a perturbed pigmentation and an important decrease of aggressiveness and DON production. Whereas one could think that the *FgVeA*

protein in this strain is not functional, no evidence is available at the moment to prove this. Furthermore, this strain shows a high-rate of perithecia production - this result is going against the phenotype expected for a non-functional protein in other fungi like *A. nidulans*. Therefore, an alternative thought would be that the protein is functional but plays different roles. More evidence is now needed to shed light on the molecular mechanism underlying this polymorphism.

### **CONCLUSION**

The present study successfully associated a non-synonymous mutation in the *FgVeA* gene with the variation of aggressiveness of *F. graminearum* on wheat, and highlighted the use of QTL detection and comparative genomic analysis to study pathogenicity of fungal species. In our case, the aggressiveness variation reflected important differences of visual symptoms on spike, fungal growing and DON production. The strain showing this never described allele of *FgVeA* is a milestone to characterize the function of Velvet protein in *F. graminearum* species and in fungi. The new question is to know whether or not the haplotypes presenting this version of *FgVeA* genes are frequent in the field. The ecological importance of this allele has yet to be defined, as it could help understand *F. graminearum* evolution.

## Materials and Methods

### Parental strains information and progeny construction

A description of the parental strains and of the construction of progeny is described in a Laurent, Palaiokostas, *et al.*, 2016. In summary, the two parental strains INRA-156 $\Delta$ *mat* and INRA-171 are French isolates of the DON/15-ADON chemotypes. INRA-156 $\Delta$ *mat* was constructed from INRA-156 strain of which the gene *mat-1-2-1* (FGRRES\_08893) was replaced by a gene bringing the resistance to hygromycin. The short read sequences of the wild type parental strain genomes (INRA-156 and INRA-171) are available at the GenBank Sequence Read Archive under the accession SRP064374. Ninety four strains from the progeny were used for phenotyping trials among 88 were used for linkage map construction. Previous inoculation of INRA-156 and INRA-156 $\Delta$ *Mat* strains was conducted on wheat spikes and showed no significant effects caused by the *Mat1-2-1* deletion (data not shown).

### Phenotyping

For each trait assessment, two experiments were conducted, referred hereafter to as E1 and E2.

*Visual symptom measurement:* a conidial suspension of 100 spores adjusted in 10 $\mu$ L of CMC medium was inoculated in two opposite pair of floral caveats per spike. Three to four spikes (when possible and depending on the plants) were inoculated per wheat plant (Roysac variety). Plants were grown under greenhouse conditions and were inoculated just after anthesis of wheat flower. Plant duplicates were conducted per strain according to the protocol described above. The visual symptom corresponded to the number of spikelet infected by spike and normalized by total number of spikelet per spike, as previously suggested (Talas *et al.*, 2012b). Visual symptoms were measured at 22 days post inoculation (dpi) and spikes were collected following notation in order to be dried and used later on for TCTB and ergosterol quantification.

*In planta TCTB assay:* TCTB were extracted from 1g of dried spike powder. Powder was diluted in 5 mL of Milli-Q water and the solution was vortexed for 10 min, and then centrifuged 12 min at 5000rpm. 1mL of this solution was diluted in 3 mL of Milli-Q water. 7.5 $\mu$ L of this solution was suspended in 742.5 $\mu$ L of a solution containing 78% of H<sub>2</sub>O and 22% of methanol, before filtration through 0.22  $\mu$ m-pore-size filters (Phenomenex, Torrance, USA). Quantification of trichothecenes was performed using a QTrap 2000 LC-MS/MS

system (Applied Biosystems) equipped with a 1100 Series HPLC system (Agilent), a reverse phase Kinetex™ 2.6 µm XB-C18 column (150 × 4.6 mm, Phenomenex; France) maintained at 45 °C and a TurboIonSpray ESI source. Solvent A consisted of methanol/water (10/90, v/v) and solvent B consisted of methanol/water (90/10, v/v). The flow rate was kept at 700 µL/min and was split so that 350 µL/min went to the electrospray source. Gradient elution was performed with the following conditions: min with a linear gradient from 85 % to 5 % A, 4 min held at 5 % A, 1 min linear gradient from 5 % to 85 % A and 85 % A for 8 min post run reconditioning. The injection volume was 5 µL. The electrospray interface, heated at 400°C, was used in the negative ion mode with the following settings: curtain gas, 20 p.s.i.; nebulizer gas, 30 p.s.i.; auxiliary gas, 70 p.s.i.; ion spray voltage, -4200 V. Quantification was performed using external calibration with standard solutions, ranging from 10 to 1000 ng/ml (Romer Labs, Getzersdorf, Austria). TCTB concentration was measured in µg/g of spike powder.

*In planta ergosterol assay:* extraction and quantification of ergosterol was performed according from Touati-Hattab *et al.*, 2016 with small modifications: dried samples were suspended in 100 µl of methanol and filtered through a 0.2 µm filter (Phenomenex, Torrance, USA) – quantification was performed on a Shimadzu Prominence UFLC chain. Ergosterol was measured in µg/g of spike powder.

*In vitro TCTB assay:* conidial suspension of 125,000 spores (0.1 - 0.6mL) were incubated in a final volume of 8mL completed with Minimum Synthetic medium (previously described by Boutigny *et al.*, 2009) for 14days at 25°C, in dark conditions. Triplicates of cultures were conducted per strain. The cultures of liquid media were transferred in Corning tubes (15 mL) and centrifuged at 5000rpm at 4°C. Supernatants were retrieved and 250µL per sample were collected, diluted in 250 µL of MeOH and filtered using 0.22 µm-pore-size filters (Phenomenex, Torrance, USA). Quantification of TCTB was performed on a Shimadzu Prominence UFLC chain, equipped with two pumps LC-20 AD, a degasser DGU-20A3R, an auto sampler SIL-30 AC and a diode-array detector SPD-M20A (Shimadzu Scientific Instruments, France). Separation was achieved on a Kinetex 2,6U XB-C18 - 100A column (150 x 4.6 mm; 2.6 µm) (Phenomenex, France) maintained at 45°C. Mobile phase consisted of water acidified with ortho-phosphoric acid to reach pH 2.6 (solvent A) and acetonitrile (solvent B).The following gradient was used for elution: 7% B for 1 min, 7-30% B in 6 min, 30-90% B in 2 min, 90% B for 2 min, 90-7% B in 1 min, and 6 min post-run equilibration with initial conditions. The flow was kept at 0.7 mL/min for a total run time of 18 min. The

injection volume was 0.5  $\mu$ L. The UV-VIS spectra were recorded from 220 to 550 nm and peak areas were measured at 230 nm. Quantification was performed using external calibration with standard solutions (Romer Labs, Austria) and measured in  $\mu$ g/mg of dried fungal biomass.

*Radial Growing:* mycelium was grown on Potatoes Dextrose Agar (PDA, Difco-France) at 25°C in dark condition. This phenotyping was conducted using automated image analysis procedure. Briefly, strains were grown on triplicate, and pictures were taken every day using standardized camera settings and lighting environments. Images were analyzed using ImageJ software (Schneider *et al.*, 2012). Pictures were stacks and distance were scaled manually using 5cm tracks disposed during picture acquisition. All pictures were treated simultaneously using the “montage” option and fungal surfaces were automatically selected using saturation and brightness thresholds. The delimited surfaces, representing fungal radial growing, were integrated using ROI manager and areas were measured in  $\text{cm}^2$ . Radial growing was measured in  $\text{cm}^2$  per day.

*In vitro biomass:* fungal biomasses resulted from MS medium cultures and *in vitro* production of mycotoxin assay presenting above. Triplicates were conducted for each strains and fungal biomass were dried during 5 days at 80°C. Biomasses were measured in mg.

*Conidia production:* 4 plugs of mycelium grown previously on PDA medium (~ 0.5cm diameters each) were incubated in 12mL of CMC liquid medium (Carboxymethyl cellulose 15%, yeast extract 1%,  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$  0.5%,  $\text{NH}_4\text{NO}_3$  1%,  $\text{KH}_2\text{PO}_4$  1%), for three days at 25°C and under permanent shake (180rpm). Singleton of conidial suspensions was produced. After incubation, conidial suspensions were filtered through Sefar Nitex 03-100 (100  $\mu$ m, SEFAR AG - Switzerland). Filtered solutions were measured for optical density at 500nm. Quantification was performed using internal conidial solutions of which concentrations of spores were counted using Thoma cell.

### **Statistical analysis and QTL detection**

Correlations between phenotypes and between experiments have been tested using Spearman rank order test and considered as significant for  $p$ -value  $< 0.05$ . Shapiro Wilk test was used to test the normality of the distribution of the trait in the progeny and considered as normal for  $p$ -value  $> 0.05$ . Broad sense heritability were estimated considering combined data according to the model:  $Y = \mu + G + E + G \times E + \varepsilon$ , with G, the genotypic effect, E, the experiment

effect,  $G \times E$  the interaction between genotype and experiment effect and  $\varepsilon$  the residual effect. The linkage map used for QTL mapping was previously described in Laurent, Palaiokostas, *et al.*, 2016. QTL detection was conducted using R/qtl (Broman et al., 2003) and the composite interval mapping analysis. The Haley-Knott regression method was set (Haley et Knott, 1992), with two markers as cofactors and a window size of 1 cM. Missing marker genotype data were filled with the imputation method proposed by Rqtl, using the Kosambi map function and an estimate genotyping error rate of 0.001. QTL were accepted if the logarithm of odd (LOD) score at the position was greater than the LOD threshold calculated with 500 permutations and for a significance level  $\alpha$  of 0.001. Confidence intervals corresponded to the positions (in cM) with the maximal LOD score of the QTL - 1 and the maximal LOD score +1, expanded to the closest markers at the outer edge of this interval. Markers were aligned on the reference genome of *F. graminearum* RRES v4.0 (King *et al.*, 2015).

### ***In silico* analysis of candidate genes**

Genomic sequences and genetic variants of parental strains were retrieved from (Laurent, Moinard, *et al.*, 2016). In order to reduce the number of candidate genes, an analytical process was proposed and summarized in Figure 3. This process proposes to identify polymorphic genes (Stage 1), affected by non-synonymous mutations (stage 2). Variant annotation was conducted using SnpEff software, as previously described (Laurent, Moinard, *et al.*, 2016). To predict if an amino-acid substitution or InDels have an impact on the biological function of the protein (stage 3) Provean algorithm was used (Choi *et al.*, 2012). Authors suggested that score below -2.5 indicates that the polymorphic protein has great chance to show altered activity. We re-evaluated this threshold to -2 to limit the chance to eliminate false negative variants. In stage 4, the remaining candidates were filtered out on the basis of the presence of additional deleterious mutation found for both parental strains and compared to the genotypic information from strains, previously characterized (Laurent, Moinard, *et al.*, 2016). Gene information and orthologues from other species were retrieved from the FungiDB database (Stajich *et al.*, 2012).

### ***FgVeA* validation and investigation**

*FgVeA* locus from the INRA-156 was first amplified using the P1 primer (AAGAAACATCTCCGCCATTTTA) and the P2 primer (TGAAACAAAAATTCTGCCAATG). Amplicon length was 3.4 kb and corresponded to

*FgVeA* coding and flanking sequences (FGRRES\_11955). Kapa Hifi™ DNA polymerase (KapaBiosystems) was used for amplification, following these conditions: 95°C 2min, 35 cycles of [98°C 20s , 61.8°C 15s, 72°C 1min45s], 72°C 5min. PCR product was then purified using the Wizard® SV Gel and PCR Clean-up System (Promega) and quantified using NanoDrop™ technology. The PCR product was mixed with plasmid pBC HygromycinR to co-transform protoplasts of *F. graminearum* strain INRA-171. Protocol for protoplast production and transformation was used according to Montibus *et al.*, 2013. Transformed strains were selected on hygromycin B containing PDA medium (60 µg/mL) and single germinating conidia were isolated and transferred on new PDA plates. Integration of the INRA-156*mat-1-2-1* allelic version of the gene *FgVeA* into INRA-171 genome was verified by Kompetitive Allele Specific PCR markers genotyping (KASP™, Supplemental file 2 Table S4). The genetic background of the transformed strains was verified using 4 additional KASP markers distributed on other chromosomes (Supplemental file 2 Table S4). Phenotypes of the transformed strains were tested as previously described in “Phenotyping” section. Two- and three-dimensional structures of the velvet proteins were predicted from the corresponding amino-acid sequence of the two allelic versions of Velvet gene using Phyre2 (Kelley *et al.*, 2015). Protein structures were visualized using Jmol (Hanson *et al.*, 2010).

### **Acknowledgement:**

BL received a PhD fellowship from the French Research Ministry. Authors would like to gratefully thank Cyrille Saintenac for helpful conversations and for assistance during *in planta* experimentations; Ross Houston and Christos Palaiokostas for remarkable advices during work conduction and along manuscript writing; Pierre Chesnot for providing a valuable help during English writing.

## References

- Ahmed, Y.L., Gerke, J., Park, H.-S., et al.** (2013) The Velvet family of fungal regulators contains a dna-binding domain structurally similar to NF- $\kappa$ B stock, A.M., éd. *PLoS Biol.* **11**, e1001750.
- Akinsanmi, O.A., Backhouse, D., Simpfendorfer, S. et Chakraborty, S.** (2006) Genetic diversity of Australian *Fusarium graminearum* and *F. pseudograminearum*. *Plant Pathol.* **55**, 494-504.
- Alexander, N.J., McCormick, S.P., Waalwijk, C., Lee, T. van der et Proctor, R.H.** (2011) The genetic basis for 3-ADON and 15-ADON trichothecene chemotypes in *Fusarium*. *Fungal Genet Biol* **48**, 485-495.
- Anderson, C., Khan, M.A., Catanzariti, A.-M., et al.** (2016) Genome analysis and avirulence gene cloning using a high-density RADseq linkage map of the flax rust fungus, *Melampsora lini*. *BMC Genomics* **17**, 667.
- Anderson, J.P., Gleason, C.A., Foley, R.C., Thrall, P.H., Burdon, J.B. et Singh, K.B.** (2010) Plants versus pathogens: an evolutionary arms race. *Funct. Plant Biol.* **37**, 499-512.
- Andrivon, D., Pilet, F., Montarry, J., Hafidi, M., Corbière, R., Achbani, E.H., Pellé, R. et Ellissèche, D.** (2007) Adaptation of *Phytophthora infestans* to partial resistance in potato: evidence from French and Moroccan populations. <http://dx.doi.org/10.1094/PHTO-97-3-0338>.
- Bai, G. et Shaner, G.** (2004) Management and resistance in wheat and barley to *Fusarium* head blight. *Annu Rev Phytopathol* **42**, 135-161.
- Bai, G.H., Desjardins, A.E. et Plattner, R.D.** (2002) Deoxynivalenol-nonproducing fusarium graminearum causes initial infection, but does not cause disease spread in wheat spikes. *Mycopathologia* **153**, 91-98.
- Bayram, O., Krappmann, S., Ni, M., et al.** (2008) VelB/VeA/LaeA complex coordinates light signal with fungal development and secondary metabolism. *Science* **320**, 1504-6.
- Boutigny, A.-L., Barreau, C., Atanasova-Penichon, V., Verdal-Bonnin, M.-N., Pinson-Gadais, L. et Richard-Forget, F.** (2009) Ferulic acid, an efficient inhibitor of type B trichothecene biosynthesis and *Tri* gene expression in *Fusarium* liquid cultures. *Mycol. Res.* **113**, 746-753.
- Broman, K.W., Wu, H., Sen, S. et Churchill, G.A.** (2003) R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**, 889-890.
- Brown, N.A., Antoniwi, J. et Hammond-Kosack, K.E.** (2012) The predicted secretome of the plant pathogenic fungus *Fusarium graminearum*: a refined comparative analysis. *PLoS One* **7**, e33731.
- Brown, N.A., Urban, M., Meene, A.M. van de et Hammond-Kosack, K.E.** (2010) The infection biology of *Fusarium graminearum*: defining the pathways of spikelet to spikelet colonisation in wheat ears. *Fungal Biol* **114**, 555-571.
- Buerstmayr, H., Ban, T. et Anderson, J.A.** (2009) QTL mapping and marker-assisted selection for *Fusarium* head blight resistance in wheat: A review. *Plant Breed.* **128**, 1-26.
- Buerstmayr, H. et Lemmens, M.** (2015) Breeding healthy cereals: genetic improvement of *Fusarium* resistance and consequences for mycotoxins. *World Mycotoxin J.* **8**, 591-602.
- Caffier, V., Cam, B. Le, Rifai, M. al, et al.** (2016) Slow erosion of a quantitative apple resistance to *Venturia inaequalis* based on an isolate-specific Quantitative Trait Locus. *Infect. Genet. Evol.*



- Caffier, V., Lasserre-Zuber, P., Giraud, M., et al.** (2014) Erosion of quantitative host resistance in the apple×*Venturia inaequalis* pathosystem. *Infect. Genet. Evol.* **27**, 481-489.
- Calvo, A.M., Lohmar, J.M., Ibarra, B. et Satterlee, T.** (2016) 18 Velvet regulation of fungal development. in growth, differentiation and sexuality., p. 475-497. Cham: Springer International Publishing.
- Choi, Y., Sims, G.E., Murphy, S., Miller, J.R. et Chan, A.P.** (2012) Predicting the functional effect of amino acid substitutions and indels. *PLoS One* **7**, e46688.
- Christians, J.K., Cheema, M.S., Vergara, I.A., Watt, C.A., Pinto, L.J., Chen, N. et Moore, M.M.** (2011) Quantitative trait locus (QTL) mapping reveals a role for unstudied genes in *Aspergillus virulence*. *PLoS One* **6**, e19325.
- Cumagun** (2004) Genetic mapping of pathogenicity and aggressiveness of *Gibberella zeae* (*Fusarium graminearum*) toward wheat. *Genet. Resist.*
- Cumagun, C.J.R. et Miedaner, T.** (2004) Segregation for aggressiveness and deoxynivalenol production of a population of *Gibberella zeae* causing head blight of wheat. *Eur. J. Plant Pathol.* **110**, 789-799.
- Cumagun, C.J.R. et Miedaner, T.** (2003) Aggressiveness of 42 isolates of *Gibberella zeae* (*Fusarium graminearum*) in wheat under field and greenhouse conditions,.
- Cuomo, C.A., Guldener, U., Xu, J.R., et al.** (2007) The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. *Science* (80-. ). **317**, 1400-1402.
- Delmas, C.E.L., Fabre, F., Jolivet, J., Mazet, I.D., Richart Cervera, S., Delière, L. et Delmotte, F.** (2016) Adaptation of a plant pathogen to partial host resistance: selection for greater aggressiveness in grapevine downy mildew. *Evol. Appl.* **9**, 709-25.
- Dowkiw, A., Voisin, E. et Bastien, C.** (2010) Potential of Eurasian poplar rust to overcome a major quantitative resistance factor. *Plant Pathol.* **59**, 523-534.
- Ferrigo, D., Raiola, A. et Causin, R.** (2016) *Fusarium* toxins in cereals: occurrence, legislation, factors promoting the appearance and their management. *Molecules* **21**, 627.
- Gale, L.R., Ward, T.J., Balmas, V. et Kistler, H.C.** (2007) Population subdivision of *Fusarium graminearum sensu stricto* in the Upper Midwestern United States. *Phytopathology* **97**, 1434-1439.
- Gilbert, J., Abramson, D., McCallum, B. et Clear, R.** (2002) Comparison of Canadian *Fusarium graminearum* isolates for aggressiveness, vegetative compatibility, and production of ergosterol and mycotoxins. *Mycopathologia* **153**, 209-215.
- Grünwald, N.J., McDonald, B.A. et Milgroom, M.G.** (2016) Population genomics of fungal and oomycete pathogens. *Annu. Rev. Phytopathol.* **54**, 323-346.
- Guo, L., Zhao, G., Xu, J.-R., Kistler, H.C., Gao, L. et Ma, L.-J.** (2016) Compartmentalized gene regulatory network of the pathogenic fungus *Fusarium graminearum*. *New Phytol.* **211**, 527-541.
- Haley, C.S. et Knott, S.A.** (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity (Edinb)*. **69**, 315-324.
- Hanson, R.M., IUCr, J., W., et al.** (2010) *Jmol* – a paradigm shift in crystallographic visualization. *J. Appl. Crystallogr.* **43**, 1250-1260.

- Harris, L.J., Balcerzak, M., Johnston, A., Schneiderman, D. et Ouellet, T.** (2016) Host-preferential *Fusarium graminearum* gene expression during infection of wheat, barley, and maize. *Fungal Biol.* **120**, 111-123.
- Jansen, C., Wettstein, D. von, Schäfer, W., Kogel, K.-H., Felk, A. et Maier, F.J.** (2005) Infection patterns in barley and wheat spikes inoculated with wild-type and trichodiene synthase gene disrupted *Fusarium graminearum*. *Proc Natl Acad Sci U S A* **102**, 16892-16897.
- Jiang, J., Liu, X., Yin, Y. et Ma, Z.** (2011) Involvement of a velvet protein *FgVeA* in the regulation of asexual development, lipid and secondary metabolisms and virulence in *Fusarium graminearum*. *PLoS One* **6**, e28291.
- Jiang, J., Yun, Y., Liu, Y. et Ma, Z.** (2012) *FgVELB* is associated with vegetative differentiation, secondary metabolism and virulence in *Fusarium graminearum*. *Fungal Genet. Biol.* **49**, 653-62.
- Kafer, E.** (1962) Translocations in stock strains of *Aspergillus nidulans*. **33**, 59-68.
- Kang, M.S.** (2002) Quantitative Genetics, Genomics and Plant Breeding, Cabi-publishing.org.
- Keller, M., Bergstrom, G. et Shields, E.** (2014) The aerobiology of *Fusarium graminearum*. *Aerobiologia (Bologna)*. **30**, 123-136.
- Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N. et Sternberg, M.J.E.** (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845-858.
- Kim, H.-K., Lee, S., Jo, S.-M., McCormick, S.P., Butchko, R.A.E., Proctor, R.H. et Yun, S.-H.** (2013) Functional roles of *FgLaeA* in controlling secondary metabolism, sexual development, and virulence in *Fusarium graminearum*. *PLoS One* **8**, e68441.
- King, R., Urban, M., Hammond-Kosack, M., Hassani-Pak, K. et Hammond-Kosack, K.** (2015) The completed genome sequence of the pathogenic ascomycete fungus *Fusarium graminearum*. *BMC Genomics* **16**, 544.
- Lannou, C.** (2012) Variation and selection of quantitative traits in plant pathogens. *Annu Rev Phytopathol* **50**, 319-338.
- Laurent, B., Balesdent, M.H., Carlier, J., Frey, P., Maupetit, A., Picard, L., Soyer, J. et Foulongne-Oriol, M.** (2016) Quantitative genetics in fungi: moving forward. *Crit. Rev. Microbiol.* **In prep.**
- Laurent, B., Moinard, M., Spataro, C., Ponts, N., Barreau, C. et Foulongne-Oriol, M.** (2016) Landscape of genomic diversity and host adaptation in *Fusarium graminearum*. *BMC Genomics* **Submitted.**
- Laurent, B., Palaiokostas, C., Spataro, C., Moinard, M., Enric, Z., Houston, R.D. et Foulongne-Oriol, M.** (2016) High resolution mapping of the recombination landscape of the phytopathogen *Fusarium graminearum* suggests two-speed genome evolution. *Mol. Plant Pathol.* **Submitted.**
- Lee, T., Han, Y.K., Kim, K.H., Yun, S.H. et Lee, Y.W.** (2002) Tri13 and Tri7 determine deoxynivalenol- and nivalenol-producing chemotypes of *Gibberella zeae*. *Appl Env. Microbiol* **68**, 2148-2154.
- Lee, T. Van der, Zhang, H., Diepeningen, A. van et Waalwijk, C.** (2015) Biogeography of *Fusarium graminearum* species complex and chemotypes: a review. *Food Addit Contam Part A Chem Anal Control Expo Risk Assess* **32**, 453-460.
- Lv, C., Song, Y., Gao, L., Yao, Q., Zhou, R., Xu, R. et Jia, J.** (2014) Integration of QTL detection

and marker assisted selection for improving resistance to *Fusarium* head blight and important agronomic traits in wheat. *Crop J.* **2**, 70-78.

- Mackay, T.F., Stone, E.A. et Ayroles, J.F.** (2009) The genetics of quantitative traits: challenges and prospects. *Nat Rev Genet* **10**, 565-577.
- Maier, F.J., Miedaner, T., Hadel, B., Felk, A., Salomon, S., Lemmens, M., Kassner, H. et Schäfer, W.** (2006) Involvement of trichothecenes in fusarioses of wheat, barley and maize evaluated by gene disruption of the trichodiene synthase (Tri5) gene in three field isolates of different chemotype and virulence. *Mol. Plant Pathol.* **7**, 449-461.
- McDonald, B. et Linde, C.** (2002) The population genetics of plant pathogens and breeding strategies for durable resistance. *Euphytica* **124**, 163-180.
- Merhej, J., Urban, M., Dufresne, M., Hammond-Kosack, K.E., Richard-Forget, F. et Barreau, C.** (2012) The velvet gene, *FgVel1*, affects fungal development and positively regulates trichothecene biosynthesis and pathogenicity in *Fusarium graminearum*. *Mol Plant Pathol* **13**, 363-374.
- Montibus, M., Ducos, C., Bonnin-Verdal, M.-N., Bormann, J., Ponts, N., Richard-Forget, F. et Barreau, C.** (2013) The bZIP transcription factor *fgap1* mediates oxidative stress response and trichothecene biosynthesis but not virulence in *Fusarium graminearum*. *PLoS One* **8**, e83377.
- Osborne, L.E. et Stein, J.M.** (2007) Epidemiology of *Fusarium* head blight on small-grain cereals. *Int. J. Food Microbiol.* **119**, 103-108.
- Pariaud, B., Ravigné, V., Halkett, F., Goyeau, H., Carlier, J. et Lannou, C.** (2009) Aggressiveness and its role in the adaptation of plant pathogens. *Plant Pathol.* **58**, 409-424.
- Paul, P.A., Lipps, P.E. et Madden, L. V** (2005) Relationship between visual estimates of *Fusarium* head blight intensity and deoxynivalenol accumulation in harvested wheat grain: a meta-analysis. *Phytopathology* **95**, 1225-36.
- Pinson-Gadais, L., Foulongne-Oriol, M., Ponts, N., Barreau, C. et Richard-Forget, F.** (2013) The French *Fusarium* Collection: a living resource for mycotoxin research. *Fungal Genet. Reports*, **60(Suppl)**, 295.
- Plank, J.E. Van der** (1963) Plant diseases : epidemics and control. *New York, USA Acad. Press.*
- Prussin, A.J., Li, Q., Malla, R., Ross, S.D. et Schmale, D.G.** (2013) Monitoring the long-distance transport of *Fusarium graminearum* from field-scale sources of inoculum. *Plant Dis.* **98**, 504-511.
- Salvi, S. et Tuberosa, R.** (2005) To clone or not to clone plant QTLs: present and future challenges. *Trends Plant Sci* **10**, 297-304.
- Schmale Iii, D.G., Leslie, J.F., Zeller, K.A., Saleh, A.A., Shields, E.J. et Bergstrom, G.C.** (2006) Genetic structure of atmospheric populations of *Gibberella zeae*. *Phytopathology* **96**, 1021-1026.
- Schneider, C.A., Rasband, W.S. et Eliceiri, K.W.** (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671-675.
- Sieber, C.M.K., Lee, W., Wong, P., et al.** (2014) The *Fusarium graminearum* genome reveals more secondary metabolite gene clusters and hints of horizontal gene transfer. *PLoS One* **9**, e110311.
- Stack, R.W.** (2000) Return of an Old Problem: *Fusarium* Head Blight of Small Grains. *Plant Heal. Prog.*

- Stajich, J.E., Harris, T., Brunk, B.P., et al.** (2012) FungiDB: an integrated functional genomics database for fungi. *Nucleic Acids Res.* **40**, D675-D681.
- Stein, J.M., Osborne, L.E., Bondalapati, K.D., Glover, K.D. et Nelson, C.A.** (2009) *Fusarium* head blight severity and deoxynivalenol concentration in wheat in response to *Gibberella zeae* inoculum concentration. *Phytopathology* **99**, 759-64.
- Stewart, E.L., Croll, D., Lendenmann, M.H., Sanchez-Vallet, A., Hartmann, F.E., Palma-Guerrero, J., Ma, X. et McDonald, B.** (2016) QTL mapping reveals complex genetic architecture of quantitative virulence in the wheat pathogen *Zymoseptoria tritici*. Cold Spring Harbor Labs Journals.
- Studer, A.J. et Doebley, J.F.** (2011) Do large effect QTL fractionate? A case study at the maize domestication QTL teosinte branched1. *Genetics* **188**, 673-81.
- Talas, F., Kalih, R. et Miedaner, T.** (2012a) Within-field variation of *Fusarium graminearum* isolates for aggressiveness and deoxynivalenol production in wheat head blight. *Phytopathology* **102**, 128-134.
- Talas, F., Kalih, R., Miedaner, T. et McDonald, B.A.** (2016) Genome-wide association study identifies novel candidate genes for aggressiveness, deoxynivalenol production and azole sensitivity in natural field populations of *Fusarium graminearum*. *Mol. Plant-Microbe Interact.*
- Talas, F., Würschum, T., Reif, J., Parzies, H. et Miedaner, T.** (2012b) Association of single nucleotide polymorphic sites in candidate genes with aggressiveness and deoxynivalenol production in *Fusarium graminearum* causing wheat head blight. *Bmc Genet.* **13**, 1-10.
- Terzi, V., Tumino, G., Stanca, A.M. et Morcia, C.** (2014) Reducing the incidence of cereal head infection and mycotoxins in small grain cereal species. *J. Cereal Sci.* **59**, 284-293.
- Touati-Hattab, S., Barreau, C., Verdal-Bonnin, M.-N., Chereau, S., Richard-Forget, F., Hadjout, S., Mekliche, L. et Bouznad, Z.** (2016) Pathogenicity and trichothecenes production of *Fusarium culmorum* strains causing head blight on wheat and evaluation of resistance of the varieties cultivated in Algeria. *Eur. J. Plant Pathol.* **145**, 797-814.
- Trail, F.** (2009) For blighted waves of grain: *Fusarium graminearum* in the postgenomics era. *Plant Physiol* **149**, 103-110.
- Varga, E., Wiesenberger, G., Hametner, C., et al.** (2015) New tricks of an old enemy: isolates of *Fusarium graminearum* produce a type A trichothecene mycotoxin. *Env. Microbiol* **17**, 2588-2600.
- Walkowiak, S., Bonner, C.T., Wang, L., Blackwell, B., Rowland, O. et Subramaniam, R.** (2015) Intraspecies interaction of *Fusarium graminearum* contributes to reduced toxin production and virulence. *Mol Plant Microbe Interact* **28**, 1256-1267.
- Walter, S., Nicholson, P. et Doohan, F.M.** (2010) Action and reaction of host and pathogen during *Fusarium* head blight disease. *New Phytol.* **185**, 54-66.
- Ward, T.J., Clear, R.M., Rooney, A.P., et al.** (2008) An adaptive evolutionary shift in *Fusarium* head blight pathogen populations is driving the rapid spread of more toxigenic *Fusarium graminearum* in North America. *Fungal Genet. Biol.* **45**, 473-484.
- Zhang, Y., Lubberstedt, T. et Xu, M.** (2013) The genetic and molecular basis of plant resistance to pathogens. *J. Genet. Genomics* **40**, 23-35.

**Zhu, M. et Zhao, S.** (2007) Candidate gene identification approach: progress and challenges. *Int. J. Biol. Sci.* **3**, 420-7.

# II. Un caractère quantitatif et un gène majeur : un paradoxe ?

La distribution continue et normale de l'agressivité dans les populations de *F. graminearum* suggère un contrôle polygénique. N'identifier qu'un seul QTL peut donc paraître paradoxal. Notre étude est basée sur une source limitée de polymorphisme. A l'échelle d'une population, on peut supposer que le QTL majeur identifié, en combinaison avec d'autres QTL, participe à la variation continue et normale de l'agressivité observée dans les populations naturelles (Talas et al. 2016). Ces autres QTLs n'ont pas été identifiés dans notre analyse, soit parce qu'ils ne ségrégaient pas dans notre population, soit parce que la puissance de notre analyse n'était pas suffisante. La distribution de la plupart des caractères mesurés dans notre expérience n'était pas normale mais plutôt bimodale, traduisant l'action d'un locus majeur. La distribution continue de ces caractères peut être expliquée par l'effet de l'environnement (Falconer & Mackay 1996).

## II.1. *Quelle est la fréquence de l'allèle identifié au gène FgVeA dans les souches de F. graminearum?*

Après avoir identifié la mutation causale dans le gène *FgVeA*, la fréquence de cette mutation a été étudiée dans la centaine de souches de la collection de diversité du laboratoire, incluant celles présentées en Chapitre 1. La souche INRA-171 est la seule des différentes souches testées à présenter cet allèle. La fréquence allélique de cet allèle dans l'échantillonnage est donc d'environ 1%. En supposant que la fréquence calculée dans la collection du laboratoire soit représentative de la situation naturelle, ce résultat suggère que l'allèle *FgVeA* est un allèle à fréquence mineure (MAF pour minor allele frequency), couramment défini pour les allèles avec une fréquence inférieure à 5% (Gorlov et al. 2008).

### II.2. *Est-ce que d'autres QTL contrôlent la variation d'agressivité observée dans la population recombinante ?*

Pour aller un peu plus loin dans l'analyse de détection de QTL de la population recombinante originelle (INRA-156 x INRA-171), et essayer d'identifier d'autres facteurs génétiques impliqués dans la variation d'agressivité dans cette population, l'analyse a été reconduite en utilisant des méthodes de détection décrites comme plus sensibles pour la détection de QTL mineurs (Li et al. 2007; Yi & Shriner 2008). Des analyses de cartographie multi-QTL ont été conduites, en utilisant le logiciel R/qtl (Broman et al. 2003), et en intégrant le QTL majeur lorsqu'il été identifié en cofacteur. Ces études n'ont pas permis d'identifier des QTLs supplémentaires de façon fiable et robuste. Quelques *loci* ont pu être détectés mais avec des LOD score inférieurs au seuil de LOD estimé après 500 permutations (résultats non montrés). Ces *loci* mineurs détectés n'étaient pas stables d'un essai à l'autre. Ces résultats interrogent sur le niveau de confiance que l'on peut accorder à leur détection et sur leur réelle implication dans la variation du caractère. Cette observation illustre l'importance de réaliser plusieurs essais lors de la mise en place d'une démarche de génétique quantitative. Dans un deuxième temps, l'analyse a été reconduite en utilisant le logiciel QTL cartographer (Basten et al. 1994). Les résultats obtenus étaient similaire à ceux détectés avec le logiciel R/qtl. L'emploi de deux logiciels de détection de QTL utilisant des algorithmes de calcul différents a permis de valider notre procédure de détection utilisée tout en consolidant les résultats. Dans un troisième temps, la population a été divisée par rapport à son génotype au locus *FgVeA* et les deux sous populations ont été analysées de façon indépendante. Aucun QTL n'a été identifié. Avec cette stratégie, les positions précédemment observées sur population entière n'étaient plus détectées. Enfin, l'épistasie entre la position au locus *FgVeA* et le reste du génome a été testée mais aucune interaction n'a pu être mise en évidence une quelconque interaction ( $p < 1.04 \times 10^{-4}$  après correction de Bonferroni).

Plusieurs raisons peuvent expliquer pourquoi aucun QTL supplémentaire n'a été identifié.

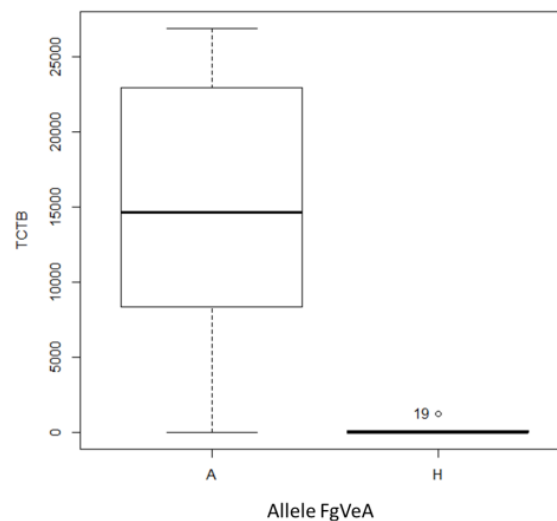
- La puissance statistique de notre analyse est limitée par la taille de la population (n=88) et par l'effet du QTL majeur. Nous ne pouvons pas exclure totalement l'existence d'autres QTL. Pour tester cette hypothèse, il faudrait augmenter la taille de la population génotypée et phénotypée, et réitérer l'analyse (Mackay et al. 2009).
- Le polymorphisme au locus *FgVeA* est l'unique responsable de la variation d'agressivité observée entre les souches parentales. En effet, la contribution du QTL est proche de l'héritabilité calculée pour chaque caractère. Par exemple, l'effet moyen du QTL impliqué dans la capacité à provoquer des symptômes visuels est de 89.4% ; cette valeur est proche de l'héritabilité moyenne calculée pour ce caractère (92.5%).

### II.3. *D'autres polymorphismes affectant le gène FgVeA ?*

L'allèle identifié dans la souche INRA-171 semble être rare. Ceci n'exclue pas que d'autres polymorphismes affectent le gène *FgVeA* chez d'autres souches. Cette hypothèse a été testée en utilisant les génomes séquencés lors du Chapitre I. La souche INRA-195 de notre sous-échantillon présente également une mutation non-synonyme au niveau du gène *FgVeA*, située aussi dans le domaine Velvet de la protéine et modifiant la glycine à la position 118 en asparagine. Cette mutation, différente de celle retrouvée chez la souche INRA-171, est tout aussi rare au sein de la collection du laboratoire puisqu'elle n'a été identifiée que chez la souche INRA-195. Pour répondre de manière préliminaire à la possible implication de cet autre allèle dans la variation du niveau d'agressivité (la souche INRA-195 présente un phénotype ressemblant à celui de la souche INRA-171), une nouvelle population expérimentale a été créée à partir du croisement de la souche INRA-156 $\Delta$ *Mat* avec la souche INRA-195. Une trentaine de souches recombinantes a été isolée, génotypée et phénotypée. Il semblerait que la région chromosomique au locus *FgVeA* soit impliquée dans la variation de la



production de TCTB *in vitro* (Figure CIII\_1). Toutefois, la faible résolution de cette analyse ne permet pas d'affirmer que la mutation identifiée chez la souche INRA-195 est responsable de cette variation. Une centaine de souche issue de ce croisement sont disponibles et pourraient être utilisées pour une nouvelle approche de cartographie génétique afin de confirmer ou non le rôle de cette nouvelle mutation dans la variation d'agressivité de cette nouvelle population.

