



Analyse bioinformatique de l'organisation structurale et fonctionnelle des génomes de deux micro-organismes pathogènes, la bactérie opportuniste *Acinetobacter baumannii* et le champignon *Fusarium graminearum*

Tarek Alouane

► To cite this version:

Tarek Alouane. Analyse bioinformatique de l'organisation structurale et fonctionnelle des génomes de deux micro-organismes pathogènes, la bactérie opportuniste *Acinetobacter baumannii* et le champignon *Fusarium graminearum*. Bactériologie. Université Clermont Auvergne; Université Mohammed V (Rabat), 2022. Français. NNT : 2022UCFAC036 . tel-03998274

HAL Id: tel-03998274

<https://theses.hal.science/tel-03998274>

Submitted on 21 Feb 2023

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.

Université Clermont-Auvergne

ÉCOLE DOCTORALE SCIENCES DE LA VIE, SANTÉ,
AGRONOMIE, ENVIRONNEMENT

THÈSE

Présentée pour obtenir le grade de :

DOCTEUR DE L'UNIVERSITÉ CLERMONT AUVERGNE
DOCTEUR DE L'UNIVERSITÉ MOHAMMED V

Spécialité : Bioinformatique

Soutenue publiquement le 19 Juillet 2022 par

Tarek ALOUANE

**Analyse bioinformatique de l'organisation structurale et fonctionnelle des
génomés de deux micro-organismes pathogènes, la bactérie opportuniste
Acinetobacter baumannii et le champignon *Fusarium graminearum***

Membres du jury :

Marc-Henri LEBRUN, Directeur de Recherche CNRS, UMR BIOGER, Thiverval-Grignon.

Diana FERNANDEZ, Directrice de Recherche IRD, UMR PHIM, Montpellier.

Hanae ABDELMOUMEN, Professeur Faculté des Sciences, Rabat.

Laila SBABOU Professeur Faculté des Sciences, Rabat.

Mouna OUADGHIRI, Professeur Faculté de Médecine et de Pharmacie, Rabat.

Azeddine IBRAHIMI, Professeur Faculté de Médecine et de Pharmacie, Rabat.

Directeurs de Thèse :

Ludovic BONHOMME, Professeur Université Clermont-Auvergne, UMR GDEC, Clermont Ferrand.

Thierry LANGIN, Directeur de Recherche CNRS, UMR GDEC, Clermont Ferrand.

**Unité Mixte de Recherche Génétique, Diversité et Écophysiologie des Céréales UMR 1095
INRAE – Université Clermont Auvergne Équipe Maladies Des Céréales « MDC »**

Résumé

Les micro-organismes pathogènes sont responsables, chaque année, de maladies entraînant de graves conséquences sanitaires et économiques, et ce, en dépit des efforts importants déployés pour les contrôler. Parmi ceux-ci, la bactérie *Acinetobacter baumannii*, un pathogène nosocomial hautement résistant aux antibiotiques, conduisant parfois à une impasse thérapeutique, et le champignon *Fusarium graminearum*, responsable de la fusariose de l'épi (FHB), une maladie dévastatrice des céréales et en particulier du blé tendre, provoquant des pertes sévères de rendement et un impact particulier sur la qualité des grains via la production in planta de mycotoxines nocives pour la santé humaine et animale. En dépit leur importance, les questions sur l'ampleur de la variabilité génomique intraspécifique des génomes de ces deux micro-organismes, et la relation structure/fonction restaient encore peu documentées. Dans le cadre de cette Thèse, nous avons développé une approche de génomique comparative intra-espèce, basée sur le séquençage et l'analyse bioinformatique des génomes complets de souches européennes de *F. graminearum*, présentant différents niveaux d'agressivité et de souches marocaines d'*A. baumannii* multi-résistantes. Cette stratégie nous a permis de construire le premier pangénome « ouvert » de l'espèce *F. graminearum*, élargissant (de 47 %) le répertoire de gènes connus dans le génome de référence. En comparant la diversité génétique et phénotypique des souches, nous avons trouvé que plus des deux tiers du total des gènes codant pour les protéines sécrétées, contribuant potentiellement à la pathogénicité, étaient systématiquement présents dans un core sécrétome, sans relation avec les différents niveaux d'agressivité des souches. La surreprésentation des motifs d'adressage au chloroplaste et au noyau des cellules végétales, identifiés dans les effecteurs fongiques, associée à leur localisation préférentielle dans les régions génomiques en évolution rapide de ce pathogène, suggèrent fortement que ces deux organites sont des cibles végétales clefs du processus de manipulation parasitaire. En parallèle, chez *A. baumannii*, nos analyses ont révélé la présence de multiples clones coexistant en milieu hospitalier marocain, avec la prédominance d'une lignée correspondant au clone international II. La diversité phénotypique de résistance aux antibiotiques des souches était cohérente avec la diversité de leurs résistomes reconstitués in silico, dont celui de la souche résistante à l'antibiotique de dernier recours.

Les résultats obtenus au cours de ce travail de Thèse apportent un éclairage nouveau et original (1) sur la relation entre variabilité génomique intraspécifique et déterminisme génétique de la pathogénicité, et (2) sur le potentiel évolutif et adaptatif de ces deux micro-organismes pathogènes.

Mots clés : *Fusarium graminearum* ; *Acinetobacter baumannii*; Pangénome; Génomique comparative ; Diversité génomique ; Diversité phénotypique.

Abstract

Pathogenic micro-organisms are responsible, every year, for diseases with serious health and economic consequences, despite the major efforts made to control them. Among these, the bacterium *Acinetobacter baumannii*, a nosocomial pathogen highly resistant to antibiotics, could lead to a therapeutic impasse, and the fungus *Fusarium graminearum*, responsible for Fusarium head blight (FHB), a devastating disease of cereals and particular soft wheat, causing severe yield losses and a particular impact on grain quality through the in planta production of mycotoxins harmful to human and animal health. Despite their importance, questions about the extent of intraspecific genomic variability of the genomes of these two microorganisms, and the structure/function relationship still remained poorly documented. As part of this thesis, we have developed an intra-species comparative genomics approach, based on the sequencing and bioinformatics analysis of the complete genomes of European strains of *F. graminearum*, presenting different levels of aggressiveness and Moroccan *A. baumannii* strains multi-resistant. This strategy allowed us to construct the first "open" pangenome of the *F. graminearum* species, expanding (by 47%) the repertoire of known genes in the reference genome. By comparing the genetic and phenotypic diversity of strains, we found that more than two-thirds of the total genes coding for secreted proteins, potentially contributing to pathogenicity, were consistently present in a core secretome, unrelated to the different levels of strain aggressiveness. The over-representation of chloroplast and plant cell nucleus targeting motifs identified in fungal effectors, together with their preferential localization in the rapidly evolving genomic regions of this pathogen, strongly suggest that these two organelles are key plant targets of the parasitic manipulation process. In parallel, in *A. baumannii*, our analyzes revealed the presence of multiple clones coexisting in Moroccan hospitals, with the predominance of a line corresponding to the international clone II. The phenotypic diversity of antibiotic resistance of the strains was consistent with the diversity of their resistomes reconstituted in silico, including that of the strain resistant to the antibiotic of last resort. The results obtained during this thesis work shed new and original light (1) on the relationship between intraspecific genomic variability and genetic determinism of pathogenicity, and (2) on the evolutionary and adaptive potential of these two pathogenic microorganisms.

Key words: *Fusarium graminearum*; *Acinetobacter baumannii*; Pangenome; Comparative Genomics; Genomic diversity; Phenotypic diversity.

Dédicace

Je dédie ce travail

A mes chères parents

« La source inépuisée d'amour et de bonheur »

C'est grâce à votre tendresse, votre affection et votre soutien tout au long
de mon parcours, que j'ai pu accomplir ce travail.

Rien au monde ne pourrait égaler l'ampleur de vos sacrifices 

A mon frère et ma sœur

Merci pour vos conseils et le soutien que vous m'apportez au quotidien.

Que ce travail soit l'expression de mon estime pour vous !!

A tous les membres de ma famille !

A mes amies

Merci pour tous les moments heureux passés ensemble,
je vous souhaite plein de bonheur, de santé, de réussite et de chance.

Remerciements

Je souhaiterais adresser mes remerciements les plus sincères à l'ensemble des personnes qui, par leurs connaissances, leurs compétences, leurs qualités humaines ou tout à la fois, ont contribué à l'aboutissement de mes travaux de recherche.

En premier lieu, Je voudrais exprimer mes profonds remerciements à mes directeurs de Thèse :

Dr. Thierry Langin de m'avoir permis de travailler au sein de son équipe "Maladies des Céréales" (MDC). Je lui suis reconnaissant d'avoir apporté son regard d'expert à ce travail, de ses conseils avisés et de ses remarques constructives pour la rédaction de mes travaux et surtout du temps qu'il m'a consacré malgré ses nombreuses responsabilités. Nos longues discussions lorsque nous partageons le trajet en bus ou à pied, qu'elles portent sur la Fusarium ou sur le sport en général, resteront un souvenir impérissable !!

Pr. Ludovic Bonhomme pour avoir suivi de près l'évolution de mes travaux, pour son aide précieuse, sa disponibilité, sa grande patience et pour ses qualités pédagogiques et scientifiques qui ont grandement contribué à la réalisation de ma Thèse. Je lui suis reconnaissant également pour les heures consacrées à relire et corriger mes travaux et pour son soutien indéfectible durant la période très difficile de la rédaction de ce manuscrit. J'ai beaucoup apprécié travailler à vos côtés tant sur le plan scientifique que sur le plan humain. Vous avez des capacités exceptionnelles pour diriger et encadrer des doctorants et nous vous en sommes tous reconnaissants. Je ne vous remercierai jamais assez !

Pr. Azeddine Ibrahimi, de m'avoir offert la possibilité d'effectuer cette thèse sous sa direction. C'est avec lui que j'ai commencé mes premiers pas dans le monde passionnant de la recherche scientifique au sein de son laboratoire de « Biotechnologie Médicale » (MedBiotech). Je le suis très reconnaissant de m'avoir accordé sa confiance « depuis mes travaux de Master 2 » et de m'avoir appris à être plus autonome durant ces travaux de recherche. Sa capacité d'analyse et son enthousiasme m'ont montré que le monde de la recherche pouvait être un univers passionnant. MedBiotech More Than a Lab !

*Mes très vifs remerciements vont aux membres du jury et notamment au **Dr. Marc-Henri Lebrun**, **Dr. Diana Fernandez**, **Pr. Leïla Sbabou** et **Pr. Mouna Oudghiri** pour m'avoir fait l'honneur d'être les rapporteurs de mon travail de Thèse et d'avoir accordé du temps à une lecture attentive de mon manuscrit. J'associe à ces remerciements **Pr. Hanae Abdelmoumen** pour avoir eu la gentillesse d'accepter de présider le jury de cette Thèse.*

*Les résultats présentés dans ce travail de Thèse sont le fruit de nombreuses collaborations. Un grand merci au **Pr. Mostafa Elouennass** de l'hôpital militaire de Rabat, **Dr. Joerg Bormann** et **Pr. Wilhelm Schäfer** de l'Université de Hambourg en Allemagne, **Pr. Michael Freitag** et **Pr. Sandra Loesgen** de l'Université de l'Oregon aux États-Unis, et la **plateforme Gentyane** de l'unité GDEC / INRAE, Clermont Ferrand.*

Il est temps pour moi de remercier les personnes qui ont été présentes dans le quotidien car durant mes études doctorales, certains arrivent, d'autres partent... mais je n'oublierai jamais les personnes qui comptaient pour moi :

*Je tiens à remercier mes collègues du Laboratoire MedBiotech. Merci à **Lamiaa Lahlou**, à l'époque thésard au labo alors que je n'étais qu'un étudiant naïf de Master2. Je la remercie de m'avoir accueilli au bureau et de m'avoir intégré dans l'équipe. Merci à **Meriem Laamarti** pour sa précieuse contribution à la préparation de librairie des génomes bactériens et pour les nuits blanches que nous avons passées par tél, viso-conf pour avancer sur le projet du COVID-19. Parler de cette période épidémique sans évoquer **Souad Kartti**, **Walid Chemao** et **Abdelmunim Essabbar**, c'est comme une PCR sans enzyme Taq, un grand merci à vous. Merci également au **Pr. Tarik Aanniz**, et tous les autres membres du Labo qui sont toujours présents ou partis pour de nouveaux défis : **Kaoutar El Khattabi**, **Younes Allou**, **Yazid Rouahi**, **Najib Lakhili**, **Fatima zahra Gabbouj**, **Oussama Benhrif**, **Mohammed Hakmi**, **Mehdi Bouricha**, **Houda Bendani** et **Nasma Boumajdi**.*

*Je tiens également à remercier tous les membres de l'équipe MDC et plus particulièrement **Francis Fabre** et **Florian Rocher** pour leur aide plus que précieuse dans mes analyses de phénotypage et pour avoir été des colocataires de bureau super agréables. **Florian**, ta motivation, ton énergie et ton envie de faire bouger les choses m'auront toujours impressionné. Merci à **Géraldine Philippe** et*

Florence Cambon pour leur sympathie et leur dévouement envers les stagiaires et les doctorants, je suis très heureux d'avoir fait leur connaissance. Merci **Olivier** pour la bonne humeur que tu communiquais à tout le laboratoire, merci aussi d'être venu me soutenir au Match de Handball, les photos professionnelles que tu as prises resteront toujours comme de très bons souvenirs pour moi. Merci à **Cyrille Saintenac** et à tous les autres membres de l'équipe toujours présents ou partis d'avoir créé cet environnement enrichissant non seulement scientifiquement mais également humainement : **David Lopez, Melissa Battache, Andrea Garavito-Espejo, Aline Ballot, Mélanie Sicard et Pauline Lasserre-Zuber.**

Je voudrais adresser quelques lignes de ces remerciements à tous les stagiaires que j'ai encadré durant ma Thèse dans le cadre de leur projet de fin d'études **Walid Chemao, Mouad Ballaoui, Zineb Kabbaj, Douae El Ghouali, Sarah Fadel** pour les six mois qui ont partagé avec moi en présentiel ou à distance. Encadrer des personnes aussi sympathiques que vous a été très facile et ce fut un réel plaisir de vous transmettre mes connaissances. Je vous souhaite le meilleur pour vos Thèses !

Production scientifique

Articles publiés dans des revues internationales à comité de lecture :

✚ **Alouane T et al.** Comparative Genomics of Eight *Fusarium graminearum* Strains with Contrasting Aggressiveness Reveals an Expanded Open Pangenome and Extended Effector Content Signatures. (2021). Int J Mol Sci. 10 ;22(12) :6257. doi:10.3390/ijms22126257. PMID : 34200775 ; PMCID : PMC8230406.

✚ **Alouane T et al.** First Whole-Genome Sequences of Two Multidrug-Resistant *Acinetobacter baumannii* Strains Isolated from a Moroccan Hospital Floor. Genome Announc. 4 ;5(18) : e00298-17. doi:10.1128/genomeA.00298-17. PMID : 28473391 ; PMCID : PMC5477196.

✚ **Alouane T et al.** Genome Sequence of *Fusarium graminearum* Strain MDC_Fg1, Isolated from Bread Wheat Grown in France. (2018). Microbiol Resour Announc. 15 ;7(19) : e01260-18. doi:10.1128/MRA.01260-18. PMID : 30533795 ; PMCID : PMC6256482.

✚ **Alouane T et al.** Genomic Diversity and Hotspot Mutations in 30,983 SARS-CoV-2 Genomes: Moving Toward a Universal Vaccine for the "Confined Virus"? (2020). Pathogens. 10;9(10):829. doi:10.3390/pathogens9100829. PMID : 33050463 ; PMCID : PMC7600297.

✚ **Alouane T^(co) et al.** Large scale genomic analysis of 3067 SARS-CoV-2 genomes reveals a clonal geo-distribution and a rich genetic variations of hotspots mutations. (2020). PLoS One. 10 ;15(11) : e0240345. doi: 10.1371/journal.pone.0240345. PMID : 33170902 ; PMCID : PMC7654798.

✚ Rocher F, **Alouane T et al.** *Fusarium graminearum* Infection Strategy in Wheat Involves a Highly Conserved Genetic Program That Controls the Expression of a Core Effectome. (2022). Int J Mol Sci. 23(3) :1914. doi:10.3390/ijms23031914. PMID : 35163834 ; PMCID : PMC8836836.

✚ Essabbar A, Kartti S, **Alouane T et al.** IDbSV: An Open-Access Repository for Monitoring SARS-CoV-2 Variations and Evolution. (2021). Front Med. 13 ;8 :765249. doi:10.3389/fmed.2021.765249. PMID : 34966754 ; PMCID : PMC8710592.

✚ Uwingabiye J, Lemnouer A, Roca I, **Alouane T et al.** Clonal diversity and detection of carbapenem resistance encoding genes among multidrug-resistant *Acinetobacter baumannii* isolates recovered from patients and environment in two intensive care units in a Moroccan hospital.

Antimicrob Resist Infect Control. 2017 Sep 26 ;6 :99. doi:10.1186/s13756-017-0262-4. PMID : 28959441 ; PMCID : PMC5615474.

✚ Laamarti M, El Mrimar N, **Alouane T** *et al* . Whole-Genome Sequencing of Two Moroccan Mycobacterium tuberculosis Strains. (2021). Microbiol Resour Announc. 10 (1), e00635-20. DOI: 10.1128/MRA.00635-20 . PMID: 33414280 PMCID: PMC8407682.

✚ Chemaou-Elfihri MW, Manni A, Laamarti M, Kartti S, Essabbar A, **Alouane T** *et al*. Draft Genome Sequence of *Stenotrophomonas maltophilia* MDMC339, Isolated from Soil of Merzouga Desert in Morocco. (2020). Microbiol Resour Announc. 6 ;9(32) : e00634-20. doi:10.1128/MRA.00634-20. PMID : 32763932 ; PMCID : PMC7409849.

✚ Laamarti M, Chemaou-Elfihri MW, Kartti S, Laamarti R, Allam L, Ouadghiri M, Smyej I, Rahoui J, Benrahma H, Diawara I, **Alouane T** *et al*. Genome Sequences of Six SARS-CoV-2 Strains Isolated in Morocco, Obtained Using Oxford Nanopore MinION Technology. (2020). Microbiol Resour Announc. ;9(32) : e00767-20. doi:10.1128/MRA.00767-20. PMID : 32763945 ; PMCID : PMC7409862.

Google Scholar : <https://scholar.google.com/citations?user=hbz4ucMAAAAJ&hl=fr>

Abréviations

AA	Activités auxiliaires
AAC	Aminosides acétyltransférases
ADN	Acide désoxyribonucléique
AED	Distance d'édition d'annotation
AME	Enzymes modifiant les aminoglycosides
ANT	Aminosides nucléotidyltransférase
APH	Aminosides phosphotransférases
ARNnc	Acide ribonucléique non codant
ARNr	Acide ribonucléique ribosomique
ARNt	Acide ribonucléique de transfert
BAM	<i>Binary Alignment Map</i>
BLAST	<i>Basic Local Alignment Search Tool Program</i>
BLSE	Bêtalactamases à spectre élargi
BUSCO	<i>Benchmarking Universal Single-Copy Orthologs</i>
CAZymes	Enzymes de dégradation des carbohydrates
CBM	Modules de liaison aux glucides
CDS	Séquences codantes
CE	Carbohydate estérase
CHDLs	<i>Carbapenem-hydrolysing classe D β-lactamases</i>
CMI	Concentrations minimales inhibitrices
DBJ	Graphe de <i>Bruijn</i>
ESKAPE	<i>Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumannii, Pseudomonas aeruginosa et Enterobacter spp</i>
EST	<i>Expressed Sequence Tag</i>
FGSC	Complexe d'espèces <i>Fusarium graminearum</i>
FHB	Fusariose de l'épi ou <i>Fusarium head blight</i>
Gb	Gégabase
GC	Guanine-Cytosine
GH	Glycoside hydrolase
GO	<i>Gene Ontology</i>
GPI	Glycosylphosphatidylinositol
hpi	Heures post-inoculation

HGT	Transfert horizontal de gènes
IC	Clone internationale
Insertions/Délétions	InDels
IP	Inhibiteur de protéase
Kb	Kilobase
KL	Locus K
LPS	Lipopolysaccharides
MAP kinase	Kinase activée par un mitogène
Mb	Mégabase
MDR	<i>Multidrug-resistant</i> ou Multirésistant
MIPS	<i>Munich Information Services for Protein Sequences</i>
miRNA-Seq	<i>MicroRNA sequencing</i>
MLST	<i>MultiLocus Sequence Typing</i>
NCBI	National Center for Biotechnology Information
NGS	Séquençage de Nouvelle Génération
NLS	Signal de localisation nucléaire
OLC	<i>Overlap-Layout-Consensus</i>
OMS	Organisation mondiale de la santé
PDR	<i>Pandrug-resistant</i>
PE	<i>Paired-end</i>
PHI-base	<i>Pathogen–Host Interactions database</i>
PLP	Protéines liant la pénicilline
QUAST	<i>Quality Assessment Tool for Genome Assemblies</i>
RGI	<i>Resistance Gene Identifies</i>
RI	Îlots de résistance
SAM	<i>Sequence Alignment Map</i>
SE	<i>Single-end</i>
SNPs	Polymorphismes mononucléotidiques
STs	Types de séquences
SVs	Variants Structuraux
Tns	Transposons
UPS	Sécrétion non conventionnelle des protéines
USI	Unités de soins intensifs
XDR	<i>Extensively drug-resistant</i> ou Ultrarésistant