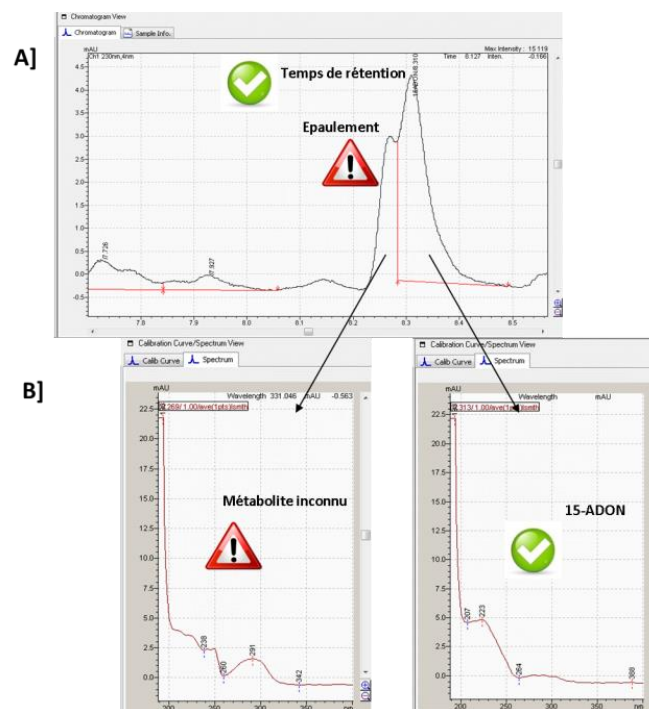


**Figure CIII\_1 :** Représentation en boîte à moustache de la production de TCTB (en µg/mL) en fonction de l'allèle au locus *FgVeA*. « A » est l'allèle sauvage et « H » est l'allèle alternatif présenté par la souche INRA-195.

## III. Des indices supplémentaires sur le rôle du polymorphisme de *FgVeA*.

### III.1. *L'allèle rare de FgVeA est lié à l'accumulation d'un métabolite intermédiaire des TCTB ?*

Lors du dosage des TCTBs des milieux de culture *in vitro* de la population recombinante (INRA-156 x INRA-171), un pic présentant un temps de rétention proche du temps de rétention des pics correspondant au 15-ADON a été identifié (Figure CIII\_2). Ce pic était observé uniquement pour les échantillons des souches présentant l'allèle *FgVeA* rare, suggérant que ces souches produisent un métabolite présentant une polarité proche du 15-ADON. Bien que spéculatif, ce pic pourrait correspondre à un métabolite intermédiaire de la voie des TCTB. Afin de tester cette hypothèse, il serait intéressant d'isoler et purifier cette molécule par chromatographie. La solution purifiée de cette molécule pourrait être ensuite analysée en RMN (résonance magnétique nucléaire), afin d'en déterminer la composition et la structure.

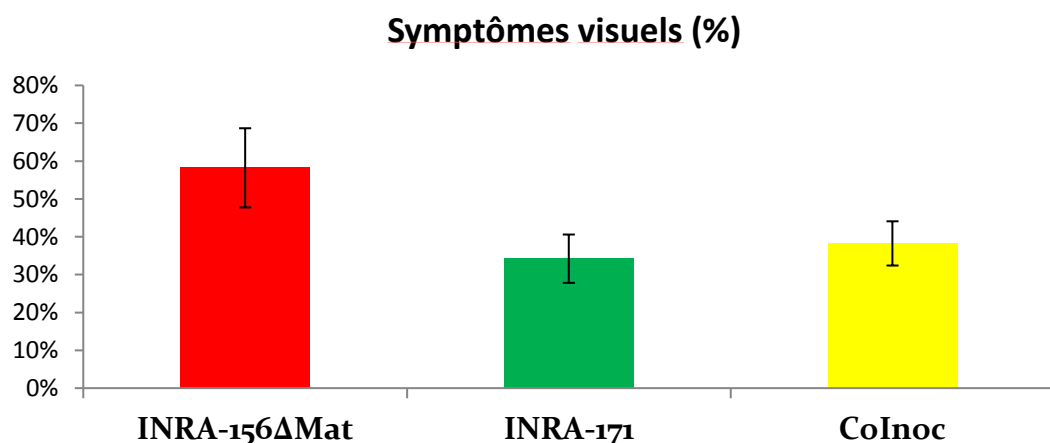


**Figure CIII\_2 :** Illustration du pic inconnu identifié en spectrométrie de masse (détection UV). A. Profils obtenus en chromatographie en mélangeant les deux composés. B. Spectre ultraviolet de chaque sous-composé.

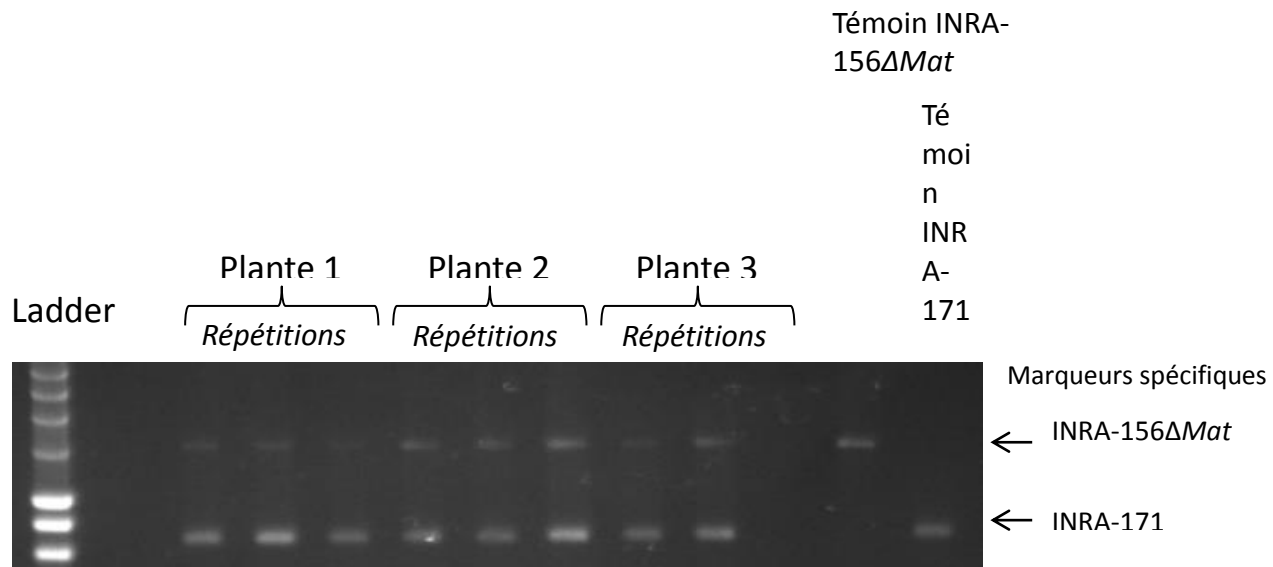
### III.2. Est-ce que la souche INRA-156 et INRA-171 sont compétitives ?

Les niveaux de diversité révélés à l'échelle du champ suggèrent que plusieurs haplotypes peuvent coexister durant l'infection des épis (Talas et al. 2012a). Au vu d'un tel scénario, on peut logiquement se demander quel sera l'interaction entre les deux souches parentales utilisées dans notre analyse et qu'elle sera le résultat de cette interaction ? Cette question a motivé la conduite d'expérimentation exploratoire, réalisée au cours de la deuxième campagne de phénotypage de la descendance, et visant à tester l'interaction entre les deux souches parentales lors de l'infection de l'épi de blé. Pour cela les épis de blés ont été co-inoculés par ces deux souches.

Après 22 jours d'infection, les symptômes visuels sur épis co-inoculés étaient inférieurs à ceux observés lors de l'inoculation de la souche INRA-156 $\Delta$ Mat seule (agressive), et comparables aux épis inoculés par la souche INRA-171 seule (Figure CIII\_3). La présence des deux souches dans les épis récoltés en fin d'essai a été confirmée à l'aide de marqueurs moléculaires spécifiques de chaque souche (Figure CIII\_3).



**Figure CIII\_3 :** Symptômes visuels sur épis de blés (3 plantes x 3 épis / souche) en inoculation souche seule ou en co-inoculation. Significatif entre les symptômes produits par INRA-156 $\Delta$ Mat et INRA-171 ( $p$ -value < 0.01).  $p$ -value = 0.054 entre les symptômes produits par INRA-156 $\Delta$ Mat et produit après la co-inoculation.



**Figure CIII\_4:** Confirmation de la présence de chaque souche dans les épis après amplification par PCR de marqueurs spécifiques.

Ce résultat confirme des observations faites par Walkowiak et al. en 2015 qui ont montré que la co-inoculation entre une souche de chémotype 3-ADON et une souche de chémotype 15-ADON de *F. graminearum* entraîne une diminution des capacités de production de mycotoxine *in vitro* et une diminution des symptômes sur épi, illustrant une compétition intra-spécifique au sein de l'épi (Walkowiak et al. 2015). Nos résultats, bien que préliminaire, laisse penser que la souche moins agressive INRA-171 ne serait pas forcément éliminée au terme de cette interaction et qu'elle serait donc en mesure de transmettre son matériel génétique à la génération suivante.

### IV. Matériels et méthodes

*Sont décrits ici le matériel et méthode utilisés dans les paragraphes II et III.*

**Génotypage des allèles identifiés dans le gène *FgVeA* :** Deux marqueurs Kasp™ (pour Kompetitive allele Specific PCR) ont été développés à partir de l'information des génomes séquencés en Chapitre 1 afin de discriminer pour le gène *FgVeA* l'allèle de la souche INRA-156 (allèle identique à la référence) avec d'une part l'allèle de la souche INRA-171, et d'autre part l'allèle de la souche INRA-195. Quatre-vingt-seize souches de la collection ont été testées. L'amplification est réalisée à l'aide du kit KASP V4.0 4 High Rox (LGC) (1 µl d'ADN, 2 µl 2x KASP master mix, 0.05 µl du mélange d'amorces à 12 µM chacun, qsp 4 µl), dans un thermocycleur de type LightCycler486 (Roche) et selon le programme PCR touch down: (94°C 15 min)x 1; (94°C 20 s, de 65°C à 57°C 1 min -0.8 °C/cycle)x 11; (94°C 20 s, 57°C 1 min)x 32; 15°C 20s. L'attribution des allèles est faite avec le logiciel LightCycler® 480 SW 1.5.1.

**Croisement de la souche INRA-156Δ*Mat* et INRA-195 :** Le croisement de la souche INRA-156Δ*Mat* et INRA-195 a été conduit comme indiqué dans le Chapitre II, à l'exception des ascospores qui ont été récupéré en écrasant les périthèces dans un tube 2mL contenant 1,5mL d'eau stérile. Cette solution a été étalée sur un milieu nutritif PDA et les spores en germinations ont été repiquées sur une nouvelle boîte de culture. Une centaine de souches ont été récupérées et 30 ont été phénotypées pour la production de TCTB *in vitro* (voir protocole Chapitre I), et génotypées sur 12 marqueurs Kasp™ développés grâce aux séquences des génomes parentaux. Le génotypage est réalisé comme décrit ci-dessus.

**Tableau CIII\_1** : les amorces utilisées pour le génotypage Kasp™.

Locus	Chrom	localisation	Amorces allèles spécifiques*	Amorce complémentaire
KSNP40	1	FGRRES_17745	AGAGCTGCTTGACCTACTCCTCCT/G	GCGAGGTCAATCTGTCTTGATCCAT
KSNP53	4	FGRRES_06478	GGTCTTGTCGGATCGCATCG/C	CTGTTCCGTGACACCCCTTGCTTC
KSNP81	1	FGRRES_11955	CAACAACCCGCCTATTCTTACTGG/A	CTTCGGATGGTCGGTCAAGATAGG
KSNP82	1	FGRRES_11955	ATGGAGAGCACGCAGCTCAG/A	AGCCCTGGTCAGCGACTGTCTT
KSNP83	1	FGRRES_01454	CAGCTTCTCTGATGAGGTTGG/A	TAGATCCTCTCCACCAAAGGCAACA
KSNP85	2	intergénique	AGCTTCAGAACCTCATCGTTACCAGTA/C	GCGGGGGCTGAGACCTTCTC
KSNP86	2	intergénique	TTCTGGTGGTTGTTGTCTGATGAT/C	TCAACGCACAAGCTGAAAGGATTCT
KSNP87	2	intergénique	CGACCTGCCCTTTGGTATTATTA/G	GCTCTTTGATCATCGCTAGCTTGAA
KSNP88	2	FGRRES_03446	CATAGGCCATCATACGCTGAAGTG/T	CGGTCCGTACCCCGACTTCAGTA
KSNP89	2	FGRRES_12251	TCAGTTGCTCTGGATCCGTCTTC/T	TTGGCTCTACTCTCGTCGGAGACTC
KSNP95	4	intergénique	GCACGGAAGAGGTGCTAGGG/A	TGATCTAGGGTCTCCAAACTCGATCC
KSNP98	4	intergénique	AGAATCCCAAGCTACTTCTAAGAATCATTG/C	CGTTCCAGGCATGTCATTTTCTAG
KSNP99	4	intergénique	AATAACAAGCAACCACGACAACAAC/A	CCAAATGTATGTGAAGGCACACTGC

\*une séquence spécifique d'un fluorochrome est ajoutée en 5' de chaque amorce allèle spécifique. cette séquence est ensuite reconnue par des oligonucléotides marqués par le fluorochrome correspondant présent dans le KASP master mix.

**Analyse QTL** : L'analyse de détection de QTL multiples proposée par R/qtl a été utilisée en utilisant une fenêtre de 10cM, 3 cofacteurs et en implémentant manuellement la position du QTL majeur identifié précédemment en CIM. L'analyse conduite sur QTL cartographier a été faite avec le module 'CIM' en utilisant les paramètres par défaut. Jusqu'à 5 cofacteurs peuvent être pris en compte. Les cofacteurs sont définis par le logiciel avec le module 'SR'. La taille de la fenêtre est fixée à 10 cM. La méthode 'Forward-backward' a été choisi pour la régression. Pour chaque caractère, le seuil statistique de détection a été déterminé par permutations (500, probabilité associée 0.05). L'épistasie est testée par une ANOVA à deux facteurs selon le modèle :  $Y = \mu + M + m_j + M \times m_j + e$ , où Y est la valeur du caractère, M le marqueur associé au QTL,  $m_j$  les j autres marqueurs du reste du génome,  $M \times m_j$ . L'interaction est considérée comme significative pour  $p < 1,04 \times 10^{-4}$ , (probabilité corrigée par la méthode de Bonferroni).

**Test de co-inoculation** : 3 plantes ont été inoculées (3 épis par plante) en suivant le protocole indiqué dans le paragraphe I de ce chapitre, en utilisant une solution conidiale contenant des concentrations similaires de spores de la souche INRA-

156 $\Delta$ Mat (présentant l'allèle sauvage de *FgVeA*) et la souche INRA-171 (présentant l'allèle rare de *FgVeA*). Les 3 épis par plantes ont été mélangés, lyophilisés et broyés. L'ADN total a été extrait à partir des poudres d'épis de blé. La présence de chaque souche dans les solutions d'ADN issues des épis a été testée en amplifiant le locus *Mat*, correspondant à la cassette de résistance chez la souche INRA-156 $\Delta$ Mat (amorces UpMat2F et NestedHyg\_R) et le gène *Mat 1-2-1* chez la souche INRA-171 (amorces UpMat2F et NestedMat2\_R). La réaction de PCR a été conduite en utilisant l'enzyme GoTaq® suivant ces conditions : 95°C 2min x 1 ; (95°C 30s, 60°C 30s, 72°C 45s pour amplifier le gène *Mat 1-2-1* et 1min30 pour la cassette de résistance à l'hygromycine) x 35 ; 72°C 5 min. Les réactions témoins ont été réalisées à partir de solutions d'ADN obtenues à partir de cultures mycéliennes de la souche INRA-156 $\Delta$ Mat et de la souche INRA-171, ajustées à 10 ng/ $\mu$ L.

**Tableau CIII\_2** : amorces utilisées pour tester la présence de souche dans l'épi.

<b>Nom de l'amorce</b>	<b>Séquence (5' -&gt; 3')</b>
UpMat2_F	5GTCAGTGATCGCCCTTCCAT
NestedHyg_R	TCCGCAACAGTTCTACGCTT
NestedMat2_R	CTAGCGACCTGATTGCCGAT

## V. Conclusion du chapitre 3

L'objectif principal de ces travaux était d'étudier les bases génétiques de l'agressivité au sein d'une population recombinante de *F. graminearum*. Cet objectif a été réalisé avec succès et nos résultats valident l'hypothèse initialement proposée : des variations d'agressivité existent entre les souches de *F. graminearum* et sont sous le contrôle de facteurs génétiques. Dans le cas de la population recombinante construite et étudiée durant ces travaux, ce facteur génétique semble être unique et correspondre à une mutation non-synonyme et rare localisée à l'intérieur du gène *FgVeA*. De nouveaux travaux sont dorénavant nécessaires pour caractériser l'implication moléculaire et évolutive de ce polymorphisme.



# **Discussion générale et perspectives**



### I. Un premier bilan

Les travaux de cette thèse portaient sur l'étude du déterminisme génétique de l'agressivité de *F. graminearum*, en interaction avec le blé. A l'heure actuelle, aucun gène d'agressivité (*Avr*) n'a été identifié chez *F. graminearum*, et aucun gène de résistance majeur (*R*) n'a été identifié chez le blé. En d'autres termes, aucune variété de blé, ou d'autres céréales, ne montre de résistance totale à ce pathogène. La relation qualitative régit par l'interaction *Avr/R* est fréquente pour les pathogènes biotrophes stricts alors qu'elle est beaucoup plus rare pour les pathogènes nécrotrophes (Poland et al. 2009; Zhang et al. 2013). *F. graminearum* est supposé alterner ces deux formes de parasitisme durant l'infection des céréales, et il est souvent décrit comme hémibiotrophe (Kazan et al. 2012). Est-ce que cette forme de parasitisme peut expliquer pourquoi aucune relation de type gène pour gène *Avr/R* n'a été identifiée à ce jour chez ce pathosystème? *Magnaporthe oryzae*, un des pathogènes majeurs du riz, est aussi décrit comme hémibiotrophe et pourtant 9 gènes *Avr* chez ce champignon et 19 gènes *R* chez le riz ont déjà été clonés (Huang et al. 2014). Pour expliquer ces différences, il est fort probable que derrière ce statut généraliste d'hémibiotrophe se cachent des différences importantes de mécanismes d'infection, à l'origine de réactions de défenses variées chez la plante.

De manière intéressante, une grande variabilité des niveaux d'agressivité est observée dans les populations de *F. graminearum* (Talas et al. 2012a; Gilbert et al. 2002). Des études antérieures à ces travaux de thèse ont montré que ce caractère est sous contrôle génétique et dans certains cas, que l'évolution de ce caractère dans les populations semble traduire une adaptation de ce pathogène à son environnement (Talas et al. 2012a; Talas et al. 2016; Ward et al. 2008). Alors qu'une compréhension des bases génétiques de l'agressivité est nécessaire pour comprendre l'évolution de ce pathogène, celles-ci restent mal connues. Pour compléter cette caractérisation, nous avons développé une approche de génétique quantitative et plus précisément de cartographie de QTL sur population expérimentale.

## Discussion générale et perspectives

Cette approche a été construite autour de 3 objectifs : i) caractérisation de la diversité d'une collection de souches du laboratoire et identification des parents potentiels ii) création d'une population et développement d'une carte génétique à haute densité, iii) estimation de l'héritabilité des caractères liés à l'agressivité des souches et localisation sur la carte génétique des *loci* ou QTL impliqués dans la variation d'agressivité et de la production de TCTB observée au sein de la population recombinante.

Ces trois objectifs ont été atteints avec succès :

i) Six souches de *F. graminearum* ont été sélectionnées comme représentatives de la diversité d'une collection de souches françaises. Ces six souches présentaient des capacités d'infection variables, allant de la non-propagation dans l'épi, jusqu'à la nécrose complète de l'épi. Les niveaux de propagation dans l'épi respectaient les potentiels de production en TCTB. Le séquençage du génome de ces six souches a permis une caractérisation à l'échelle du génome de la diversité génétique observée entre ces individus. Il a révélé 243 000 variations génétiques comparées au génome de référence dont 160 000 ont été identifiées entre souches françaises.

ii) Deux de ces six souches ont été sélectionnées pour leur diversité génétique et phénotypique (hautement agressive *versus* modérément agressive ; 63 500 SNPs) et ont été croisées pour permettre la création d'une population recombinante. Le génotypage de cette population à partir de ~1 300 marqueurs moléculaires a permis la construction de la première carte génétique à haute densité de *F. graminearum*, prérequis indispensable pour la détection de QTL.

iii) La population recombinante a été phénotypée pour différents caractères en lien avec l'agressivité des souches. Ces caractères, dont la croissance radiale *in vitro*, les symptômes visuels sur épis ou la production de TCTB, étaient fortement corrélés entre eux et montraient une héritabilité modérée à très forte. Un QTL pléiotropique à effet majeur a été identifié comme impliqué dans la plupart de ces caractères. L'étude de la région génomique contenant le QTL a permis l'identification d'une mutation sur le gène *FgVeA*, probablement rare, comme responsable d'une forte diminution d'agressivité et de production de mycotoxines de *F. graminearum* chez le blé.

## Discussion générale et perspectives

L'utilisation des technologies de séquençage tout au long de cette stratégie d'étude a apporté de nombreuses données de génomiques. Ces données ont été valorisées par une étude approfondie de la dynamique des génomes de *F. graminearum*. Ainsi, nous avons montré que le génome de *F. graminearum* pouvait être divisé en deux types de compartiments : des compartiments hyper polymorphiques contenant les gènes notamment en lien avec la spécificité d'hôte et enrichis en effecteurs, *versus* des compartiments très conservés et contenant les gènes en lien avec le fonctionnement plutôt basal de la cellule. Le lien entre la localisation des gènes et le polymorphisme amène à des questionnements : est-ce la localisation de certains gènes qui génère de la diversité par le biais de la sélection ou est-ce la localisation des gènes dans le génome qui s'accorde en fonction de l'organisation du polymorphisme sur les chromosomes ? La forte ressemblance des profils de polymorphisme observée entre individus suggère que cette organisation est tout d'abord la conséquence de l'architecture des chromosomes plutôt que des forces évolutives associées aux fonctions des gènes. Nous avons apporté des arguments supplémentaires en faveur de cette hypothèse en montrant que la recombinaison méiotique joue probablement un rôle moteur dans cette organisation si particulière, en favorisant le mélange et la diversité des régions hyper polymorphiques, alors que les régions conservées recombinent très peu. Nous supposons que cette organisation du génome permettrait à *F. graminearum* de répondre à ses forts besoins évolutifs tout en protégeant ses fonctions vitales d'effets délétères.

Une telle dichotomie du potentiel évolutif des génomes de champignons pathogènes de plantes a récemment été évoquée et fait l'objet de la théorie des génomes à deux vitesses (Dong et al. 2015). Plusieurs mécanismes moléculaires ont été proposés pour expliquer cette dichotomie, comme le rôle des régulations épigénétiques, ou la mobilité des transposons. Chez *F. graminearum*, ces mécanismes restent très largement inconnus.

Au terme de ces travaux, nous n'avons pas identifié de facteurs génétiques qui pourraient expliquer une hausse des niveaux actuels d'agressivité dans les populations de *F. graminearum*. Ceci n'implique pas que de tels facteurs ne soient pas présents dans les populations naturelles. Une analyse plus large de cette diversité est maintenant

nécessaire pour répondre à cette question. Nous avons toutefois démontré que des facteurs génétiques existent et peuvent avoir un effet fort sur l'agressivité de *F. graminearum*. Ce résultat valide d'autant plus la méthode développée au laboratoire pour l'étude des caractères d'histoire de vie de *F. graminearum*, ainsi que pour d'autres pathogènes. De plus, les connaissances apportées sur le génome de *F. graminearum* auront démontré l'importance de son fort potentiel adaptatif. Ces éléments alertent sur le devenir des populations de ce pathogène dans les agrosystèmes actuels.

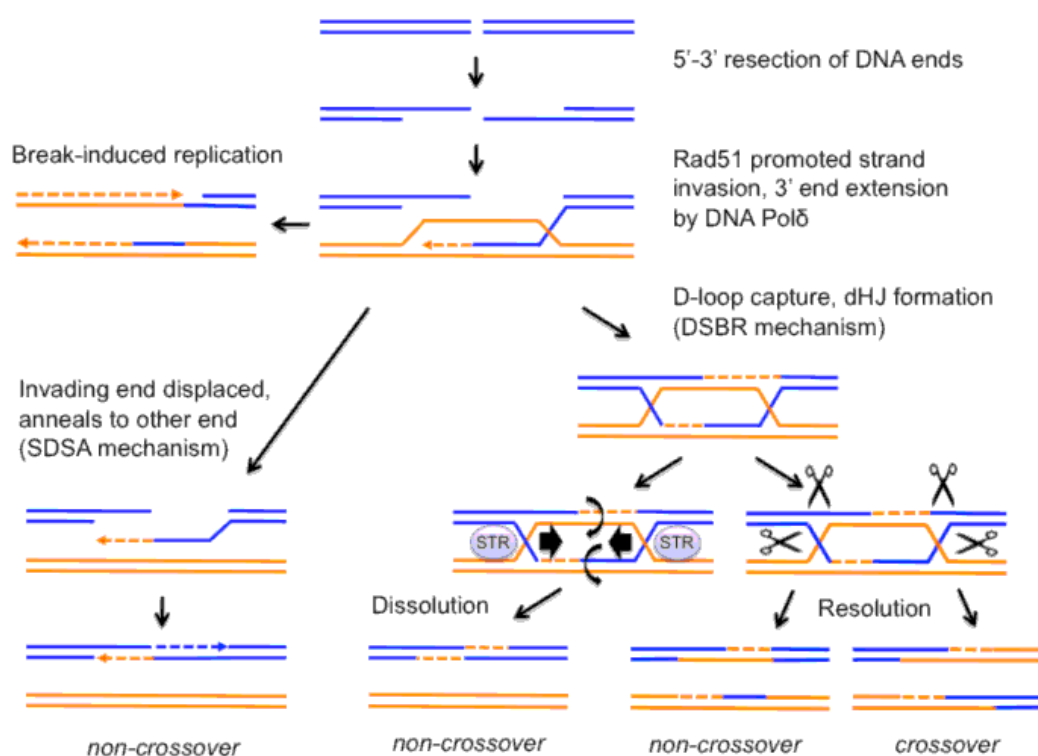
Fort des résultats et des connaissances acquises, de nouvelles hypothèses se dessinent. Je vous propose par la suite d'en détailler quelques-unes, et d'apporter un regard critique sur les approches développées durant ces trois années de travaux.

## **II. Quels sont les éléments régulateurs de la recombinaison chez *F. graminearum* ?**

En plus des résultats sur les bases génétiques de l'agressivité chez *F. graminearum*, nos travaux ouvrent des perspectives pour des études plus fondamentales sur la régulation des génomes eucaryotes, notamment concernant la régulation de la recombinaison méiotique et des crossovers. Les crossovers (COs) correspondent à un échange de chromatide entre paires de chromosomes homologues et se manifestent durant la recombinaison méiotique des génomes eucaryotes. Ils sont issus de la réparation des ruptures des doubles brins (DSB pour double strand break) et permettent la ségrégation des chromosomes homologues durant la méiose, mais aussi une redistribution des allèles dans les générations suivantes. Ce deuxième phénomène joue un rôle moteur pour l'adaptation rapide et efficace des organismes dans un environnement changeant. Les COs ne sont pas les seuls produits de réparation des DSBs puisque ces derniers peuvent être réparés en non-crossover (NCO), amenant à un échange d'ADN beaucoup plus limité (Figure D\_1).

## Discussion générale et perspectives

La localisation des COs est régulée dans les génomes, amenant à des profils spécifiques entre génomes eucaryotes (Mezard et al. 2015 pour revue). Nous avons démontré que chez *F. graminearum*, les crossovers étaient localisés préférentiellement dans les régions télomériques et subtélomériques des chromosomes ainsi que dans des régions centrales. Notre hypothèse est que cette répartition joue un rôle très important dans l'organisation du polymorphisme dans le génome. Des avancées majeures sur la régulation de la recombinaison ont été obtenues chez des espèces modèles (Mercier et al., 2015 et Mezard et al., 2015 pour revues). Il serait donc intéressant d'identifier des homologues des éléments régulateurs de ces espèces chez *F. graminearum* et de caractériser leur implication pour la recombinaison par des approches de mutagenèses classiques.



**Figure D\_1:** Les différences entre la formation des crossovers et les non-crossovers (<http://www.microbiology.columbia.edu/symington/resolvases.html>).

## Discussion générale et perspectives

- Le contrôle CO/NCO

Durant nos travaux, nous n'avons pas pu étudier la localisation des NCO dans le génome, car la taille des fragments d'ADN échangés entre les génomes parentaux était plus petite que la distance entre les marqueurs moléculaires. Il serait pourtant intéressant de tester l'implication de la distribution des DSBs dans le génome sur la distribution des COs chez *F. graminearum*. En d'autres termes, est-ce que la distribution non-aléatoire des COs est un reflet de la distribution des DSBs sur le génome, ou est-ce due à une régulation variable du ratio CO/NCO dans le génome. Chez la levure, certains acteurs moléculaires responsables du contrôle CO/NCO ont déjà été identifiés. Parmi ces régulateurs, le régulateur FANCM a été décrit comme favorisant la formation de DSB en NCO plutôt qu'en CO (Lorenz et al. 2012), et présentant donc une activité anti-CO. Un homologue de ce gène a été identifié chez *F. graminearum* (FGRRES\_17603, données personnelles), et les mutants pour ce gène chez la levure sont viables et fertiles (Lorenz et al. 2012). Une caractérisation fonctionnelle de FGRRES\_17603 semble envisageable chez *F. graminearum*.

La stratégie de recherche consisterait donc à supprimer l'expression de ce gène chez les deux souches parentales utilisées dans notre analyse. La descendance serait ensuite génotypée et une carte génétique pourrait être construite afin d'étudier la localisation des COs dans le génome. Les profils de recombinaison obtenus durant la période doctorale pourraient servir de témoins.

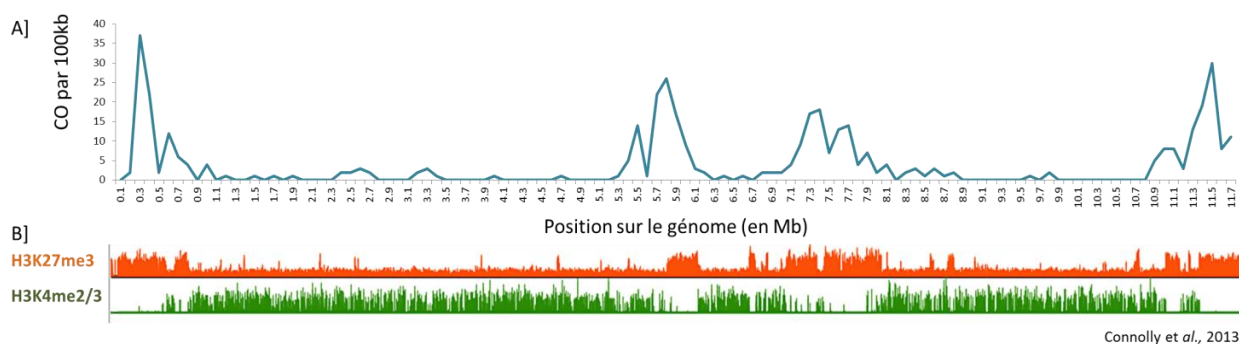
- Le contrôle des DSBs

Dans un deuxième temps, il serait intéressant de tester directement la distribution des DSBs dans le génome. A ce jour, seuls l'homme et la souris disposent de cartes génétiques complètes localisant à la fois les NCOs et les COs (Mezard et al. 2015), permettant d'étudier la fréquence des DSBs dans le génome. Chez ces deux organismes, la localisation des DSBs corrèle en partie avec la localisation des COs, suggérant que la régulation des DSBs joue un rôle important dans la localisation des COs chez ces espèces. D'autres mécanismes ont donc été proposés et étudiés pour expliquer la variation des DSBs. Parmi ces mécanismes, l'état chromatinien semble jouer un rôle majeur (Mezard et al. 2015). L'hypothèse est que les DSBs se forment



## Discussion générale et perspectives

majoritairement dans les régions accessibles de la chromatine (*i.e.* l'euchromatine) alors que l'état plus compacté des autres régions (*i.e.* l'hétérochromatine) limite la formation des COs. Chez de nombreux organismes, H3K4me3 est une marque caractéristique de l'état euchromatinien. Ainsi chez l'homme ou la souris, mais aussi chez la levure, la tri-méthylation des lysines 4 des histones 3 (H3K4me3) est corrélée avec une hausse importante de l'apparition de DSBs, et des COs (Mezard et al. 2015). Chez *F. graminearum*, la première étude visant à étudier les mécanismes épigénétiques de cet organisme (Connolly et al. 2013) montre des résultats inattendus face aux données de recombinaison que nous avons proposé durant la thèse. En effet, comme observé en Figure D\_2, la localisation des marques H3K4me3 dans le génome ne corrèle pas avec une hausse du taux de CO mais l'inverse. C'est la marque H3K27 triméthylée (H3K27me3) qui est enrichie dans les régions recombinantes. Cette dernière est pourtant décrite chez de nombreux organismes comme une marque non-constitutive de l'état hétérochromatinien et elle est communément associée à des régions moins recombinantes (Galazka & Freitag 2014).



**Figure D\_2** : Représentation du chromosome 1. A] le nombre de CO par fenêtre différentes de 100kb observé durant notre analyse B] Résultat (en séquence par kb) du séquençage après immunoprécipitation de la chromatine (ChIP-seq) ciblant les marques H3K27me3 (en orange) et les marques H3K4me2 (en vert), proposé par Connolly et al. (2013).

Pour étudier sur le rôle des modifications épigénétiques dans la régulation de la recombinaison chez *F. graminearum*, le même type d'approche que proposé pour l'étude de l'homologue de FANCM pourrait être développée. Elle serait basée là aussi

sur la mutagenèse d'homologue de régulateurs épigénétiques et de la chromatine chez *F. graminearum*, et l'étude de la recombinaison *via* l'utilisation de cartes génétiques.

L'étude de la recombinaison chez *F. graminearum* apportera d'une part des éléments de réponses théoriques sur la régulation de ce processus chez les eucaryotes, d'autre part une meilleure compréhension des génomes à deux vitesses des champignons phytopathogènes et éclaircira également les mécanismes en lien avec l'organisation du génome de *F. graminearum*. A plus long terme, ce genre d'études permettra de mieux anticiper l'émergence de nouvelles populations de *F. graminearum*. Aussi, ce type d'approche amène à des applications intéressantes puisqu'une augmentation du nombre de COs par recombinaison permettrait, en augmentant le nombre de marqueurs informatifs, la construction de cartes génétiques « ultra » denses et donc la cartographie fine de QTL.

### **III. La cartographie génétique de QTL, une approche pertinente pour l'étude de l'agressivité chez *F. graminearum* ?**

Dans notre cas, un seul QTL majeur a été identifié. Ce dernier est responsable de 90% de la variation des symptômes observés sur les épis. L'héritabilité du caractère calculé dans cette population approchait cette valeur et il est donc fort probable que ce QTL soit le facteur génétique majoritairement impliqué dans la variation d'agressivité observée. Notre analyse de détection de QTL suggère l'existence d'autres *loci* putatifs mais les seuils statistiques utilisés ne nous ont pas permis d'en garantir la fiabilité. Augmenter la taille de la population permettrait d'augmenter la puissance statistique (Mackay et al. 2009). Il serait d'ailleurs pertinent de sélectionner uniquement des souches présentant l'allèle favorable au locus *FgVeA* afin de limiter les biais produits par l'effet fort du QTL. Toutefois, cette approche ne permettrait pas d'identifier, s'il en existe, les QTL mineurs localisés dans la région du locus *FgVeA* dû au déséquilibre de liaison introduit par la sélection systématique de l'allèle favorable.

## Discussion générale et perspectives

La diversité génétique exploitée dans le cadre de la détection de QTL par cartographie génétique utilise une source limitée de la diversité naturelle des populations. Il est certain que d'autres facteurs génétiques contrôlant l'agressivité de *F. graminearum* existent dans les populations naturelles, comme en attestent les travaux de Talas et al. (2016) qui ont identifié 50 QTN associés à des variations d'agressivité observées au sein de 119 souches de *F. graminearum*. Le développement de GWAS, l'autre stratégie majeure pour l'étude des bases génétiques de caractère quantitatif, peut donc apparaître au premier abord comme plus pertinente pour l'étude de caractère aux conséquences épidémiologiques fortes comme l'agressivité (Talas et al. 2016).