Liste des Figures et Tables

Figure 1 : Quelques dates marquantes de la recherche sur les pangénomes	3
Figure 2 : Construction du pangénome.....	5
Figure 3 : Nature du pangénome « ouverte et fermée »	7
Figure 4. Le cluster Tri et quatre autres gènes - Tri1, Tri16, Tri15 et Tri101 - impliqués dans la biosynthèse des trichothécènes	22
Figure 5. Workflow représentant les grandes étapes de l'étude.....	29
Figure 6 : Prétraitement des données brutes d' <i>A. baumannii</i> et <i>F. graminearum</i> issues de NGS	33
Figure 7 : Pipeline utilisé pour la détection de SNPs et InDels.....	34
Figure 8 : Les outils utilisés pour l'assemblage <i>de novo</i> des <i>reads</i> courts (d'Illumina) et longs (de PacBio)	37
Figure 9 : Pipeline d'annotation adapté au génome de <i>F. graminearum</i>	39
Figure 10 : Pipeline d'annotation adapté au génome d' <i>A. baumannii</i>	41
Figure 11 : Pipeline utilisé pour l'identification du sécrétome putatif.....	45
Figure 12 : Résumé du workflow d'analyse génomique des deux micro-organismes <i>A. baumannii</i> et <i>F. graminearum</i>	48
Tableau 1 : Liste des déterminants de pathogénicité et de virulence de <i>F. graminearum</i>	24
Tableau 2 : Origine, site et date d'isolement de treize et huit nouvelles souches d' <i>A. baumannii</i> et <i>F. graminearum</i> , respectivement, utilisées dans ces travaux Thèses	32

Table des matières

Introduction générale	1
I – De la notion de génome au concept de pangénome : un changement de paradigme	2
1.1. Génome et génomique comparative à l'ère du séquençage à haut débit	2
1.2. Concept de pangénome	3
1.3. La taille du pangénome : “open” or “closed” ?	6
1.4. Pangénome : Au-delà du génome de référence	8
II – La pathogénomique microbienne : chez <i>Acinetobacter baumannii</i> et <i>Fusarium graminearum</i>	9
2.1. <i>A. baumannii</i> : bactérie pathogène opportuniste	10
2.1.1. Structure de la population d' <i>A. baumannii</i>	10
2.1.2. Caractéristiques du génome d' <i>A. baumannii</i>	11
2.1.3. Déterminants de résistances aux antibiotiques	11
2.1.4. Déterminants de virulence	15
2.2. <i>F. graminearum</i> : agent principal de la FHB	17
2.2.1. Caractéristiques du génome de <i>F. graminearum</i>	18
2.2.2. Effectome protéique : éléments clé du pouvoir pathogène	18
2.2.3. Les trichothécènes : facteurs de virulence spécialisés	21
2.2.4. Gènes de virulence et de pathogénicité vérifiés expérimentalement : PHI-base	22
2.2.5. Localisation subcellulaire des effecteurs fongiques dans les cellules végétales	25
Contexte et objectifs de la Thèse	26
Chapitre 1 : Workflow d'analyse génomique utilisé pour les deux micro-organismes	31
I. Origines des souches d' <i>A. baumannii</i> et de <i>F. graminearum</i>	31
II. Analyse génomique des données issues de NGS	33
2.1. Prétraitement des <i>reads</i> issus de NGS	33
2.2. Détection et annotation des SNPs et InDels	34
2.2.1. Alignement des <i>reads</i> sur la séquence de référence	35
2.2.2. Traitement post-alignement	35
2.2.3. Détection et filtrage des SNP et InDels	36
2.2.4. Annotation fonctionnelle des SNP et InDels	36
2.3. Assemblage de novo et annotation des génomes	36
2.3.1. Assemblage de novo : des <i>reads</i> prétraités aux génomes assemblés	37
2.3.2. Évaluation de l'exhaustivité des génomes assemblés	38
2.3.3. Annotation des génomes eucaryotes fongiques : <i>F. graminearum</i>	38

2.3.4. Annotation des génomes bactériens : <i>A. baumannii</i>	40
2.3.5. Évaluation de l'exhaustivité de gènes annotés	41
2.4. Construction du pangénome et analyse phylogénomique	42
2.5. Prédiction du pan-résistome et du pan-virulome d' <i>A. baumannii</i>	43
2.6. Prédiction du pan-sécrétome de <i>F. graminearum</i>	43
2.7. Localisation subcellulaire des protéines sécrétées dans la cellule végétale	47
2.8. Résumé du workflow d'analyse génomique	48
2.9. Défis et orientations futures	49
Chapitre 2 : Résultats – <i>F. graminearum</i>	51
Introduction	52
Article 1 : <i>Genome Sequence of Fusarium graminearum Strain MDC_Fg1, Isolated from Bread Wheat Grown in France</i>	53
Article 2 : <i>Comparative Genomics of Eight Fusarium graminearum Strains with Contrasting Aggressiveness Reveals an Expanded Open Pangenome and Extended Effector Content Signatures</i>	55
Conclusion	79
Chapitre 3 : Résultats – <i>Acinetobacter baumannii</i>	80
Introduction	81
Article 1: <i>First Whole-Genome Sequences of Two Multidrug-Resistant Acinetobacter baumannii Strains Isolated from a Moroccan Hospital Floor</i>	82
Article 2: <i>Comparative genomics of multidrug-resistant Acinetobacter baumannii strains isolated from patients and environment in the intensive care unit of a Moroccan hospital</i>	84
Conclusion	107
Discussion générale et perspectives	108
I – Construction du premier pangénome de <i>Fusarium graminearum</i>	109
II – Réévaluation à la hausse du pangénome d' <i>Acinetobacter baumannii</i>	112
III – Quelles sont les relations entre organisation structurale et organisation fonctionnelle des génomes de ces deux micro-organismes ?	114
Conclusion générale	119
Références bibliographiques	122
Annexe A – Technologies de séquençage actuelles	161
Annexe B – Préparation de librairie pour le séquençage Illumina MiSeq	165
Annexe C – Assemblage et annotation du génome	168

Introduction générale

Introduction générale

I – De la notion de génome au concept de pangénome : un changement de paradigme

1.1. Génome et génomique comparative à l'ère du séquençage à haut débit

Jusqu'au développement des technologies de séquençage de nouvelle génération (NGS), il fallait énormément d'efforts et de temps pour séquencer un seul génome (Fleischmann *et al.*, 1995; Venter *et al.*, 2001) en raison du faible débit du séquençage de première génération. La révolution de la génomique est étroitement liée aux développements des technologies de séquençage (Annexe A) et des outils/algorithmes bioinformatiques nécessaires à l'analyse des séquences. Cependant, ces progrès de séquençage et d'analyse génomique ouvrent à la fois la possibilité d'accéder à des informations précieuses contenues dans le génome et de nouvelles perspectives pour étudier la diversité et l'évolution du génome. Parmi les premières questions que l'on pourrait se poser en essayant de faire le point sur le contenu des génomes serait de savoir quelle est leur taille, leurs niveaux de ploïdie et d'hétérozygotie ? Les premières réflexions ont soutenu que la taille du génome devrait être directement liée au nombre de gènes qu'il contient. Cela a été réfuté par la similitude du nombre de gènes codant pour des protéines dans des génomes de tailles très différentes. Par exemple, le génome de l'homme ou de la souris ($\approx 20,000$ gènes), possède un nombre de gènes codant pour des protéines très similaire à celui de *Caenorhabditis elegans* (un nématode) même si sa taille est environ 30 fois plus petite (Kim *et al.*, 2018). De nos jours, les génomes sont séquencés à une vitesse si rapide que cela devient une routine dans de nombreux domaines. En conséquence, il y a un intérêt croissant pour comprendre la signification des informations stockées et codées dans ces génomes. Cette richesse d'informations a ouvert la porte à la génomique comparative dont l'objectif est d'identifier les éléments génomiques structuraux et fonctionnels conservés dans différentes espèces. Un concept important dans la comparaison de génomes entiers est la synténie, c'est-à-dire la préservation de l'ordre des gènes. En d'autres termes, nous dirons qu'une région "m" du génome A est synténique avec une région "n" du génome B si chaque gène présent dans la région "m" a un orthologue correspondant en "n", et si tous les gènes de "m" apparaissent dans le même ordre que les gènes de la région "n". On dit que les régions "m" et "n" forment un bloc synténique. Lors de la comparaison d'un ou plusieurs génomes, il est important de rechercher cette synténie, car les blocs synténiques représentent généralement des blocs conservés, c'est-à-dire des sections du génome qui n'ont pas subi de réarrangements internes depuis l'ancêtre commun le plus récent des génomes comparés (D. Liu *et al.*, 2018). Cette conservation, à son tour, peut suggérer des contraintes évolutives fonctionnelles qui méritent d'être connues. Les études comparatives *in silico* menées sur les séquences de micro-organismes ont notamment permis de progresser dans de nombreux aspects de la recherche, dont la diversité structurale et fonctionnelle des génomes pathogènes et en particulier

l'identification des déterminants de la pathogénicité (Quainoo *et al.*, 2017). L'objectif fondamental de la génomique comparative reste d'identifier toutes les différences entre les génomes, puis d'élucider quelles différences de séquence sont responsables des changements phénotypiques et des mécanismes génétiques sous-jacents. Néanmoins, certaines des limites de cette approche ont également été démontrées. La capacité de séquençage haut débit a déplacé les verrous méthodologiques de la génération des données de séquence à leur analyse, tels que la vitesse et la complexité de l'analyse, l'évaluation de la qualité des données, ainsi que la visualisation et l'interprétation des résultats pour répondre à des questions biologiques.

1.2. Concept de pangénome

Il y a près de 17 ans, Tettelin *et al.* (2005) ont proposé le concept de pangénome (pan, dérivé du mot grec (παν), signifiant « entier ») pour tenter de décrire et de modéliser la totalité génomique d'un taxon d'intérêt. En effet, la notion de pangénome a été utilisée pour la première fois chez les bactéries en 2005 (Medini *et al.*, 2005; Tettelin *et al.*, 2005). Le séquençage et l'analyse génomique de huit souches bactériennes de *Streptococcus agalactiae* avaient révélé l'existence de 80 % des gènes communs à toutes les souches analysées, les 20 % restant étant absents dans au moins une des souches analysées. Depuis lors, le concept de pangénome s'est étendu à un grand nombre d'organismes, couvrant d'autres règnes (Figure 1).

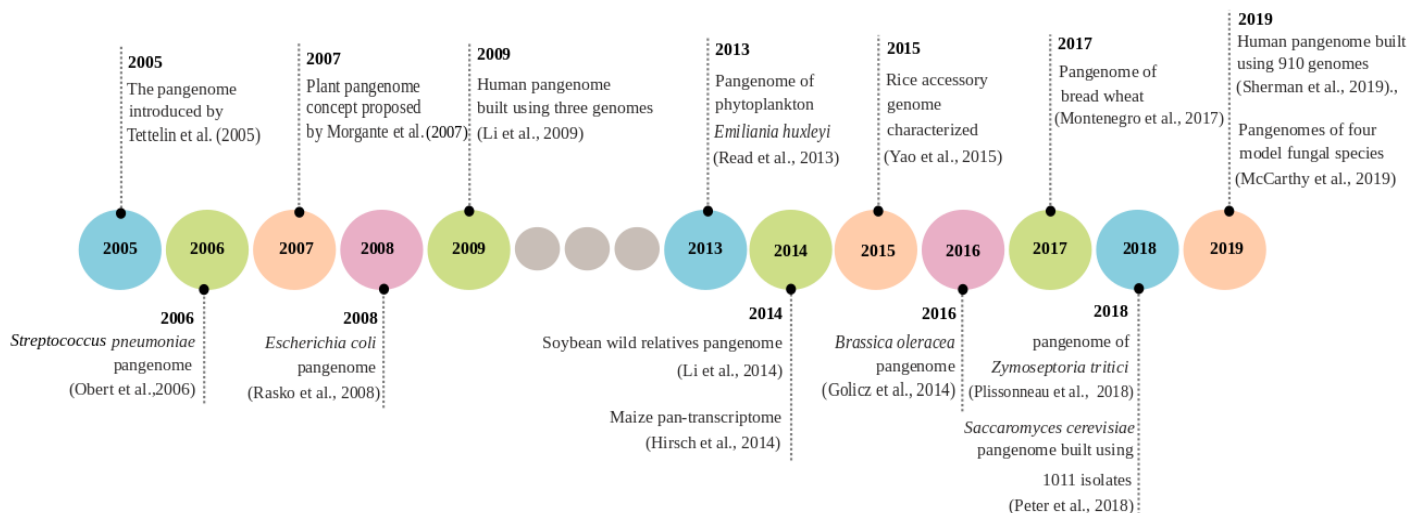


Figure 1 : Quelques dates marquantes de la recherche sur les pangénomes (adapté de Golicz *et al.*, 2020).

Bien que le concept du pangénome soit bien établi en génomique comparative des procaryotes, il n'a été appliqué que récemment aux études comparatives intraspécifiques des génomes eucaryotes, malgré l'observation répétée de la variation de leur contenu génomique. Par exemple, les premiers pangénomes fongiques ont été publiés en 2018, ce sont des phytopathogènes *Zymoseptoria tritici* (Plissonneau *et al.*, 2018), *Parastagonospora spp.* (Syme *et al.*, 2018), et l'espèce modèle *Saccharomyces cerevisiae* (Peter *et al.*, 2018). La plus grande étude fongique à ce jour a porté sur

1011 génomes de *Saccharomyces cerevisiae*, représentant un panel mondial de diversité génétique et phénotypique (Peter *et al.*, 2018 ; Figure 1).

Les définitions et les objectifs du concept du pangénome ont fluctué entre diverses interprétations, notamment (i) le nombre total de gènes non redondants présents dans un ensemble de données, (ii) le répertoire complet des gènes d'une espèce, (iii) le résultat de la comparaison génomique de différents organismes de la même espèce ou du même genre, ou (iv) la représentation basée sur la similitude de l'ensemble total de gènes, qui sont présents dans un groupe d'espèces ou de souches étroitement apparentées (Alcaraz *et al.*, 2010; Guimarães *et al.*, 2015; Tettelin *et Medini*, 2020). Ces multiples définitions mettent en évidence la flexibilité du concept de pangénome et les niveaux de résolution possibles par rapport à la taxonomie (genre, espèce, sous-espèce). En termes simples, le core génome représente l'ensemble des séquences présentes chez tous les membres des taxons d'intérêt, le génome dispensable (également connu sous le nom de génome accessoire) capture un sous-ensemble de séquences partagées par certains membres du groupe, tandis que le pangénome est simplement l'addition du core génome et des génomes dispensables (décrivant la totalité des taxons au niveau des ensembles de données de séquence).

Le core génome pour un ensemble de génomes apparentés est un concept qui implique l'identification de gènes orthologues communs à une espèce ou à un genre. Dans certains cas, la pertinence biologique du core génome tend à diminuer si davantage de génomes sont ajoutés à la comparaison (Barajas *et al.*, 2019). Cependant, il fournit un ensemble de gènes qui sont probablement responsables d'une cohésion évolutive dans le genre. Par exemple, 20 souches englobant 13 espèces du genre *Bacillus* ont été déterminées comme partageant 814 core gènes qui définissent des caractéristiques spécifiques de ce genre telle que la capacité à construire des endospores (Alcaraz *et al.*, 2010). Un autre rapport basé sur les pangénomes de quatre espèces de champignons (*Saccharomyces cerevisiae*, *Candida albicans*, *Cryptococcus neoformans var. grubii* et *Aspergillus fumigatus*) a montré qu'entre 80 et 90 % des gènes de chacune de ces espèces étaient attribués au core génome, dont la plupart de ces gènes conservés sont impliqués dans les processus de survie essentiels chez les quatre espèces pathogènes (McCarthy *et Fitzpatrick*, 2019).

Il a été démontré que le core génome est principalement constitué de gènes associés au maintien des aspects fondamentaux de la biologie de l'organisme : (i) maintien des fonctions de base de l'organisme qui incluent la réplication de l'ADN (pour acide désoxyribonucléique), la traduction et le maintien de l'homéostasie cellulaire, et (ii) participation aux processus cellulaires essentiels (ex. glycolyse) (Gordon *et al.*, 2017; Guimarães *et al.*, 2015). Cependant, certaines études ont proposé que le core génome se compose de deux sous-compartiments, l'un essentiel et l'autre « persistant » (Collins *et al.*, 2012). Le sous-compartiment persistant comprend des gènes qui étaient peut-être nécessaires à un moment donné chez l'organisme, mais qui ont perdu leur nécessité et n'ont pas encore été éliminés par la dérive génétique « *genetic drift* ». La taille du core génome fournit alors une mesure de la

conservation et, en même temps, de la diversité des génomes parmi les organismes étudiés. En effet, les gènes qui codent pour des fonctions extrachromosomiques ou qui sont acquis horizontalement sont généralement sous-représentés dans le core génome et présentent une source de variabilité génique au sein des micro-organismes. Si le core génome devient plus petit à mesure que la diversité augmente parmi les organismes, les génomes phylogénétiquement apparentés ont tendance à partager plus de gènes et ont donc un core génome plus grand.

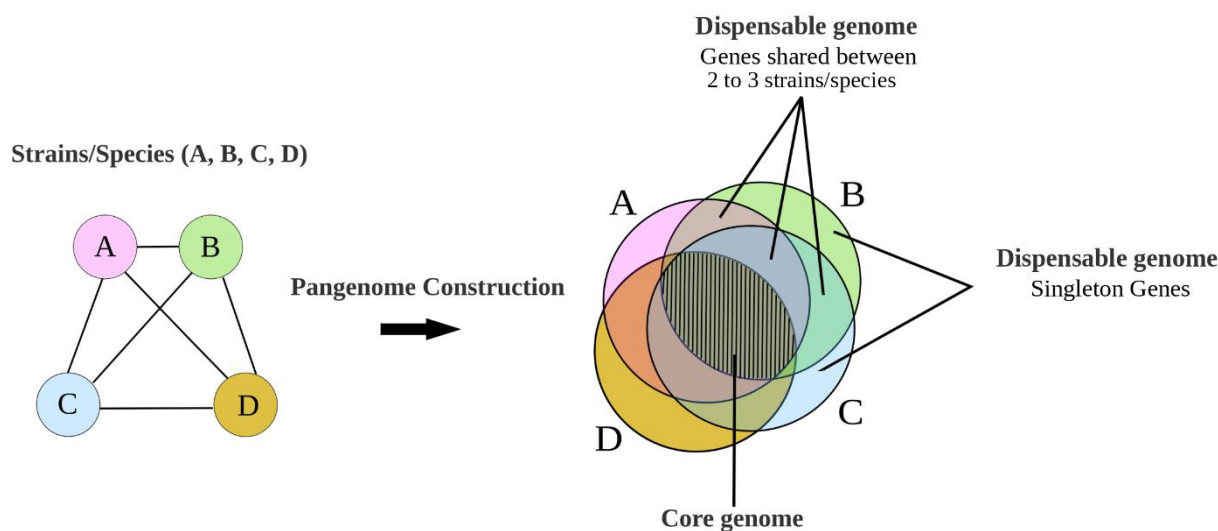


Figure 2 : Construction d'un pangénome (adapté de Golicz *et al.*, 2020)

Le génome dispensable peut en outre être subdivisé en gènes présents chez deux individus ou plus et en gènes singletons (également appelés uniques ou spécifiques) – présents chez un seul individu (Figure 2). Ils sont peu susceptibles d'être impliqués dans les fonctions métaboliques essentielles, mais constituent une base importante pour la variabilité génétique. Le génome dispensable d'une espèce est souvent impliqué dans l'adaptation à une niche spécifique ou dans la manifestation d'un phénotype spécifique, comme l'adaptation à l'hôte ou la pathogénicité (Lee *et al.*, 2021). Dans les études qui incluent des phylums supérieurs, tels que des genres ou des familles, le génome dispensable comprend également des gènes qui sont spécifiques à tout sous-taxon, par exemple des gènes spécifiques à une espèce (Gupta *et al.*, 2015). Un rapport similaire basé sur 13 génomes d'*Haemophilus influenzae* a montré qu'environ 25 % des gènes uniques déterminés sont homologues aux gènes associés aux phages (Hogg *et al.*, 2007). Il est donc devenu largement admis qu'une fraction importante de ces gènes uniques proviennent de transfert horizontal de gènes (HGT). Ce dernier est considéré comme la force motrice principale façonnant les pangénomes bactériens. Il se produit à des taux beaucoup plus faibles chez les eucaryotes et est plus difficile à détecter (Keeling *et Palmer*, 2008; Van Etten *et Bhattacharya*, 2020). Outre le HGT, la variation du contenu génétique est influencée par un continuum de changements allant de polymorphismes mononucléotidiques (SNPs) à des variants structuraux (SVs ; constituées de large délétions, insertions, inversions, duplications, etc.), dont la taille peut varier de quelques kilobases (Kb) à des millions de bases (Mb).