Toutefois les approches de GWA présentent aussi des limites, dont la principale est certainement leur faible puissance de détection pour les allèles à faible fréquence dans les populations (allèles MAF, Korte et Farlow, 2013). En effet, la faible fréquence des allèles MAF complique leur étude dans les populations naturelles, due à la faible puissance statistique associée à leurs détections (Gorlov et al. 2008). De fait, ces allèles sont souvent éliminés lors des analyses de GWA. Pourtant, le fait qu'un allèle soit rare dans une population n'implique pas que celui-ci ne soit pas important pour la variation d'un caractère (Gorlov et al. 2008; Mackay et al. 2009; Maher, 2008; Marian, 2012). Chez l'homme par exemple, le nombre d'allèle MAF est important dans le génome (beaucoup de variants génétiques sont peu partagés entre les individus d'une population) et lorsque ces derniers sont liés à la variation d'un caractère, ils apportent généralement des effets plus forts (Gorlov et al. 2011). Leur rôle dans la variation des caractères quantitatifs n'est d'ailleurs pas anodin et des études chez *Drosophila melanogaster* ou chez l'homme montrent que le rôle des allèles MAF dans le contrôle génétique des caractères quantitatifs est important (Mackay et al. 2009). L'héritabilité cachée réfère à la part de la variance génotypique qui n'est pas expliquée par les QTL identifiés (Maher 2008; Marian 2012). La non-détection des allèles MAF peut expliquer une part de cette héritabilité cachée. L'exemple le plus connu est certainement celui de la taille chez l'homme, où 80% de sa variation est estimée comme étant sous le contrôle de facteurs génétiques, alors que les différents facteurs génétiques identifiés par GWAS n'en expliquent qu'environ 5% (Maher 2008). Les allèles MAF ne sont pas

## Discussion générale et perspectives

non plus les seuls éléments responsables d'héritabilité cachée, et d'autres éléments, comme les régulations épigénétiques, l'implication des variations génétiques du nombre de copies (non prises en compte dans les analyses de GWA) ou les biais directement liés au calcul de l'héritabilité, peuvent aussi en expliquer une autre partie (Marian 2012; Koch 2014; Golan et al. 2014).

Nous ne pouvons pas généraliser l'importance des allèles MAF dans la variation des caractères, pour tous les caractères et pour tous les organismes, mais notre étude illustre bien les problèmes liés à leur détection. En effet, en considérant que la fréquence allélique calculée pour l'allèle INRA-171 au gène *FgVeA* dans la collection soit représentative de la fréquence allélique dans les populations naturelles, cet allèle aurait été certainement filtré avant l'analyse. Connaissant le rôle fort que peuvent jouer les allèles MAF dans le contrôle d'un caractère (Gorlov et al. 2011), la mise en évidence du polymorphisme au locus *FgVeA* et de son implication a peut-être été favorisée par le choix de souches parentales aux phénotypes extrêmes.

Pour conclure, dans le cadre de la caractérisation de l'architecture génétique de la pathogénicité de *F. graminearum*, il est certainement inapproprié de penser que l'approche de GWA serait plus adaptée que l'approche de cartographie de QTL classique, ou *vice versa* puisque ces deux approches se complètent. Ainsi, la combinaison de ces deux approches apparaît comme la solution la plus aboutie et la plus prometteuse.

### IV. Quel est le rôle des technologies de séquençage dans le regain d'intérêt de la cartographie de QTL?

D'une manière générale, les études de génétique quantitative (GWA et détection de QTL classique) bénéficient grandement de l'arrivée des outils de séquençage de nouvelles générations pour les étapes de génotypage. Cette amélioration se traduit par une hausse drastique du débit de marqueurs moléculaires génotypés ainsi que la réduction des temps de manipulations (Grover & Sharma 2016; Davey et al. 2011). A titre d'exemple, et alors que 2 à 3 années étaient généralement nécessaires pour construire une carte génétique, 6 mois auront suffi pour construire la carte génétique de *F. graminearum*.

Dans le cadre de l'identification de gènes candidats par cartographie de QTL, une approche de cartographie fine est classiquement réalisée pour réduire l'intervalle de confiance associé au QTL et limité le nombre de gènes candidats (Salvi & Tuberosa 2005). Le principe de la cartographie fine est basée sur l'amélioration de la résolution de la carte, en augmentant le nombre d'événements de recombinaison dans la région (une population plus large et/ou des croisements successifs), et en saturant la région du QTL en marqueurs moléculaires (Salvi & Tuberosa 2005). Un tel type d'approche peut prendre plusieurs mois voire quelques années pour certains modèles biologiques aux temps de génération longs. Cette stratégie peut être grandement facilitée et guidée si on dispose des séquences des génomes parentaux, afin de ne garder comme candidats que les gènes polymorphiques (Christians et al. 2011; Lendenmann et al. 2015; Stewart et al. 2016). Nous avons pu bénéficier d'un tel avantage. L'originalité de notre approche a consisté à utiliser, en plus des séquences parentales, d'autres séquences issues de souches de notre sous-collection afin de diminuer le nombre de gènes candidats potentiellement impliqués dans le caractère étudié. L'identification des gènes candidats par l'annotation du polymorphisme des génomes a été possible en quelques semaines. Cette stratégie, telle que nous l'avons développée, était toutefois risquée car non sans *a priori*. Par exemple, il est important de noter que la mutation

causale n'aurait pas été détectée si elle correspondait à une mutation inter-génique ou synonyme - et affectant par exemple un promoteur ou modulant la stabilité d'un ARNm respectivement (Sauna & Kimchi-Sarfaty 2011). De plus, seul l'allèle porté par le gène *FgVeA* a fait l'objet d'une étude approfondie par des approches de génétique fonctionnelle. Ce choix présentait lui aussi une certaine part de risque. On ne peut d'ailleurs toujours pas exclure que les autres gènes candidats identifiés aient un rôle dans la variation d'agressivité observée. Leur implication potentielle devra être vérifiée. Chez *F. graminearum*, de nombreuses ressources et techniques sont disponibles comme par exemple la disponibilité d'outils de phénotypages adaptés ainsi que de la maîtrise de la mutagenèse, ou la disponibilité d'un génome de référence annoté. Ces éléments ont aussi participé au succès de notre analyse.

En conclusion, et bien que cette approche ne soit pas encore envisageable pour tous les modèles d'études, l'identification de régions génomiques d'intérêt par une stratégie de cartographie génétique classique, et la mise en relation avec des données de génomique, apparaît comme une stratégie très efficace et prometteuse pour identifier rapidement les gènes candidats impliqués dans le contrôle des traits d'histoire de vie des champignons.

## **V. L'étude des génomes : un résultat parallèle ou complémentaire?**

Les approches de génétique quantitative et de génomique apportent toutes les deux des informations sur la base génétique d'un caractère et de son potentiel d'évolution. L'approche de génétique quantitative permet notamment d'estimer l'importance de la variance génétique dans un caractère, le nombre d'éléments impliqués dans sa variation et leurs possibles interactions (Falconer & Mackay 1996). Le re-séquençage des génomes permet d'estimer leur dynamique, notamment dans leurs parties fonctionnelles, de rechercher des traces éventuelles de sélection, et d'identifier des gènes candidats (Grünwald et al. 2016). Outre la possibilité d'identifier les bases moléculaires sous-jacentes aux QTL, la combinaison des deux approches apporte une vraie plus-value pour l'estimation du potentiel évolutif d'un caractère.

Cette estimation est permise sur la base de la localisation des QTL dans le génome, en lien avec la dynamique évolutive associée à ces régions.

Cette interprétation ne prend toutefois pas en compte les différentes forces qui peuvent opérer en plus sur l'évolution de certains gènes. Globalement, il semblerait que *F. graminearum* ne soit pas soumis à de fortes pressions de sélections (Talas & McDonald 2015; Von der Ohe & Miedaner 2010), bien que certains gènes puissent montrer localement des traces de sélections (Sperschneider et al. 2015). Ainsi il n'est pas impossible qu'un gène, même localisé dans une région dynamique, évolue très peu car sous forte sélection purifiante. C'est le cas par exemple des gènes du cluster *Tri* dans les souches analysées durant ces travaux, qui sont localisés dans une région très polymorphique (9,4 variants avec la souche de référence en moyenne par gènes dans la région de 100kb contenant le cluster, dont 3 non-synonyme) et qui sont pour leur part très conservés (2,4 variants avec la souche de référence en moyenne par gène dont 0,8 non-synonyme). Ce résultat est cohérent avec des données proposées par Sperschneider et al. (2015) qui suggère que ces gènes sont sous sélection purifiante. A l'inverse, il n'est pas impossible qu'un gène, même localisé dans un compartiment peu dynamique des chromosomes, puisse évoluer sous l'action de la sélection diversifiante, malgré le fait que cette sélection soit peu efficace du fait des larges déséquilibres de liaisons (Sperschneider et al. 2015).

## **VI. Quel est le rôle de *FgVeA* chez *F. graminearum* ?**

La protéine VeA est un important régulateur du développement et du métabolisme secondaire des champignons (voir Calvo et al. 2016 ou Bayram et al. 2008 pour revues). Le gène *VeA* a été identifié pour la première fois chez *Aspergillus nidulans* (Kafer, 1962) et le modèle d'action des protéines Velvet reste à ce jour le mieux documenté chez cette espèce. En effet, cette protéine n'agit pas seule mais en complexe. Chez cette espèce, 4 protéines Velvet sont décrites: VeA, VelB (Velvet like B), VosA (viability of spores A) et VelC (Velvet like C). Ces protéines sont capables d'interagir et présentent toutes un domaine caractéristique conservé d'environ 150 acides aminés : le domaine

## Discussion générale et perspectives

velvet. La fonction de ce domaine n'est pas très claire mais il semblerait qu'il soit impliqué dans l'interaction protéine-protéine permettant la formation du complexe, ainsi que dans l'interaction ADN-protéine pour l'activation de certains gènes (Bayram & Braus 2012; Ahmed et al. 2013). Chez *A. nidulans*, VeA interagit avec VelB. A l'obscurité, l'hétérodimère VeA-VelB est transporté dans le noyau grâce à l'action de KapA. Dans le noyau VeA-VelB interagit avec LaeA, une méthyltransférase formant ainsi le complexe Velvet (Figure D\_3). Chez *A. nidulans*, LaeA semblerait empêcher la triméthylation des lysines 9 des histones 3, une marque caractéristique d'hétérochromatine (Sarıkaya-Bayram et al. 2015). L'absence de cette marque engendrerait un relâchement de la chromatine et l'accès de la machinerie de transcription aux gènes. Par ce mécanisme, le complexe VeA-VelB-LaeA activerait de nombreux gènes en lien au métabolisme secondaire et en lien avec le développement fongique. L'hétérodimère VeA-VelB est lui aussi actif dans le noyau et favorise l'expression de gènes impliqués dans la formation des organes fructifères, activant la reproduction sexuée (Figure D\_3). VelB est aussi capable d'interagir avec VosA, permettant la répression de la reproduction asexuée. A la lumière ; VeA-VelB n'est plus transporté dans le noyau, amenant à l'induction de la reproduction asexuée au dépend de la reproduction sexuée (Figure D\_3) (Bayram et al. 2008).





aux mutants pour *FgVeA*. De plus, la protéine FgLaeA semble interagir avec la protéine FgVeA (Kim et al. 2013), alors que l'interaction entre FgVeA et FgVelB n'a pas été validée (Jiang et al. 2012).

D'une manière générale, le rôle des protéines Velvet et leurs modes d'action moléculaires sont moins clairs chez *F. graminearum*. En ce sens, la souche INRA-171 disposant de l'allèle muté pour *FgVeA* constitue un matériel biologique de choix pour clarifier le rôle et le fonctionnement de ce gène chez *F. graminearum*.

## **VII. Le polymorphisme de *FgVeA*, quelles implications moléculaires ?**

Durant ces travaux de thèse, nous avons étudié le rôle de la mutation identifiée dans le gène *FgVeA* en remplaçant l'allèle « rare » présent dans la souche INRA-171 par l'allèle « commun » de la souche INRA-156. Après transformation, la souche INRA-171 complétementée montre un phénotype semblable à la souche INRA-156. L'introduction de l'allèle rare de *FgVeA* dans la souche INRA-156 pourrait aussi être envisagée à court terme afin d'apporter une preuve supplémentaire du rôle du polymorphisme associé à *FgVeA* pour le phénotype des souches.

La mutation identifiée chez la souche INRA171 affecte la séquence codante de ce gène. La mutation a été prédite comme étant non-synonyme et modifierait la serine localisée à la position 207 de la protéine en asparagine. La mutation identifiée est intéressante car les sérines sont les cibles fréquentes de modifications post-traductionnelles, notamment par l'ajout de groupement phosphate (Dickman & Yarden 1999). Cette substitution protéique pourrait donc entraîner une dérégulation des modifications post-traductionnelles de FgVeA. De plus, cette mutation est localisée dans le domaine Velvet (prédit), et plus particulièrement dans l'hélice  $\alpha$  dont le rôle dans l'interaction ADN-protéine a été proposé (Ahmed et al. 2013). La première étape serait de vérifier que le gène *FgVeA* est exprimé chez la souche INRA171. Ensuite, si le gène est exprimé, il serait donc intéressant de savoir si la protéine FgVeA codée est fonctionnelle. Les résultats obtenus chez *A. nidulans* présentés ci-dessus suggèrent que

## Discussion générale et perspectives

la protéine VeA est présente de manière constitutive dans la cellule et que son activité diffère *via* des localisations subcellulaires variées, ainsi que la mise en place de multiples complexes. Un premier axe d'étude serait donc de localiser la protéine FgVeA dans les cellules de la souche présentant l'allèle « commun » et dans les cellules de la souche présentant l'allèle « rare », dans différentes conditions (lumière *versus* jour). Il serait également intéressant d'identifier les interactants de ces deux protéines dans ces 2 conditions. L'utilisation des souches modifiées pour produire les protéines FgVeA avec des petites étiquettes protéiques permettraient de mener à bien ce type d'études. L'utilisation d'anticorps « anti-étiquette » permettrait ainsi d'étudier la localisation de la protéine dans la cellule, la purification et la caractérisation des complexes protéiques impliquant FgVeA. La purification des protéines FgVeA permettra de plus de tester les niveaux de phosphorylation des différents allèles.

De manière complémentaire, il serait nécessaire de caractériser l'implication du polymorphisme de *FgVeA* pour le phénotype des souches de *F. graminearum*. Ce point a été étudié à la fin de la période doctorale mais nécessite d'être poursuivi. Plusieurs phénotypes, en lien avec le rôle de *FgVeA* chez *F. graminearum* et d'autres champignons pourront être envisagés (Merhej et al. 2012; Jiang et al. 2011) : la perméabilité des membranes, la sensibilité aux fongicides et autres stress (osmotique par exemple), la concentration en glycérol intracellulaire, la caractérisation des stades sexuels et asexuels en lien avec l'alternance jour/nuit, la pigmentation, la morphologie des conidies et leurs potentiels germinatifs... En plus de ces phénotypes « ciblés », le phénotypage de ces souches pourrait être réalisé de manière plus globale. Cette caractérisation pourrait concerner la souche INRA-171 contenant l'allèle rare de *FgVeA* et sa version complétementée avec l'allèle commune de *FgVeA*, construite pendant la durée de la thèse. L'étude du transcriptome de ces deux souches, à la lumière et à l'obscurité pourrait être envisagée par des études globales de séquençage des ARNm totaux (RNAseq). L'étude du métabolome est également une approche pertinente. La comparaison des données de métabolome et de transcriptome permettront de retracer des réseaux de régulations qui pourront être associés par la suite à des phénotypes caractéristiques. La construction de ces réseaux pourra être menée avec l'appui des

modèles de régulations de l'expression des gènes déjà proposés chez *F. graminearum* (Guo et al. 2016).

A terme, ces différents travaux permettront une caractérisation en profondeur des conséquences moléculaires et phénotypiques du polymorphisme génétique associé à *FgVeA*, et donc du rôle et des mécanismes moléculaires impliquant les protéines *Velvet* chez ce pathogène et chez les champignons.

### **VIII. Un allèle rare de *FgVeA*, un impact sur la fitness des souches ?**

L'allèle nouvellement identifié semble être rare. Quatre hypothèses peuvent être proposées pour expliquer sa faible fréquence dans la collection :

- La fréquence allélique n'est pas représentative de la fréquence allélique dans la nature. On peut en effet imaginer que les échantillonnages réalisés aux champs ont favorisé le prélèvement de souches agressives (compétentes pour la phase biotique) puisqu'ils ont été réalisés sur épis symptomatiques (Talas et al. 2012b).
- L'allèle a été créé durant le stockage de la souche au laboratoire. En effet, le stockage des souches ne prévient pas de l'évolution de celle-ci. Par exemple, King et al. (2015) ont identifié 2 500 mutations lors du re-séquençage de la souche de référence PH-1, qu'ils ont attribué à des mutations spontanées apparues lors du stockage de la souche.
- L'allèle est apparu récemment dans les populations naturelles.
- L'allèle n'est pas récent. Sa fréquence reflète une sélection purifiante ou une élimination par des effets aléatoires de dérive génétique.

Des informations complémentaires sur l'effet du polymorphisme associé au gène *FgVeA* sur le fitness des souches de *F. graminearum* pourraient clarifier la situation. Le fitness, ou la valeur sélective, peut être défini par le succès de survie et reproductif d'un allèle, d'un individu ou d'un groupe (Van der Plank 1963). Il est défini comme sa capacité de croître, de se reproduire et de survivre pendant plusieurs cycles

## Discussion générale et perspectives

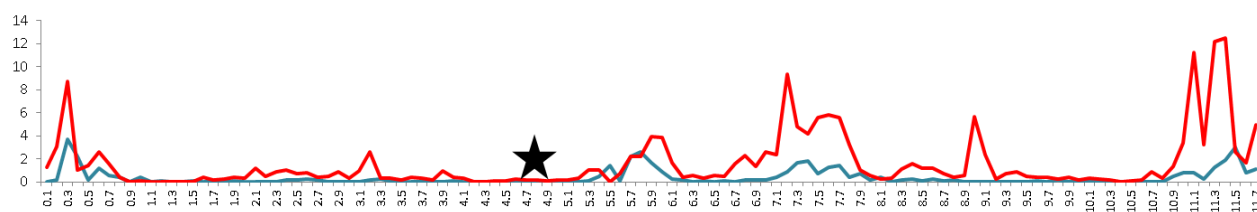
de vie (Corio-Costet et al. 2011). La réalisation de telles études est d'autant plus intéressante qu'elle est appuyée par des résultats préliminaires plutôt surprenants, montrant que la souche INRA-171 modérément agressive, n'est pas éliminée de l'épi à la suite de la mise en interaction avec la souche INRA-156 agressive. De plus, l'agressivité de cette dernière est sensiblement réduite dans ces conditions. Ce phénomène pourrait traduire de la compétition entre les deux souches pour occuper la niche (Walkowiak et al. 2015). Bien que ce résultat soit préliminaire, il demande plus d'investigations.

Pour ce faire, de nouvelles expérimentations pourraient être menées et consisteraient à co-inoculer la souche INRA-171 et sa version complétée par l'allèle commun de *FgVeA* sur des épis de blé. Il faudrait de plus s'intéresser à des stades de vie du champignon autres que la phase biotique, et notamment la phase nécrotrophique hivernale. Une partie des épis co-inoculés pourrait donc être laissée à l'état de débris et servir de source d'inoculum pour une nouvelle génération de culture de blés. Aux différents stades de vie de la plante, l'abondance relative fongique dans les épis et/ou sur les débris pourrait-être mesurée par l'utilisation de marqueurs spécifiques de type Kasp™ (discriminant l'absence ou la présence du SNP dans le gène *FgVeA*). Alors que la technologie Kasp™ lors de ces travaux de thèse a été utilisée pour du génotypage (réponse qualitative), cette dernière peut aussi être employée de façon quantitative, comme démontré lors de l'étude de l'abondance relative de deux génotypes de *Gaeumannomyces graminis* var. *tritici*, un champignon du sol (Gazengel et al. 2013). Enfin, l'étude de la fitness des souches de *F. graminearum* pourrait aussi être entrepris *in vitro*, et en testant des facteurs abiotiques tels que la résistance aux fongicides, ou la tolérance aux températures extrêmes.

*In fine*, ces études apporteraient des informations précieuses et novatrices sur les interactions interspécifiques de souches de *F. graminearum* et sur l'importance des phases hivernales pour l'épidémiologie de ce pathogène, ces deux éléments étant à ce jour très peu documentés.

### IX. Quel est le potentiel d'évolution de *FgVeA*

Le gène *FgVeA* se situe dans une région très conservée (0,58 variant/kb entre isolats français) et non-recombinante du chromosome I (Figure D\_4). Il sera donc hérité aux générations suivantes avec de larges fragments chromosomiques d'un même haplotype parental. Indépendamment des différentes pressions de sélection qui peuvent agir sur ce gène, sa localisation dans le génome ne favorise pas une évolution rapide. D'ailleurs, à la vue du rôle clé que ce gène joue dans le développement de *F. graminearum*, ce résultat est en accord avec notre postulat selon lequel les gènes aux fonctions vitales sont préférentiellement localisés dans des régions conservées et peu dynamiques du génome.



**Figure D\_4 :** Localisation du locus *FgVeA* sur le chromosome I (étoile noire, 4 786 462 – 4 788 117 pb), en rouge le nombre de variants/kb moyen calculé par fenêtre de 100kb, en bleu le nombre de CO (en dizaine) détecté tous les 100 kb.

Paradoxalement, une deuxième mutation non-sens a été identifiée dans le gène *FgVeA* d'une des 5 autres souches étudiées. Même si ce résultat reste préliminaire et nécessite d'être validé, les premières indications suggèrent que cette mutation pourrait aussi être liée à une diminution de l'agressivité observée chez cette souche. S'il est validé, le fait de retrouver une deuxième souche affectée sur le gène *FgVeA* pourrait être un biais lié à la sélection de parents aux phénotypes extrêmes. Une explication plus spéculative serait que *FgVeA* soit soumis à une forte sélection diversifiante. Connaissant le rôle carrefour que joue *FgVeA* pour le développement et le métabolisme des souches de *F. graminearum*, la mutation de ce gène permettrait de répondre rapidement à un nouvel environnement. Le polymorphisme sur ce gène pourrait donc

être favorisé (*i.e.* sélectionné). Alors que le taux moyen de variants localisés dans les gènes contenus dans la même région de 100kb que *FgVeA* est de 1,1 variant par gène comparé à la référence (dont 0,44 non-synonymes), *FgVeA* contient, dans notre échantillonnage, 7 variants comparé à la référence (dont 2 non-synonymes). Des sources de polymorphismes multiples affectant un même gène ne sont pas rares, comme montré pour le gène *PCSK9*, impliqué dans la variation quantitative du niveau de cholestérol chez l'homme (Kotowski et al. 2006). En effet, un spectre de variation nucléotidique affecte ce gène, avec une fréquence de 0,2% à 34% dans la population testée et expliquant 3% à 49% de la variation en cholestérol dans le plasma (Kotowski et al. 2006).

Pour tester l'hypothèse que *FgVeA* soit fréquemment associé à des variations d'agressivités par le biais de mutations variées, il serait intéressant de développer une approche de type gène-candidat telle que proposée précédemment par Talas et al. (2012b) pour l'étude de l'agressivité chez *F. graminearum*. Les séquences codantes pour *FgVeA* pourraient-être amplifiées par PCR chez l'ensemble des souches de la collection, séquencées et comparées aux données déjà disponibles de production TCTB, alors que l'agressivité des souches *in planta* pourrait-être envisagée. Cette approche est d'autant plus intéressante que des souches présentant des phénotypes similaires aux souches INRA-171 et INRA-195 ont été observées dans la collection de travail du laboratoire (données personnelles) et pourraient aussi être affectées par des mutations sur le gène *FgVeA*.

## **X. La génétique de l'agressivité chez *F. graminearum* : un bilan actualisé des connaissances**

Le QTL identifié dans cette analyse se rajoute à la liste des éléments déjà caractérisés comme contrôlant l'agressivité des souches de *F. graminearum* (Figure D\_5). Ces différentes études illustrent bien le fait que l'agressivité chez *F. graminearum* est un caractère extrêmement variable et que cette variation est modérément à

## Discussion générale et perspectives

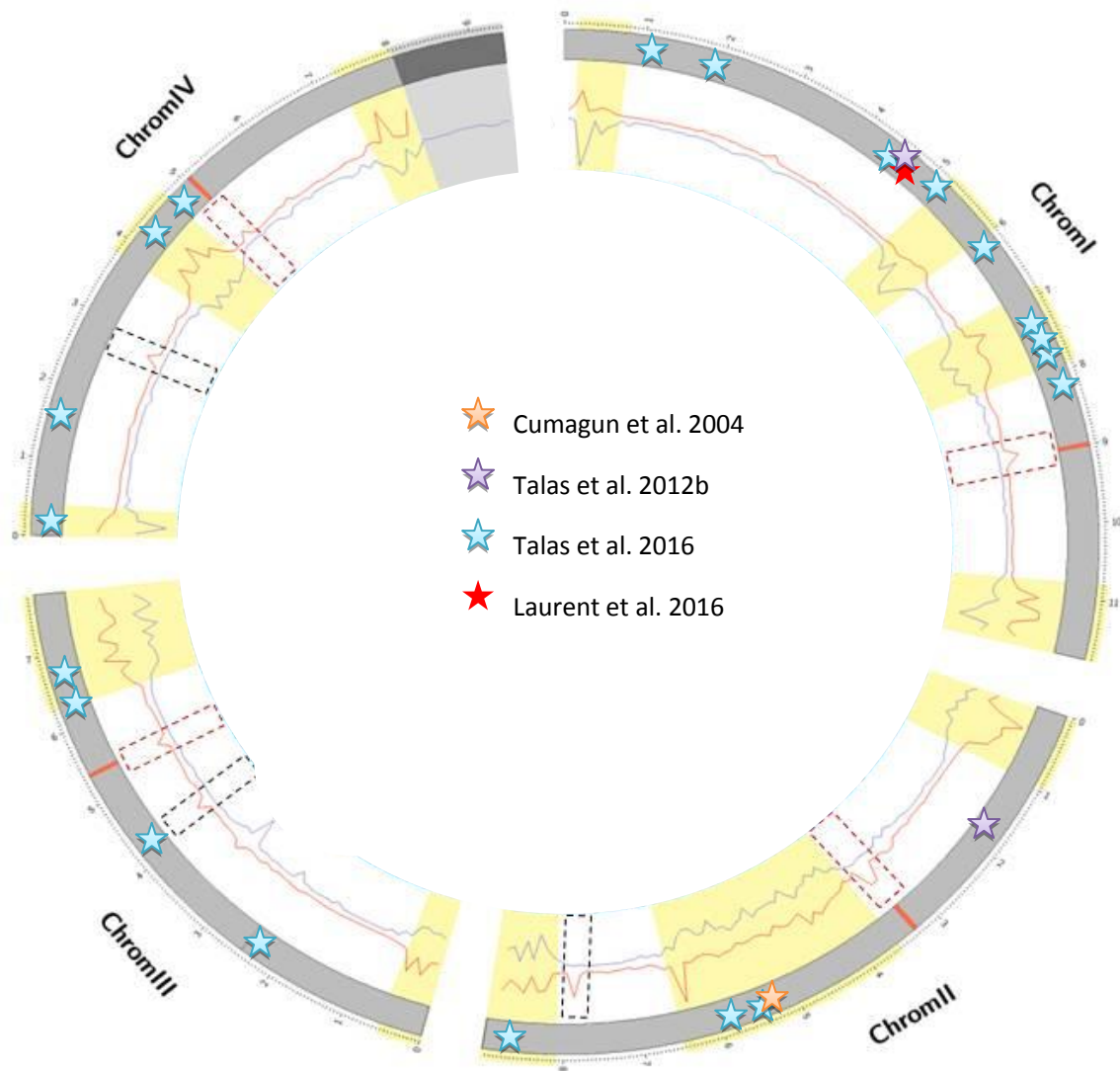
fortement héritable. L'existence de facteurs génétiques suggère donc que l'agressivité de *F. graminearum* est susceptible de répondre à des forces évolutives (dérive génétique ou sélection) et donc d'évoluer dans le temps. Le brassage génétique de ces éléments *via* la reproduction sexuée pourrait notamment créer de nouvelles combinaisons alléliques et amener à des phénotypes transgressifs (Voss et al. 2010).

Les variations d'agressivités sont dans certains cas liées à la variation de production de TCTB (Cumagun, 2004, notre étude). L'implication des mycotoxines est soit qualitative (les souches produisant du DON sont plus agressives chez le blé que les souches produisant du NIV (Cumagun 2004)) ou alors quantitative (la concentration en DON est fortement corrélée aux niveaux d'agressivité (notre étude)). Dans ce dernier cas, la variation du potentiel toxigène ne semble pas être liée à un polymorphisme allélique des gènes *Tri*, impliqués dans la biosynthèse des TCTB. Parfois, la corrélation entre production de mycotoxines et d'agressivité n'est pas observée. Par exemple, Gale et al. (2010) ont rapporté le cas de souches agressives ne produisant pas de mycotoxines. Il a été montré plus tard que ces souches produisaient en réalité une variante des TCTB, et non caractérisée à l'époque (Varga et al. 2015).

A la vue des gènes associés à ce jour dans l'agressivité de *F. graminearum*, il apparaît que les facteurs de transcriptions et autres régulateurs cellulaires, tels que des protéines RAS-GTPase, sont importants pour le contrôle de ce caractère (Talas et al. 2016; Talas et al. 2012b, notre étude, 2016). Les fréquences de ces allèles peuvent être mineures (notre étude, 2016) ou à l'équilibre (Talas et al. 2016). Les facteurs génétiques précédemment identifiés sont majoritairement (55%) localisés dans les régions dynamiques du génome (Talas et al. 2016; Cumagun 2004), mais ce phénomène n'exclue pas que d'autres facteurs soient localisés dans des régions plus conservées (notre étude, Talas et al. 2012b; Talas et al. 2016). Notons par exemple que 2 autres facteurs associés à l'agressivité de *F. graminearum* sont localisés à proximité, et de fait dans la même région « non-dynamique », que *FgVeA* (Figure D\_5). Plus globalement, en contenant 11 des 24 gènes précédemment identifiés, le chromosome I semble jouer un rôle important pour le contrôle de ce caractère (Figure D\_5).



## Discussion générale et perspectives



**Figure D\_5 :** Localisation des différents gènes candidats sur les chromosomes associés à l'agressivité de *F. graminearum* identifiés lors des études précédentes et dans cette étude (échelle en Mb). Les rectangles jaunes sont les régions chromosomiques identifiés comme hautement recombinants (courbe bleue) et polymorphiques (courbe rouge). Les encadrés en pointillés marquent la position des centromères (identifiés aussi par les traits oranges) et les pseudo-centromères.

# XI. Les prochains défis pour une meilleure compréhension de l'agressivité et de son évolution

- Poursuivre le développement de la génétique quantitative chez *F. graminearum*

La détection de QTL sur population expérimentale pour l'étude de la pathogénicité n'avait jamais été réalisée au laboratoire. Les résultats obtenus illustrent la pertinence de cette approche tout en apportant des outils et un recul qui pourront permettre de réitérer l'expérience à plus haut-débit dans le futur. A court terme multiplier le nombre de populations expérimentales pour augmenter les chances de détecter d'autres facteurs génétiques est envisageable. Le choix de souches parentales ne présentant aucun polymorphisme au locus *FgVeA* permettra d'optimiser la détection de nouveaux facteurs génétiques. Comme précisé précédemment, le développement d'une approche combinée de GWAS et de cartographie génétique sur population recombinante est très certainement l'approche la plus complète et la plus prometteuse. Par exemple, une étude réalisée chez le maïs a identifié des QTL par le biais de population expérimentale, dont la position a ensuite été affinée par une approche de GWAS (Mammadov et al. 2015). Dans ce cas-là, la puissance de détection obtenue sur population expérimentale a été complétée par la résolution apportée par l'étude des populations naturelles. Alternativement, nous pouvons donc imaginer qu'une étude de GWAS puisse être menée dans un premier temps pour permettre l'identification de facteurs génétiques. Dans le cas où certains individus présenteraient des phénotypes intéressants mais ne contiendraient aucun des QTN identifiés, ces derniers pourraient être utilisés dans un deuxième temps dans une approche de cartographie génétique sur population expérimentale. Signalons cependant que le développement de GWAS nécessite d'utiliser des populations naturelles de *F. graminearum*, ressources dont ne dispose pas actuellement le laboratoire.

## Discussion générale et perspectives

- Elargir le spectre du phénotypage

Les symptômes visuels sont classiquement mesurés pour étudier l'agressivité des souches de *F. graminearum* (Talas et al. 2016; Talas et al. 2012b; Cumagun 2004). Cependant, ce type de mesure apporte une vision restreinte de la pathogénicité de *F. graminearum*, ne reflétant pas toutes les étapes de l'interaction avec la plante. Ainsi, les variations de symptômes visuelles refléteraient des aptitudes développementales et métaboliques intrinsèques au champignon, variables entre souches, plutôt qu'une capacités d'interaction avec la plante. Hormis le cas des variations observées entre souches produisant du DON ou du NIV (Cumagun 2004), aucun effecteur classiquement décrit comme impliqué lors des interactions du pathogène avec son hôte n'a été à ce jour associé à des variations du niveau d'agressivité. L'apport de phénotypages plus globaux, de type métabolomique et transcriptomique, apporterait une interprétation plus complète de la réelle interaction entre *F. graminearum* et le blé. Alors que l'intégration de ce type de phénotypage dans des études de génétiques quantitatives est très prometteuse chez les espèces fongiques, elle n'a été faite à ce jour que chez la levure (Brem et al. 2002; Breunig et al. 2014; Ambroset et al. 2011). Dans le cadre de l'étude de la pathogénicité, elle fait face à un challenge majeur qui réside à isoler le transcriptome et métabolome fongique à partir des tissus végétaux.

- Comprendre les forces évolutives associées à l'agressivité?

L'identification des facteurs génétiques est une première étape indispensable pour pouvoir estimer l'évolution d'un caractère. Comme nous l'avons montré, la localisation de ces gènes dans le génome permet d'appréhender leur potentiel d'évolution. Toutefois, l'interprétation de ces données n'est pas complète sans une connaissance détaillée des forces évolutives qui s'appliquent effectivement sur ces gènes. Il est notamment important d'estimer si les fréquences alléliques associées à ces gènes reflètent un équilibre, ou à l'inverse, traduisent un effet des forces évolutives. Des approches de génétique de population nécessitant des centaines de souches et une couverture homogène de marqueurs moléculaire sur le génome permettrait d'étudier ces fréquences alléliques et tester l'hypothèse d'équilibre (Talas & McDonald 2015). Alternativement, des approches de génome-scan, nécessitant le séquençage complet (à

## **Discussion générale et perspectives**

faible couverture) de plusieurs dizaines de souches, pourraient être développées pour identifier des empreintes de sélection diversifiante dans la séquence codante des gènes (Persoons et al. 2014). Avantageusement, l'information globale apportée sur les génomes permettrait d'identifier d'autres gènes que ceux identifiés par détection de QTL. Il est également important de compléter la caractérisation de l'agressivité avec l'étude d'autres caractères qui pourraient apporter un meilleur fitness des souches (Lannou 2012), tel que la production de périthèce et d'inoculum, le taux de germination des spores, la résistance aux températures extrêmes, ou au fongicides, la compétition intra et inter spécifique... Ainsi, ces études pourraient révéler des avantages sélectifs ou des compromis entre caractères, et permettraient de mieux comprendre l'évolution des populations de *F. graminearum*.

# Conclusion générale:

Ces travaux de thèse se sont concentrés sur l'interaction se produisant entre *F. graminearum* et un de ses hôtes, le blé. Ils auront démontré que les souches de *F. graminearum* présentent un niveau de diversité génétique et phénotypique remarquable, suggérant un fort potentiel d'adaptation rapide de ces pathogènes dans leur environnement. Cette capacité d'adaptation est optimisée par une architecture chromosomique originale, favorisant l'évolution de gènes agissant à l'interface de l'organisme avec l'environnement, tout en protégeant les gènes aux fonctions vitales d'événements délétères.

Notre étude défend l'existence d'un contrôle génétique important des variations d'agressivité entre souches. Dans le cas des souches étudiées durant ces travaux, la variation d'agressivité a été attribuée majoritairement à un polymorphisme génétique affectant le gène *FgVeA*. Cette découverte amène des perspectives intéressantes sur l'étude des protéines Velvet chez *F. graminearum*.

Le fort contrôle génétique de l'agressivité révélé dans cette analyse et dans les précédentes suggère que ce caractère est susceptible de répondre à des phénomènes de sélection et de dérive génétique et donc d'évoluer dans le temps. De nouvelles études sont dorénavant nécessaires pour évaluer les forces de sélection agissant sur les gènes identifiés et donc prédire l'évolution de l'agressivité de ce pathogène. De plus, il apparaît important de compléter l'étude de l'agressivité par une description plus exhaustive des différents phénotypes pouvant être en lien avec la survie des souches sur le long terme. Cette caractérisation pourrait révéler des compromis entre caractères et/ou avec des stratégies de parasitisme contrastées entre souches, et donc apporter des indications épidémiologiques précieuses pour, à long terme, mieux gérer et anticiper le développement et l'évolution de la fusariose sur les céréales.



## **Bibliographie**





# Bibliographie

- 2002/32/CE Directive 2002/32/CE du Parlement européen et du Conseil du 7 mai 2002 sur les substances indésirables dans les aliments pour animaux,.
- 2006/576/CE COMMISSION RECOMMENDATION of 17 August 2006 on the presence of deoxynivalenol, zearalenone, ochratoxin A, T-2 and HT-2 and fumonisins in products intended for animal feeding,.
- Ahmed, Y.L., Gerke, J., Park, H.-S., et al.** (2013) The Velvet family of fungal regulators contains a dna-binding domain structurally similar to NF- $\kappa$ B stock, A.M., éd. *PLoS Biol.* **11**, e1001750.
- Alexander, N.J., McCormick, S.P., Waalwijk, C., Lee, T. van der et Proctor, R.H.** (2011) The genetic basis for 3-ADON and 15-ADON trichothecene chemotypes in *Fusarium*. *Fungal Genet Biol* **48**, 485-495.
- Ambroset, C., Petit, M., Brion, C., et al.** (2011) Deciphering the molecular basis of wine yeast fermentation traits using a combined genetic and genomic approach. *G3 (Bethesda)*. **1**, 263-81.
- Anderson, J.P., Gleason, C.A., Foley, R.C., Thrall, P.H., Burdon, J.B. et Singh, K.B.** (2010) Plants versus pathogens: an evolutionary arms race. *Funct. Plant Biol.* **37**, 499-512.
- Andrивon, D., Pilet, F., Montarry, J., Hafidi, M., Corbière, R., Achbani, E.H., Pellé, R. et Ellissèche, D.** (2007) Adaptation of *Phytophthora infestans* to partial resistance in potato: evidence from French and Moroccan populations. <http://dx.doi.org/10.1094/PHTO-97-3-0338>.
- Aoki, T., O'Donnell, K. et Geiser, D.** (2014) Systematics of key phytopathogenic *Fusarium* species: current status and future challenges. *J. Gen. Plant Pathol.* **80**, 189-201.
- Attanayake, R.N., Tennekoon, V., Johnson, D.A., Porter, L.D., Rio-Mendoza, L. del, Jiang, D. et Chen, W.** (2014) Inferring outcrossing in the homothallic fungus *Sclerotinia sclerotiorum* using linkage disequilibrium decay. *Heredity (Edinb)*. **113**, 353-363.
- Audenaert, K., Callewaert, E., Höfte, M., Saeger, S. De et Haesaert, G.** (2010)

- Hydrogen peroxide induced by the fungicide prothioconazole triggers deoxynivalenol (DON) production by *Fusarium graminearum*. *BMC Microbiol.* **10**, 1-14.
- Audenaert, K., Vanheule, A., Höfte, M. et Haesaert, G.** (2014) Deoxynivalenol: a major player in the multifaceted response of *Fusarium* to its environment. *Toxins (Basel)*. **6**, 1-19.
- Backhouse, D.** (2014) Global distribution of *Fusarium graminearum*, *F. asiaticum* and *F. boothii* from wheat in relation to climate. *Eur. J. Plant Pathol.* **139**, 161-173.
- Bai, G. et Shaner, G.** (2004) Management and resistance in wheat and barley to *Fusarium* head blight. *Annu Rev Phytopathol* **42**, 135-161.
- Bai, G.H., Desjardins, A.E. et Plattner, R.D.** (2002) Deoxynivalenol-nonproducing *Fusarium graminearum* causes initial infection, but does not cause disease spread in wheat spikes. *Mycopathologia* **153**, 91-98.
- Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A. et Johnson, E.A.** (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* **3**, e3376.
- Basler, R.** (2016) Diversity of *Fusarium* species isolated from UK forage maize and the population structure of *F. graminearum* from maize and wheat. *PeerJ* **4**, e2143.
- Basse, C.W.** (2010) Mitochondrial inheritance in fungi. *Curr Opin Microbiol* **13**, 712-719.
- Basten, C.J., Weir, B.S. et Zeng, Z.-B.** (1994) Zmap-a QTL cartographer. *Proc. 5th World Congr. Genet. Appl. to Livest. Prod. Comput. Strateg. Softw.* **22**, 65-66.
- Baudin, B.** (2014) Malnutrition et sous-alimentation. *Rev. Francoph. des Lab.* **2014**, 25-37.
- Baxter, S.W., Davey, J.W., Johnston, J.S., Shelton, A.M., Heckel, D.G., Jiggins, C.D. et Blaxter, M.L.** (2011) Linkage mapping and comparative genomics using next-generation RAD sequencing of a non-model organism. *PLoS One* **6**.
- Bayram, Ö. et Braus, G.H.** (2012) Coordination of secondary metabolism and development in fungi: the velvet family of regulatory proteins. *FEMS Microbiol. Rev.* **36**.
- Bayram, O., Krappmann, S., Ni, M., et al.** (2008) VelB/VeA/LaeA complex coordinates light signal with fungal development and secondary metabolism.

*Science* **320**, 1504-6.

**Bebber, D.P., Holmes, T., Smith, D. et Gurr, S.J.** (2014) Economic and physical determinants of the global distributions of crop pests and pathogens. *New Phytol.* **202**, 901-910.

**Becher, R., Miedaner, T. et Wirsal, S.G.R.** (2013) 8 Biology, diversity, and management of FHB-Causing *Fusarium* species in small-grain cereals. In *Agricultural Applications.*, p. 199-241. Berlin, Heidelberg: Springer Berlin Heidelberg.

**Beyer, M., Pogoda, F., Pallez, M., Lazic, J., Hoffmann, L. et Pasquali, M.** (2014) Evidence for a reversible drought induced shift in the species composition of mycotoxin producing *Fusarium* head blight pathogens isolated from symptomatic wheat heads. *Int. J. Food Microbiol.* **182**, 51-56.

**Bhat, R. V, Beedu, S.R., Ramakrishna, Y. et Munshi, K.L.** (1989) Outbreak of trichothecene mycotoxicosis associated with consumption of mould-damaged wheat production in Kashmir Valley, India. *Lancet (London, England)* **1**, 35-7.

**Blount, W.** (1961) Turkey « X » disease. *Turkey* **9**.

**Boenisch, M.J., Schäfer, W., Bai, S., et al.** (2011) *Fusarium graminearum* forms mycotoxin producing infection structures on wheat. *BMC Plant Biol.* **11**, 110.

**Bottalico, A. et Perrone, G.** (2002) Toxigenic *Fusarium* species and mycotoxins associated with head blight in small-grain cereals in Europe. *Eur. J. Plant Pathol.* **108**, 611-624.

**Boutigny, A.-L., Ward, T., Ballois, N., Iancu, G. et Ioos, R.** (2014) Diversity of the *Fusarium graminearum* species complex on French cereals. *Eur. J. Plant Pathol.* **138**, 133-148.

**Bowden, R.L. et Leslie, J.F.** (1992) Nitrate-nonutilizing mutants of *Gibberella zeae* (*Fusarium graminearum*) and their use in determining vegetative compatibility. *Exp. Mycol.* **16**, 308-315.

**Bowden, R.L. et Leslie, J.F.** (1999) Sexual Recombination in *Gibberella zeae*. *Phytopathology* **89**, 182-188.

**Bowers, E., Hellmich, R. et Munkvold, G.** (2014) Comparison of fumonisin contamination using HPLC and ELISA methods in *Bt* and near-isogenic maize

- hybrids infested with european corn borer or western bean cutworm. *J. Agric. Food Chem.* **62**, 6463-6472.
- Brakhage, A.A.** (2013) Regulation of fungal secondary metabolism. *Nat Rev Micro* **11**, 21-32.
- Brem, R.B., Yvert, G., Clinton, R. et Kruglyak, L.** (2002) Genetic dissection of transcriptional regulation in budding yeast. *Science* (80-. ). **296**.
- Breunig, J.S., Hackett, S.R., Rabinowitz, J.D., et al.** (2014) Genetic basis of metabolome variation in yeast copenhagen, G.P., éd. *PLoS Genet.* **10**, e1004142.
- Broman, K.W., Wu, H., Sen, S. et Churchill, G.A.** (2003) R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**, 889-890.
- Brown, N.A., Antoniw, J. et Hammond-Kosack, K.E.** (2012) The predicted secretome of the plant pathogenic fungus *Fusarium graminearum*: a refined comparative analysis. *PLoS One* **7**, e33731.
- Brown, N.A., Urban, M., Meene, A.M. van de et Hammond-Kosack, K.E.** (2010) The infection biology of *Fusarium graminearum*: defining the pathways of spikelet to spikelet colonisation in wheat ears. *Fungal Biol* **114**, 555-571.
- Buerstmayr, H., Ban, T. et Anderson, J.A.** (2009) QTL mapping and marker-assisted selection for *Fusarium* head blight resistance in wheat: A review. *Plant Breed.* **128**, 1-26.
- Buerstmayr, H. et Lemmens, M.** (2015) Breeding healthy cereals: genetic improvement of *Fusarium* resistance and consequences for mycotoxins. *World Mycotoxin J.* **8**, 591-602.
- Caffier, V., Lasserre-Zuber, P., Giraud, M., et al.** (2014) Erosion of quantitative host resistance in the apple×*Venturia inaequalis* pathosystem. *Infect. Genet. Evol.* **27**, 481-489.
- Calvo, A.M., Lohmar, J.M., Ibarra, B. et Satterlee, T.** (2016) 18 Velvet regulation of fungal development. in growth, differentiation and sexuality., p. 475-497. Cham: Springer International Publishing.
- Carapito, R., Hatsch, D., Vorwerk, S., Petkovski, E., Jeltsch, J.M. et Phalip, V.** (2008) Gene expression in *Fusarium graminearum* grown on plant cell wall. *Fungal Genet. Biol.* **45**, 738-748.

- Carapito, R., Vorwerk, S., Jeltsch, J.-M. et Phalip, V.** (2013) Genome-wide transcriptional responses of *Fusarium graminearum* to plant cell wall substrates. *FEMS Microbiol. Lett.* **340**, 129-134.
- Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A. et Cresko, W.A.** (2013) Stacks: an analysis tool set for population genomics. *Mol Ecol* **22**, 3124-3140.
- Cavinder, B., Sikhakolli, U., Fellows, K.M. et Trail, F.** (2012) Sexual development and ascospore discharge in *Fusarium graminearum*. *J Vis Exp*.
- Chakraborty, S. et Newton, A.C.** (2011) Climate change, plant diseases and food security: an overview. *Plant Pathol.* **60**, 2-14.
- Charmet, G., Bony, S., Fardet, A., Abecassis, J., Lullien-pellerin, V. et Richard-Forget, F.** (2016) Ouvrage collectif groupe filière INRA, Agriculture et alimentation durables dans la filière céréales. Trois enjeux clés : teneur en protéines, qualités sanitaire et nutritionnelle, Rédactrice du chapitre « qualité sanitaire. *Ed. QUAE*.
- Christians, J.K., Cheema, M.S., Vergara, I.A., Watt, C.A., Pinto, L.J., Chen, N. et Moore, M.M.** (2011) Quantitative trait locus (QTL) mapping reveals a role for unstudied genes in *Aspergillus* virulence. *PLoS One* **6**, e19325.
- Colot, H. V, Park, G., Turner, G.E., Ringelberg, C., Crew, C.M., Litvinkova, L., Weiss, R.L., Borkovich, K.A. et Dunlap, J.C.** (2006) A high-throughput gene knockout procedure for *Neurospora* reveals functions for multiple transcription factors. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 10352-7.
- Connolly, L.R., Smith, K.M. et Freitag, M.** (2013) The *Fusarium graminearum* histone H3 K27 methyltransferase KMT6 regulates development and expression of secondary metabolite gene clusters. *PLoS Genet* **9**, e1003916.
- Corio-Costet, M.-F., Dufour, M.-C., Cigna, J., Abadie, P. et Chen, W.-J.** (2011) Diversity and fitness of *Plasmopara viticola* isolates resistant to QoI fungicides. *Eur. J. Plant Pathol.* **129**, 315-329.
- Cowger, C. et Mundt, C.C.** (2002) Aggressiveness of *Mycosphaerella graminicola* isolates from susceptible and partially resistant wheat cultivars. *Phytopathology* **92**, 624-630.
- Cumagun** (2004) Genetic mapping of pathogenicity and aggressiveness of *Gibberella*

- zeae* (*Fusarium graminearum*) toward wheat. *Genet. Resist.*
- Cumagun, C.J.R. et Miedaner, T.** (2004) Segregation for aggressiveness and deoxynivalenol production of a population of *Gibberella zeae* causing head blight of wheat. *Eur. J. Plant Pathol.* **110**, 789-799.
- Cumagun, C.J.R. et Miedaner, T.** (2003) Aggressiveness of 42 isolates of *Gibberella zeae* (*Fusarium graminearum*) in wheat under field and greenhouse conditions,.
- Cuomo, C.A., Guldener, U., Xu, J.R., et al.** (2007) The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. *Science* (80-. ). **317**, 1400-1402.
- Davey, J.W., Cezard, T., Fuentes-Utrilla, P., Eland, C., Gharbi, K. et Blaxter, M.L.** (2013) Special features of RAD Sequencing data: implications for genotyping. *Mol. Ecol.* **22**, 3151-3164.
- Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M. et Blaxter, M.L.** (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat Rev Genet* **12**, 499-510.
- Dean, R., Kan, J.A.L. Van, Pretorius, Z.A., et al.** (2012) The Top 10 fungal pathogens in molecular plant pathology. *Mol. Plant Pathol.* **13**, 414-430.
- Delmas, C.E.L., Fabre, F., Jolivet, J., Mazet, I.D., Richart Cervera, S., Delière, L. et Delmotte, F.** (2016) Adaptation of a plant pathogen to partial host resistance: selection for greater aggressiveness in grapevine downy mildew. *Evol. Appl.* **9**, 709-725.
- Delmotte, F., Mestre, P., Schneider, C., Kassemeyer, H.-H., Kozma, P., Richart-Cervera, S., Rouxel, M. et Delière, L.** (2014) Rapid and multiregional adaptation to host partial resistance in a plant pathogenic oomycete: Evidence from European populations of *Plasmopara viticola*, the causal agent of grapevine downy mildew. *Infect. Genet. Evol.* **27**, 500-508.
- Dickman, M.B. et Yarden, O.** (1999) Serine/threonine protein kinases and phosphatases in filamentous fungi. *Fungal Genet. Biol.* **26**, 99-117.
- Dong, S., Raffaele, S. et Kamoun, S.** (2015) The two-speed genomes of filamentous pathogens: waltz with plants. *Curr. Opin. Genet. Dev.* **35**, 57-65.
- Dowkiw, A., Voisin, E. et Bastien, C.** (2010) Potential of Eurasian poplar rust to

- overcome a major quantitative resistance factor. *Plant Pathol.* **59**, 523-534.
- Dubin, H.J., International Maize and Wheat Improvement Center., L., Reeves, J. et McNab, A.** (1997) *Fusarium* head scab : global status and future prospects : proceedings of a workshop held at CIMMYT, El Batan, Mexico, 13-17 October, 1996, CIMMYT.
- EFSA** (2013) Scientific opinion on risks for animal and public health related to the presence of nivalenol in food and feed. *EFSA J.* **11**, 3262.
- Eudes, F., Comeau, A., Rioux, S. et Collin, J.** (2000) Phytotoxicité de huit mycotoxines associées à la fusariose de l'épi chez le blé. *Can. J. Plant Pathol.* **22**, 286-292.
- Fabre, F., Rousseau, E., Mailleret, L. et Moury, B.** (2012) Durable strategies to deploy plant resistance in agricultural landscapes. *New Phytol.* **193**, 1064-1075.
- Falconer, D.S. et Mackay, T.** (1996) Introduction to quantitative genetics. **Ed. 4.**
- FAO** (2009) FAO's Director-general on how to feed the world in 2050. *Popul. Dev. Rev.* **35**, 837-839.
- FAO, F. et P.** (2015) L'état de l'insécurité alimentaire dans le monde 2015. Objectifs internationaux 2015 de réduction de la faim: des progrès inégaux. Rome, FAO.
- Ferrigo, D., Raiola, A. et Causin, R.** (2016) *Fusarium* Toxins in Cereals: Occurrence, Legislation, Factors Promoting the Appearance and Their Management. *Molecules* **21**, 627.
- Flood, J.** (2010) The importance of plant health to food security. *Food Secur.* **2**, 215-231.
- Flor, H.H.** (1971) Current status of the gene-for-gene concept. *Annu. Rev. Phytopathol.* **9**, 275-296.
- Foroud, N.A., McCormick, S.P., MacMillan, T., Badea, A., Kendra, D.F., Ellis, B.E. et Eudes, F.** (2012) Greenhouse studies reveal increased aggressiveness of emergent canadian *Fusarium graminearum* chemotypes in wheat. *Plant Dis.* **96**, 1271-1279.
- Foulongne-Oriol, M.** (2012) Genetic linkage mapping in fungi: current state, applications, and future trends. *Appl. Microbiol. Biotechnol.* **95**, 891-904.
- Galazka, J.M. et Freitag, M.** (2014) Variability of chromosome structure in pathogenic fungi--of « ends and odds ». *Curr Opin Microbiol* **20**, 19-26.

- Gale, L., Ward, T.J. et Kistler, H.** (2010) A subset of the newly discovered Northland population of *Fusarium graminearum* from the U.S. does not produce the B-type trichothecenes DON, 15ADON, 3ADON or NIV). *Meet. abstract.*
- Gale, L.R., Bryant, J.D., Calvo, S., et al.** (2005) Chromosome complement of the fungal plant pathogen *Fusarium graminearum* based on genetic and physical mapping and cytological observations. *Genetics* **171**, 985-1001.
- Gandon, S. et Michalakakis, Y.** (2000) Evolution of parasite virulence against qualitative or quantitative host resistance. *Proc. R. Soc. London B Biol. Sci.* **267**.
- Garel, J.-R.** (2015) Contexte socio-culturel et domestication des céréales au Proche-Orient. Phd thesis.
- Gauthier, L., Atanasova-Penichon, V., Chereau, S. et Richard-Forget, F.** (2015) Metabolomics to decipher the chemical defense of cereals against *Fusarium graminearum* and deoxynivalenol accumulation. *Int J Mol Sci* **16**, 24839-24872.
- Gazengel, K., Lebreton, L., Guillerm-Erckelboudt, A.-Y., Marchi, M., Sarniguet, A. et Daval, S.** (2013) Simultaneous monitoring of two fungal genotypes on plant roots by single nucleotide polymorphism quantification with an innovative KASPar quantitative PCR. *Mycol. Prog.* **12**, 657-666.
- Gilbert, J., Abramson, D., McCallum, B. et Clear, R.** (2002) Comparison of Canadian *Fusarium graminearum* isolates for aggressiveness, vegetative compatibility, and production of ergosterol and mycotoxins. *Mycopathologia* **153**, 209-215.
- Glenn, A.E.** (2007) Mycotoxigenic *Fusarium* species in animal feed. *Anim. Feed Sci. Technol.* **137**, 213-240.
- Golan, D., Lander, E.S. et Rosset, S.** (2014) Measuring missing heritability: inferring the contribution of common variants. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E5272-81.
- Gonen, S., Lowe, N.R., Cezard, T., Gharbi, K., Bishop, S.C. et Houston, R.D.** (2014) Linkage maps of the Atlantic salmon (*Salmo salar*) genome derived from RAD sequencing. *BMC Genomics* **15**, 1-17.
- Gorlov, I., Gorlova, O., Frazier, M., Spitz, M. et Amos, C.** (2011) Evolutionary evidence of the effect of rare variants on disease etiology. *Clin. Genet.* **79**, 199-206.
- Gorlov, I.P., Gorlova, O.Y., Sunyaev, S.R., Spitz, M.R. et Amos, C.I.** (2008) Shifting



- paradigm of association studies: value of rare single-nucleotide polymorphisms. *Am. J. Hum. Genet.* **82**, 100-12.
- Goswami, R.S.** (2012) Targeted gene replacement in fungi using a split-marker approach. *Methods Mol. Biol.* **835**, 255-69.
- Goswami, R.S. et Kistler, H.C.** (2004) Heading for disaster: *Fusarium graminearum* on cereal crops. *Mol Plant Pathol* **5**, 515-525.
- Grover, A. et Sharma, P.C.** (2016) Development and use of molecular markers: past and present. *Crit. Rev. Biotechnol.* **36**, 290-302.
- Grünwald, N.J., McDonald, B.A. et Milgroom, M.G.** (2016) Population genomics of fungal and oomycete pathogens. *Annu. Rev. Phytopathol.* **54**, 323-346.
- Guenther, J.C., Hallen-Adams, H.E., Bücking, H., Shachar-Hill, Y. et Trail, F.** (2009) Triacylglyceride metabolism by *Fusarium graminearum* during colonization and sexual development on wheat. *Mol. Plant-Microbe Interact.* **22**, 1492-1503.
- Guo, L., Zhao, G., Xu, J.-R., Kistler, H.C., Gao, L. et Ma, L.-J.** (2016) Compartmentalized gene regulatory network of the pathogenic fungus *Fusarium graminearum*. *New Phytol.* **211**, 527-541.
- Hallen-Adams, H.E., Wenner, N., Kuldau, G.A. et Trail, F.** (2011) Deoxynivalenol biosynthesis-related gene expression during wheat kernel colonization by *Fusarium graminearum*. *Phytopathology* **101**, 1091-1096.
- Harris, L.J., Balcerzak, M., Johnston, A., Schneiderman, D. et Ouellet, T.** (2016) Host-preferential *Fusarium graminearum* gene expression during infection of wheat, barley, and maize. *Fungal Biol.* **120**, 111-123.
- Houston, R.D., Davey, J.W., Bishop, S.C., et al.** (2012) Characterisation of QTL-linked and genome-wide restriction site-associated DNA (RAD) markers in farmed Atlantic salmon. *BMC Genomics* **13**.
- Huang, J., Si, W., Deng, Q., et al.** (2014) Rapid evolution of avirulence genes in rice blast fungus *Magnaporthe oryzae*. *BMC Genet.* **15**, 45.
- Hussein, H.S. et Brasel, J.M.** (2001) Toxicity, metabolism, and impact of mycotoxins on humans and animals. *Toxicology* **167**, 101-134.
- Jia, L.-J. et Tang, W.-H.** (2015) The omics era of *Fusarium graminearum* :

- opportunities and challenges. *New Phytol.* **207**, 1-3.
- Jiang, J., Liu, X., Yin, Y. et Ma, Z.** (2011) Involvement of a Velvet protein *FgVeA* in the regulation of asexual development, lipid and secondary metabolisms and virulence in *Fusarium graminearum*. *PLoS One* **6**, e28291.
- Jiang, J., Yun, Y., Liu, Y. et Ma, Z.** (2012) *FgVelB* is associated with vegetative differentiation, secondary metabolism and virulence in *Fusarium graminearum*. *Fungal Genet. Biol.* **49**, 653-662.
- Jiao, F., Kawakami, A. et Nakajima, T.** (2008) Effects of different carbon sources on trichothecene production and *Tri* gene expression by *Fusarium graminearum* in liquid culture. *FEMS Microbiol. Lett.* **285**, 212-219.
- Jones, J.D.G. et Dangl, J.L.** (2006) The plant immune system. *Nature* **444**, 323-329.
- Jorge, V., Dowkiw, A., Faivre-Rampant, P. et Bastien, C.** (2005) Genetic architecture of qualitative and quantitative *Melampsora larici-populina* leaf rust resistance in hybrid poplar: genetic mapping and QTL detection. *New Phytol.* **167**, 113-127.
- Jurgenson, J.E., Bowden, R.L., Zeller, K.A., Leslie, J.F., Alexander, N.J. et Plattner, R.D.** (2002) A genetic map of *Gibberella zeae* (*Fusarium graminearum*). *Genetics* **160**, 1451-1460.
- Kadis, S., Ciegler, A. et AJL, S.J.** (1971) Preface. In algal and fungal toxins.
- Kaffer, E.** (1962) Translocations in stock strains of *Aspergillus nidulans*. **33**, 59-68.
- Kazan, K., Gardiner, D.M. et Manners, J.M.** (2012) On the trail of a cereal killer: recent advances in *Fusarium graminearum* pathogenomics and host resistance. *Mol Plant Pathol* **13**, 399-413.
- Keller, M., Bergstrom, G. et Shields, E.** (2014) The aerobiology of *Fusarium graminearum*. *Aerobiologia (Bologna)*. **30**, 123-136.
- Keller, N.P., Turner, G. et Bennett, J.W.** (2005) Fungal secondary metabolism — from biochemistry to genomics. *Nat. Rev. Microbiol.* **3**, 937-947.
- Kim, H.K., Cho, E.J., Lee, S., Lee, Y.S. et Yun, S.H.** (2012) Functional analyses of individual mating-type transcripts at *MAT* loci in *Fusarium graminearum* and *Fusarium asiaticum*. *FEMS Microbiol Lett* **337**, 89-96.
- Kim, H.-K., Jo, S.-M., Kim, G.-Y., Kim, D.-W., Kim, Y.-K. et Yun, S.-H.** (2015) A

- large-scale functional analysis of putative target genes of mating-type loci provides insight into the regulation of sexual development of the cereal pathogen *Fusarium graminearum*. *PLoS Genet.* **11**, e1005486.
- Kim, H.-K., Lee, S., Jo, S.-M., McCormick, S.P., Butchko, R.A.E., Proctor, R.H. et Yun, S.-H.** (2013) Functional roles of *FgLaeA* in controlling secondary metabolism, sexual development, and virulence in *Fusarium graminearum*. *PLoS One* **8**, e68441.
- Kimura, M., Tokai, T., O'Donnell, K., Ward, T.J., Fujimura, M., Hamamoto, H., Shibata, T. et Yamaguchi, I.** (2003) The trichothecene biosynthesis gene cluster of *Fusarium graminearum* F15 contains a limited number of essential pathway genes and expressed non-essential genes. *FEBS Lett.* **539**, 105-110.
- King, R., Urban, M., Hammond-Kosack, M., Hassani-Pak, K. et Hammond-Kosack, K.** (2015) The completed genome sequence of the pathogenic ascomycete fungus *Fusarium graminearum*. *BMC Genomics* **16**, 544.
- Koch, L.** (2014) Epigenetics: An epigenetic twist on the missing heritability of complex traits. *Nat Rev Genet.*
- Korte, A. et Farlow, A.** (2013) The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods* **9**, 29.
- Kotowski, I.K., Pertsemlidis, A., Luke, A., Cooper, R.S., Vega, G.L., Cohen, J.C. et Hobbs, H.H.** (2006) A spectrum of *PCSK9* alleles contributes to plasma levels of low-density lipoprotein cholesterol. *Am. J. Hum. Genet.* **78**, 410-422.
- Kuo, H.F., Olsen, K.M. et Richards, E.J.** (2006) Natural variation in a subtelomeric region of *Arabidopsis*: implications for the genomic dynamics of a chromosome end. *Genetics* **173**, 401-417.
- Lannou, C.** (2012) Variation and selection of quantitative traits in plant pathogens. *Annu Rev Phytopathol* **50**, 319-338.
- Lee, J., Jurgenson, J.E., Leslie, J.F. et Bowden, R.L.** (2008) Alignment of genetic and physical maps of *Gibberella zeae*. *Appl Env. Microbiol* **74**, 2349-2359.
- Lee, J., Lee, T., Lee, Y.-W., Yun, S.-H. et Turgeon, B.G.** (2003) Shifting fungal reproductive mode by manipulation of mating type genes: obligatory heterothallism of *Gibberella zeae*. *Mol Microbiol* **50**, 145-152.
- Lee, J., Myong, K., Kim, J.-E., Kim, H.-K., Yun, S.-H. et Lee, Y.-W.** (2012) FgVelB

- globally regulates sexual reproduction, mycotoxin production and pathogenicity in the cereal pathogen *Fusarium graminearum*. *Microbiology* **158**, 1723-33.
- Lee, K.-M., Cho, W.K., Yu, J., Son, M., Choi, H., Min, K., Lee, Y.-W. et Kim, K.-H.** (2014) A comparison of transcriptional patterns and mycological phenotypes following infection of *Fusarium graminearum* by four mycoviruses. *PLoS One* **9**, e100989.
- Lee, T., Han, Y.K., Kim, K.H., Yun, S.H. et Lee, Y.W.** (2002) *Tri3* and *Tri7* determine deoxynivalenol- and nivalenol-producing chemotypes of *Gibberella zeae*. *Appl Env. Microbiol* **68**, 2148-2154.
- Lee, T. Van der, Zhang, H., Diepeningen, A. van et Waalwijk, C.** (2015) Biogeography of *Fusarium graminearum* species complex and chemotypes: a review. *Food Addit Contam Part A Chem Anal Control Expo Risk Assess* **32**, 453-460.
- Lendenmann, M.H., Croll, D. et McDonald, B.A.** (2015) QTL mapping of fungicide sensitivity reveals novel genes and pleiotropy with melanization in the pathogen *Zymoseptoria tritici*. *Fungal Genet. Biol.* **80**, 53-67.
- Lendenmann, M.H., Croll, D., Palma-Guerrero, J., Stewart, E.L. et McDonald, B.A.** (2016) QTL mapping of temperature sensitivity reveals candidate genes for thermal adaptation and growth morphology in the plant pathogenic fungus *Zymoseptoria tritici*. *Heredity (Edinb)*. **116**, 384-394.
- Lendenmann, M.H., Croll, D., Stewart, E.L. et McDonald, B.A.** (2014) Quantitative trait locus mapping of melanization in the plant pathogenic fungus *Zymoseptoria tritici*. *G3 Genes|Genomes|Genetics* **4**, 2519-2533.
- Leslie, J.F., Anderson, L.L., Bowden, R.L. et Lee, Y.-W.** (2007) Inter- and intra-specific genetic variation in *Fusarium*. *Int. J. Food Microbiol.* **119**, 25-32.
- Li, F.Q., Luo, X.Y. et Yoshizawa, T.** (1999) Mycotoxins (trichothecenes, zearalenone and fumonisins) in cereals associated with human red-mold intoxications stored since 1989 and 1991 in China. *Nat. Toxins* **7**, 93-7.
- Li, H., Ye, G. et Wang, J.** (2007) A modified algorithm for the improvement of composite interval mapping. *Genetics* **175**, 361-74.
- Liu, X., Han, S., Wang, Z., Gelernter, J. et Yang, B.Z.** (2013) Variant callers for next-

- generation sequencing data: a comparison study. *PLoS One* **8**, e75619.
- Lorenz, A., Osman, F., Sun, W., et al.** (2012) The fission yeast *FANCM* ortholog directs non-crossover recombination during meiosis. *Science* (80-. ). **336**, e299-e299.
- Lowe, R.G.T., McCorkelle, O., Bleackley, M., Collins, C., Faou, P., Mathivanan, S. et Anderson, M.** (2015) Extracellular peptidases of the cereal pathogen *Fusarium graminearum*. *Front. Plant Sci.* **6**.
- Lysoe, E., Seong, K.Y. et Kistler, H.C.** (2011) The transcriptome of *Fusarium graminearum* during the infection of wheat. *Mol Plant Microbe Interact* **24**, 995-1000.
- Ma, L.-J., Does, H.C. van der, Borkovich, K.A., et al.** (2010) Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* **464**, 367-373.
- Mackay, T.F., Stone, E.A. et Ayroles, J.F.** (2009) The genetics of quantitative traits: challenges and prospects. *Nat Rev Genet* **10**, 565-577.
- Maher, B.** (2008) Personal genomes: The case of the missing heritability. *Nature* **456**, 18-21.
- Maier, F.J., Miedaner, T., Hadel, B., Felk, A., Salomon, S., Lemmens, M., Kassner, H. et Schäfer, W.** (2006) Involvement of trichothecenes in fusarioses of wheat, barley and maize evaluated by gene disruption of the trichodiene synthase (*Tri5*) gene in three field isolates of different chemotype and virulence. *Mol. Plant Pathol.* **7**, 449-461.
- Maldonado-Ramirez, S.L., Schmale Iii, D.G., Shields, E.J. et Bergstrom, G.C.** (2005) The relative abundance of viable spores of *Gibberella zeae* in the planetary boundary layer suggests the role of long-distance transport in regional epidemics of *Fusarium* head blight. *Agric. For. Meteorol.* **132**, 20-27.
- Mammadov, J., Sun, X., Gao, Y., et al.** (2015) Combining powers of linkage and association mapping for precise dissection of QTL controlling resistance to gray leaf spot disease in maize (*Zea mays L.*). *BMC Genomics* **16**, 916.
- Marian, A.J.** (2012) Elements of « missing heritability ». *Curr. Opin. Cardiol.* **27**, 197-201.

- Massman, J., Cooper, B., Horsley, R., et al.** (2011) Genome-wide association mapping of *Fusarium* head blight resistance in contemporary barley breeding germplasm. *Mol. Breed.* **27**, 439-454.
- McDonald, B. et Linde, C.** (2002) The population genetics of plant pathogens and breeding strategies for durable resistance. *Euphytica* **124**, 163-180.
- McKenna, A., Hanna, M., Banks, E., et al.** (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-1303.
- Mefford, H.C. et Trask, B.J.** (2002) The complex structure and dynamic evolution of human subtelomeres. *Nat Rev Genet* **3**, 91-102.
- Menke, J., Weber, J., Broz, K. et Kistler, H.C.** (2013) Cellular development associated with induced mycotoxin synthesis in the filamentous fungus *Fusarium graminearum*. *PLoS One* **8**, e63077.
- Mercier, R., Mézard, C., Jenczewski, E., Macaisne, N. et Grelon, M.** (2015) The molecular biology of meiosis in plants. *Annu. Rev. Plant Biol.* **66**, 297-327.
- Merhej, J., Richard-Forget, F. et Barreau, C.** (2011) Regulation of trichothecene biosynthesis in *Fusarium*: recent advances and new insights. *Appl. Microbiol. Biotechnol.* **91**, 519-528.
- Merhej, J., Richard-Forget, F. et Barreau, C.** (2011b) The pH regulatory factor *Pac1* regulates *Tri* gene expression and trichothecene production in *Fusarium graminearum*. *Fungal Genet Biol* **48**, 275-284.
- Merhej, J., Urban, M., Dufresne, M., Hammond-Kosack, K.E., Richard-Forget, F. et Barreau, C.** (2012) The velvet gene, *FgVe1*, affects fungal development and positively regulates trichothecene biosynthesis and pathogenicity in *Fusarium graminearum*. *Mol Plant Pathol* **13**, 363-374.
- Mezard, C., Jahns, M.T. et Grelon, M.** (2015) Where to cross? New insights into the location of meiotic crossovers. *Trends Genet* **31**, 393-401.
- Miedaner, T., Reinbrecht, C. et Schilling, A.G.** (2000) Association among aggressiveness, fungal colonization, and mycotoxin production of 26 isolates of *Fusarium graminearum* in winter rye head blight. *Zeitschrift Fur Pflanzenkrankheiten Und Pflanzenschutz-Journal Plant Dis. Prot.* **107**, 124-134.

- Montibus, M., Ducos, C., Bonnin-Verdal, M.-N., Bormann, J., Ponts, N., Richard-Forget, F. et Barreau, C.** (2013) The bZIP Transcription Factor *Fgap1* Mediates Oxidative Stress Response and Trichothecene Biosynthesis But Not Virulence in *Fusarium graminearum*. *PLoS One* **8**, e83377.
- Morgavi, D.P. et Riley, R.T.** (2007) *Fusarium* and their toxins: Mycology, occurrence, toxicity, control and economic impact. *Anim. Feed Sci. Technol.* **137**, 199-200.
- Nganje, W., Bangsund, D. et Leistriz, F.** (2002) Estimating the economic impact of a crop disease: the case of *Fusarium* head blight in US wheat and barley. *Natl. Fusarium Head Blight Forum Proc.*, 275-281.
- Nimchuk, Z., Eulgem, T., Holt III, B.F. et Dangl, J.L.** (2003) Recognition and Response in the Plant Immune System. *Annu. Rev. Genet.* **37**, 579-609.
- O'Donnell, K., Ward, T.J., Geiser, D.M., Corby Kistler, H. et Aoki, T.** (2004) Genealogical concordance between the mating type locus and seven other nuclear genes supports formal recognition of nine phylogenetically distinct species within the *Fusarium graminearum* clade. *Fungal Genet Biol* **41**, 600-623.
- Oerke, E.-C.** (2006) Crop losses to pests. *J. Agric. Sci.* **144**, 31-43.
- Ohe, C. Von der et Miedaner, T.** (2010) Competitive aggressiveness in binary mixtures of *Fusarium graminearum* and *F. culmorum* isolates inoculated on spring wheat with highly effective resistance QTL. *J. Phytopathol.* **159**, no-no.
- Osborne, L.E. et Stein, J.M.** (2007) Epidemiology of *Fusarium* head blight on small-grain cereals. *Int. J. Food Microbiol.* **119**, 103-108.
- Osborn, A.** (2010) Secondary metabolic gene clusters: evolutionary toolkits for chemical innovation. *Trends Genet.* **26**, 449-457.
- Pan, D., Mionetto, A., Calero, N., Reynoso, M.M., Torres, A. et Bettucci, L.** (2016) Population genetic analysis and trichothecene profiling of *Fusarium graminearum* from wheat in Uruguay. *Genet. Mol. Res.* **15**.
- Pariaud, B., Ravigné, V., Halkett, F., Goyeau, H., Carlier, J. et Lannou, C.** (2009) Aggressiveness and its role in the adaptation of plant pathogens. *Plant Pathol.* **58**, 409-424.
- Pariaud, B., Robert, C., Goyeau, H. et Lannou, C.** (2009) Aggressiveness components and adaptation to a host cultivar in wheat leaf rust. *Phytopathology*

99, 869-878.

**Passions Céréales** (2012) « Nourrir 9 milliards d'hommes en 2050 ». *Doss. thématique*.

**Paul, P.A., Lipps, P.E. et Madden, L. V** (2005) Relationship between visual estimates of *Fusarium* head blight intensity and deoxynivalenol accumulation in harvested wheat grain: a meta-analysis. *Phytopathology* **95**, 1225-36.

**Pernaci, M., Mita, S. De, Andrieux, A., Petrowski, J., Halkett, F., Duplessis, S. et Frey, P.** (2014) Genome-wide patterns of segregation and linkage disequilibrium: the construction of a linkage genetic map of the poplar rust fungus *Melampsora larici-populina*. *Front. Plant Sci.* **5**.

**Persoons, A., Morin, E., Delaruelle, C., Payen, T., Halkett, F., Frey, P., Mita, S. De et Duplessis, S.** (2014) Patterns of genomic variation in the poplar rust fungus *Melampsora larici-populina* identify pathogenesis-related factors. *Front. Plant Sci.* **5**.

**Pestka, J.J.** (2007) Deoxynivalenol: Toxicity, mechanisms and animal health risks. *Anim. Feed Sci. Technol.* **137**, 283-298.

**Petes, T.D.** (2001) Meiotic recombination hot spots and cold spots. *Nat Rev Genet* **2**, 360-369.

**Picot, A., Hourcade-Marcolla, D., Barreau, C., Pinson-Gadais, L., Caron, D., Richard-Forget, F. et Lannou, C.** (2012) Interactions between *Fusarium verticillioides* and *Fusarium graminearum* in maize ears and consequences for fungal development and mycotoxin accumulation. *Plant Pathol.* **61**, 140-151.

**Pinson-Gadais, L., Foulongne-Oriol, M., Ponts, N., Barreau, C. et Richard-Forget, F.** (2013) The French *Fusarium* Collection: a living resource for mycotoxin research. *Fungal Genet. Reports*, **60**(Suppl), 295.

**Plank, J.E. Van der** (1963) Plant diseases : epidemics and control. *New York, USA Acad. Press.*

**Poland, J.A., Balint-Kurti, P.J., Wisser, R.J., Pratt, R.C. et Nelson, R.J.** (2009) Shades of gray: the world of quantitative disease resistance. *Trends Plant Sci* **14**, 21-29.

**Ponts, N.** (2015) Mycotoxins are a component of *Fusarium graminearum* stress-response system. *Front Microbiol* **6**, 1234.