Les gènes dispensables sont généralement soumis à une pression de sélection plus réduite diversifiante, leur permettant d'accumuler des mutations plus fréquemment que les gènes appartenant au core génome, qui sont, quant à eux, soumis à une forte pression de sélection favorisant la conservation de leur séquence (Guimarães *et al.*, 2015). D'autre part, certaines mutations (comme par exemple, un SNP provoquant l'apparition d'un codon stop dans la phase codante) peuvent conduire à la création de pseudogènes (gènes non fonctionnels) qui, au cours de l'évolution, peuvent être éliminés des génomes (Lapierre *et* Gogarten, 2009).

1.3. La taille du pangénome : “open” or “closed” ?

Pour chaque analyse de pangénome, des questions récurrentes se posent : (i) combien de génomes faut-il séquencer pour prendre en compte de façon la plus exhaustive possible la diversité génétique et génomique au sein d'une espèce ? (ii) Existe-t-il toujours le même ensemble de séquences partagées par tous les membres d'un groupe même lorsque de nouveaux individus séquencés sont ajoutés ? (iii) De nouvelles séquences spécifiques seront-elles encore découvertes avec un séquençage individuel supplémentaire ? De façon à répondre à ces différentes questions, Tettelin *et al.* (2005) ont proposé de représenter l'évolution du nombre total de séquences distinctes après l'addition de chaque génome séquencé illustrant la notion du pangénome ouvert “*open pangenome*” ou fermé “*closed pangenome*” (Figure 3).

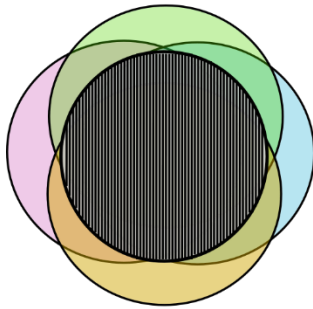
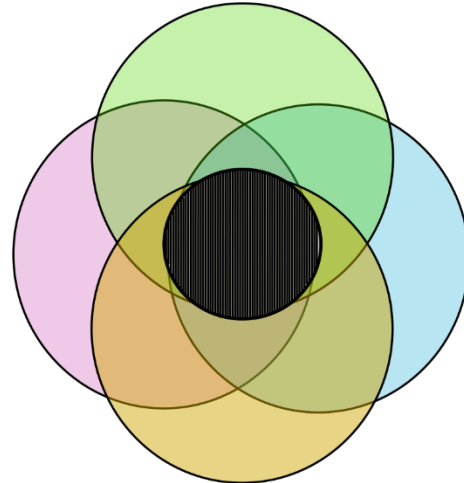
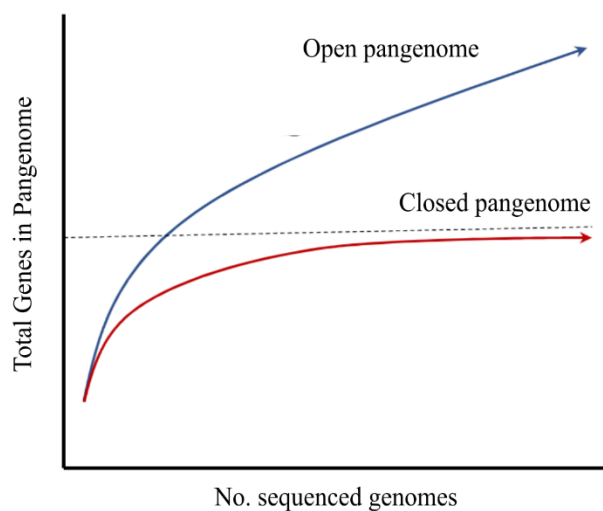
A. Closed Pangenome**B. Open Pangenome****C. Pangenome size**

Figure 3 : Pangénome « ouvert et fermé » (adapté de Tranchant-Dubreuil *et al.*, 2019). **(A)** Le pangénome fermé est caractérisé par un large core génome et un petit génome dispensable. **(B)** Le pangénome ouvert a tendance à avoir un petit core génome et un large génome dispensable. **(C)** La taille du pangénome ouvert a tendance à augmenter avec chaque génome ajouté, tandis que la taille du pangénome fermé a tendance à être constante malgré l'ajout de de génomes supplémentaires.

Dans le premier cas, il semble qu'il n'y ait pas de nombre fini de gènes/protéines dans l'espèce ; chaque nouvelle souche ou individu ajouté accumule de nouveaux gènes dans le pangénome. Un pangénome ouvert est caractéristique des espèces bactériennes qui vivent dans des habitats très différents et nécessitent donc une adaptabilité élevée face à diverses conditions environnementales, ce qui peut se traduire par des niveaux élevés de HGT entre les organismes (Brockhurst *et al.*, 2019). Au contraire, les pangénomes fermés ne changent pas par l'ajout de nouveaux génomes, un nombre minimal de représentants sera suffisant pour décrire la totalité de la variabilité du pangénome. De telles espèces ont un contenu en gènes constant et il est possible de déterminer le répertoire complet

de ces gènes en séquençant un nombre limité de génomes. Un pangénome fermé ne signifie pas nécessairement que toutes les souches présentent le même phénotype, car comme évoqué dans la partie ci-dessus des polymorphismes nucléotidiques peuvent conférer des caractéristiques spécifiques des souches. Par exemple, *Buchnera aphidicola* est l'endosymbiote intracellulaire obligatoire des pucerons. Il habite une niche isolée et limitée qui entraverait la capacité d'acquérir des gènes externes, et de plus, ils ne disposent pas de mécanismes d'échange de gènes (Guimarães *et al.*, 2015). Des analyses du pangénome de cette espèce révèlent un pangénome fermé avec un nombre estimé d'environ 2600 gènes (Guimarães *et al.*, 2015; Snipen *et al.*, 2009) alors qu'une tolérance au stress thermique peut être modifiée par une seule mutation nucléotidique dans une région promotrice (Dunbar *et al.*, 2007). Chez les bactéries intracellulaires libres et facultatives qui habitent plusieurs environnements et soumises à de nombreux stress externes, elles montrent une grande capacité à acquérir de nouveaux gènes ((Guimarães *et al.*, 2015). Par conséquent, le mode de vie est l'un des paramètres dictant fortement la typologie du pangénome. Les bactéries au mode de vie sympatrique (en contact avec d'autres organismes) ont tendance à avoir des pangénomes ouverts, avec une proportion beaucoup plus faible de core gènes, tandis que les espèces bactériennes allopatriques (vivant isolées ; y compris les bactéries intracellulaires obligatoires) ont tendance à avoir des pangénomes fermés, avec un plus petit nombre de gènes dispensables (Rouli *et al.*, 2015).

1.4. Pangénome : Au-delà du génome de référence

Les analyses génomiques et autres approches -omiques nécessitent souvent un génome de référence ; un modèle de séquences nucléotidiques pour un individu « typique » de chaque espèce. Ce génome est utilisé comme étalon de comparaison dans la recherche fondamentale, la recherche clinique, etc. La création de chaque référence était, et est toujours, un objectif majeur de la communauté génomique. Différents travaux montrent que le génome de référence dérivé d'une seule souche est inadéquat à de nombreuses fins. Une étude de Sherman *et al.* (2019) a consisté à séquencer un ensemble de génomes de 910 individus humains d'ascendance africaine dans le but d'assembler un pangénome humain. Les auteurs ont estimé que jusqu'à 10 % des séquences génomiques totales sont absentes de la référence. Le fait d'avoir un pangénome humain ou le pangénome d'un sous-ensemble de la population humaine, permet la découverte de gènes absents du génome de référence et qui peuvent être associés à des phénotypes spécifiques (Sherman *et Salzberg*, 2020). Un rapport similaire basé sur le riz a souligné que 41,6 % des SNPs associés à certains traits n'étaient pas présents dans la séquence du génome de référence (Yao *et al.*, 2015). De plus, sur la base de 18 cultivars de blé, Montenegro *et al.* (2017) ont révélé l'absence de 12,150 gènes dans la séquence de référence *Chinese Spring* mais présents dans tous les autres cultivars analysés. L'utilisation du pangénome de référence dans les analyses qui nécessite des séquences de référence est donc importante pour améliorer la précision de l'analyse d'association de traits.

II – La pathogénomique microbienne : chez *Acinetobacter baumannii* et *Fusarium graminearum*

L'analyse systématique des séquences génomiques de microbes pathogènes, connue sous le terme de « pathogénomique », est une approche qui cherche à identifier les déterminants potentiels de virulence, de pathogénicité, et de résistance (aux antibiotiques) à l'échelle de génomes entiers et leurs contributions à la maladie ou au changement phénotypique (Murugan *et al.*, 2016; Pallen *et Wren*, 2007; Rampersad, 2020). Définir la virulence n'est peut-être pas simple, car elle s'appréhende assez naturellement, et sa définition varie selon les domaines scientifiques, voire dans un même domaine. Tant en pathologie animale qu'en biologie évolutive, la virulence est généralement comprise comme étant liée aux dommages que l'infection parasitaire cause à l'hôte et, plus précisément, a été définie comme l'effet quantitatif de l'infection parasitaire sur l'aptitude de l'hôte (Read, 1994). Par exemple, la capacité des bactéries pathogènes à provoquer une maladie chez un hôte sensible est déterminée par de multiples facteurs de virulence agissant individuellement ou ensemble à différents stades de l'infection. Elle est définie comme la capacité d'une bactérie à pénétrer, se multiplier et persister dans un site hôte (Peterson, 1996).

En phytopathologie, la définition de la virulence proposée par Plank *et Van Der Plank*, (1968) persiste encore aujourd'hui. Il définit la virulence comme la capacité d'un agent pathogène à infecter un hôte spécifique, et est donc un caractère qualitatif. Cette définition correspond à la définition du terme « pathogénicité » : une variable discontinue, c'est-à-dire s'il y a ou non pathogénicité (Sacristán *et García-Arenal*, 2008). Ici, les questions qui se sont posées autour de l'interaction hôte-pathogène et les réflexions qui ont suivi pour tenter de répondre aux questions biologiques, nous ont amenés à décrire la virulence comme une variable continue, c'est-à-dire qu'elle est définie par l'ampleur de la maladie et sa variable quantitative est appelée agressivité (Pariaud *et al.*, 2009; Sacristán *et García-Arenal*, 2008).