- Ponts, N., Couedelo, L., Pinson-Gadais, L., et al.** (2009) *Fusarium* response to oxidative stress by H<sub>2</sub>O<sub>2</sub> is trichothecene chemotype-dependent. *FEMS Microbiol. Lett.* **293**, 255-62.
- Postic, J., Cosic, J., Vrandecic, K., Jurkovic, D., Saleh, A.A. et Leslie, J.F.** (2012) Diversity of *Fusarium* species isolated from weeds and plant debris in Croatia. *J. Phytopathol.* **160**, 76-81.
- Prandini, A., Sigolo, S., Filippi, L., Battilani, P. et Piva, G.** (2009) Review of predictive models for *Fusarium* head blight and related mycotoxin contamination in wheat. *Food Chem. Toxicol.* **47**, 927-931.
- Prussin, A.J., Li, Q., Malla, R., Ross, S.D. et Schmale, D.G.** (2013) Monitoring the long-distance transport of *Fusarium graminearum* from field-scale sources of inoculum. *Plant Dis.* **98**, 504-511.
- Pusztahelyi, T., Holb, I.J. et Pácsi, I.** (2015) Secondary metabolites in fungus-plant interactions. *Front. Plant Sci.* **6**, 573.
- Quinlan, A.R. et Hall, I.M.** (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-2.
- Ramirez, M.L., Reynoso, M.M., Farnochi, M.C. et Chulze, S.** (2006) Vegetative compatibility and mycotoxin chemotypes among *Fusarium graminearum* (*Gibberella zeae*) isolates from wheat in Argentina. *Eur. J. Plant Pathol.* **115**, 139-148.
- Rampitsch, C., Day, J., Subramaniam, R. et Walkowiak, S.** (2013) Comparative secretome analysis of *Fusarium graminearum* and two of its non-pathogenic mutants upon deoxynivalenol induction *in vitro*. *Proteomics* **13**, 1913-1921.
- Reddy, K., Salleh, B., Saad, B., Abbas, H., Abel, C. et Shier, W.** (2010) An overview of mycotoxin contamination in foods and its implications for human health. <http://dx.doi.org/10.3109/15569541003598553>.
- Rep, M.** (2005) Small proteins of plant-pathogenic fungi secreted during host colonization. *FEMS Microbiol. Lett.* **253**, 19-27.
- Reverberi, M., Ricelli, A., Zjalic, S., Fabbri, A.A. et Fanelli, C.** (2010) Natural functions of mycotoxins and control of their biosynthesis in fungi. *Appl. Microbiol. Biotechnol.* **87**, 899-911.

- Rice, P., Longden, I. et Bleasby, A.** (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* **16**, 276-7.
- Salvi, S. et Tuberosa, R.** (2005) To clone or not to clone plant QTLs: present and future challenges. *Trends Plant Sci.* **10**, 297-304.
- Sarikaya-Bayram, Ö., Palmer, J.M., Keller, N., Braus, G.H. et Bayram, Ö.** (2015) One Juliet and four Romeos: VeA and its methyltransferases. *Front. Microbiol.* **6**, 1.
- Sarwar, H., Sarwar, M.F., Sarwar, M., Qadri, N.A. et Moghal, S.** (2013) The importance of cereals (*Poaceae: Gramineae*) nutrition in human health: A review. *J. Cereal. Oilseeds* **4**, 32-35.
- Sauna, Z.E. et Kimchi-Sarfaty, C.** (2011) Understanding the contribution of synonymous mutations to human disease. *Nat. Rev. Genet.* **12**, 683-691.
- Schmale Iii, D.G., Leslie, J.F., Zeller, K.A., Saleh, A.A., Shields, E.J. et Bergstrom, G.C.** (2006) Genetic structure of atmospheric populations of *Gibberella zeae*. *Phytopathology* **96**, 1021-1026.
- Schwendemann, A.B., Decombeix, A.-L., Taylor, T.N., Taylor, E.L. et Krings, M.** (2011) Morphological and functional stasis in mycorrhizal root nodules as exhibited by a Triassic conifer. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 13630-4.
- Selin, C., Kievit, T.R. de, Belmonte, M.F. et Fernando, W.G.D.** (2016) Elucidating the role of effectors in plant-fungal interactions: progress and challenges. *Front Microbiol* **7**.
- Seong, K.Y., Pasquali, M., Zhou, X., Song, J., Hilburn, K., McCormick, S., Dong, Y., Xu, J.R. et Kistler, H.C.** (2009) Global gene regulation by *Fusarium* transcription factors *Tri6* and *Tri10* reveals adaptations for toxin biosynthesis. *Mol Microbiol* **72**, 354-367.
- Shailia, F.Z., Barreau, C. et Foulongne-Oriol, M.** (2012) Etude de la diversité génétique et phénotypique de *Fusarium graminearum*, producteur de toxines dans les céréales.
- Sieber, C.M.K., Lee, W., Wong, P., et al.** (2014) The *Fusarium graminearum* genome reveals more secondary metabolite gene clusters and hints of horizontal gene transfer. *PLoS One* **9**, e110311.

- Simpson, D.R., Weston, G.E., Turner, J.A., Jennings, P. et Nicholson, P.** (2001) Differential control of head blight pathogens of wheat by fungicides and consequences for mycotoxin contamination of grain. *Eur. J. Plant Pathol.* **107**, 421-431.
- Sims, D., Sudbery, I., Ilott, N.E., Heger, A. et Ponting, C.P.** (2014) Sequencing depth and coverage: key considerations in genomic analyses. *Nat Rev Genet* **15**, 121-132.
- Smith, M.-C., Madec, S., Coton, E. et Hymery, N.** (2016) Natural co-occurrence of mycotoxins in foods and feeds and their in vitro combined toxicological effects. *Toxins (Basel)*. **8**, 94.
- Spencer, C.C.A., Deloukas, P., Hunt, S., Mullikin, J., Myers, S., Silverman, B., Donnelly, P., Bentley, D. et McVean, G.** (2006) The influence of recombination on human genetic diversity. *PLoS Genet.* **2**, e148.
- Sperschneider, J., Gardiner, D.M., Thatcher, L.F., Lyons, R., Singh, K.B., Manners, J.M. et Taylor, J.M.** (2015) Genome-wide analysis in three *Fusarium* pathogens identifies rapidly evolving chromosomes and genes associated with pathogenicity. *Genome Biol. Evol.*
- Spolti, P., Ponte, E.M. Del, Cummings, J.A., Dong, Y. et Bergstrom, G.C.** (2013) Fitness attributes of *Fusarium graminearum* isolates from wheat in new york possessing a 3-adon or 15-adon trichothecene genotype. *Phytopathology* **104**, 513-519.
- Stack, R.W.** (2000) Return of an old problem: *Fusarium* head blight of small grains. *Plant Heal. Prog.*
- Starkey, D.E., Ward, T.J., Aoki, T., et al.** (2007) Global molecular surveillance reveals novel *Fusarium* head blight species and trichothecene toxin diversity. *Fungal Genet Biol* **44**, 1191-1204.
- Stewart, E.L., Croll, D., Lendenmann, M.H., Sanchez-Vallet, A., Hartmann, F.E., Palma-Guerrero, J., Ma, X. et McDonald, B.** (2016) QTL mapping reveals complex genetic architecture of quantitative virulence in the wheat pathogen *Zymoseptoria tritici*. Cold Spring Harbor Labs Journals.

- Talas, F., Kalih, R. et Miedaner, T.** (2012a) Within-field variation of *Fusarium graminearum* isolates for aggressiveness and deoxynivalenol production in wheat head blight. *Phytopathology* **102**, 128-134.
- Talas, F., Kalih, R., Miedaner, T. et McDonald, B.A.** (2016) Genome-wide association study identifies novel candidate genes for aggressiveness, deoxynivalenol production and azole sensitivity in natural field populations of *Fusarium graminearum*. *Mol. Plant-Microbe Interact.*
- Talas, F. et McDonald, B.** (2015) Genome-wide analysis of *Fusarium graminearum* field populations reveals hotspots of recombination. *BMC Genomics* **16**, 996.
- Talas, F., Parzies, H.K. et Miedaner, T.** (2011) Diversity in genetic structure and chemotype composition of *Fusarium graminearum* sensu stricto populations causing wheat head blight in individual fields in Germany. *Eur. J. Plant Pathol.* **131**, 39-48.
- Talas, F., Wurschum, T., Reif, J.C., Parzies, H.K. et Miedaner, T.** (2012b) Association of single nucleotide polymorphic sites in candidate genes with aggressiveness and deoxynivalenol production in *Fusarium graminearum* causing wheat head blight. *BMC Genet* **13**, 14.
- Terzi, V., Tumino, G., Stanca, A.M. et Morcia, C.** (2014) Reducing the incidence of cereal head infection and mycotoxins in small grain cereal species. *J. Cereal Sci.* **59**, 284-293.
- Trail, F.** (2009) For blighted waves of grain: *Fusarium graminearum* in the postgenomics era. *Plant Physiol* **149**, 103-110.
- Trail, F. et Common, R.** (2000) Perithecial Development by *Gibberella zeae*: A Light Microscopy Study. *Mycologia* **92**, 130.
- Tsai, H.Y., Robledo, D., Lowe, N.R., Bekaert, M., Taggart, J.B., Bron, J.E. et Houston, R.D.** (2016) Construction and annotation of a high density SNP linkage map of the atlantic salmon (*Salmo salar*) genome. *Genes|Genomes|Genetics* **6**, 2173-2179.
- Ueno, Y.** (1983) Trichothecenes : chemical. biological and toxicological and aspects.
- Varga, E., Wiesenberger, G., Hametner, C., et al.** (2015) New tricks of an old enemy: isolates of *Fusarium graminearum* produce a type A trichothecene mycotoxin. *Env.*

*Microbiol* **17**, 2588-2600.

- Walkowiak, S., Bonner, C.T., Wang, L., Blackwell, B., Rowland, O. et Subramaniam, R.** (2015) Intraspecies interaction of *Fusarium graminearum* contributes to reduced toxin production and virulence. *Mol Plant Microbe Interact* **28**, 1256-1267.
- Walter, S., Nicholson, P. et Doohan, F.M.** (2010) Action and reaction of host and pathogen during *Fusarium* head blight disease. *New Phytol.* **185**, 54-66.
- Ward, T.J., Clear, R.M., Rooney, A.P., et al.** (2008) An adaptive evolutionary shift in *Fusarium* head blight pathogen populations is driving the rapid spread of more toxigenic *Fusarium graminearum* in North America. *Fungal Genet. Biol.* **45**, 473-484.
- Wegulo, S.N.** (2012) Factors influencing deoxynivalenol accumulation in small grain cereals. *Toxins (Basel)*. **4**, 1157-80.
- Wegulo, S.N., Baenziger, P.S., Hernandez Nopsa, J., Bockus, W.W. et Hallen-Adams, H.** (2015) Management of *Fusarium* head blight of wheat and barley. *Crop Prot.* **73**, 100-107.
- Wong, P., Walter, M., Lee, W., Mannhaupt, G., Münsterkötter, M., Mewes, H.-W., Adam, G. et Güldener, U.** (2011) FGDB: revisiting the genome annotation of the plant pathogen *Fusarium graminearum*. *Nucleic Acids Res.* **39**, D637-D639.
- World Health Organization** (2012) Prevention and reduction of food and feed contamination.
- Yi, N. et Shriner, D.** (2008) Advances in Bayesian multiple quantitative trait loci mapping in experimental crosses. *Heredity (Edinb)*. **100**, 240-252.
- Yli-Mattila, T., Gagkaeva, T., Ward, T.J., Aoki, T., Kistler, H.C. et O'Donnell, K.** (2009) A novel Asian clade within the *Fusarium graminearum* species complex includes a newly discovered cereal head blight pathogen from the Russian Far East. *Mycologia* **101**, 841-52.
- Yun, S.H., Arie, T., Kaneko, I., Yoder, O.C. et Turgeon, B.G.** (2000) Molecular organization of mating type loci in heterothallic, homothallic, and asexual *Gibberella/Fusarium* species. *Fungal Genet Biol* **31**, 7-20.

- Zeller, K.A., Bowden, R.L. et Leslie, J.F.** (2004) Population differentiation and recombination in wheat scab populations of *Gibberella zeae* from the United States. *Mol. Ecol.* **13**, 563-571.
- Zhang, X., Halder, J., White, R.P., Hughes, D.J., Ye, Z., Wang, C., Xu, R., Gan, B. et Fitt, B.D.L.** (2014) Climate change increases risk of *Fusarium* ear blight on wheat in central China. *Ann. Appl. Biol.* **164**, 384-395.
- Zhang, Y., Lubberstedt, T. et Xu, M.** (2013) The genetic and molecular basis of plant resistance to pathogens. *J. Genet. Genomics* **40**, 23-35.

## **Annexes**





# I. Rapport de stage de Fatima Zahra SHAILIA



Master 2 Professionnel

Bio protection et Microbiologie de L'Environnement.

Université de Pau et des Pays de l'Adour

Année universitaire 2011/2012



Rapport de stage:

Etude de la diversité génétique et phénotypique de  
*Fusarium graminearum*, producteur de toxines dans les  
céréales.

Fatima Zahra SHAILIA

Maitres de stage : Marie Foulongne-Oriol et Christian Barreau.

Unité MycSA, INRA Centre Bordeaux Aquitaine



## Résumé

*Fusarium graminearum* est l'agent responsable de la fusariose de l'épi le plus représenté en Europe et est susceptible de produire différentes familles de mycotoxines. Les Trichothécènes de type B (TCTB) et la Zéaralénone (ZEA) sont les deux mycotoxines majoritairement produites par *F.graminearum*. Il n'existe aucune stratégie de maîtrise efficace limitant la présence de mycotoxines dans les récoltes et donc il faut limiter sa biosynthèse au champ. Une approche de génétique quantitative pour étudier le déterminisme génétique du niveau de biosynthèse des mycotoxines a été initiée au laboratoire MycSA de l'INRA de Bordeaux. La connaissance de la diversité génétique et phénotypique est un prérequis indispensable au développement de ce type d'approche. Une mycothèque regroupant des souches de diverses espèces de *Fusarium* dont *F. graminearum* est disponible au laboratoire.

La présente étude a consisté à évaluer la diversité génétique au sein de 90 souches de *F. graminearum* à l'aide des marqueurs microsatellites(SSR) disponibles dans la littérature. Une étape de mise au point de 20 marqueurs SSR avec optimisation pour une révélation sur séquenceur automatique a été préalablement nécessaire. En parallèle, afin de compléter les données déjà disponibles au laboratoire sur une collection de *F. graminearum*, un phénotypage de 23 souches du niveau de production de TCTB et de ZEA a été réalisé par dosage par HPCL/DAD et/ou HPLC/MS/MS.

Une grande diversité génétique a été démontrée au sein de la collection disponible au laboratoire. A partir des 90 souches, 60 génotypes différents ont été identifiés. Ceci nous permet d'estimer la redondance de la collection à environ 33 %.

L'analyse du niveau de production de toxine parmi 23 souches de *F. graminearum* montre également une grande diversité phénotypique. Les résultats ont montré que 9% des souches produisent uniquement du déoxynivalénol (DON), et/ou sa forme acétylée (15 ADON), 26% produisent de la ZEA, 52% produisent les deux toxines, et 13% ne produisent aucune de ces toxines. Six souches ont pu être identifiées comme étant fortement productrice, 9 souches sont non productrices ou faiblement productrices, et 8 souches sont classées intermédiaires.

## SOMMAIRE

<b>Remerciements</b>	
<b>Résumé</b>	
<b>Présentation de lieu de stage</b>	1
<b>Introduction</b>	3
<b>I Contexte bibliographique</b>	5
I.1. Fusarium : le champignon et le pathogène	5
I.2. <i>Fusarium graminearum</i> producteur de mycotoxines	6
I.2.1. Les trichothécènes de type B (TCTB)	6
I.2.2. La Zéaralénone	7
I.2.3. Toxicologie et Réglementation	7
I. 2.4. Biosynthèse de trichothécènes	7
I.3. Diversité génétique de <i>F. graminearum</i>	8
<b>II Matériels et méthodes</b>	10
II.1. Matériel biologique	10
II.2. Caractérisation de la diversité génétique au sein des souches de <i>F. graminearum</i> disponibles	10
II.2.1. Extraction d'ADN	10
II.2.2 Mise au point des marqueurs microsatellites (SSR)	10
II.2.3. Génotypage des souches de <i>F. graminearum</i> .	13
II.2.4. Analyse de la diversité génétique	13
II.3. Phénotypage complémentaire : dosage de toxines	14
II.3.1. Cultures sur grains	14
II.3.2. Culture sur milieu liquide : Milieu synthétique (MS)	15
II.3.4. Extraction des mycotoxines (TCTB et ZEA)	15
3.4.1. Extraction à partir des grains	15
3.4.2. Extraction à partir du milieu MS	15
II.3.5. Analyse par Chromatographie liquide Haute Pression HPLC/DAD	16
II.3.6. Analyse par Chromatographie Liquide Haute Pression couplée à la spectrométrie de masse (HPLC/MS/MS)	16
<b>III Résultat et discussion</b>	17
III.1. Diversité génétique	17
III.1.1. Mise au point et sélection des marqueurs SSR les plus polymorphes.	17
III.1.2. Paramètres de la diversité génétique	19
III.2 Evaluation du niveau de production de mycotoxines	23
III.2.1. Souches de la Mycothèque de l'Unité MycSa, INRA.	23
III.2.2 Souches de Fungal Genetics Stock Center (FGSC), USA.	24
<b>IV Conclusion générale et perspectives</b>	27
<b>V Références bibliographiques</b>	28
<b>Annexes</b>	31

# SOMMAIRE

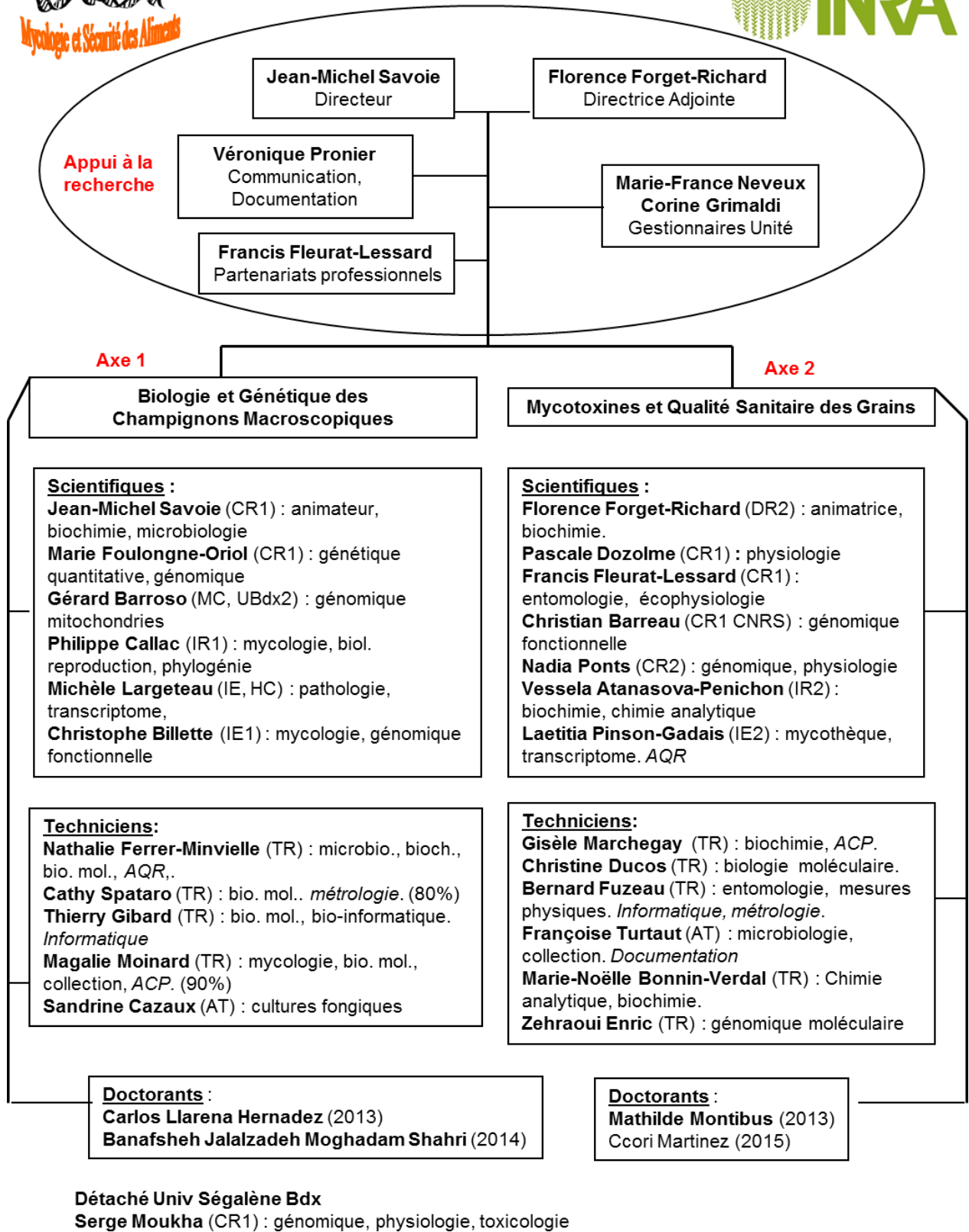


Figure 1 : Organigramme de MycSA

## Présentation de lieu de stage

L'*Institut National de la Recherche Agronomique (INRA)*, est un *Etablissement Public Français* à caractère *Scientifique et Technologique (EPST)* créé en 1946 et réparti en **19 Centres Régionaux** (<http://www.inra.fr/>). Il est placé sous la double tutelle du *Ministère de la Recherche et du Ministère de l'Agriculture et de la Pêche*, qui orientent sa stratégie, lui allouent si nécessaire des moyens, contrôlent sa gestion et évaluent ses résultats. Il dispose de l'autonomie financière, bénéficie de subventions mais aussi de ressources propres et recrute son personnel.

L'*INRA* a plusieurs fonctions. Tout d'abord, il maintient l'équilibre entre les exigences de la *recherche* et les demandes de la *société* afin de prendre en compte l'intérêt public. Ensuite, en agissant principalement dans les domaines de l'agriculture, de l'alimentation et de l'environnement, l'*INRA* produit et diffuse des connaissances scientifiques et des innovations. Par ailleurs, il propose quelques autres services, par exemple, l'expertise, la formation, la promotion de la culture scientifique et technique, les débats *science/société*.

La mission principale de l'*INRA* Bordeaux Aquitaine est la recherche ([www.bordeaux-aquitaine.inra.fr](http://www.bordeaux-aquitaine.inra.fr)). Il se compose de *15 implantations* sur **5 départements**. Autour de la recherche, se détachent également d'autres missions y compris la formation, l'Enseignement ainsi que les relations internationales et industrielles. Il collabore avec les Universités, l'Enseignement Supérieur et les autres organismes de recherche. En région Aquitaine, cette collaboration est reflétée par la création d'*UMR (Unité Mixte de la Recherche)* avec les *Universités de Bordeaux I et II, l'Université de Pau et des Pays de l'Adour, l'ENITA (Ecole Nationale d'Ingénieurs des Travaux Agricoles)*, le *CNRS (Centre National de Recherche Scientifique)*, l'*INSERM (l'Institut National de la Santé et de la Recherche médicale)* et l'*IFREMER (l'Institut Français de Recherche pour l'Exploitation de la Mer)*.

Parmi toutes les *Unités de Recherche en Aquitaine* concernant les végétaux, leurs écosystèmes et leurs environnements, nous nous intéressons plus particulièrement à l'*Unité de Recherche 1264 « Mycologie et Sécurité des Aliments » (MycSA)*. Celle-ci est la composante du *Pôle de Recherche QUALIS*, soutenu par la *Région Aquitaine* en partenariat avec des organismes interprofessionnels et les entreprises du secteur agroalimentaire concerné par la production des céréales, tel qu'Arvalis, Syngenta, Monsanto, BASF, Bayer, et Biogemma.

Les missions de l'Unité **MycSA** sont l'acquisition et l'intégration de connaissances sur les champignons comestibles d'une part et sur les agents d'altération (fongiques incluant les agents mycotoxinogènes et les insectes) des produits végétaux, principalement grains, fruits, légumes secs d'autre part. Elle est constituée de **deux équipes** (voir Figure 1): *l'équipe Mycotoxine et Qualité Sanitaire des Grains* et *l'équipe Biologie et Génétique des Champignons Macroscopiques* (<http://www.bordeaux-aquitaine.inra.fr/mycsa>). J'ai effectué mon stage au sein de l'équipe *MQSG*, sous la co-tutelle de Marie Foulongne-Oriol et Christian Barreau.

# **INTRODUCTION**

Les céréales (blé, maïs ...) sont l'un des aliments de base de la population mondiale. En France, les céréales et en particulier le blé tendre représentent plus de 9 millions d'hectares de cultures, soit le tiers de la surface cultivée totale. Avec une production moyenne de 68 Mt par an, la France est le premier producteur et exportateur européen de céréales, le deuxième exportateur mondial pour le blé, et le premier exportateur mondial pour le malt. (<http://www.arvalis-infos.fr>; <http://www.franceexportcereales.org>). Mais depuis quelques années les différentes cultures céréalières sont confrontées à une pression parasitaire croissante, exercée par de nombreux agents pathogènes. Différentes maladies sont recensées, les plus dommageables sont celle d'origine fongique.

Parmi les maladies fongiques, la « fusariose de l'épi » est l'une des plus préjudiciables aux récoltes de céréales. Cette maladie induit des pertes considérables de rendement et est fréquemment associé à une contamination des grains par des métabolites fongiques toxiques pour l'homme et les animaux, des mycotoxines. En effet ces mycotoxines bloquent la synthèse protéique chez les eucaryotes et causent des graves mycotoxicoses. (Cumagun *et al*, 2004). De ce fait, la réglementation européenne s'est précisée en 2005 en matière de fusariotoxines (mycotoxines de *Fusarium*) sur les céréales et leurs produits dérivés (règlement n° 856/2005 du juin 2005, modification votée par le CPCASA le 20/07/07, en vigueur le 01/10/07) et des teneurs maximales autorisées ont été fixées.

La « fusariose de l'épi » peut être causée par différentes espèces du genre *Fusarium* dont la plus représentée en Europe est *Fusarium graminearum* (anamorphe de *Gibberella zeae*). Au cours du processus infectieux, *F.graminearum* est susceptible de produire différentes familles de mycotoxines dont essentiellement, les trichothécènes de type B « TCTB » (DON, déoxynivalénol ; ADON acétyldeoxynivalenol ; NIV, Nivalénol ; FX, fusarénone X et la Zéaralénone (ZEA). Les TCTB se trouvent être les mycotoxines les plus fréquemment rencontrées dans les grains fusariés, dont majoritairement la DON et ses formes acétylées (3- ADON et 15-ADON) (Windels, 2000). Afin de maîtriser les problèmes économiques et sanitaires évoqués ci-dessus, une solution consisterait à limiter la biosynthèse de ces mycotoxines au champ, cela nécessite une connaissance approfondie du pathosystème afin d'ajuster au mieux les méthodes de lutte.

La voie de biosynthèse des TCTB est bien connue chez *F.graminearum* à travers l'étude des gènes *tri*. (Kimura *et al*. 2007). Si le chémotype (DON/ADON versus NIV/FX) d'une souche est déterminé de façon qualitative, peu de choses sont actuellement connues sur

les éléments génétiques qui permettent, parmi les souches d'un même chémotype, d'expliquer la très grande variabilité observée pour le niveau de production de toxine, en particulier pour les TCTB.

Afin de répondre à cette problématique, un projet de recherche basé sur une approche de génétique quantitative pour étudier le déterminisme génétique du niveau de biosynthèse des mycotoxines est initié à l'unité MycSa. Dans un premier temps, l'utilisation de deux descendances existantes (Gale *et al.* 2005 ; Jurgenson *et al.* 2002) est privilégiée. Ces descendances sont disponibles au Fungal Genetics Stock Center (FGSC), USA (<http://www.fgsc.net/>). Il s'agira de caractériser phénotypiquement un sous ensemble de chacune de ces deux descendances pour divers caractères liés à la production de toxines (niveau de production de toxine, niveau d'expression de certains gènes impliqué dans la biosynthèse). Cela permettra d'évaluer la faisabilité de l'outil QTL (Quantitative Trait Locus) appliqué à *F.graminearum*. En parallèle, et dans le but d'identifier des parents potentiels pour créer une descendance au laboratoire, les données phénotypiques disponibles sur les souches de *F.graminearum* en collection de la mycothèque de l'unité MycSa seront complétées. La diversité génétique existante au sein de la collection sera également évaluée à l'aide de marqueurs moléculaires.

Le travail réalisé au cours de mon stage s'intègre dans ce projet. Il se décline en deux objectifs :

- 1) Caractérisation moléculaire de 90 souches de *F. graminearum* de la collection à l'aide de marqueurs microsatellites afin d'évaluer la diversité génétique existante. Les souches parentales des deux descendances existantes ont été incluses afin de les positionner génétiquement par rapport à notre collection. Nous avons au préalable mis au point le génotypage de *F. graminearum* à l'aide de marqueurs microsatellites disponibles dans la littérature.
- 2) Caractérisation phénotypique basée sur le niveau de production de mycotoxines (dosage de TCTB et ZEA) sur 23 souches de la collection afin de compléter les données déjà disponibles au laboratoire. Le dosage de toxines sur les souches parentales des deux descendances existantes ainsi qu'un sous ensemble d'une de ces deux descendances a également été réalisé.



# I-CONTEXTE BIBLIOGRAPHIQUE

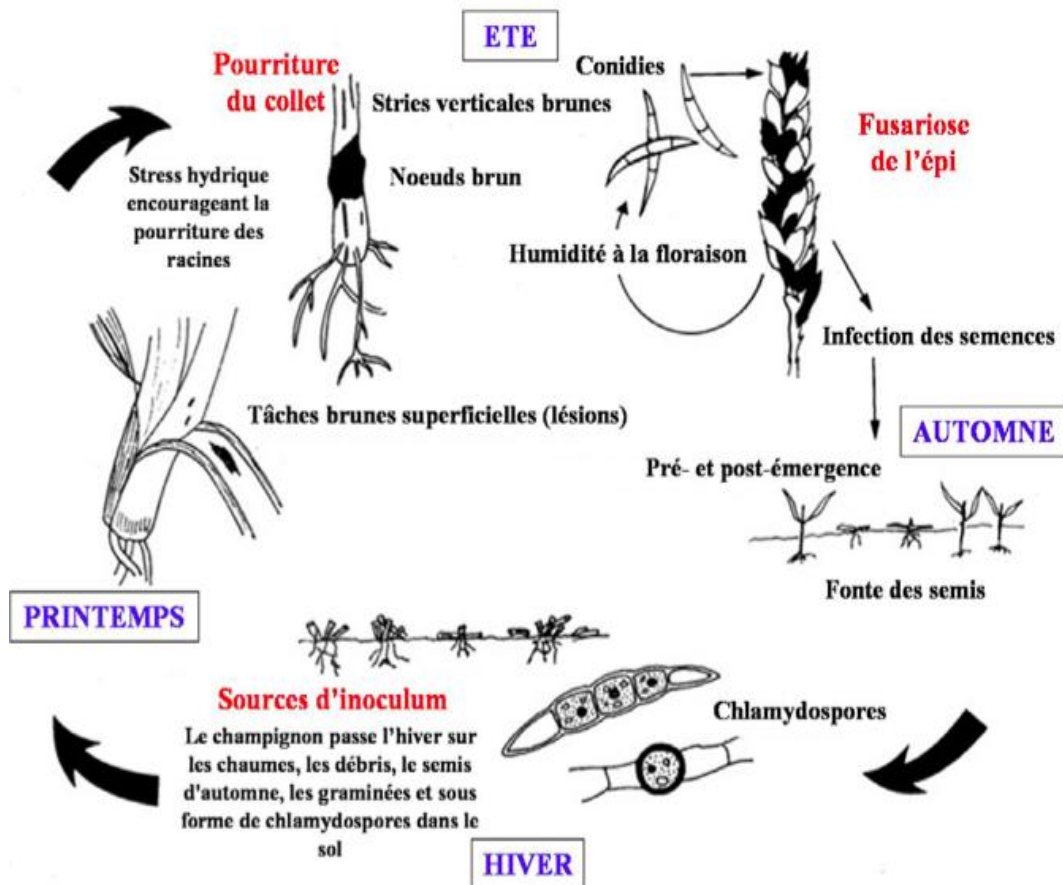
## **I.1. Fusarium : le champignon et le pathogène**

Les champignons du genre *Fusarium* sont des ascomycètes, qui appartiennent à la classe des *Sordariomycetes*, l'ordre des *Hypocreales* et la famille des *Nectriaceae*. Beaucoup sont des champignons phytopathogènes qui peuvent infecter les céréales et provoquer une maladie que l'on appelle fusariose de l'épi ou "Fusarium Head Blight" (FHB) (Figure 2). La plupart des espèces responsables de la fusariose produisent des métabolites secondaires toxiques ou mycotoxines que l'on appelle des "fusariotoxines" (Goswami and Kistler, 2004).



**Figure 2:** Exemple d'un épi de blé (à gauche) et de maïs (à droite) atteint de fusariose

En Europe, le blé est principalement infecté par l'espèce *Fusarium graminearum* (anamorphe de *Gibberella zeae*). Dans des conditions défavorables, *F. graminearum* se comporte comme un saprophyte et ses spores restent dans le sol. Ces dernières sont capables de survivre dans le sol et forment la source majeure d'inoculum. Les résidus des années précédentes servent aussi de réservoir et le champignon est capable sur ces résidus d'effectuer un cycle de reproduction sexuée, les ascospores produites au printemps servant souvent d'inoculum primaire (Figure 3). Chez le blé, au moment de l'épiaison pendant la croissance végétative et lorsque les conditions deviennent favorables, les spores parviennent jusqu'aux épis et causent l'infection. La période critique pour l'infection des épis débute à l'épiaison et s'étend sur les quelques jours suivants. Pendant cette période, les facteurs climatiques comme la pluie ou l'humidité, associés à la chaleur auront l'impact le plus important sur le niveau d'infection et sur l'accumulation de toxines. L'infection chez le blé a lieu principalement pendant une très courte période, au moment de la sortie des étamines. A ce stade du développement, la fleur du blé est largement ouverte et sujette à l'invasion par le champignon. L'infection à ce stade a le plus d'impact sur le rendement en grains (Merhej, 2010).



**Figure 3:** Cycle biologique de *Fusarium graminearum* sur céréales (d'après Parry, 1995)

## I.2. *Fusarium graminearum* producteur de mycotoxines

*F. graminearum* est le principal producteur de trichothécènes plus particulièrement de type B (TCTB) (Champeil *et al*, 2004) et de Zéaralénone (Bennett and Klich, 2003) qui peuvent s'accumuler au niveau du grain et ne peuvent pas être éliminés des grains récoltés.

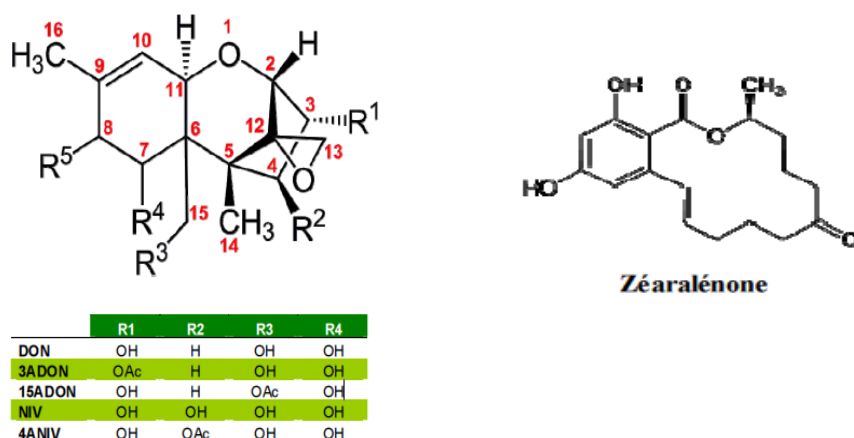
### I.2.1. Les trichothécènes de type B (TCTB)

Les TCTB font partie de la famille des sesquiterpénoïdes cycliques, et contiennent tous un squelette commun (12,13 époxytrichodiène). Le caractère amphiphile des TCT ainsi que la présence du groupement époxy sont à l'origine de leur toxicité. Les TCTB sont caractérisés par la présence d'une fonction cétone en C8. Les chaînes latérales sont variables, ce qui permet de les diviser en quatre groupes: le déoxynivalénol (DON), le 3- et 15-déoxynivalénol (3-ADON et 15-ADON), le nivalénol (NIV) et la fusarénone X (FX) (Bennett and Klich, 2003) (Figure 4).

Une souche de *Fusarium* est capable de produire un seul type de TCTB avec sa (ses) forme (s) acétylé(s). Cela détermine le chémotype de la souche en question. En effet, certaines souches sont à chémotype DON/ADON et d'autres souches à chémotype NIV/FX.

### I.2.2. La Zéaralénone :

La zéaralénone est une lactone macrocyclique dérivée de l'acide résorcyclique (Resorcyclic Acid Lacton RAL), de formule brute C<sub>18</sub>H<sub>22</sub>O<sub>5</sub> (Gaumy *et al.* 2001) (Figure 4). Elle est également classée comme mycoestrogène. En effet, elle présente suffisamment de ressemblance avec une hormone ovarienne humaine pour se fixer sur le récepteur d'œstrogènes d'où sa toxicité (Elisavet *et al.* 2010). Elle se trouve majoritairement sur les céréales comme le blé et le maïs. Les gènes impliqués dans sa voie de biosynthèse sont décrits récemment (Kim *et al.* 2005 ; Lysoe *et al.* 2006).



**Figure 4 :** Structure chimique des trichothécènes de type B (à gauche) et de Zéaralénone (ZEA) (d'après Merhej, 2010, Bennett and Klich, 2003).