Différents travaux ont révélé que les déterminants de la virulence ou de pathogénicité sont expliqués par l'architecture unique du génome (Fernandes *et al.*, 2011; Guimarães *et al.*, 2015; Soanes *et al.*, 2008). Les comparaisons entre différents génomes des micro-organismes pathogènes conduisent souvent à une identification plus rapide des mécanismes distincts sous-jacents à la pathogénicité et la virulence. Parmi les micro-organismes pathogènes, la bactérie *Acinetobacter baumannii* (*A. baumannii*) et le champignon *Fusarium graminearum* (*F. graminearum*) représentent un défi sanitaire et économique important. Le premier (*A. baumannii*), est un pathogène opportuniste responsable d'infections nosocomiales sévères, causant de réelles difficultés thérapeutiques du fait de sa capacité à développer plusieurs mécanismes de résistances aux antibiotiques. Il a été reconnu comme un agent de pneumonie, de septicémie, de méningite, d'infections des voies urinaires et de plaies, et est associé à une morbidité et une mortalité élevée (Eliopoulos *et al.*, 2008; Zhou *et al.*, 2019). *F. graminearum* est un pathogène opportuniste des céréales où il provoque de sévères pertes de rendement et un impact particulier sur la qualité des grains via leur contamination par des

mycotoxines incompatibles avec l'alimentation humaine et animale (Goswami *et al.*, 2004; Karlsson *et al.*, 2021; Zain, 2011). Ce pathogène est la principale cause de la fusariose de l'épi ou *Fusarium head blight* (FHB) du blé (*Triticum aestivum*).

2.1. *A. baumannii* : bactérie pathogène opportuniste

Acinetobacter appartient à la famille des *Moraxellaceae* dans la classe des *Gammaproteobacteria* (Bergogne-Bérézin *et al.*, 1996). Des espèces d'*Acinetobacter* ont été récupérées dans pratiquement tous les échantillons de sol et d'eau, elles peuvent faire partie de la flore normale et certaines ont des rôles d'agents pathogènes humains (Al Atrouni *et al.*, 2016; Baumann, 1968; Weber *et al.*, 2016). L'espèce *A. baumannii* est à l'origine de la majorité des infections cliniques causées par les espèces d'*Acinetobacter* et est un agent pathogène nosocomial important qui provoque des infections opportunistes (Joly-Guillou, 2005). Ces infections surviennent fréquemment chez les patients immunodéprimés ou dans les unités de soins intensifs (USI). *A. baumannii*, fait partie des organismes ESKAPE (*Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *A. baumannii*, *Pseudomonas aeruginosa* et *Enterobacter spp.*), qui représentent une menace globale pour la santé humaine et un défi thérapeutique du fait de l'émergence de la constante augmentation de la résistance (Santajit *et al.*, 2016). Le 27 février 2017, l'organisation mondiale de la santé (OMS) a publié une liste de bactéries contre lesquelles il est urgent d'avoir de nouveaux antibiotiques (WHO). Parmi les douze familles de bactéries qui devraient recevoir la plus haute priorité, *A. baumannii* était en tête de cette liste, constituant une grande menace pour la santé humaine. L'importance clinique de cet agent pathogène est largement due à une capacité remarquable à acquérir des déterminants de résistance avec un large éventail d'enzymes d'hydrolyse d'antibiotiques, des changements de pompe à efflux, une imperméabilité et des mutations d'antibiotiques cibles, ce qui en fait l'un des organismes multirésistants menaçant l'antibiothérapie actuelle.

2.1.1. Structure de la population d'*A. baumannii*

Généralement, le terme « clone » est souvent utilisé pour décrire les isolats épidémiologiquement indépendants mais phénotypiquement et génotypiquement liés (Karah *et al.*, 2012; Orskov *et al.*, 1983). Des enquêtes épidémiologiques moléculaires ont révélé la présence de plusieurs lignées clonales génétiquement distinctes parmi les populations d'*A. baumannii* (Diancourt *et al.*, 2010; Zarrilli *et al.*, 2013). Trois de ces lignées, ont été initialement définies comme des clones européens I à III et plus tard nommés comme des clones internationaux (IC ; également appelés clones globaux) en raison de leur propagation dans d'autres continents. En effet, IC-I et IC-II se sont avérés être les plus largement distribués dans plusieurs pays sur tous les continents (Hamidian *et al.*, 2019; Martins *et al.*, 2016; Matsui *et al.*, 2018; Skerniškytė *et al.*, 2019). IC-III est également considéré comme un clone majeur d'*A. baumannii* en raison de sa détection dans des pays européens,

aux États-Unis, en Amérique latine, au Liban et en Irak (Fedrigo *et al.*, 2022; Giannouli *et al.*, 2013; Huang *et al.*, 2011; Levy-Blitchtein *et al.*, 2018; Rafei *et al.*, 2015). En plus de ces trois lignées clonales majeures, d'autres clones ont été identifiés dans plusieurs régions du monde. Une étude précédente examinant la clonalité de 496 isolats d'*A. baumannii* a montré que 283 souches appartenaient aux 26 clones, dont 18 étaient répartis sur plus d'un continent (Karah *et al.*, 2012). Cela montre le succès de certains clones à persister et à se répandre dans différentes zones géographiques. Malgré l'existence de diverses techniques de typage, telles que le polymorphisme de longueur de fragments amplifiés (AFLP), la plupart des clones décrits ci-dessus ont été déterminés via la méthode MLST (pour *MultiLocus Sequence Typing*) ; une méthode de référence pour fournir une classification universelle des agents pathogènes microbiens tels que *A. baumannii*. Dans cette méthode, une combinaison d'allèles de sept locus ou gènes de ménage, générant différents types de séquences (STs) qui sont comparés à ceux contenus dans une base de données internationale (Bartual *et al.*, 2005; Diancourt *et al.*, 2010). La base de données MLST (<http://pubmlst.org/abaumannii>; consulté le 22/02/2022) contient actuellement plus de 2000 STs d'*A. baumannii*.

2.1.2. Caractéristiques du génome d'*A. baumannii*

Le premier génome d'*A. baumannii* a été séquencé en 2006 (Fournier *et al.*, 2006), et à l'avènement de ce projet, il y avait 17 génomes complets, dont 9 appartenaient aux IC-II. La taille du génome d'*A. baumannii* varie de 3,7 Mb à 4,3 Mb, avec un pourcentage de Guanine-Cytosine (GC) d'environ 39%. La densité des régions codant pour les protéines est relativement homogène avec 87% du génome totale (Adams *et al.*, 2008; Hamidian *et al.*, 2020). Les tentatives pour définir le core génome d'*A. baumannii* variaient en fonction du nombre des souches analysées, mais les preuves globales suggèrent qu'il se situe entre 1455 et 2688 gènes (L. C. S. Antunes *et al.*, 2014; Hassan *et al.*, 2016). Cela représente environ 60% du contenu moyen du génome de la souche et code principalement pour des protéines impliquées dans les processus métaboliques et cellulaires. Alors que, le pangénome, est d'une taille impressionnante avec plus de 7000 gènes (Hassan *et al.*, 2016) et continuera de croître à mesure que davantage de séquences seront ajoutés. La plasticité génomique présentée par *A. baumannii* est un facteur important dans la capacité de cette bactérie à s'adapter à une large gamme d'hôte (Gallagher *et al.*, 2015).

2.1.3. Déterminants de résistances aux antibiotiques

Les souches résistantes d'*A. baumannii* peuvent être classés en trois catégories (Magiorakos *et al.*, 2012) : (i) MDR (multirésistant ou *multidrug-resistant*), définie comme une résistante acquise à au moins un agent dans trois catégories d'antimicrobiens ou plus, la (ii) XDR (ultrarésistants ou *extensively drug-resistant*) a été définie comme une résistance à au moins un agent dans toutes les catégories d'antimicrobiens sauf deux ou moins (c'est-à-dire que les souches bactériennes restent

sensibles à seulement une ou deux catégories) et la PDR (*pandrug-resistant*) a été définie comme une résistance à tous les agents de toutes les catégories d'antimicrobiens (Magiorakos *et al.*, 2012). *A. baumannii* a acquis, au fil des années, de multiples mécanismes de résistance à un large éventail d'antibiotiques.

2.1.3.1. Déterminants de résistance aux β -lactamines

La résistance aux β -lactamines est principalement médiée par une dégradation enzymatique avec des β -lactamases. Dans la classification d'Ambler (Hall *et Barlow*, 2005), les β -lactamases sont regroupées en quatre classes A, B, C et D en fonction des motifs de séquence et des différences de mécanisme hydrolytique.

Les enzymes appartenant à la classe A se caractérisent par leur capacité à hydrolyser l'amide cyclique lié à la molécule de β -lactame, sont principalement identifiées chez les entérobactéries ainsi que chez *Pseudomonas aeruginosa*, mais dans une bien moindre mesure chez *A. baumannii* (Hakemi Vala *et al.*, 2014). PER-1 a été la première enzyme des β -lactamases à spectre large (BLSE) décrite chez *A. baumannii* (Sechi *et al.*, 2004). Ainsi, d'autres carbapénémases appartenant à cette classe comme KPC-2 et GES-14, ont également été identifiées dans des souches d'*A. Baumannii* (Bonnin *et al.*, 2011). Les β -lactamases de **classe B** (dénommées ainsi métallo- β -lactamases) possèdent un ou deux métaux de Zinc dans leur site actif (Hakemi Vala *et al.*, 2014). Parmi les différentes enzymes appartenant à cette classe « *IMP*, *VIM*, *NDM* et *SIM* ». Par exemple, le gène *NDM* a été largement disséminé chez les Enterobacteries mais aussi chez *A. baumannii*. Actuellement, 8 variants de *NDM* ont été rapportés (NDM-1 à NDM-8) dont deux (NDM-1 et NDM-2) décrits chez *A. baumannii* (Kaase *et al.*, 2011; Villacís *et al.*, 2019). Pour les β -lactamases de **classe C**, *A. baumannii* possède naturellement un gène codant pour une céphalosporinase *AmpC*. La présence des séquences d'insertion (ISs) telle que l'ISAbal en amont du gène *AmpC* augmente son expression, ce qui confère la résistance aux BLSE (Lin *et Lan*, 2014). Plusieurs variants alléliques de ces enzymes ont été précédemment identifiés chez *A. baumannii* et sont ainsi nommés céphalosporinase dérivée d'*Acinetobacter* (ADC). Enfin, les β -lactamases de **classe D** connues sous le nom d'oxallinases (OXA), sont les carbapénémases les plus répandues, avec plus de 400 enzymes de type OXA identifiées (Antunes *et al.*, 2014). Certaines enzymes de cette classe possèdent une activité carbapénémase d'où le nom CHDLs (pour *carbapenem-hydrolysing classe D β -lactamases*). Les CHDLs sont les carbapénémases les plus répandues chez *Acinetobacter spp*, dont OXA-51, OXA-23, OXA-24, et OXA-58 (Coelho *et al.*, 2006; Y. Zhao *et al.*, 2019). Le premier (OXA-51) est un groupe de β -lactamases intrinsèques de type OXA identifié avec plusieurs variants alléliques. Il a été initialement détecté en 1996 chez *A. baumannii* récupéré en Argentine (Brown *et al.*, 2005), ensuite le gène est devenu un marqueur important utilisé pour l'identification de l'organisme au niveau de l'espèce (Howard *et al.*, 2012). La présence de ce gène n'entraîne pas de résistance aux carbapénèmes

à moins d'être associée à un IS en amont (ISAbal) qui agirait comme un puissant promoteur (Evans et Amyes, 2014). Alors que les autres groupes, dont OXA-23, OXA-24, et OXA-58 sont considérés comme des déterminants de résistances acquises et leur présence est suffisante pour conférer une résistance aux carbapénèmes (Leungtongkam et al., 2018).

2.1.3.2. Déterminants de résistance aux fluoroquinolones

L'émergence d'une résistance aux fluoroquinolones (ou quinolones) chez *A. baumannii* résulte de mutations des enzymes cibles des fluoroquinolones, l'ADN gyrase et l'ADN topoisomérase IV, respectivement codées par les gènes *gyrA* et *parC* (Hamouda et Amyes, 2004). Les substitutions courantes d'acides aminés étant Ser83 et Gly81 dans *gyrA*, et Ser80 et Glu84 dans *parC*. Une résistance cliniquement significative aux fluoroquinolones peut être obtenue avec une seule mutation dans *gyrA*. Cependant, les mutations doubles de *gyrA* et de *parC* produisent une résistance accrue par rapport aux mutations uniques (Nouri et al., 2016). Ainsi, il n'y a pas eu de preuve de résistance avec des mutations dans *parC* sans mutation concomitante dans *gyrA*, ce qui suggère que l'ADN topoisomérase IV pourrait être une cible complémentaire pour les fluoroquinolones. Les gènes de résistance aux quinolones à médiation plasmidique comme *qnrA*, *qnrB* et *qnrS*, sont également retrouvés chez des souches d'*A. baumannii* (Moosavian et al., 2020; Yang et al., 2016).

2.1.3.3. Déterminants de résistance aux aminosides

Les aminosides ou aminoglycosides sont de puissants antibiotiques à large spectre qui se lient au site A de l'ARNr (pour *acide ribonucléique ribosomique*) 16S de la sous-unité 30S, entraînant une perturbation de la liaison protéique chez les bactéries aérobies à Gram négatif (Krause et al., 2016; Viehman et al., 2014). Les aminosides sont en retour inactivés par les enzymes modifiant les aminoglycosides (AME). En effet, trois groupes fonctionnels différents d'AME sont connus, qui comprennent (i) les aminosides acétyltransférases (AAC), par exemple *AAC (6')-Ih* (qui confère également une résistance à la gentamicine et à l'amikacine), (ii) les aminosides phosphotransférases (APH), pour exemple *APH (3')-IA* (qui confère une résistance à la gentamicine), et (iii) les aminosides nucléotidyltransférase (ANT), par exemple *ANT (2'')-IA* (Atasoy et al., 2015; Tahbaz et al., 2019). Un autre mécanisme de résistance aux aminosides est la production de gènes de méthylase d'ARNr 16S ; *armA*, *rmtA*, *rmtB*, *rmtC* et *rmtD*, induisant une résistance de haut niveau dans tous les aminosides cliniquement utiles (Jouybari et al., 2021).

2.1.3.4. Déterminants de résistance à la colistine

En tant qu'antibiotique "ancien", la colistine était à l'origine utilisée pour traiter les infections bactériennes cliniques. Mais il a été remplacé par d'autres médicaments en raison d'une néphrotoxicité et d'une neurotoxicité sévères dans les années 1970 (Falagas et Kasiakou, 2006). La colistine est réapparue comme dernière ligne pour le traitement des infections par des bacilles Gram négatifs

multirésistants, en particulier les espèces hautement résistantes comme *A. baumannii*, *Klebsiella pneumoniae*, et *Pseudomonas aeruginosa* (Choi et Ko, 2014; J. Li et al., 2006). L'acquisition de résistances est principalement liée à des modifications des voies de biosynthèse des lipopolysaccharides (LPS) ; un composant majeur de la membrane externe des bactéries Gram négatif. Chez *A. baumannii*, deux mécanismes majeurs de résistance à la colistine ont été rapportés : (i) le premier implique la modification du lipide A, un composant la membrane externe lipopolysaccharide qui est médiée par l'opéron *pmrCAB* (Adams et al., 2009; Arroyo et al., 2011). Le système à deux composants *pmrAB* régule l'expression de *pmrC*, qui est responsable de l'ajout de phosphoéthanolamine (pEtN) au lipide A du LPS (par la pEtN transférase). Cela réduit la charge négative nette de la membrane externe, affectant ainsi la liaison de la colistine et empêchant la perte d'intégrité et la perturbation de la membrane cellulaire. Des mutations dans *pmrAB* augmentaient l'expression de *pmrC*, qui est associée à une augmentation des concentrations minimales inhibitrices de colistine (CMI) (Gerson et al., 2019). De plus, il a été suggéré que l'insertion de la séquence *ISAbal* en amont du gène affectent l'expression de l'*eptA* un homologue de *pmrC*, augmentant ainsi les CMI de la colistine (Gerson et al., 2019). (ii) la seconde est la perte complète du lipide A. Cette perte est la conséquence de mutations au sein d'un des gènes impliqués dans la biosynthèse du lipide A : *lpxA*, *lpxC* et *lpxD* (Moffatt et al., 2010). En plus de ces deux mécanismes, des délétions dans *lpsB*, un gène codant pour une glycosyltransférase impliquée dans la synthèse de la biotine, ont été associées à la résistance à la colistine (Hood et al., 2013). Ces mécanismes sont à l'origine chromosomique, ce qui limite sa distribution et sa dissémination rapides. En revanche, les déterminants de la résistance à la colistine à médiation plasmidique (gènes *mcr*), ont été identifiés chez *Escherichia coli* d'origine animale, humaine et environnementale de Chine en 2016 (Liu et al., 2016). Cependant, ces gènes a été signalés chez une souche d'*A. baumannii* d'origine nosocomiale (*mcr-1* ; Hameed et al., 2019) et animale (*mcr-4.3* ; Ma et al., 2019).

2.1.3.5. Autres déterminants de résistance

Les pompes à efflux intrinsèques ou acquises sont d'autres déterminants importants qui confèrent une résistance à plusieurs classes d'antibiotiques. Ils exportent activement de nombreux agents antimicrobiens de la cellule bactérienne (Blanco et al., 2016). Le système d'efflux AdeABC appartenant à la superfamille RND (résistance-nodulation-division) présente dans la plupart des souches d'*A. baumannii* (Abdi et al., 2020; Yoon et al., 2013). La surexpression d'*AdeABC* est associée à un niveau accru de résistance à l'imipénème et au méropénème, en particulier lorsqu'elle est associée à l'expression d'enzymes de type CHDL (Rumbo et al., 2013). Ainsi, AdeIJK est une pompe spécifique d'*A. baumannii* qui contribue à la résistance intrinsèque aux β -lactamines et autres antibiotiques (Damier-Piolle et al., 2008). Outre les pompes à efflux, *A. baumannii* est intrinsèquement résistant

aux macrolides, mais des gènes de résistance acquis conférant une résistance à cette classe d'antibiotiques ont été détectés. Ceux-ci incluent les gènes *msrE* et *mphE* (Blackwell *et al.*, 2017).

2.1.3.6. Transfert horizontal de gènes

Les éléments génétiques mobiles tels que les plasmides, les transposons (Tns), les cassettes/intégrons et les îlots génomiques (IG) abritent des gènes de résistance aux antibiotiques (Partridge *et al.*, 2018). Ensemble, ces éléments jouent un rôle central dans la facilitation des échanges génétiques horizontaux et favorisent donc l'acquisition et la propagation des gènes de résistance. Par exemple, les IG sont de grands éléments génétiques (plus de 10 kb) qui contribuent à la fitness et à l'adaptation des bactéries à diverses conditions environnementales (Juhas *et al.*, 2009). Ils peuvent également être porteurs de gènes de résistance aux antibiotiques (îlots de résistance ; RI) et contribuer à l'évolution rapide de bactéries pathogènes. L'analyse du séquençage du génome montre que presque tous les RI chez *A. baumannii* sont un composite de plusieurs éléments génétiques mobiles tels que les Tns et les Iss (Cerezales *et al.*, 2020; Leal *et al.*, 2020). Au cours d'une étude comparative du génome de la souche MDR *A. baumannii* AYE et de la souche sensible SDF, un groupe d'environ 45 déterminants de la résistance aux antimicrobiens a été trouvé en tant que RI de 86 kb sur le chromosome de la souche MDR (Fournier *et al.*, 2006). Cet îlot étant le premier îlot identifié chez *A. baumannii*, il a été nommé *A. baumannii* îlot de résistance 1 (AbaR1). À ce jour, divers îlots de type AbaR ont été signalés chez de nombreuses souches d'*A. baumannii* et ils ont été numérotés de AbaR0 à AbaR27 (Godeux *et al.*, 2020; Post *et al.*, 2010; Šeputienė *et al.*, 2012).

2.1.4. Déterminants de virulence

Bien que des analyses génomiques récentes aient identifié plusieurs déterminants de virulence qui contribuent à la pathogénicité, relativement peu de déterminants ont été identifiés chez *A. baumannii* par rapport à ceux d'autres agents pathogènes à Gram négatif (Aliramezani *et al.*, 2019; Wongsuk *et al.*, 2022). Parmi eux, les porines, les phospholipases, les polysaccharides capsulaires, les lipopolysaccharides, et d'autres déterminants impliqués dans la formation de biofilms.