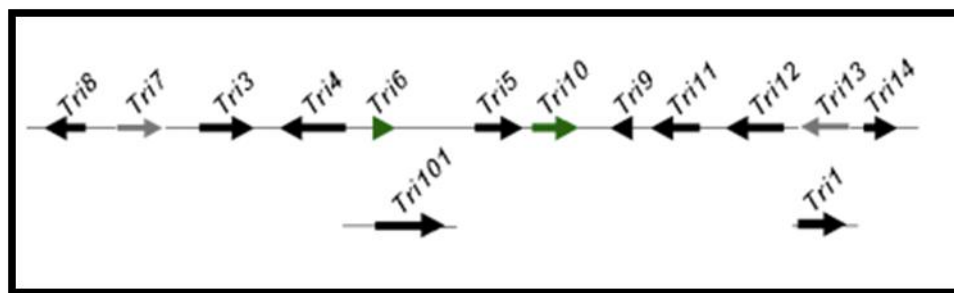
### I.2.3. Toxicologie et Réglementation

Les contaminations aux TCT et plus particulièrement au DON seraient responsables à dose élevée de toxicité aigüe chez l'homme. Nommées Aleucies Toxiques Alimentaire (ATA), elles se traduisent par différents symptômes tels de fortes diarrhées, des vomissements et des modifications hématologiques (Grevet, 2004). A plus faible dose, les mycotoxines seraient responsables d'une toxicité chronique. En effet, leur consommation régulière pourrait s'avérer délétère pour la santé même si actuellement il est difficile d'établir quels sont les risques réels. Cependant par principe de précaution, la réglementation européenne a imposé des teneurs maximales pour le DON et de la Zéaralénone dans les céréales et produits dérivés destinés à l'alimentation humaines (Annexe I).

### I. 2.4. Biosynthèse de trichothécènes

Les TCT dérivent du farnesyl pyrophosphate (FPP) (Kimura *et al.* 2007). Le FPP est un intermédiaire commun à l'isoprénylation des protéines et à la biosynthèse des stérols. L'ensemble des gènes impliqués dans la voie de biosynthèse des trichothécènes et de sa régulation spécifique se nomme gène *tri*. Ces gènes sont au nombre de 15 chez *F.*

*graminearum* dont 12 sont regroupés en un cluster (cluster *tri*) et 3 sont présents sur des locis particuliers. (Kimura *et al.* 2007) (Figure 5).



**Figure 5 :** Représentation schématique des gènes *Tri* de *F. graminearum*. La ligne supérieure représente les gènes regroupés dans le « Cluster Tri » sur le chromosome 2 et la ligne inférieure les autres gènes de la voie de biosynthèse positionnée ailleurs dans le génome

Si le chémotype d'une souche est déterminé de façon qualitative, par la présence à l'état fonctionnel ou non du gène *Tri13* (Kim *et al.* 2003), peu de données sont actuellement disponibles sur les éléments génétiques qui permettent, parmi des souches d'un même chémotype, d'expliquer la très grande variabilité observée pour le niveau de production de toxine. L'hérédité de ce potentiel toxigène est mal connue. Pour répondre à cette question, la génétique quantitative semble prometteuse. Il s'agit de déterminer le nombre de loci impliqués dans la régulation de la toxinogénèse mais aussi leurs effets, leur position sur le génome et leurs éventuelles interactions. Une première étude utilisant basée sur l'étude d'une descendance a pu mettre en évidence un locus impliqué dans le niveau de production de toxine (*Tox1*) localisé sur le génome en dehors du cluster *Tri* (Jurgenson *et al.* 2002). Pour développer ce type d'approche, le choix de la descendance est primordiale (Cumagan *et al.* 2007). Dans cette optique, disposer de l'information sur la diversité génétique existante entre les souches parentales est un prérequis indispensable (Foulongne-Oriol, 2012).

### **I.3. Diversité génétique de *F. graminearum***

Plusieurs études ont montrés la présence d'une grande diversité génétique au sein de l'espèce de *F. graminearum* (Brygoo et Gautier, 2007 ; Toth *et al.* 2005 ; Alvarez *et al.* 2011). Cette diversité génétique est le résultat de plusieurs phénomènes dont la sélection, les mutations, la migration et la dérive génétique. Ces forces évolutives ont toutes des effets sur l'importance et l'organisation de la variabilité génétique. L'étude de Bourdages *et al.* (2005) sur la caractérisation de la diversité génétique chez plusieurs populations de *Fusarium* de Québec a montré que la diversité intra-spécifique qui règne au sein du complexe *F. graminearum* est plus importante que la diversité génétique entre les populations testées. Les souches appartenant à l'espèce *F.graminearum* sont défini comme des *F. graminearum* au

sens strict ou « *sensu stricto* » (O'Donnell *et al.*, 2000). Certaines souches peuvent être incluses dans un complexe avec *F. graminearum* (Yli-Mattila *et al.* 2009, Wang *et al.* 2011). Ces souches sont génétiquement très proches. Dans l'étude menée par Brygoo et Gautier (2007) sur des souches de *Fusarium* isolées sur des épis de blé et de maïs en France a montré que les souches *F. mesoamericanum*, *F. asiaticum*, *F. boothii* et *F. austroamericanum* sont des souches très proches des *F. graminearum* sensu stricto.

L'estimation de cette diversité est basée sur le choix du type de marqueur susceptible de la traduire le plus fidèlement possible. Chez *F. graminearum*, différents marqueurs moléculaires sont utilisés comme RFLP (Restriction Fragment Length Polymorphism), SSR (Simple Sequence Repeats), RAPD (Random Amplified Polymorphic DNA) et AFLP (Amplified Fragment Length Polymorphism) (Wang *et al.* 2011)

Plusieurs études ont montrés que les microsatellites ou SSR pour Single Sequence Repeat sont les plus informatifs et les adaptés pour l'étude de la diversité génétique de *Fusarium*. (Vogelgsang *et al.*, 2009).

Les microsatellites sont constitués de répétitions en tandem de motifs mono-, di-, tri-, ou tétra-nucléotidiques répétés en plusieurs exemplaires sur l'ensemble du génome d'une espèce et présentent un taux de polymorphisme élevé (Morgante and Olivieri, 1993). De tels motifs sont très nombreux dans le génome des organismes eucaryotes. Les SSR présentent un taux de polymorphisme extrêmement élevé qui repose sur la variation du nombre d'unités de répétition (n) constituant le microsatellite. Chaque locus microsatellite est encadré par de séquences qui lui sont spécifiques et qui permettent d'y définir des amorces pour en faire l'amplification par PCR («Polymerase Chain Reaction»). Les microsatellites sont des marqueurs codominants, extrêmement polymorphes, faciles à manipuler et permettant le multiplexage. Ces marqueurs sont très reproductibles et moins lourds d'utilisation que les RFLP ou AFLP avec une bonne résolution analytique. Ces marqueurs ont été très utiles pour des études de diversité génétique, phylogénétiques et la construction de cartes génétiques.

Chez *F. graminearum*, plusieurs marqueurs microsatellites ont été développés (Vogelgsang *et al.*, 2009 ; Naef and Défago, 2006 ; Giraud *et al.*, 2002 ; Brygoo et Gautier 2007). Il a été démontré qu'ils étaient aléatoirement répartis sur l'ensemble du génome de *F. graminearum* (Karaoglu *et al.* 2005).

L'objectif de mon travail était d'évaluer la diversité génétique au sein de *F. graminearum* de la mycothèque et de compléter la caractérisation phénotypique en dosant le niveau de production des TCTB et de la ZEA de certaines souches. Les méthodes utilisées et les résultats obtenus seront illustrés par la suite.

## **II- MATERIELS ET METHODES**

### **II.1. Matériel biologique**

Nos travaux sont réalisés avec des souches de *Fusarium graminearum* provenant de la mycothèque de l'unité MycSa et des souches provenant de Fungal Genetics Stock Center (FGSC) <http://www.fgsc.net/>, USA. Ces souches sont maintenues à 4°C sur un milieu Potato Dextrose Agar (PDA) incliné (infusât de pomme de terre 4% W/V, Agar Agar 15% W/V, glucose 20% W/V, autoclavage 15minutes à 120 °C). L'Annexe II regroupe toutes les souches utilisées pour les deux parties de cette étude.

Une première culture (repiquage) est réalisée classiquement sur gélose inclinée PDA et sur des petites boîtes. L'incubation a lieu à l'étuve à 25 °C pendant 7 jours. L'inoculum est obtenu en ajoutant 6 ml d'eau MilliQ® stérile aux cultures sur PDA. Après homogénéisation, la suspension de spores prélevée constitue l'inoculum de toutes les cultures utilisées.

### **II.2. Caractérisation de la diversité génétique au sein des souches de *F. graminearum* disponibles**

La première partie de cette étude était de caractériser le polymorphisme génétique existant au sein de génotypes des souches de *F. graminearum* de la mycothèque, unité MycSa à l'aide de marqueurs de types microsatellites (SSR) décrits dans la littérature (Vogelsgang *et al*, 2009 ; Naef and Défago, 2006 ; Giraud *et al*, 2002).

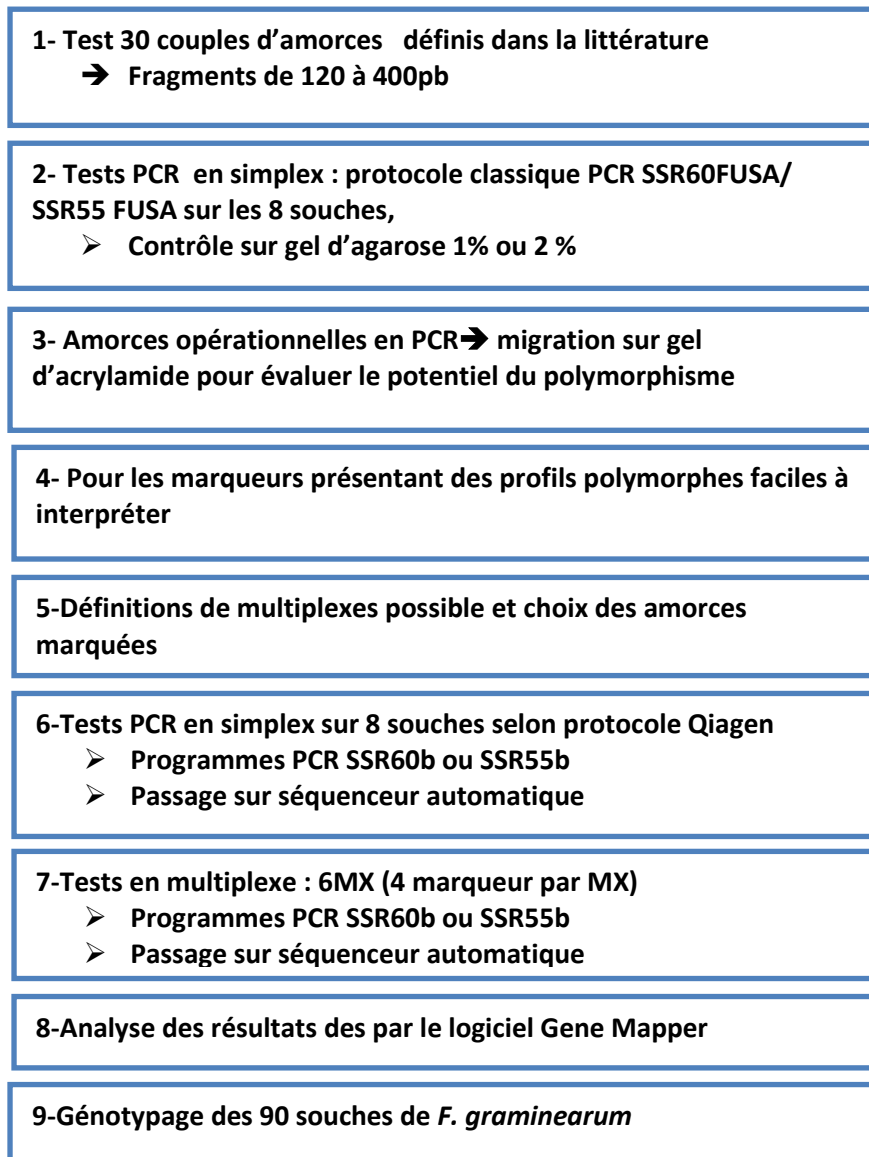
#### **II.2.1. Extraction d'ADN**

A partir de mycélium âgé de 7 jours, préalablement lyophilisé (Lyophilisateur Flexi-dryTMMP) et broyé à l'aide d'un TissueLyser System® (Qiagen, France), l'extraction d'ADN est réalisée à l'aide du kit Dneasy®Plant Mini Kit (Qiagen, France) selon les instructions du fournisseur. La concentration d'ADN a été estimée par une mesure d'absorbance à 260 nm à l'aide d'un spectrophotomètre Nanodrop®ND1000 (Labtech, France). Les ADN ont été dilués à une concentration de 25ng/μL. Les ADN de l'ensemble des souches génotypés dans le cadre de cette étude étaient disponibles avant mon arrivée. J'ai réalisé l'extraction d'ADN de 5 des souches provenant de FGSC.

#### **II.2.2 Mise au point des marqueurs microsatellites (SSR)**

Trente marqueurs ont été choisis pour leur niveau de polymorphisme décrits et pour la taille des fragments amplifiés (entre 120 et 400 pb) (Vogelsgang *et al*, 2009 ; Naef et Défago, 2006 ; Giraud *et al*, 2002 ; Brygoo et Gautier 2007). Mon travail a consisté dans un premier temps à mettre au point l'utilisation de ces marqueurs microsatellites en routine au

laboratoire. La Figure 6, présente les différentes étapes de la mise au point réalisée avec ces marqueurs. Huit souches supposées génétiquement différentes selon leur origine ou leur chémotype ont été sélectionnées pour réaliser la mise au point.



**Figure 6** : Protocole de mise au point des marqueurs SSR au laboratoire

Une amplification PCR a été faite pour tester les 30 couples d'amorces sur ces 8 souches (Etapes 1 et 2, Fig.6). La réaction d'amplification est réalisée avec 0.2 µl de la GoTaq polymérase à 5 U/µl, dans un milieu réactionnel contenant 4µl du tampon 5X ; 3.5 µl de dNTP (déoxyribonucléotides triphosphate) à 1.2 mM ; 8.3 µl d'H<sub>2</sub>O ; 1µl de chaque amorce à 10 µM et 2 µl d'ADN à 25 ng/µl. Deux programmes d'amplification ont été utilisés, SSR55FUSA et SSR60FUSA ayant une température d'hybridation de 55°C et 60°C respectivement en fonction des amorces utilisées. L'amplification PCR est réalisée dans un

thermocycler Eppendorf avec une étape de dénaturation initiale de 1min à 95°C suivi par 35 cycles de 30s à 94°C de dénaturation, 30s à 55/60°C d'hybridation selon le  $T_m^\circ$  spécifique des amorces utilisées, et 30s à 72°C d'élongation, et après une élongation finale de 10 min à 72°C. Pour vérifier l'amplification, les produits de PCR ont été séparés selon leur taille par électrophorèse sur gel d'agarose 2% dans le tampon TAE 1X (Tris- Acétate- EDTA) avec du BET (Bromure d'Ethidium) à une concentration finale de 0.5 µg/ml. Le marqueur de taille utilisé est le 1KB DNA, Promega®). La visualisation des fragments est faite sous UV avec un Imager Vilber®, après 45 minutes de migration à 90 Volts.

Pour les couples d'amorces opérationnelles en PCR (présentant une seule bande sur gel d'agarose), une migration sur un gel d'acrylamide à 6% (Etapes 3 et 4-Fig. 6) a été réalisée afin d'évaluer le potentiel de polymorphisme. La visualisation est faite par la méthode non-radioactive de coloration au nitrate d'argent. Les marqueurs SSR présentant du polymorphisme sur les 8 souches testées, ainsi qu'un profil facile à interpréter ont été choisis pour les étapes suivantes.

Il s'agit de définir des multiplexes opérationnels pour l'électrophorèse sur séquenceur automatique (3130 Genetic Analyzer ABI®, Applied Biosystems) (Etapes 5- Fig 6). Le choix de ces multiplexes est fait manuellement en fonction des tailles des fragments amplifiés, de la température d'hybridation, et des fluorochromes disponibles. Six multiplexes ont pu être définis, chacun constitué de 4 couples d'amorces. Pour chaque marqueur microsatellite, les amorces forward sont marquées à l'aide des fluorochromes PET, NED, VIC et 6-FAM. (Annexe III). Les couples d'amorces marquées ont été testées d'abord individuellement en simplexe, avant d'être testées en multiplexes (Etapes 6. Fig. 6). Le mélange réactionnel utilisé pour les amplifications en simplexe de l'ADN des 8 souches est le suivant : 8 µl de Mix Taq constitué de 5 µL de Master Mix PCR à 2X (Qiagen Multiplex Master Mix kit) (contient la Taq polymérase et les dNTPs) ; 0.15µl de chaque amorce à 10 µM, 2.7 µl d'H<sub>2</sub>O (RNase free water) et 2 µl d'ADN à 25 ng/µl. Le Master Mix PCR Qiagen a été développé spécifiquement pour permettre l'amplification de plusieurs fragments au sein d'une même PCR. L'utilisation de ce Master Mix à cette étape de PCR en simplex permet de vérifier que l'amplification indépendante est opérationnelle.

Pour l'étape de multiplexage (Etape 7. Fig.6), un mélange d'amorces (10x Mix Multiplexes) est préalablement préparé, constitué de 1.5µl de chaque amorce marqué et non marqué à 10µM de chaque multiplexe et 88µl d'H<sub>2</sub>O (RNase free water). Le mélange réactionnel utilisé pour le multiplexage est composé de 10 µl de Mix Taq contenant 5 µl de Master Mix PCR Qiagen à 2X ; 1 µl de 10x Mix multiplexes ; 4 µl d'H<sub>2</sub>O (RNase free water)



et 2 µl d'ADN à 25 ng/µl. Les programmes PCR utilisés sont identiques aux programmes précédents avec une étape de dénaturation initiale de 15 min à 95°C afin d'assurer le fonctionnement de la Taq hot start du Master mix. Le nombre de cycle a également été réduit à 28 cycles au lieu de 35 pour limiter la saturation du signal lors du passage sur séquenceur automatique.

Les produits de PCR sont préalablement contrôlés sur gel d'agarose selon la même procédure que celle décrite précédemment. Une dilution a été faite en fonction de l'intensité des bandes obtenues sur gel d'agarose, pour passer au séquenceur automatique (Etape 8. Fig6) afin d'ajuster l'intensité du signal des fluorochromes. Une dilution de 1/100000 est réalisée et un volume de 1.5 µl de chaque dilution est transférée dans une plaque 96 MicroAmp (Optical 96-Well Reaction Plate, ABI®) spécifique au séquenceur automatique. A chaque échantillon est ajouté 10µl d'un mix constitué de 10 µl d'une solution dénaturante de formamide (Hi-Di™ Formamide, Applied Biosystems) et 1.7 µl du marqueur de taille (GeneScan™-600 LIZ™ Size) marqué avec le fluorochrome LIZ™.

Les allèles amplifiés des différents marqueurs sont séparés par électrophorèse dans des capillaires contenant un polymère d'acrylamide (POP7, Applied Biosystems) à l'aide du séquenceur multi-capillaires ABIPRISM® 3130x genetic analyser (Applied Biosystems situé au sein de l'UMR santé végétal, INRA. Bordeaux). La fluorescence associée à chaque marqueur permet leur détection au moyen d'un laser et d'une caméra. Les profils d'électrophorèse sont capturés par le logiciel ABI 3130 xL Data Collection. La lecture des données issues du séquenceur (taille des différents allèles amplifiés) a été réalisée à l'aide du logiciel GeneMapper v4.0 (Applied Biosystems).

### **II.2.3. Génotypage des souches de *F. graminearum*.**

Le génotypage a été réalisé sur un ensemble des 90 souches de *F. graminearum* (85 souches de la collection + 5 souches du FGSC) en utilisant les 6 multiplexes définis. L'amplification, l'électrophorèse sur séquenceur automatique et l'analyse des profils ont été faits de la même manière que lors de la mise au point (Etape 9. Fig6).

### **II.2.4. Analyse de la diversité génétique**

Les paramètres génétiques que sont le nombre d'allèles par locus, la fréquence allélique, et la distance génétique ont été calculés à l'aide d'un logiciel d'analyse génétique Power Marker, version 3.25. La distance génétique *Shared Allele Distance (DAS)* entre les différentes souches est calculée selon la formule suivante  $DAS=1-(\text{somme du nombre d'allèle en commun})/(\text{nombre de loci analysés})$ .

A partir de la matrice de distances, un arbre de distance génétique est construit selon la méthode UPGMA (*Unweighted Pair Group Method with Arithmetic mean*) avec le logiciel Power Marker et visualisé à l'aide du logiciel Treeview version 1.6.6.

### II.3. Phénotypage complémentaire : dosage de toxines

Les souches disponibles dans la mycothèque ont été partiellement caractérisées pour leur niveau de production de toxine. Mon travail a consisté à compléter ces données. Ainsi, 23 souches ont été caractérisées dans le cadre du stage.

#### II.3.1. Cultures sur grains

Les grains secs de blé (Grand moulin- Paris) sont humidifiés par de l'eau MilliQ® et placés à 4 °C pendant 3 jours (pour se situer dans des conditions d'Activité de l'eau (AW) optimales à *Fusarium*, c'est-à-dire proche de 1). Cent g de grains sont placés dans un erlenmeyer de 500 ml et stérilisés 25 minutes à 110 °C à deux reprises. Le contrôle de l'activité de l'eau «AW» ou l'humidité est effectué à l'aide d'un Awmètre (Jules Richard instruments). Les grains sont inoculés avec 0.5 ml de la suspension de spores, et sont incubés 21 jours à 25 °C et à l'obscurité. Deux répétitions biologiques ont été faite pour chaque souche. Les cultures sont stoppées à 21 jours et ont été mises à sécher pendant 4 jours dans une étuve à 50 °C. Une étape de broyage a été effectuée par la suite à l'aide d'un broyeur centrifuge Retsch® (grille 0.5mm). L'humidité des broyats (farines) obtenues a été mesurée, en mettant 5 g de farine dans une étuve Chopin pendant 2 h à 130 °C et les pesés par la suite la Figure 7 présente une culture de *F.graminearum* sur grains de blé.



**Figure 7** : Culture de *F.graminearum* de 21 jours sur grains de blé incubée à 25°C. Un témoin négatif est présenté sur la droite de la photo.

### **II.3.2. Culture sur milieu liquide : Milieu synthétique (MS)**

Les cultures des souches provenant de FGSC (USA) ont été effectuées sur un milieu synthétique (MS), qui a été mis au point au laboratoire (Boutigny, 2007). Ce milieu est constitué de  $\text{KH}_2\text{PO}_4$  à 50 g/L,  $\text{K}_2\text{HPO}_4$  à 60 g/L,  $\text{MgSO}_4$  à 1,7 g/L,  $(\text{NH}_4)_2\text{SO}_4$  à 100 g/L, biotine, à 100 mg/L, 20g/L de glucose, et une solution des oligos éléments à 100 $\mu\text{g/L}$  (Vogel, 1956). (la Composition de la solution des oligos présentée en annexe).

Des petites boîtes de Pétri stériles (55 mm de diamètre) contenant 6 mL de milieu MS ont été inoculées avec des suspensions de spores pour atteindre une concentration finale de  $4 \times 10^6$  spores/mL. L'ajustement de concentration finale a été fait par numération sur cellule malesses. Les cultures ont été faites en triplicatas et sont incubées pendant 12 à 16 jours dans une étuve à 25°C et à l'obscurité. Deux arrêts de cultures ont été faites à 12 jours et 16 jours. Le mycélium a été récupéré par centrifugation et lyophilisé (lyophilisateur Cryotec®) pendant 48h, pour déterminer la biomasse sèche et le surnageant utilisé pour le dosage de toxine.

### **II.3.4. Extraction des mycotoxines (TCTB et ZEA)**

Plusieurs études antérieures ont montré que *Fusarium graminearum* produit essentiellement, et le plus souvent les trichothécènes de type B (TCT B) et la Zéaralénone (ZEA) (Bakan et al. 2002; Bottalico et Perrone. 2002)). Notre analyse a été faite uniquement sur le dosage de ces deux mycotoxines.

#### **3.4.1. Extraction à partir des grains :**

Les TCTB et la Zéaralénone sont extraits par 25 mL d'un mélange acétonitrile/eau à partir de 5 g de grains broyés. Une agitation pendant 1h est faite puis une centrifugation 5 min à 5000 rpm. Cinq mL de surnageant sont prélevés et filtrés sur une colonne clean up (R-Biopharm Rhône. LTD). Par la suite 3 ml sont récupérés et évaporés à sec sur bloc chauffant sous flux d'azote à 50 °C. Les extraits secs sont repris avec un volume de 400  $\mu\text{l}$  d'un mélange méthanol/eau (1/1 v/v) et filtrés sur un filtre de 0.45  $\mu\text{m}$ , avant dosage par HPLC/DAD ou spectrométrie de masse.

#### **3.4.2. Extraction à partir du milieu MS**

Quatre mL de milieu de culture ont été extraits avec 8 mL d'acétate d'éthyle puis sont vortexer pendant 1min et décantés. Après décantation 4 mL de la phase organique ont été récupérés et évaporés à sec à 50 °C sous flux d'azote. L'extrait sec a été repris avec un volume de 500  $\mu\text{l}$  de méthanol/eau (1 :1, v/v) puis filtré à l'aide d'un filtre de 0.45 $\mu\text{m}$  avant analyse par HPLC/DAD et spectrométrie de masse (MS/MS).

### II.3.5. Analyse par Chromatographie liquide Haute Pression HPLC/DAD

Toutes les analyses ont été réalisées sur une chaîne HPLC Hewlett-Packard ChemStation Series 1100 (Agilent Technologie, Palo Alto, USA) couplée à un détecteur à barrette de diodes DAD (Diode Array Detector). L'identification et la quantification des composés ont été réalisées par une méthode d'étalonnage externe. La calibration externe est réalisée grâce à une gamme étalon pour chacun des composés analysés. La séparation des TCTB et Zéaralénone est réalisée sur une colonne Kinetex (C18, 150x 4.6mm) maintenue à 45 °C. La phase mobile est composée d'eau acidifiée par l'acide ortho phosphorique à pH 2.6 (Solvant A) et d'acétonitrile 100% (Solvant B). Le débit est fixé à 1 ml/min. Les conditions du gradient d'élution sont décrites dans le tableau 1. Le volume d'injection est de 5 µL. La détection a été faite à 230 nm et la quantification est réalisée par étalonnage externe en utilisant des composés de référence (Sigma-Aldrich Co, St Louis, USA).

### II.3.6. Analyse par Chromatographie Liquide Haute Pression couplé à la spectrométrie de masse (HPLC/MS/MS)

Les analyses ont été réalisées sur une chaîne HPLC Hewlett-Packard ChemStation Series 1100 (Agilent) Applied Biosystems couplé à un system Q-TRAP LC/MS/MS (Applied Biosystems MDS Sciex) équipé d'une source d'ionisation par électro nébulisation (ESI) et d'un analyseur de triple quadripôle. L'ensemble est contrôlé par un logiciel Analyst 1.4.2. La séparation est faite sur une colonne Zorbax XDB C18 (2.1 x 150 mm, 5 µm, Agilent). La phase mobile est composée du solvant A Méthanol/Eau (90:10, v/v) et solvant B Méthanol/Eau (10:90, v/v). Le débit est de 300 µL.min<sup>-1</sup>. Le gradient d'élution est réalisé sous les conditions présentées dans le tableau 5. L'analyse par spectrométrie de masse a été réalisée en mode négatif de type MRM (Multiple Reaction Monitoring). La quantification a été effectuée par étalonnage externe avec des standards des TCTB (DON ADON NIV ET FX) et de la ZEA. (Sigma-Aldrich Co, St Louis, USA). Le gradient d'élution est décrit dans le tableau 2

**Tableau 1** : Gradient d'élution utilisé pour l'analyse des TCTB et ZEA par HPLC/DAD

Temps en min	% Solvant A	% Solvant B
0	93	7
10	70	30
15	10	90
20	10	90
22	93	7
27	93	7

**Tableau 2** : Gradient d'élution utilisé pour l'analyse des TCTB et ZEA par HPLC/MS/MS.

Temps en min	% Solvant A	% Solvant B
8	10	90
10	70	30
12	100	0
19	100	0
20	10	90

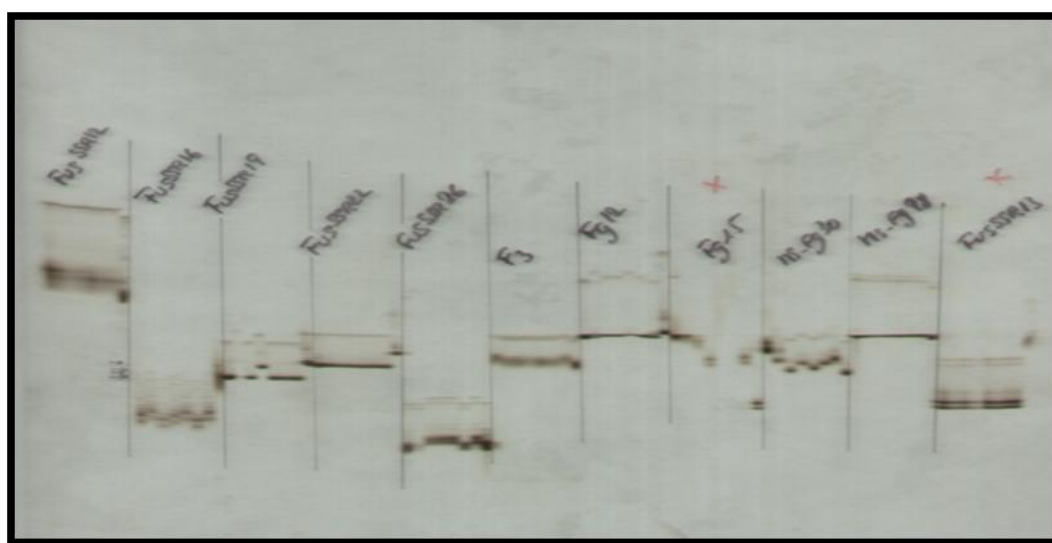
## III-RESULTATS ET DISCUSSION

### III.1. Diversité génétique

#### III.1.1. Mise au point et sélection des marqueurs SSR les plus polymorphes.

Les trente couples d'amorces testés sont opérationnels en PCR. La taille du fragment observée après électrophorèse sur gel d'agarose correspond à la taille du fragment attendue. Les conditions d'amplification décrites dans la littérature (Vogelgsang *et al*, 2009 ; Naef and Défago, 2006 ; Giraud *et al*, 2002) ont donc été transposables au laboratoire, de façon efficace et reproductible.

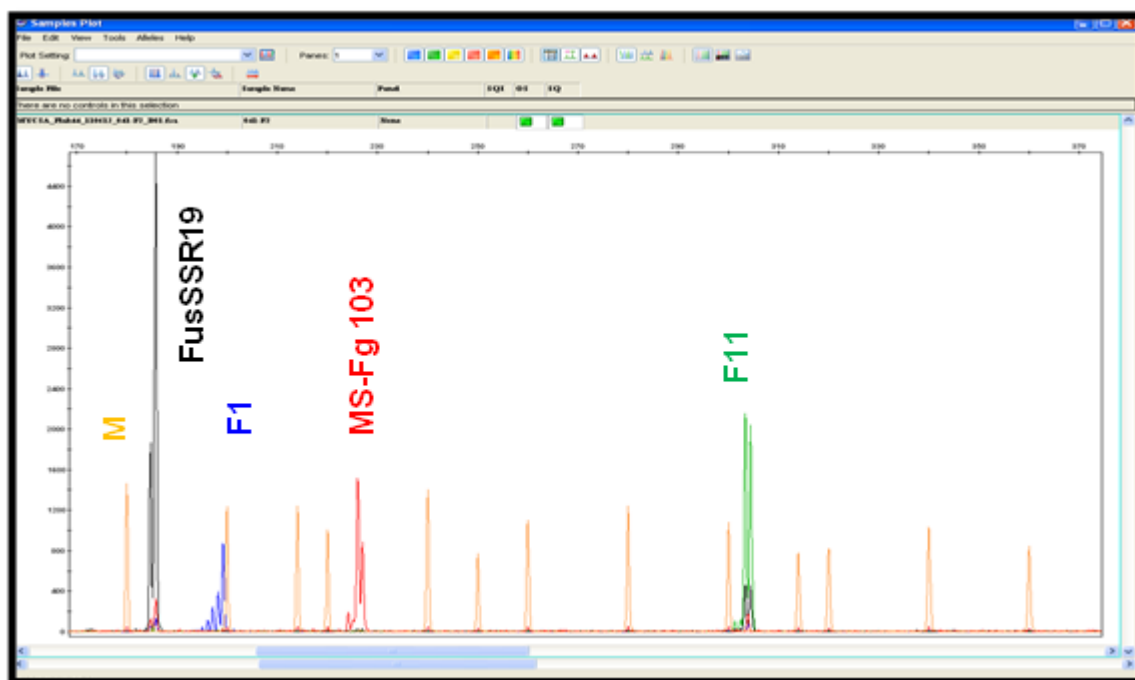
Les 30 couples d'amorces ont été testés pour leur niveau de polymorphisme par électrophorèse sur gel d'acrylamide à 6% (Figure 8). Pour chaque couple, la lecture des profils est basée sur la taille et le nombre d'allèles obtenus parmi les 8 souches testées (Annexe IV). Le marqueur F7 n'a montré aucun profil et cela provient certainement d'un problème lors de la PCR. Ce manque de reproductibilité nous a conduit à l'éliminer pour la suite. Le marqueur FusSSR 13 ne présente qu'un seul allèle et est donc considéré comme étant non polymorphe. Sur la base des 8 souches testées, les marqueurs les plus polymorphes sont (FusSSR23, Fg16, F1), (FusSSR27) et (MS-Fg30, FusSSR18, FusSSR20, F11, FusSSR17) avec respectivement 7, 6 et 5 allèles. Les marqueurs Fg15, FusSSR16, MS-Fg90 et MS-Fg 97 montraient un profil complexe sur gel d'acrylamide, difficilement interprétable. Les marqueurs utilisés pour la suite de l'étude sont sélectionnés en fonction de la facilité de lecture des profils sur le gel d'acrylamide et du niveau de polymorphisme obtenu.



**Figure 8 :** Profils de polymorphisme obtenu sur gel d'acrylamide (6%) pour 11 marqueurs sur les 8 souches testées.

Au final, 24 couples d'amorces ont été retenus pour le multiplexage. Le génotypage avec ces marqueurs est réalisé en six multiplexes, chacun constitué de 4 locus. Les six multiplexes ont été définis en fonction 1) de la température d'hybridation lors du programme PCR, 2) de la taille des fragments amplifiés, 3) du fluorochrome (voir Annexe III).

Ces multiplexes ont été testés par électrophorèse sur séquenceur automatique. La lecture des résultats de génotypage obtenus est réalisée à l'aide du logiciel GeneMapper, la taille des allèles pour chaque locus est défini par rapport au marqueur de taille 600LIZ® utilisé. La Figure 9 illustre un des multiplexes sur un génotype.



**Figure 9 :** Profil électrophorétique obtenu avec le multiplexe MX2 (FusSSR19,F1,MS-Fg103 etF11) sur le génotype INRA 041. M est le marqueur de taille LIZ600.

Les profils obtenus en multiplexe sont conformes à ceux obtenus en simplexe et cohérents avec les observations faites en électrophorèse avec révélation au nitrate d'argent. Presque tous les multiplexes se sont révélés opérationnels. Au sein des multiplexes MX1 et MX3 respectivement, les marqueurs Fg13 et Fg14 ont un signal faible. Ceci peut s'expliquer par une compétition trop forte entre les amorces lors de la PCR (Guichoux *et al.* 2011) ou par une intensité trop faible du fluorochrome associé. Ces deux marqueurs pourront être passés indépendamment en simplexe, ou un ajustement des concentrations en amorces marquées pourraient être testé (Guichoux *et al.* 2011). Certains marqueurs (FusSSR23-MX3 et MS-Fg75-MX5) ont été éliminés car les profils électrophorétiques étaient trop complexes. La définition de nouvelles amorces pourrait améliorer la qualité du signal (Guichoux *et al.* 2011).

Vingt marqueurs seront finalement utilisés pour le génotypage des 90 souches de la collection, soit une perte d'un tiers du potentiel de marqueurs initial. Le facteur limitant n'est pas tant le niveau de polymorphisme que l'obtention d'un profil reproductible et facilement interprétable. Le nombre de marqueurs opérationnels est cependant largement suffisant pour évaluer la diversité et donc aucune optimisation supplémentaire n'a été entreprise.

### **III.1.2. Paramètres de la diversité génétique**

Les paramètres génétiques ont été estimés pour les 20 marqueurs génotypés sur les 90 souches.

Un total de 205 allèles a été détecté pour l'ensemble des 20 loci. Le nombre d'allèles par locus varie de 4 (FusSSR19) à 17 (FusSSR18), pour une moyenne de 10.25 (Tableau 3). Cela est cohérent avec les données de la littérature (Vogelgsang et *al*, 2009 ; Naef et Défago, 2006 ; Giraud et *al*, 2002). Le marqueur FusSSR19 a montré une diversité allélique faible avec 5 allèles, dont l'allèle majeur est celle à 186 pb avec une fréquence très élevée de 0.73. Le marqueur FusSSR24 montre une grande diversité allélique avec 11 allèles mais cependant, l'allèle le plus représenté (121 pb) apparaît dans près de 67 souches. Les 10 autres allèles à ce locus sont donc distribués dans les 23 souches restantes.

Si le nombre d'allèle n'est pas identique à ceux décrits dans ces publications, les ordres de grandeurs restent les mêmes avec les marqueurs décrits comme étant les plus polymorphes sont également ceux qui présentent le plus grand nombre d'allèle dans notre étude. Pour illustration, le marqueur FusSSR18 qui montre le plus d'allèles (17) dans notre étude a été initialement décrit pour avoir 12 allèles. La gamme de taille d'allèle observée dans notre étude est également plus large, variant de 371 pb à 426 pb, contre 380 p à 418 pb (Vogelgsang et *al*, 2009). Ce n'est pas étonnant de trouver ces différences compte tenu du fait que nous n'avons pas travaillé avec le même échantillonnage. Vogelgsang et *al*, 2009 a travaillé principalement sur des souches européennes, mais hors France. Cela nous laisse supposer une grande variabilité des souches françaises.

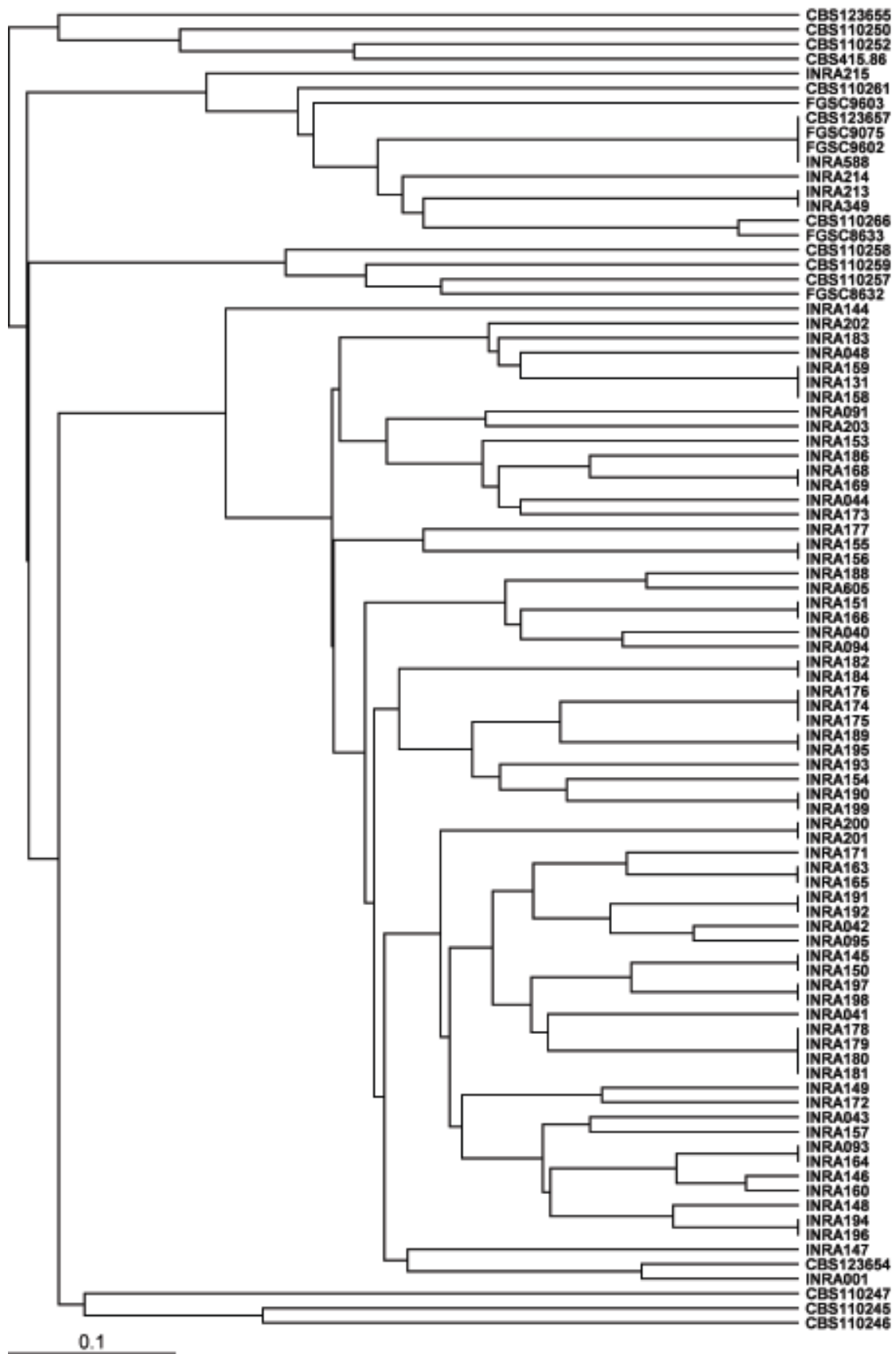
Pour un locus donné, les différences de taille d'allèles correspondent soit à une différence du nombre de répétition du motif, soit à un polymorphisme de type SNP ou indel existant en dehors du motif, dans les régions flanquantes. Par exemple, pour le marqueur F3, les allèles 188, 190, 192, 194, 196 correspondent bien à une incrémentation d'un motif dinucléotidique, mais les allèles 201, 203 et 211, 219, laissent supposer la présence d'un indel.

**Tableau 3 :** Paramètres génétiques pour les 20 marqueurs SSR estimés sur les 90 souches de *F. graminearum* à l'aide de logiciel PowerMarker : nombre d'allèle par locus, allèle majeur et sa fréquence (en jaune les marqueurs les plus polymorphes).

Marqueur	Nbre d'allèles	Allèle majeur	Fréquence de l'allèle majeur
FusSSR19	4	186	0.76
F1	12	199	0.29
MS-Fg103	12	227	0.52
F11	16	319	0.22
MS-Fg60	6	150	0.41
FusSSR17	12	206	0.52
FusSSR26	6	148, 152	0.40
FusSSR27	12	182	0.26
Ms-Fg98	6	232	0.81
FusSSR12.	6	272	0.83
FusSSR16	12	162	0.28
FusSSR22	5	200	0.77
FusSSR14	8	239	0.57
FusSSR24	11	121	0.75
F3	9	192	0.64
MS-Fg30	14	197	0.21
FG12	14	227	0.40
FusSSR21	9	157	0.57
Fg16	14	188	0.48
FusSSR18	17	397	0.20

Une matrice de distance génétique a été établie entre les 90 génotypes. Cette matrice a permis la construction de l'arbre de distance génétique selon la méthode UPGMA (Figure 10). L'arbre obtenu montre une grande diversité génétique au sein de la collection de *F. graminearum* de la mycothèque.





**Figure10** : Arbre de distance génétique (selon la méthode UPGMA) des 90 souches testées avec 20 marqueurs SSR.

Les souches qui présentent une distance génétique nulle entre elles sont des souches qui ne présentent aucune différence d'allèles pour tous les locus et sont considérées comme étant un même génotype. Ainsi, à partir des 90 souches incluses dans notre échantillon, 60 génotypes différents ont été obtenus. Ceci nous permet d'estimer la redondance de la collection à environ 33 %. Pour certaines souches, cette redondance était attendue. Les souches INRA588, CBS123657, FGSC9075 et FGSC9602 sont génétiquement identiques et correspondent à la souche de référence PH1 utilisé pour le séquençage du génome complet de *F. graminearum*. Il en est de même avec les souches INRA213 et INRA349 qui sont en fait deux fois la souche CBS18532 récupérées au laboratoire à plusieurs années d'intervalle. Certaines redondances peuvent également être biologiquement expliquées. Par exemple, les 3 souches INRA174, INRA175, INRA176 ont une distance génétique nulle et sont considérées comme un même génotype. Les informations disponibles sur ces souches montrent une même origine et une même campagne de prélèvement (blé tendre, France-27, 2002). Les résultats de génotypage laissent supposer que ces 3 souches sont en fait 3 isolats d'une même souche. D'autres redondances sont plus inattendues comme le regroupement de la souche INRA093 (maïs, France-45, 2001) et la souche INRA164 (blé tendre, France-76, 2002).

Malgré la redondance, une grande diversité génétique a pu être mise en évidence au sein de notre collection constituée principalement de souches françaises. Ce résultat confirme ceux précédemment obtenus par Brygoo et Gautier (2007).

Les souches qui appartiennent au complexe de *F. graminearum*, en dehors des souches *F. graminearum sensu stricto* se distinguent nettement dans l'arbre. Les espèces *F. mesoamericanum*, *F. austroamericanum* et *F. boothi* sont les plus éloignés génétiquement des autres espèces du complexe. Ceci est conforme à la littérature (Yli Mattila *et al.* 2009). Pour certains marqueurs, des allèles semblent spécifiques de l'espèce. Par exemple, pour FusSSR12, l'allèle 137 n'est retrouvée que dans les souches *F. asiaticum*, pour MS-Fg103, l'allèle 217 n'est présente que chez les représentants de *F. austroamericanum*.

Les deux couples parents de FGSC sont aussi localisés sur l'arbre. Les souches FGSC 8632 et FGSC8633 sont très éloignées, avec une distance de 0.90. La souche FGSC8632 se retrouve regroupé avec des souches de *F. asiaticum* (CBS110257, CBS110258, CBS110259), ce qui est conforme à son origine décrite dans la littérature (Jurgenson *et al.* 2002). Les souches FGSC9602 et FGSC 9603 sont génétiquement différentes, mais avec une distance génétique plus faible de l'ordre de 0.56. Ces résultats sont en accord avec les données publiées (Gale *et al.*, 2005 ; Jurgenson *et al.*, 2002). Remarquons que FG9602 est décrite comme étant une PH1 génétiquement modifiée (Gale *et al.*, 2005) et qu'au niveau des marqueurs microsatellites étudiés, aucune différence n'a été mise en évidence.

Compte tenu de la redondance observée, nous pourrions recalculer les paramètres génétiques avec les 60 génotypes différents identifiés. Les fréquences alléliques s'en trouveraient modifiées.

### **III.2 Evaluation du niveau de production de mycotoxines**

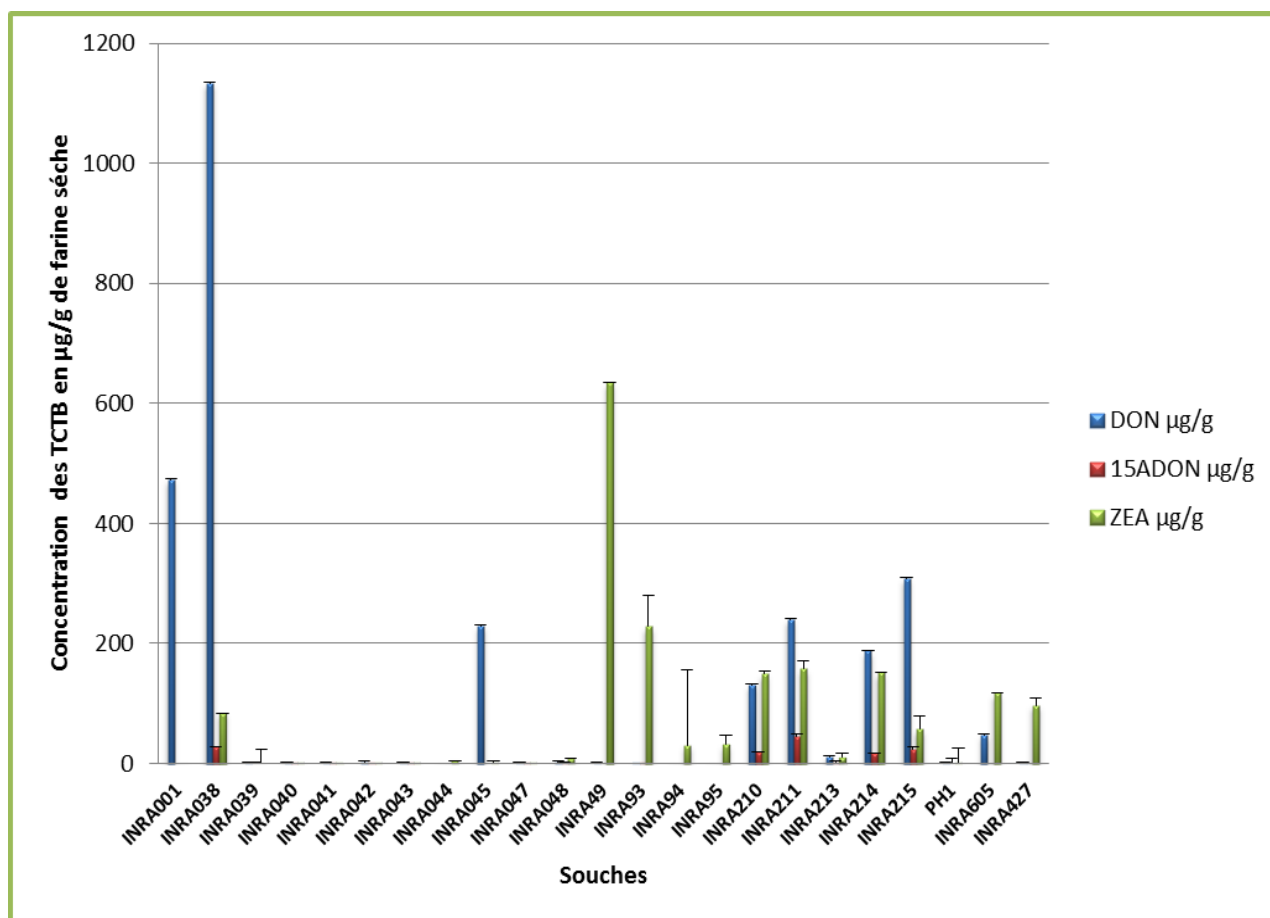
#### **III.2.1. Souches de la Mycothèque de l'Unité MycSa, INRA.**

Les résultats ont montré que 9% des souches de *F.graminearum* utilisées produisent uniquement du déoxynivalénol (DON), et (ou) sa forme acétylée (15 ADON), 26% produisent de la Zéaralénone seule (ZEA), 52% produisent les deux toxines, et 13% (3 souches) ne produisent aucune des deux toxines. Aucune souche ne s'est révélée productrice de NIV, FIX et 3 ADON, donc ces trois TCTB ne seront pas présentés sur le graphique des résultats. Toutes les souches testées produisant des trichothécènes ont présenté un chémotype (DON/15ADON), mais pas de chémotype (DON/3ADON) ni (NIV/FX). La même observation a été reportée dans plusieurs études sur *F.graminearum* (Gale *et al*, 2005; Goswami *et al*, 2004 ; Brygoo *et al*, 2007).

Les valeurs de production de TCTB et de la ZEA présenté ci-après correspondent à la moyenne des trois répétitions technologiques. Au total, les 23 souches ont une moyenne de production des TCTB (DON/15ADON) de 121 µg/g et une moyenne de 73 µg/g pour la ZEA.

Les résultats du dosage sont reportés dans la figure 11 et dans l'annexe IV « Tableau (A) ». Les résultats montrent que la souche INRA 038 est la souche la plus productrice de DON et de 15 ADON à des concentrations de (1133.4 µg/g ±60.64) et (28.3µg/g±1.45) respectivement. Elle produit également de la ZEA avec une concentration de (84.37µg/g±24.43). Les six souches suivantes ont été constatées comme productrices de DON et de 15ADON à des concentrations élevées (> 100 µg/g) par rapport à l'ensemble de 23 souches testées (INRA001, INRA215, INRA211, INRA045, INRA214, et INRA210). Les deux souches INRA001, INRA 045 ne produisent pas de 15ADON. Ces souches produisent aussi de la ZEA, sauf la souche INRA001 qui produit uniquement du DON. Les souches INRA213 et INRA605 produisent des quantités faibles mais non négligeables de toxines (>3 µg/g et <100 µg/g). Pour les trois souches INRA044, INRA094 et INRA095 aucune production du DON et de 15ADON, n'a été détectée. Elles ne produisent que la ZEA. Les neuf souches considérées comme non productrices ou faiblement productrices de TCTB (DON/15ADON) et de ZEA sont les suivantes : INRA039, INRA040, INRA041, INRA042, INRA043, INRA044, INRA047, INRA048 et INRA 588. En se basant sur les résultats

obtenus, une souche est classée faiblement productrice lorsque la concentration en TCTB est entre (0.0  $\mu\text{g/g}$  – 3  $\mu\text{g/g}$ ) et celle de ZEA entre (0.0 $\mu\text{g/g}$  – 8 $\mu\text{g/g}$ ).



**Figure 11** : Représentation graphique des concentrations de TCTB (DON, 15 ADON) et de ZEA produites par les 23 souches *F. graminearum* issues de la collection INRA sur grains de blé pendant 21 jours de culture.

### III.2.2 Souches de Fungal Genetics Stock Center (FGSC), USA.

Le dosage des TCTB et de la ZEA des souches provenant de FGSC a été réalisé sur grains de blé et sur milieu liquide MS. Les souches sont les suivantes : les 4 souches parents (FGSC 8632 / FGSC 8633) et FGSC 9602(PH-1 modifiée)/ FGSC 9603) et la souche PH-1 FGSC 9075 (Jungkwan *et al*, 2008). Cette dernière est la souche de référence.

Les résultats de dosage sont reportés dans les Figures 12 et 13 (A et B).

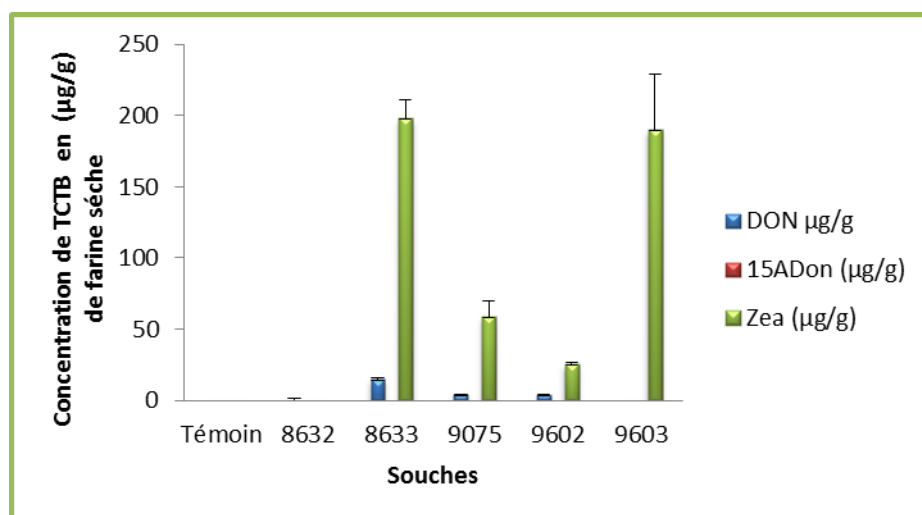
Des différences de production de TCTB et de la ZEA ont été constatées entre les deux milieux de cultures (sur grains et sur MS). La production des TCTB (DON et 15ADON) sur milieu MS est beaucoup plus élevée que sur grains de blé. La production de 15ADON est totalement absente sur grains alors qu'elle est très élevée dans le milieu MS. La différence de production entre les deux milieux de cultures est déjà observée précédemment (Boutigny, 2007). Cela peut être expliqué par la composition du milieu minimum synthétique (MS), qui

est un milieu pauvre mis au point pour provoquer une induction rapide de la production de toxines lors de la culture *in vitro* de souches de *Fusarium* (Boutigny, 2007).

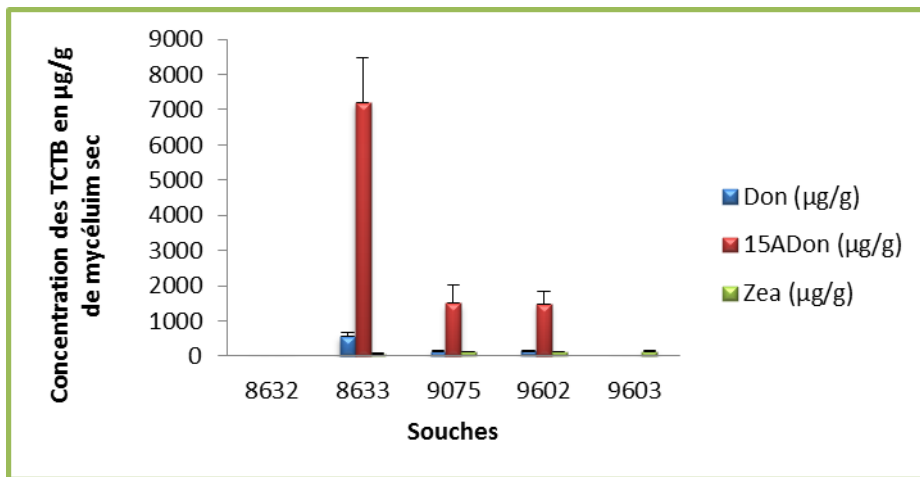
Dans le couple de parents (FGSC 8632 / FGSC 8633), les deux souches apparaissent très contrastées : la souche FGSC 8633 est productrice de TCTB alors que pour la souche FGSC8632, aucune toxine n'a été détectée. Ce résultat n'est pas étonnant compte tenu du fait que cette souche est décrite comme étant de chémotype NIV (Jurgenson *et al.* 2002). Dans notre expérience nous n'avons pas pu détecter de toxine de type NIV. Ceci peut s'expliquer par une dilution trop importante lors de la préparation des échantillons ou alors nos conditions de culture ne sont pas adaptées à la production de ce type de toxine sur cette souche.

Le deuxième couple des parents (FGSC 9602 / FGSC 9603), la FGSC 9603 ne produit pas de TCTB, alors qu'elle produit de la ZEA. L'autre souche FGSC 9602 produit du 15ADON à une concentration très élevée (1495  $\mu\text{g/g} \pm 526.4$ ) par rapport à celles du DON (304  $\mu\text{g/g} \pm 49.4$ ) et de la ZEA (107  $\mu\text{g/g} \pm 22.6$ ). Il y a très peu de différence dans les niveaux de production entre les souches FGSC9075 et FGSC9602, la souche génétiquement modifié FGSC9602 reste phénotypiquement très proche de la souche sauvage FGSC9075 (=PH1).

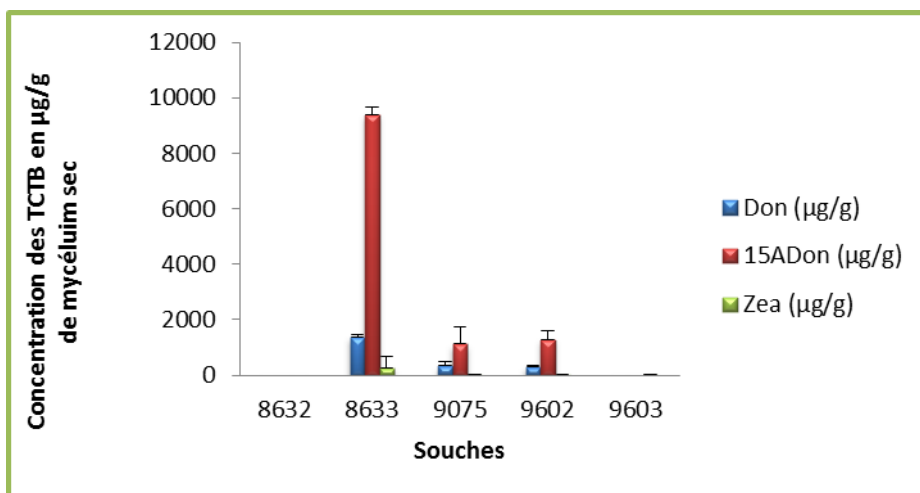
Le contraste entre les souches parentales est plus important pour le couple FGSC8632/FGSC8633 que pour le couple FGSC9602/FGSC9603. Malgré cela, nous avons choisi de caractériser un sous ensemble de la descendance (30 génotypes) issue du croisement FGSC9602xFGSC9603 car les deux souches sont du même chémotype. Malheureusement le temps ne nous a pas permis d'obtenir les résultats de dosage à temps pour les intégrer dans le rapport.



**Figure 12 :** Représentation graphique des concentrations de TCTB (DON, 15 ADON) et la ZEA produites par les 4 souches *F. graminearum* parents de FGSC sur grains de blé pendant 21 jours de culture.



(A) Culture de 12 jours sur MS



(B) Culture de 16 jours sur MS

**Figure 13 :** Représentation graphique des concentrations de TCTB (DON, 15 ADON) et la ZEA produites par les 4 souches *F. graminearum* parents de FGSC sur milieu MS. (A) culture de 12 jours, (B) culture de 16 jours.

## **IV- CONCLUSION GENERALE ET PERSPECTIVES**

Dans le cadre de ce stage, un set de 20 marqueurs moléculaires de type microsatellites a été mis au point au laboratoire afin de caractériser les souches de *F. graminearum*. Ils peuvent dès à présent être utilisés en routine afin de dresser le profil génétique de nouvelles souches rentrant en collection. Ces marqueurs pourront également être testés sur des espèces proches de *F. graminearum* telles que *F. culmorum* ou *F. cerealis*.

L'identification de la redondance existante au sein de la collection permet d'envisager une meilleure gestion de celle-ci. L'élimination des souches redondants permettra notamment un gain de place et de temps. La diversité génétique existante au sein de notre collection est également importante pour le maintien et la valorisation de ressources génétiques.

Il reste à faire le lien entre la diversité génétique et la diversité phénotypiques observées. Les données de marqueurs ou de dosage de toxine seront intégrées dans la base de données existante et comparées aux données déjà disponibles.

Les résultats acquis pendant ce stage nous permettent de proposer des couples de parents potentiels pour la construction d'une nouvelle descendance. Notre choix est guidé par une distance génétique importante et un contraste au niveau de la production de toxines. Ainsi les couples de parents que nous avons pu choisir sont les suivants : les couples (INRA 215/ INRA 94), (INRA 214/ INRA 41) et (INRA191/ INRA 44) de chémotype (DON/15ADON). Les deux souches de chaque couple sont très éloignées génétiquement avec une distance de 0.76, 0.90 et 0.60 respectivement. De plus elles sont très différentes au niveau de production de TCTB et ZEA.

## V- RÉFÉRENCES BIBLIOGRAPHIQUES

Alvarez.C.L, Somma.S, Proctor.R.H, Stea.G, Mulé.G, Logrieco.A.F, Pinto.V.F, Moretti.A, (2011) Genetic diversity in *fusarium graminearum* from a major wheat-producing region of argentina. *Toxins* **3**: p 1294-1309.

Bakan.B, Bily.A.C, Melcion.D, Cahagnier.B, Regnault-Roger.C, Philogene.B.J.R, and Richard-Molard.D, (2003) Possible role of plant phenolics in the production of trichothecenes by *Fusarium graminearum* strains on different fractions of maize kernels. *J. Agric. Food Chem.* **51**: p 2826-2831.

Bennett,J.W and Klich.M, (2003) Mycotoxins. *Clinical Microbiology Reviews.* **16**: p. 497-516.

Bourdages.J.V, Marchand. S, Rioux.S et Belzilz.F, (2005) Caractérisation moléculaire de la diversité génétique chez des *Fusarium* responsables de toxi-infections. Québec. Université Laval.

Boutigny A.L, (2007) Identification dans les grains de blé dur de composés inhibiteurs de la biosynthèse des trichothécènes B par *Fusarium*. Thèse de doctorat, Bordeaux, Université Bordeaux 1.

Bottalico. A., and Perrone.G, (2002). Toxigenic *Fusarium* species and mycotoxins associated with head blight in small-grain cereals in Europe. *Eur. J. Plant Pathol.* **108**: p 611-624.

Brygoo. Y et Gautier. A, (2007) Polymorphisme moléculaire des souches de *Fusarium* isolées sur épis de blé et de maïs en France-Récoltes 2003 et 2004. *Mycotoxines fusariennes des céréales-Arcachon.*

Champeil. A, Dore.T and Fourbet. J.F, (2004) *Fusarium* head blight: epidemiological origin of the effects of cultural practices on head blight attacks and the production of mycotoxins by *Fusarium* in wheat grains. *Plant Science.* **166**: p. 1389-1415.

Cumagun,C.J.R., Bowden, R.L., Jurgenson, J.E., Leslie, J.F., and Miedaner, T, ( 2004) Genetic mapping of pathogenicity and aggressiveness of *Gibberella Zeae* ( *Fusarium graminearum*) toward wheat. *Phytopathology* **94**: p 520-526.

Cumagan CJR, Parzies HK, Miedaner T, (2007) Genetic variation and segregation of DNA polymorphisms in *Gibberella zeae* detected with AFLP and RAPD markers. *Asian J Plant Sci* **6**: p 1174–1181.

Creste.S, Tulmann Neto.A, Figueria.A, (2001) Detection of single sequence repeat polymorphisms in denaturing polyacrylamide sequencing gels by silver staining. *Plant Molecular Biology Reporter* **19**: p 299-306.

Elisavet.V et al, (2010), Cleavage of Zearalenone by *Trichosporon mycotoxinivorans* to a novel nonestrogenic metabolite. *Applied and Environmental Microbiology*,**76**: p 2353-2359.



- Foulongne-Oriol, M, (2012) Genetic linkage mapping in fungi: current state, applications, and future trends. *Appl Microbiol Biotechnol* **95**: p 891–904.
- Gale, L.R., J. D. Bryant, S. Calvo, H. Giese, T. Katan, K. O'Donnell, H. Suga, M. Taga, T. R. Usgaard, T. J. Ward, and H. C. Kistler, (2005) Chromosome complement of the fungal plant pathogen *Fusarium graminearum* based on genetic and physical mapping and cytological observations. *Genetics* **171**: p 985-1001.
- Gaumy, J.L., Bailly, J.D., Burgat, V.et. Guerre, P, (2001) Zéaralénone : propriétés et toxicité expérimentale. *Revue Méd. Vét*, **152** : P 219-234
- Grevet. N, (2004) Modes d'action et toxicité des trichothécènes. Thèse de doctorat. Toulouse.
- Goswami. R.S. and. Kistler. H.C, (2004) Heading for disaster: *Fusarium graminearum* on cereal crops. *Molecular Plant Pathology*. **5**: p. 515-525.
- Guichoux E**, Lagache L, Wagner S, *et al*, (2011) Current trends in microsatellite genotyping. *Molecular Ecology Resources* **11**: p 591-611.
- Jungkwan. L, Jurgenson .J, Leslie.J.F, L. Bowden R, (2008) Alignment of Genetic and Physical Maps of *Gibberella zeae*. *Appl Environ Microbiol*.**74**: p 2349–2359.
- Jurgenson, J.E et al, (2002) A genetic map of *Gibberella zeae* (*Fusarium graminearum*).*Genetics* **160**: p 1451-1460.
- Karaoglu. H., Lee. C.M.Y, Meyer. W, (2005) Survey of simple sequence repeats in completed fungal genomes. *Mol. Biol. Evol.* **22**: p 639–649.
- Kim, H. S., T. Lee, *et al*, (2003) "Polymorphism of trichothecene biosynthesis genes in deoxynivalenol- and nivalenol-producing *Fusarium graminearum* isolates." *Mycological Research* **107**: p 190-197.
- Kim J.E, Han K.H., Jin J., Kim H., Kim J.C., Yun S.H., Lee Y.W, (2005) Putative polyketide synthase and laccase genes for biosynthesis of aurofusarin in *Gibberella zeae*. *Appl. Environ. Microbiol.* **71**: p 1701-1708.
- Kimura.M. *et al*, (2007) Molecular and genetic studies of *fusarium* trichothecene biosynthesis: pathways, genes, and evolution. *Biosci. Biotechnol. Biochem*, **71**: p. 2105-23.
- Lysoe E, Klemsdal SS, Bone KR, Frandsen RJN, Johansen T, Thrane U and Giese H, (2006) The pks4 gene of *Fusarium graminearum* is essential for zearalenone production. *Appl. Environ. Microbiol.* **72**: p 3924-3932.
- Mebarek. L, (2010) Analyse comparative de la diversité génétique et de la structure des populations chez l'orge (*hordeum vulgare*l.) à l'aide de marqueurs ssr, dart et du Pedigree. Thèse de doctorat. Québec, université Laval.
- Merhej. J, (2010) Mécanismes moléculaires contrôlant la biosynthèse de mycotoxines par le champignon micromycètes *Fusarium graminearum*. Thèse de doctorat. Bordeaux, Université Bordeaux 1.

Morgante. M., Olivieri. A.M, (1993) PCR-amplified microsatellites as markers in plant genetics. *Plant J.* **3**: p. 175–182.

Naef. A, Senatore. M, Défago.G, (2006) A microsatellite based method for quantification of fungi in decomposing plant material elucidates the role of *Fusarium graminearum* Don production in the saprophytic competition with *Trichoderma atroviride* in maize tissue microcosms. *FEMS Microbiol Ecol*, **55** : p 211-220.

Najimi. B, El Jaafari. S, Jlibène. M, Jacquemin. J.M , (2003) Applications des marqueurs moléculaires dans l'amélioration du blé tendre pour la résistance aux maladies et aux insectes. *Biotechnol. Agron. Soc. Environ.* **7** : p 17–35.

O'Donnell.K, Kistler.H.C, Tacke.B.K, Casper.H.H, (2000) Gene genealogies reveal global phylogeographic structure and reproductive isolation among lineages of *Fusarium graminearum*, the fungus causing wheat scab. *Proceedings of National Academy of Science* **97**: p 7905-7910.

Parry. DW, Jenkinson. P and McLeod L, (1995) *Fusarium* ear blight (scab) in small grain cereals-a review. *Plant Pathology* **44**: p207-238.

Ponts, N, (2002) Effets de H<sub>2</sub>O<sub>2</sub> sur la production de trichothécènes et de fumonisines par *Fusarium* en cultures liquides. pp. 25. ISTAB - Université Bordeaux 1, Bordeaux.

Toth.B, Mesterhazy.A, Horvath.Z, Bartok.T, Varga.M, Varga.J, (2005) Genetic variability of central European isolates of *Fusarium graminearum* species complex. *European Journal of Plant Pathology* **113**: p 35-45.

Vogel. H.J, (1956) A convenient growth medium for *Neurospora* (medium N) *Microbial Genetics Bulletin*, **13**: p 42-43.

Vogelgsang, S. Widmer, F. Jenny, E. Enkerli, J, ( 2009) Characterisation of novel *Fusarium graminearum* microsatellite markers in different *Fusarium* species from various countries. *Eur J Plant Pathol* **123**: p 477-482.

Wang.J.H, Ndoye.M, Zhang.J.B, Li.H, Liao.Y.C, (2011) Population structure and genetic diversity of the *Fusarium graminearum* species complex. *Toxins* **3**: p 1020-1037

Windels. C, (2000) Economic and social impacts of *Fusarium* head blight : changing farms and rural communities in the northern Great Plains. *Phytopathol* **90** : p 17-21.

Yli-Mattila, T., Gagkaeva, T., Ward, T. J., Aoki, T., Kistler, C. & O'Donnell, K, (2009) A novel Asian clade within the *Fusarium graminearum* species complex includes a newly discovered cereal head blight pathogen from the Far East. *Mycologia* **101**: p 841–852.

# ANNEXES

## ANNEXE I

Tableau des concentrations maximales autorisées de certaines « fusariotoxines » dans les produits alimentaires dans l'Union Européenne (règlement de la Commission européenne n°1881/2006 modifié par le règlement (CE) n°1126/2007).

<b>Produits</b>	<b>Mycotoxines</b>	<b>Valeurs µg/Kg</b>
<b>Céréales brutes autres que le blé dur, l'avoine et le maïs</b>	<b>Déoxynivalénol</b>	<b>1250</b>
<b>Blé dur et avoine bruts</b>	<b>Déoxynivalénol</b>	<b>1750</b>
<b>Maïs brut</b>	<b>Déoxynivalénol</b>	<b>1750</b>
<b>Pain</b>	<b>Déoxynivalénol</b>	<b>500</b>
<b>Pâtes (sèches)</b>	<b>Déoxynivalénol</b>	<b>750</b>
<b>Céréales brutes autres que le maïs</b>	<b>Zéaralénone</b>	<b>100</b>
<b>Maïs brut</b>	<b>Zéaralénone</b>	<b>350</b>

## ANNEXE II

Tableau regroupant toutes les souches (de la mycothèque et de FGSC) testées génétiquement et phénotypiquement durant cette étude.

Souches de la collection INRA - Mycothèque Unité MycSa				
Phénotypage	Génotypage			
Les 23 souches de la Mycothèque	les 8 Souches utilisées pour la mise au point des Marqueurs FusSSR			
INRA001	INRA091	INRA144	INRA149	INRA153
INRA038	INRA158	INRA186	INRA195	INRA349
INRA039	Les 85 Souches de la Mycothèque			
INRA040	INRA001	INRA157	INRA186	CBS110258*
INRA041	INRA040	INRA158	INRA188	CBS110259*
INRA042	INRA041	INRA159	INRA189	CBS110245*
INRA043	INRA042	INRA160	INRA190	CBS110246*
INRA044	INRA043	INRA163	INRA191	CBS110250*
INRA045	INRA044	INRA164	INRA192	CBS123655*
INRA047	INRA048	INRA165	INRA193	CBS123654
INRA048	INRA091	INRA166	INRA194	CBS110261
INRA49	INRA093	INRA168	INRA195	CBS110266
INRA93	INRA094	INRA169	INRA196	CBS123657 (PH1)
INRA94	INRA095	INRA171	INRA197	CBS415.86*
INRA95	INRA131	INRA172	INRA198	CBS110252*
INRA210	INRA144	INRA173	INRA199	CBS110247*
INRA211	INRA145	INRA174	INRA200	5 Souches FGSC
INRA213	INRA146	INRA175	INRA201	FGSC8632
INRA214	INRA147	INRA176	INRA202	FGSC8633
INRA215	INRA148	INRA177	INRA203	FGSC9602
INRA588 (PH1)	INRA149	INRA178	INRA213	FGSC9603
INRA605	INRA150	INRA179	INRA214	FGSC9075(PH1)
INRA427	INRA151	INRA180	INRA215	
	INRA153	INRA181	INRA349	
	INRA154	INRA182	INRA588	
	INRA155	INRA183	INRA605	
	INRA156	INRA184	CBS110257*	

\* souches du complexe *F. graminearum* . CBS110257-CBS110258-CBS110259 *F. asiaticum* ; CBS110245-CBS110246 *F. austroamericanum*; CBS110250 *F. boothii*, CBS123655 *F. cortaderiae*; CBS415.86 CBS110252 *F. mesoamericanum*, CBS110247 *F. meridionale*.

## ANNEXE III

Tableau regroupant les 20 marqueurs et les 6 multiplexes (MX) choisis pour le génotypage : nom du marqueur, références, tailles d'allèle décrites dans la littérature, Nombre d'allèles obtenus lors de la mise au point sur gel d'acrylamide, Température d'hybridation utilisée lors de la PCR, amorce R, amorce F marquée, fluorochrome, et numéro du multiplexe.

Locus	Référence*	Taille allèles (littérature)	T°C	amorce F marquée	Fluorochromes	amorce R	Multiplexes
<b>FusSSR21</b>	1	154-175	55	TCTTGACTGCCCAACATGGAC	FAM	GGGCAATATGAAGTCTGCGTC	<b>MX 1</b>
<b>Fg13</b>	3	146-188	55	TTTGCATGAGGATACGGTGA	NED	CATCACTTTGGTGGTGATGC	
<b>Fg16</b>	3	162-256	55	CCAAGCCTCCAATTGTTTGT	PET	GGACAAAGGCCAAATGAGAA	
<b>FusSSR18</b>	1	380-418	55	GAATACTCAAGTCCCGTAGCCG	VIC	TGCAGAGACCCATAAACGACG	
<b>F1</b>	2	176-232	60	GACAAGCAAGCGATAGGAAA	FAM	CTTGATAGCACGGACCGACG	<b>MX 2</b>
<b>FusSSR19</b>	1	184-204	60	AGCCGGACATGAGACAAAGTAG	NED	TGTTGTTCCCTCCAGTACTCG	
<b>MS-Fg103</b>	4	225-267	60	GGTATCCGTACAACCCGATG	PET	TTCTTTGATTTGGACCGAGG	
<b>F11</b>	2	294-346	60	CAGTCTTGGTCGCTCATCAG	VIC	CAGGTTGGCAGCTTCTTAA	
<b>MS-Fg60</b>	4	147-187	60	GAGCCATTACATGTACCCC	FAM	TCCTCTCGCAAGTGTGTTG	<b>MX 3</b>
<b>FusSSR23</b>	1	165-221	60	GTTGACACAGAAGAATGGCAGG	NED	CGCTAGGTACAAATTGCTGGG	
<b>FusSSR17</b>	1	203-245	60	AATCCGGTATCCGTACAACCC	PET	AAGAACTACTTCGCTGCCGTG	
<b>Fg14</b>	3	232-263	60	GCCATTACGTTGAGCAAAT	VIC	TTTTTCGGTCTGGCTATTGG	
<b>FusSSR26</b>	1	145-157	60	CTAAGTTGCCAATGGACGTCAG	FAM	AGTTTTCTGCAGCCCCAAGAG	<b>MX 4</b>
<b>FusSSR27</b>	1	157-205	60	TCACCAAAAAGTCTCCTCAGTCAAC	NED	GTGGTCTCCGTAAACGAGCC	
<b>MS-Fg98</b>	4	224-240	60	TCTAGGCCAAACCCAGACAC	PET	TGTTGCGTTCCTGAAAACAG	
<b>FusSSR12</b>	1	263-281	60	ATTAGAGAGTGCGGCAGAAAGAG	VIC	CAGACATGTTCAAGGGCAATTG	
<b>FusSSR16</b>	1	139-195	60	CTTTCTCAACCGAATTGGG	FAM	AGTACGGACCGACGACATAACTC	<b>MX 5</b>
<b>MS-Fg75</b>	4	196 (350)	60	TTGGTATGAAAGACTGCCCC	NED	GAAGCGCAAAGGTATGCTC	
<b>FusSSR22</b>	1	200-212	60	GAGGGCGATGGTTGAAGTGAC	PET	TGGGCATGAAACAAGAGAGAGAC	
<b>FusSSR14</b>	1	237-261	60	TTCAGCATCACTTGCGTCG	VIC	TACCGCCCCTCACTCTTATACC	
<b>FusSSR24</b>	1	122-152	60	CGGACATCTTCACCTGGTCC	FAM	GCCGCGAATACAGAGTGCTC	<b>MX 6</b>
<b>F3</b>	2	191-219	60	CATATTCAACCGACCCACAA	NED	TTGAATGATAAGGGCGACGG	
<b>MS-Fg30</b>	4	184-216	60	TATGACTGCACGTTTGCTCC	PET	AGTCTCTAGTGTCGCCGAGG	
<b>Fg12</b>	3	(200)-247	60	AGATGCCGATGATGGATATG	VIC	GTACAAGTGCTTCACGGAAA	

\*1) Vogelgsang *et al.* 2009 ; 2\*) Giraud *et al.* 2002 ; 3\*) Brygoo *et al.* 2007 ; 4\*) Naef & Defago 2006

## **ANNEXE IV**

Tableau du nombre d'allèles obtenus par chaque marqueur avec les 8 souches testées après migration sur gel d'acrylamide

<b>Marqueur</b>	<b>Nbre d'allèles</b>
FusSSR21	2
Fg13	3
Fg16	7
FusSSR18	5
F1	7
FusSSR19	3
MS-Fg103	5
F11	5
MS-Fg60	4
FusSSR23	2
FusSSR17	5
Fg14	4
FusSSR26	2
FusSSR27	6
MS-Fg98	2
FusSSR12	2
FusSSR16	4
MS-Fg75	4
FusSSR22	2
FusSSR14	4
FusSSR24	2
F3	3
MS-Fg30	5
Fg12	4
MS-Fg90	Lecture difficile
F7	Pb d'amplification
MS-Fg97	Lecture difficile
Fg15	Lecture difficile
FusSSR13	1

## ANNEXE V

Tableau A des concentrations de TCTB (DON, 15 ADON) et la ZEA produites par les 23 souches *F. graminearum* de la collection INRA (Mycothèque) sur grains de blé pendant 21 jours de culture.

souches	DON µg/g	15ADON µg/g	ZEA µg/g
Témoin N	0	0	0
INRA001	473±174.1	0	0
INRA038	1133±60.6	28±1.5	84±24.4
INRA039	0	0	0
INRA040	2±0.2	0	1±0.2
INRA041	0	0	1±0.2
INRA042	3±0.1	0	0
INRA043	0	0	0
INRA044	0	0	4±2.5
INRA045	229±108.4	0	1±0.3
INRA047	0	0	0
INRA048	3±0.1	1±0.4	8±0.7
INRA49	1±0.1	0	634±52.0
INRA93	0	0	125.1
INRA94	0	0	30±15.7
INRA95	0	0	32±4.5
INRA210	131±31.6	19±2.9	149±13.6
INRA211	240±22.2	45±3.3	157±7.4
INRA213	11±0.5	1±0.1	10±0.3
INRA214	188±37.0	18±4.8	151±21.1
INRA215	308±59.6	23±8.9	57±25.4
INRA 588	1±0.2	0	1±0.0
INRA605	48±5.4	0	117±11.8
INRA427	0	0	97±33.1

Tableau B des concentrations de TCTB (DON, 15 ADON) et la ZEA produites par les 4 souches *Fusarium graminearum* parents de FGSC sur grains de blé pendant 21 jours de culture.

Souches	Don µg/g	15ADon µg/g	Zea µg/g
Témoin	0	0	0
8632	0	0	0
8633	15±1.5	0	198±12.5
9075	4±0.1	0	59±10.8
9602	4±0.3	0	26±1.3
9603	0	0	190±38.7

Tableau C des concentrations de TCTB (DON, 15 ADON) et la ZEA produites par les 4 souches *Fusarium graminearum* parents de FGSC sur milieu MS pendant de 12 jours.

Souches	Don µg/g	15ADon µg/g	Zea µg/g
Témoin	0	0	0
8632	0	0	0
8633	562±93.2	7203±1258.7	63±12.4
9075	127±37.7	1495±526.4	117±23.3
9602	142±34.5	1480±369.8	107±22.6
9603	0	0	136±12.9

Tableau D des concentrations de TCTB (DON, 15 ADON) et la ZEA produites par les 4 souches *Fusarium graminearum* parents de FGSC sur milieu MS pendant de 16 jours.

Souches	Don µg/g	15ADon µg/g	Zea µg/g
Témoin	0	0	0
8632	0	0	0
8633	1372±107.9	9398±273.9	268±415.5
9075	335±155.4	1125±626.8	7±12.9
9602	305±49.5	1280±341.0	21±0.5
9603	0	0	9±15.2



## II. Les scripts utilisés pour les analyses de bio-informatique :

Les outils bioinformatiques utilisés pour analyser les données de NGS ont été exécutés sur le cluster de calcul de la plateforme Genotoul (<http://bioinfo.genotoul.fr/>) par l'intermédiaire de fichiers bash.

### II.1. La détection de variants

#### II.1.1. (Pré-) Traitement des données de séquençages

#La qualité des reads a été vérifiée en utilisant FastQC, à partir des fichiers fastq délivrés après séquençage. Deux fichiers de reads par souche sont produits, les reads R1 et les reads R2 (séquençage paired end). On prendra l'exemple de la souche INRA\_156 mais l'opération sera répétée pour toutes les souches.

```
gzip -d /Chemin/INRA_156_1.fastq.gz  
gzip -d /Chemin/INRA_156_2.fastq.gz  
fastqc /Chemin/INRA_156_1.fastq -o /Chemin/INRA_156_1.fastqc  
fastqc /Chemin/INRA_156_2.fastq -o /Chemin/INRA_156_2.fastqc
```

#Tout le long du process de nettoyage des reads, on utilise PrinSeq (<http://prinseq.sourceforge.net/manual.html>). Ce logiciel est très complet. 1) Il faut enlever les reads qui aurait pu être dupliqués lors de la construction des banques et qui représente un biais. Dans notre cas il faut aussi enlever systématiquement les 9 premiers nucléotides dû à un biais de composition en séquence (contamination par les primers utilisés par la technologie Illumina lors de la construction des banques)

```
chemin_soft/prinseq-lite.pl -fastq /Chemin/INRA_156_1.fastq -fastq2  
/Chemin/INRA_156_2.fastq -trim_left 9 -out_format 3 -derep 14 -  
derep_min 2 -log Chemin/log.I_156.dedup.txt -out_good  
Chemin/prinseq.I_156 -out_bad null
```

# -> trim\_left 9. -out\_format 3 signifie que l'output sera .fastq. -out\_bad null on ne garde pas les séquences filtrées, -out\_good on met le préfixe du fichier output et il renommera automatiquement les R1 et R2. -log permet de créer un fichier log avec quelques statistiques. derep 14 signifie qu'il va sortir les reads exactement **similaires** ainsi que les reverse-complément exactement similaires.

#Ensuite on filtre les reads trop courts ou de mauvaises qualités générales et puis on élimine les extrémités qui ont une qualité trop faible en 5' et 3' ; en calculant la qualité moyenne sur une fenêtre de 3 bases et en avançant une base par une base.

```
prinseq-lite.pl -fastq prinseq.I_156_1.fastq -fastq2
prinseq.I_156_2.fastq -trim_qual_left 20 -trim_qual_right 20 -
trim_qual_type mean -trim_qual_window 3 -trim_qual_step 1 -min_len
20 -min_qual_mean 20 -out_format 3 -log
Chemin/log.I_156.trimfilt.txt -graph_data I_156.gd -out_good
Chemin/prinseq.trimfilt.I_156 -out_bad null
```

## II.1.2. L'alignement des reads sur le génome de référence

# Pour l'alignement (*i.e.* mapping) BWA et l'algorithme « mem » a été utilisé. Le « mem » est l'algorithme recommandé pour la taille de nos reads en paired-end.

Dans notre cas on a réduit la taille de la « graine » (-k 15) pour une meilleure précision, les autres paramètres sont standards. "/Chemin/" est le chemin de l'emplacement de mon fichier. Avant utilisation, la référence (ici appelée NewRef) doit avoir été indexée.

```
bwa mem -k 15 /CheminRéférence/NewRef
/Chemin/prinseq.trimfilt.I_156_1.fastq
/Chemin/prinseq.trimfilt.I_156_2.fastq >
/Chemin/FullGenomeBWA_156.sam
```

#Le fichier SAM de sortie est compressé en fichier BAM avec les outils samtools

```
samtools view -b -S /Chemin/FullGenomeBWA_156.sam >
/Chemin/FullGenomeBWA_156.bam
```

# Le fichier bam doit être ré-ordonné pour suivre les coordonnées de la référence.

```
samtools sort /Chemin/FullGenomeBWA_156.bam
/Chemin/FullGenomeBWA_156_sorted
```

#Un index est créé.

```
java -jar -Xmx4G /usr/local/bioinfo/src/picard-
tools/current/BuildBamIndex.jar \
```

```
I=/Chemin/FullGenomeBWA_156_sorted.bam \
```

```
O/Chemin/FullGenomeBWA_156_sorted.bai
```

#L'outil samtools flagstats permet de générer de nombreuses statistiques.

```
samtools flagstat /Chemin/FullGenomeBWA_156_sorted.bam >
/Chemin/Stats/FullGenomeBWA_156.txt
```

## II.2. La détection de variants génétiques

### II.2.1. Préparer les fichiers BAM

#Pour cela il faut préalablement ajouter des « reads groups » au fichier bam, nécessaires pour utiliser GATK. ID et autres paramètres sont des désignations qui seront rajoutés au fichier et qui apportent des infos sur l'identité de la souche, le type de séquençage ... Le fichier bam ainsi complété est indexé à nouveau.

```
java -jar -Xmx4G /usr/local/bioinfo/src/picard-  
tools/current/AddOrReplaceReadGroups.jar \  
I=/Chemin/FullGenomeBWA_156_sorted.bam \  
O=/Chemin/FullGenomeBWA_156_rg.bam \  
ID=[INRA156] \  
RGLB=[INRA156] \  
PL=[Illumina] \  
PU=[Mgx] \  
SM=[INRA156]
```

```
java -jar -Xmx4G /usr/local/bioinfo/src/picard-  
tools/current/BuildBamIndex.jar \  
I=/Chemin/FullGenomeBWA_156_rg.bam \  
O=/Chemin/FullGenomeBWA_156_rg.bai
```

#Les InDels non-proprement positionnés peuvent créer de faux SNPs. Nous allons donc étudier ces InDels. Tout d'abord, il faut identifier les zones du génome où sont positionnées les InDels et les noter dans un fichier « .intervals »

```
java -jar -Xmx4G  
/usr/local/bioinfo/src/GATK/latest/GenomeAnalysisTK.jar \  
-T RealignerTargetCreator \  
-I /Chemin/FullGenomeBWA_156_rg.bam \  
-R /Chemin/NewRef.fa \  
-o /Chemin/FullGenomeBWA156.intervals \  
-mismatch 0.0
```

#Ensuite le logiciel va réaligner les reads au niveau des InDels précédemment repérés.

```
java -jar -Xmx4G  
/usr/local/bioinfo/src/GATK/latest/GenomeAnalysisTK.jar \  
-T IndelRealigner \  
-I /Chemin/FullGenomeBWA_156_rg.bam \  
-R /Chemin/NewRef/GATK/NewRef.fa \  
-targetIntervals /Chemin/FullGenomeBWA156.intervals \  
-o /Chemin/FullGenomeBWA_156_realigned.bam \  
-model USE_READS
```

```
samtools flagstat /Chemin/FullGenomeBWA_156_realigned.bam >
/Chemin/FullGenomeBWA_156_realigned_flagstat.txt
```

## II.2.2. La détection de variants

#L'utilisation de la fonction UnifiedGenotyper de GATK présente deux avantages dans notre cas : il prend en compte l'haploïdie et il permet de faire le variant calling de plusieurs souches en même temps. Ainsi des variants avec une qualité trop inférieure pour un individu peuvent être appelés parce que l'info sur les autres souches est bonne et donc le logiciel en réévalue la qualité. Les variants et informations relatives sont contenues dans un fichier de type .vcf (vcf pour variant call format):

```
java -jar -Xmx4G /usr/local/bioinfo/src/GATK/GenomeAnalysisTK-3.4-
46/GenomeAnalysisTK.jar \
-T UnifiedGenotyper \
-I /Chemin/FullGenomeBWA_156_realigned.bam \
-I /Chemin/FullGenomeBWA_159_realigned.bam \
-I /Chemin/FullGenomeBWA_164_realigned.bam \
-I /Chemin/FullGenomeBWA_171_realigned.bam \
-I /Chemin/FullGenomeBWA_181_realigned.bam \
-I /Chemin/FullGenomeBWA_195_realigned.bam \
-R /Chemin/NewRef.fa \
-o /Chemin/UGParallelQ250.vcf \
-glm BOTH \
-mbq 20 \
-gt_mode DISCOVERY \
-stand_call_conf 250 \
-stand_emit_conf 250 \
-ploidy 1
```

#glm: BOTH permet de détecter à la fois les SNPs et à la fois les InDels

#gt\_mode: détermine la façon de découvrir l'allèle alternative

#mbq: qualité minimum de la base pour appeler les variants minimum base quality to call variants

# stand\_call\_conf and stand\_emit\_conf: phred score pour appeler et accepter la détection du variant# Les variants par souche sont ensuite redistribués :

```
java -jar -Xmx4G /usr/local/bioinfo/src/GATK/GenomeAnalysisTK-3.4-
46/GenomeAnalysisTK.jar \
-T SelectVariants \
-R ${PathRef}/NewRef.fa \
--variant /Chemin/UGParallelNewRef_Q250.vcf \
```

```
-sn [INRA156] \  
-select 'vc.getGenotype("[INRA156]").isHomVar()' \  
--removeUnusedAlternates \  
-o /Chemin/156_raw.vcf
```

# Les variants sont filtrés selon une profondeur de séquençage supérieur ou égal à 5, et un ratio alternative allele/Read Depth > = à 80% :

```
java -jar -Xmx4G /usr/local/bioinfo/src/GATK/GenomeAnalysisTK-3.4-46/GenomeAnalysisTK.jar \  
-T SelectVariants \  
-R ${PathRef}/NewRef.fa \  
--variant ${PathRaw}/Souches/156_raw.vcf \  
-select 'DP >= 5' \  
-select 'vc.getGenotype("[INRA156"]').getAD().1.floatValue() / DP >= 0.80' \  
-o ${PathOut}/INRA156/INRA_156_filtered.vcf
```

#Les fichiers vcf souche-spécifiques sont ré-associés en un seul fichier :

```
java -jar -Xmx4G /usr/local/bioinfo/src/GATK/GenomeAnalysisTK-3.4-46/GenomeAnalysisTK.jar \  
-T CombineVariants \  
-R ${PathRef}/NewRef.fa \  
-o ${PathOut}/GlobalDiversity_filtered.vcf \  
--variant ${PathOut}/INRA156/INRA_156_filtered.vcf \  
--variant ${PathOut}/INRA159/INRA_159_filtered.vcf \  
--variant ${PathOut}/INRA164/INRA_164_filtered.vcf \  
--variant ${PathOut}/INRA171/INRA_171_filtered.vcf \  
--variant ${PathOut}/INRA181/INRA_181_filtered.vcf \  
--variant ${PathOut}/INRA195/INRA_195_filtered.vcf \  
-genotypeMergeOptions UNSORTED
```

#VCFtools est une suite d'outils qui permet d'obtenir des statistiques sur les fichiers vcf.

#Ts-Tv summary donne le détail des transitions vs. les transversions

#SNPdensity permet de compter le nombre de variants (SNPs et InDels) sur un intervalle donné ici: 100kb

#depth: donne la profondeur moyenne de couverture à la position variable

```
vcftools --vcf /Chemin/INRA156.vcf --out /Chemin/156Stats --TsTv-summary  
vcftools --vcf /Chemin/INRA156.vcf --out /Chemin/156Stats --SNPdensity 100000  
vcftools --vcf /Chemin/INRA156.vcf --out /Chemin/156Stats --depth
```

#Ts-Tv summary donne le détail des transitions vs. les transversions

#SNPdensity permet de compter le nombre de variants (SNPs et InDels) sur un intervalle donné ici: 100kb

#depth: donne la profondeur moyenne de couverture à la position variable

## II.3. L'annotation de variants génétiques

#Les variants sont annotés grâce à l'outil SnpEff.

#SnpEff estime l'effet des variants par rapport à leurs positions dans les gènes. L'outil nécessite donc une base de données avec l'annotation des gènes sur le génome. La base de données a été construite à partir du fichier GTF.

*wget*

```
http://sourceforge.net/projects/snpeff/files/snpEff_latest_core.zip  
unzip snpEff_latest_core.zip
```

```
java -Xmx4G -jar /Chemin/snpEff/snpEff.jar Fghm.2 -d -ud 0 -  
onlyProtein -s /Chemin/snpEff_AllVar.html \  
/Chemin/Variants6souches_filtered.vcf > /Chemin/AllVar.ann.vcf
```

# -ud 0 : les variants hors des séquences géniques ne sont pas pris en compte (upstream and downstream)

# - onlyProtein : l'annotation n'est faite que sur les gènes codant pour des protéines.

## II.4. L'analyse des données de RAD-seq et la construction de la carte génétique

L'analyse des données de RAD-seq a été réalisée grâce au logiciel dédié Stacks (Catchen et al. 2011). Stacks marche lui aussi en ligne de commande.

### II.4.1. Démultiplexer les reads

# L'outil `process_radtags` proposé par Stacks permet de démultiplexer et de réattribuer les séquences à chaque génotype à l'aide de leur étiquette ou tag.

; -e indique l'enzyme, dans notre cas PstI :

```
process_radtags -P -i gzfastq -p /Chemin/où/sont/les/reads/bruts/ -o  
/Chemin/où/sont/les/reads/démultiplexés/ -b  
/Chemin/où/sont/les/barcodes/ BarcodesMycsa.txt -e pstI -r -c -q
```

## II.4.2. Aligner les reads sur le génome de référence

#Pour l'alignement des reads, on utilise Bowtie2 plutôt que BWA-mem car le fichier de sortie est compatible avec Stacks (analyse et génotypage). Il faut au préalable créer un index du génome de référence.

```
bowtie2-build /Chemin/NewRef.fa NewRef
```

```
gzip -d /Chemin/où/sont/les/reads/démultiplexés/${P1}.1.fq.gz  
gzip -d /Chemin/où/sont/les/reads/démultiplexés/${P1}.2.fq.gz
```

```
bowtie2 -x /Chemin/NewRef -1  
/Chemin/où/sont/les/reads/démultiplexés/${P1}.1.fq -2  
/Chemin/où/sont/les/reads/démultiplexés/${P1}.2.fq -S  
/Chemin/${P1}.sam
```

```
gzip /Chemin/où/sont/les/reads/démultiplexés/${P1}.1.fq  
gzip /Chemin/où/sont/les/reads/démultiplexés/${P1}.2.fq
```

```
samtools view -b -S /Chemin/${P1}.sam > /Chemin/${P1}.bam
```

```
rm /Chemin/${P1}.sam
```

```
samtools sort /Chemin/${P1}.bam /Chemin/${P1}_sorted
```

```
rm /Chemin/${P1}.bam  
samtools index /Chemin/${P1}_sorted.bam  
samtools flagstat /Chemin/${P1}_sorted.bam  
/Chemin/${P1}_sorted_fgstat.txt
```

```
# P1="nom_de_la_souche"
```

### II.4.3. Génotypage

```
/usr/local/bioinfo/src/Stacks/stacks-1.32/bin/ref_map.pl -e
/usr/local/bioinfo/src/Stacks/stacks-1.32/bin/ -m 3 -A DH -B
FusaGram_radtags -i 1 -b 13 \
-D "CurrentVersion" \
-o /Chemin \
-p /Chemin/RADParents/INRA156dMAT_sorted.bam \
-p /Chemin/RADParents/INRA171_sorted.bam \
-r /Chemin/A1_sorted.bam \
-r /Chemin/A4_sorted.bam \
# .../...
-r /Chemin/A143_sorted.bam \
-r /Chemin/A145_sorted.bam \
-r /Chemin/A146_sorted.bam \
-r /Chemin/A148_sorted.bam \
-r /Chemin/A149_sorted.bam
```

#Paramètres: 40 threads, la profondeur de séquençage au SNP doit être de 3 au minimum, le cross est un backcross, les résultats seront visualisables sur la database « Fusagram » sur l'interface web proposé par le genotoul, sous l'identifiant « 13 ». Le génotypage se fait avec l'outil ref\_map.pl (script perl).

#Une fois sur l'interface du genotoul, il est possible de télécharger le fichier de génotypage.

### II.5. Construction de la carte génétique

```
library(qtl)
library(erer)
```

# le package erer permet d'enregistrer des listes d'éléments (dans notre cas des listes de marqueurs)

```
mapthis=read.cross("csv", "Chemin/", "Input.csv",
map.function=c("kosambi"))
summary(mapthis)
```

```
plotMissing(mapthis)
par(mfrow=c(1,2), las=1)
plot(ntyped(mapthis), ylab="No. typed markers", main="No. genotypes
by individual")
```



```
plot(ntyped(mapthis, "mar"), ylab="No. typed individuals", main="No.  
genotypes by marker")
```

```
# Tri des données de marqueurs : éliminer les marqueurs avec trop de données  
manquantes
```

```
mapthis=subset(mapthis, ind=(ntyped(mapthis)>600))  
nt.bymar=ntyped(mapthis, "mar")  
todrop=names(nt.bymar[nt.bymar < 70])  
write(todrop, "C:/Chemin/ListeMarqueursEcartés")  
mapthis <- drop.markers(mapthis, todrop)  
summary(mapthis)
```

```
#Identification des individus trop ressemblant (90%)
```

```
cg=comparegeno(mapthis)  
hist(cg[lower.tri(cg)], breaks=seq(0, 1, len=101), xlab="No.  
matching genotypes")  
rug(cg[lower.tri(cg)])
```

```
wh=which(cg > 0.9, arr=TRUE)  
wh=wh[wh[,1] < wh[,2],]  
wh  
g=pull.geno(mapthis)
```

```
#Identification des marqueurs dupliqués co-ségrégants. Ils sont sauvegardés pour une  
autre analyse
```

```
print(dup <- findDupMarkers(mapthis, exact.only=TRUE))  
write.list(dup, "C:/Users/belaurent/Desktop/DupExact")  
mapthis <- drop.markers(mapthis, unlist(dup))  
summary(mapthis)
```

```
print(dupF <- findDupMarkers(mapthis, exact.only=FALSE))  
write.list(dup, "C:/Users/belaurent/Desktop/DupNonExact")  
mapthis <- drop.markers(mapthis, unlist(dupF))  
summary(mapthis)
```

```
gt=geno.table(mapthis)  
gt[gt$P.value<0.001,]  
todrop=rownames(gt[gt$P.value < 0.001,])  
write(todrop,  
"C:/Users/belaurent/Desktop/MarqueursBiasedSegregation")  
mapthis=drop.markers(mapthis, todrop)  
summary(mapthis)
```

```
g=pull.geno(mapthis)
```

```
gfreq <- apply(g, 1, function(a) table(factor(a, levels=1:2)))
gfreq <- t(t(gfreq) / colSums(gfreq))
par(mfrow=c(1,2), las=1)
for(i in 1:2)
plot(gfreq[i,], ylab="Genotype frequency", main=c("AA", "BB")[i],
ylim=c(0,1))
summary(mapthis)
```

#Estimation du taux de recombinaison entre paire de marqueurs (max fraction de recombinaison = 0.35) et formation les groupes de liaison, ici avec un LOD de 6.

```
mapthis=est.rf(mapthis)
lg=formLinkageGroups(mapthis, max.rf=0.35, min.lod=6)
table(lg[,2])
mapthis <- formLinkageGroups(mapthis, max.rf=0.35, min.lod=6,
reorgMarkers=TRUE)
```

#Etablissement de l'ordre des marqueurs. Pour faire cette opération gourmande en ressource de calcul, et compte tenu du nombre important de marqueurs à ordonner, un fichier bash pour travailler sur le cluster de calcul a été écrit.

```
library(qtl)
mapthis=read.cross("csv", "/Chemin", "DataToOrder.csv",
map.function=c("kosambi"))
summary(mapthis)
mapthis=est.rf(mapthis)
lg=formLinkageGroups(mapthis, max.rf=0.35, min.lod=6)
table(lg[,2])
lg=formLinkageGroups(mapthis, max.rf=0.35, min.lod=6,
reorgMarkers=TRUE)
```

#Pour calculer les distance, on choisit une erreur de génotypage de 0.001 et la fonction kosambi pour transformer le taux de recombinaison en distance génétique.

```
mapthis=orderMarkers(mapthis, chr=4, use.ripple=TRUE,
error.prob=0.001, map.function=c("kosambi"))
pull.map(mapthis, chr=4)
mapthis=orderMarkers(mapthis, chr=3, use.ripple=TRUE,
error.prob=0.001, map.function=c("kosambi"))
pull.map(mapthis, chr=3)
mapthis=orderMarkers(mapthis, chr=2, use.ripple=TRUE,
error.prob=0.001, map.function=c("kosambi"))
pull.map(mapthis, chr=2)
mapthis=orderMarkers(mapthis, chr=1, use.ripple=TRUE,
error.prob=0.001, map.function=c("kosambi"))
pull.map(mapthis, chr=1)
```

```
summaryMap (mapthis)
write.cross (mapthis, "csv", /Chemin/OutputOrder")
save.image ("/Chemin/OutputOrder")
```

```
library (qt1)
mapthis2=read.cross ("csv", "Chemin", "OutputOrder.csv")
```

```
summaryMap (mapthis2)
```

```
pull.map (mapthis2, chr=1)
mapthis2=flip.order (mapthis2, chr=2)
```

#Recherche des marqueurs qui changent beaucoup la taille de la carte

```
dropone <- droponemarker (mapthis2, error.prob=0.001)
par (mfrow=c (2,1))
plot (dropone, lod=1, ylim=c (-100,0))
plot (dropone, lod=2, ylab="Change in chromosome length")
summary (dropone, lod.column=2)
```

#Les marqueurs identifiés sont éliminés s'ils ne sont pas en bout de chromosomes. Dans notre cas aucun marqueur n'est problématique.

#Les distances génétiques sont recalculées

```
newmap2 <- est.map (mapthis2, error.prob=0.001,
map.function=c ("kosambi"))
mapthis2 <- replace.map (mapthis2, newmap2)
summaryMap (mapthis2)
```

# Identification et suppression des individus où le génotype n'est pas cohérent (double crossover dans un intervalle réduit).

```
mapthis2=calc.errorlod (mapthis2, error.prob=0.001,
map.function=c ("kosambi"), version=c ("new"))
print (toperr <- top.errorlod (mapthis2, cutoff=6))
dev.off ()
plotGeno (mapthis2, chr=1, ind=toperr$id[toperr$chr==1], cutoff=6,
include.xo=FALSE)
plotGeno (mapthis2, chr=2, ind=toperr$id[toperr$chr==2], cutoff=6,
include.xo=FALSE)
plotGeno (mapthis2, chr=3, ind=toperr$id[toperr$chr==3], cutoff=6,
include.xo=FALSE)
plotGeno (mapthis2, chr=4, ind=toperr$id[toperr$chr==4], cutoff=6,
include.xo=FALSE)
```

```
mapthis2.clean <- mapthis2
```

```
for(i in 1:nrow(toperr)) {
  chr <- toperr$chr[i]
  id <- toperr$id[i]
  mar <- toperr$marker[i]
  mapthis2.clean$geno[[chr]]$data[mapthis2$pheno$ID==id, mar] <-
NA
}
summaryMap(mapthis2.clean)
```

#Recalcul des distances génétiques

```
newmap3 <- est.map(mapthis2.clean, error.prob=0.001,
map.function=c("kosambi"))
mapthis2 <- replace.map(mapthis2.clean, newmap3)
summaryMap(mapthis2)
```

```
gt <- geno.table(mapthis2, scanone.output=TRUE)
par(mfrow=c(2,1))
plot(gt, ylab=expression(paste(-log[10], " P-value")))
plot(gt, lod=3.5, ylab="Genotype frequency")
abline(h=c(0.5), lty=2, col="gray")
```

```
summaryMap(mapthis2)
```

#Recherche de marqueurs dupliqués générés suite aux données manquantes rajoutée.

```
print(dup <- findDupMarkers(mapthis2, exact.only=FALSE))
write.list(dup, "C:/Chemin/DuplicatedWithNA")
```

```
mapthis2 <- drop.markers(mapthis2, unlist(dup))
summaryMap(mapthis2)
```

```
mapthis2=est.rf(mapthis2)
newmap=est.map(mapthis2, error.prob=0.001,
map.function=c("kosambi"))
mapthis2=replace.map(mapthis2, newmap)
summaryMap(mapthis2)
```

```
write.cross(mapthis, "csv", /Chemin/LaCarteTouteBelle")
```

## II.6. La détection de QTL

```
##### Travailler sous R #####
```

```
library(qtl)  
detect=read.cross("csv", "/Chemin/Input", "Input.csv",  
map.function=c("kosambi"), error.prob=0.001)
```

#pheno.col permet d'indiquer la colonne du quel phénotype à analyser.

```
plot(detect)  
nind(detect)  
nchr(detect)  
totmar(detect)  
nmar(detect)  
nphe(detect)
```

```
plotMissing(detect)  
plotMap(detect)  
jpeg("/Chemin/HistoPhenoInPlanta.jpeg")  
par(mfrow=c(4,1))  
plotPheno(detect, pheno.col=2)
```

```
dev.off()
```

#Dans le cas du Simple Interval Mapping

```
detect <- calc.genoprob(detect, step=1, error.prob=0.001,  
map.function=c("kosambi"))  
SIMPheno <- scanone(detect, pheno.col=4, method="em")  
summary(SIMPheno, threshold=2)
```

#Dans le cas d'un phénotype discontinue avec 2 modalités : choisir model binary

```
SIMPheno <- scanone(detect, pheno.col=3, model=c("binary"),  
method="em", n.perm=500)  
summary(SIMPheno, alpha=0.001)
```

#L'option plot permet de tracer les courbes de LOD.

```
plot(SIMPheno, chr=1)  
title("Title Name")  
PERMSIMPheno=scanone(detect, pheno.col=10, method="hk", n.perm=1000)  
summary(PERMSIMPheno, alpha=0.001)  
summary(PERMSIMPheno, perms=PERMSIMPheno, alpha=0.01, pvalues=TRUE)
```

###Estimer l'intervalle de confiance associé au QTL, drop=1 signifie que l'on se fixe comme intervalle la position à LODmax -1 et LODmax +1. « expandtomarkers=TRUE » permet d'étendre la position au marqueurs le plus proches (vers l'extérieur de l'intervalle)

```
lodint(SIMPheno, chr=1, drop=1, expandtomarkers=TRUE)
bayesint(SIMtryPheno, chr=1, prob=0.99)
```

###Estimer l'effet du QTL

```
max(SIMPheno)
marDON <- find.marker(detect, chr=1, pos=95)
plotPXG(detect, pheno.col=3, marker=mar)
effectplot(detect, pheno.col=3, mname1=mar)
```

###Faire l'analyse avec tous les phénotypes en même temps

```
SIM_all_pheno=scanone(detect, method="hk", pheno.col=4:18)
summary(SIM_all_pheno, threshold=3, format="tabByCol")
```

###Two dimensionnal QTL analysis: interacting loci ou séparer un gros QTL en plusieurs :

```
detect2D=calc.genoprob(detect, step=2)
SC2 <- scantwo(detect2D, pheno.col=5, method="hk")
plot(SC2)
operm2 <- scantwo(detect2D, pheno.col=5, method="hk", n.perm=5)
```

#Dans le cas du Composite Interval Mapping

```
detect <- calc.genoprob(detect, step=1, error.prob=0.001,
map.function=c("kosambi"))
```

#Calculer le LOD score seuil pour accepter un QTL en faisant 500 permutations

```
CIMPheno=cim(detect, pheno.col=4, n.marcover=2, window=10,
method="hk", imp.method=c("imp"), error.prob=0.001,
map.function="kosambi", n.perm=500)
summary(CIMPheno, alpha=0.001)
```

#Faire la detection, en utilisant une fenêtre de 10cM et 2 covariables

```
CIMPheno=cim(detect, pheno.col=4, n.marcover=2, window=10,
method="hk", imp.method=c("imp"), error.prob=0.001,
map.function="kosambi")
summary(CIMPheno, chr=1)
```

```
lodint(CIMPheno, chr=1, drop=1)  
plot(CIMPheno, chr=1)
```

```
###Estimer l'effet du QTL
```

```
QTLmajeur <- makeqtl(detect, chr=1, pos=c(92.1), what="prob")  
out.fq_Per <- fitqtl(detect, pheno.col=4, qtl=QTLmajeur,  
method="hk")  
summary(out.fq_Per)  
summary(fitqtl(detect, qtl=QTLmajeur, pheno.col=4, method="hk",  
get.ests=TRUE, dropone=FALSE))
```







**Résumé :** Le champignon *Fusarium graminearum* est l'un des principaux agents responsables de la fusariose des épis, une maladie nécrosante des céréales associée à une contamination des grains et des aliments par des mycotoxines. De récentes observations suggèrent une évolution de l'agressivité des populations de ce pathogène, questionnant l'efficacité et la durabilité des moyens de luttés actuels. Mieux anticiper cette évolution nécessite une meilleure caractérisation de la diversité phénotypique et génotypique existante entre souches. Six nouveaux génomes de *F. graminearum* ont été séquencés et ont permis l'identification et la caractérisation de 243 000 variations génétiques. La majorité de ces variants (77%) est concentrée dans des îlots de polymorphisme, représentant 32% du génome et enrichis en probables effecteurs liés à la pathogénicité de *F. graminearum*. La construction d'une population recombinante, et son génotypage avec 1 300 marqueurs moléculaires, ont permis le développement de la première carte génétique à haute-densité de l'espèce. La corrélation entre le taux de recombinaison et le polymorphisme a mis en évidence une organisation « à deux-vitesses » du génome de cette espèce. Finalement, l'intégration de ces données dans une approche de génétique quantitative a permis l'identification d'un locus polymorphe, affectant le gène *FgVeA*, et responsable de 90% de la variation d'agressivité et de la production de mycotoxine observée. Les différents résultats obtenus durant ces travaux font l'objet d'une discussion générale sur le potentiel adaptatif et d'évolution de ce pathogène.

**Mots clés :** Champignons phytopathogène, mycotoxines, interaction hôte-pathogène, séquençage de nouvelle génération, génome à deux-vitesses, potentiel adaptatif, caractères complexes, cartographie de QTL, protéines Velvet.

---

### **Genetic basis and evolutionary potential of the pathogenicity of the fungus *Fusarium graminearum***

**Abstract:** *F. graminearum* is one of the main causal agents of the fusarium head-blight (FHB), a cereal disease leading to head necrosis, in addition to grain and food/feed contamination by stable and toxic metabolites. Recent observations refer to an increase of pathogenicity, questioning efficiency and durability of current management practices. In order to anticipate this evolution, we must bring a deeper characterization of the currently existing diversity. Six new genomes of *F. graminearum* were sequenced, and 243,000 genetic variations have been identified and characterized. Seventy seven percent of the total number of the variants was located within 32% of the genome, delineating highly polymorphic islands. These islands are enriched with probable effectors linked to *Fusarium's* pathogenicity. The construction and the genotyping on 1,300 molecular markers of a recombinant population have enabled the development of the first high-density genetic map of the species. The remarkable correlation between polymorphism and recombination rate highlighted the 'two-speed' genome organization of this pathogen. Finally, the integration of these data through a quantitative genetic approach allowed the discovery of one quantitative trait locus, likely to affect the gene *FgVeA*, and responsible for 90% of the observed variation of aggressiveness and mycotoxin production. These results are discussed in the light of *F. graminearum's* adaptive potential and evolution.

**Keywords:** phytopathogenic fungi, mycotoxins, host-pathogen interaction, next-generation-sequencing, two-speed genome, potential of adaptation, complex traits, QTL mapping, Velvet proteins.

---

Unité de recherche 1264 Mycologie et Sécurité des Aliments (MycSA)

71 Avenue Edouard Bourlaux CS20032 33883 VILLENAVE D'ORNON CEDEX