2.1.4.1. Protéines de la membrane externe

Les porines sont des protéines de la membrane externe associées à la modulation de la perméabilité cellulaire. OmpA est l'une des porines la plus abondante de la membrane externe, caractérisée avec multiples fonctions, dont l'adhésion aux cellules épithéliales humaines (Choi *et al.*, 2005; McConnell *et al.*, 2013). Cette protéine est libérée par les vésicules de la membrane externe « OMV » d' *A. baumannii* et cible les mitochondries dans les cellules hôtes, entraînant la fragmentation mitochondriale et l'apoptose par la libération de la molécule pro-apoptotique cytochrome C (Choi *et al.*, 2005; Tiku *et al.*, 2021). D'autres porines, telles que la protéine de membrane externe CarO a été signalé chez *A. baumannii* (Mussi *et al.*, 2007). Une étude précédente

a montré que les souches d'*A. baumannii* avec un gène *carO* tronqué présentait une adhérence plus faible aux cellules épithéliales pulmonaires humaines, ainsi qu'une moindre capacité à les envahir, suggèrent que *CarO* est impliqué dans l'adhésion et l'invasion d' *A. baumannii* des cellules épithéliales pulmonaires humaines (Labrador-Herrera *et al.*, 2020).

2.1.4.2. Le LPS et le polysaccharide capsulaire

Au-delà des protéines de la membrane externe, l'enveloppe d' *A. baumannii* est associée à de nombreux déterminants qui contribuent à la pathogénicité. Parmi ceux-ci, les polysaccharides capsulaires et les LPS. En effet, le LPS, jouant un rôle dans l'intégrité et la stabilité de la membrane (Friedrich *et* Whitfield, 2005). De plus, comme évoqué dans la « section 2.1.3.4 », la fraction lipidique A de LPS est la cible de la dernière ligne de traitement. De nombreuses études ont montré que des modifications ou la perte de lipide A de LPS pouvaient conduire à une résistance à la colistine en empêchant sa liaison (Lee *et al.*, 2017). Alors que le polysaccharide capsulaire englobe l'ensemble de la bactérie et la protège de toute menace extérieure, et constitue ainsi l'un des facteurs contribuant à la survie d'*A. baumannii*. C'est un facteur important où les souches dépourvues du polysaccharide capsulaire se sont avérées avirulentes (Roy *et al.*, 2015). Notamment, de nombreux isolats de patients atteints d'infections à *A. baumannii* expriment des polysaccharides capsulaires de surface et contiennent un groupe de gènes chromosomiques connu sous le nom de locus K (KL) ; une région hautement conservée suivie d'une région très variable et codante pour des protéines impliquées dans la biosynthèse des précurseurs de sucre et des transférases nécessaires aux différents types capsulaires. Des isolats d'*A. baumannii* ont été identifiés, avec plus de 100 types de capsules distincts, incorporant une grande variété de sucres (Wyres *et al.*, 2020). Il a été proposé que les polysaccharides capsulaires soient une cible thérapeutique basée sur les anticorps (Wang-Lin *et al.*, 2017).

2.1.4.3. Phospholipases

Les phospholipases sont des enzymes hydrolytiques possèdent une activité lipolytique contre les phospholipides des membranes cellulaires humaines afin de faciliter la lyse cellulaire et l'invasion bactérienne. Trois classes de phospholipases, telles que la phospholipase A « PLA », la phospholipase C « PLC » et la phospholipase D « PLD » ont été définies en fonction du site de clivage (Lee *et al.*, 2017). La PLC et la PLD ont été identifiées chez *A. baumannii* (Fiester *et al.*, 2016; Jacobs *et al.*, 2010). L'inactivation de PLC entraîne une réduction de l'effet cytotoxique d' *A. baumannii* sur les cellules épithéliales par rapport à celui de la souche parentale (Fiester *et al.*, 2016). Ainsi, la perturbation de PLD entraîne une résistance réduite au sérum humain, une capacité réduite à envahir les cellules épithéliales et une virulence réduite dans un modèle murin de pneumonie (Jacobs *et al.*, 2010).

2.1.4.4. Formation de biofilm

A. baumannii a une capacité remarquable à survivre et prospérer en milieu hospitalier en raison de ses interactions avec une large gamme d'hôte abiotiques. En outre, cette bactérie est capable de coloniser les surfaces biotiques telles que les cellules épithéliales. L'adhérence d'*A. baumannii* aux surfaces biotiques et abiotiques aboutit à la formation des biofilms (Aliramezani *et al.*, 2019). Par définition, les biofilms sont des micro-communautés de microbes enfermés dans une substance extracellulaire et rendant les microbes résistants aux stress. Par exemple, *A. baumannii* se caractérise par la production de protéines associées au biofilm (BapAb) similaires à la protéine Bap de *Staphylococcus aureus*, et qui ont la capacité de s'agréger pour constituer une matrice de biofilm en réponse à des conditions environnementales stressantes (Goh *et al.*, 2013). En plus de Bap, les csu pili sont essentiels à la formation de biofilm sur les surfaces abiotiques et jouent un rôle dans la motilité bactérienne (Vijayakumar *et al.*, 2016). La motilité de contraction, qui est utilisée par *A. baumannii* pour la locomotion, repose sur les pili de type IV ; les gènes codant pour ce type de pili sont régulés à la hausse au cours de la croissance bactérienne dans le sérum humain, indiquant son implication possible dans la bactériémie (Liu *et al.*, 2020).

2.2. *F. graminearum* : agent principal de la FHB

Le genre *Fusarium* regroupe un grand nombre d'espèces phytopathogènes responsables de maladies importantes chez un grand nombre de plantes cultivées importantes agronomiquement au niveau mondial dont la fusariose des épis (FHB) du blé (*Triticum aestivum*), de l'orge (*Hordeum vulgare*) et du maïs (*Zea mays*) (Bottalico *et Perrone*, 2002; Goswami *et Kistler*, 2004; McMullen *et al.*, 1997). La FHB est causée par un complexe de plus d'une vingtaine d'espèces appartenant pour l'essentiel au genre *Fusarium*. L'agent principal en Europe est l'espèce *F. graminearum*. Cette maladie est responsable de pertes importantes de rendement et de qualité dans les cultures de blé auxquelles s'ajoutent des conséquences délétères sur la santé humaine et animale par l'accumulation de mycotoxines (Goswami *et Kistler*, 2004; X. Xu *et Nicholson*, 2009). En effet, la présence de mycotoxines dans les denrées alimentaires humaines et les aliments pour animaux peut provoquer des maladies inflammatoires chroniques et des symptômes émetiques graves (Hägglom *et Nordkvist*, 2015; Janik *et al.*, 2020; Pierron *et al.*, 2016). *F. graminearum* dispose d'un mode de vie hémibiotrophe (Ding *et al.*, 2011), combinant les caractéristiques des phases biotrophes et nécrotrophes et présentant souvent un arsenal moléculaire complexe pour pénétrer et parasiter les cellules végétales.

2.2.1. Caractéristiques du génome de *F. graminearum*

F. graminearum, en raison de son impact sanitaire et économique mondial, fut la première espèce de *Fusarium* et le troisième champignon filamenteux pathogène dont le génome a été séquencé (Cuomo *et al.*, 2007). La souche nord-américaine PH-1 de *F. graminearum* (NRRL 31084) a été choisie pour générer le génome de référence par l'institut BROAD. La taille du génome était de 36,1 Mb pour 11,640 gènes prédits (Cuomo *et al.*, 2007). Par la suite, l'annotation du génome de PH-1 a été améliorée et les modèles de gènes ont été affinés (MIPS ; *Munich Information Services for Protein Sequences*), ce qui a permis de décrire un ensemble de 13,718 gènes codant pour des protéines (Wong *et al.*, 2011). Les séquences d'ADN qui ne codent pas pour les protéines ont été associées à des traits importants contrôlant le mode de vie, l'adaptabilité et l'évolution de nombreux organismes (Raffaele *et Kamoun*, 2012; Rech *et al.*, 2014), ce qui a suscité l'intérêt d'assembler entièrement le génome, au-delà des régions codant pour des protéines. Le re-séquençage complet de la souche PH-1 avec une couverture de 85 fois a résulté en un génome de 38 Mb répartis sur quatre "scaffolds" attribués aux quatre chromosomes attendus avec deux "gaps" restants et une teneur de GC de 48% (King *et al.*, 2015). Ce génome de référence contient 14,164 gènes codant pour des protéines (King *et al.*, 2015). En 2017, d'autres mises à jour ont été effectuées permettant de combler les deux "gaps" restant et de finaliser une annotation contenant 14,145 gènes (King *et al.*, 2017a). L'ensemble du génome PH-1 s'est avéré riche en diversité génétique mesurée par la densité de SNP dans les régions proximales des télomères. Ces régions présentaient également un niveau élevé de recombinaison (King *et al.*, 2017a; Laurent *et al.*, 2017). De plus, *F. graminearum* est connu pour avoir un génome « à deux vitesses » avec des tailles et des nombres de gènes comparables : le sous-génome « rapide » et « lent » (Q. Wang *et al.*, 2017). Le sous-génome rapide a une teneur en GC inférieure, une longueur de gène plus courte et une variation plus élevée du nombre d'exons que le sous-génome lent.

2.2.2. Effectome protéique : éléments clé du pouvoir pathogène

L'une des conditions requises pour que les agents pathogènes des plantes établissent une infection et une colonisation réussies chez l'hôte, est leur capacité à ne pas activer ou à surmonter les réponses de défense des plantes. Pour y parvenir, les bactéries, les champignons, les oomycètes et les nématodes sont capables de sécréter un ensemble de protéines extracellulaires qui constituent l'exoprotéome, aussi communément appelé son sécrétome (Brown *et al.*, 2012). Les sécrétomes jouent un rôle crucial dans les interactions plantes-pathogènes, incluant notamment des « effecteurs de pathogénie » représentant un ensemble de protéines capables d'interférer avec les réponses de l'hôte (Alfaro *et al.*, 2014; Gupta *et al.*, 2015; Guyon *et al.*, 2014).

2.2.2.1. Effecteurs protéiques fongiques

Identifier les fonctions des effecteurs protéiques de différents groupes d'espèces d'agents pathogènes permet d'élucider les processus qui sous-tendent la colonisation et la pathogénicité de l'hôte (Selin *et al.*, 2016). Généralement, les effecteurs protéiques sont définis comme, mais sans s'y limiter, de petites protéines sécrétées (≤ 300 acides aminés) riches en résidus de cystéine (stabilisant la structure tertiaire de la protéine en formant des ponts disulfures ; (Lo Presti *et al.*, 2015; Zhang *et al.*, 2016). Bien que le terme générique d'effecteur soit largement utilisé en conjonction avec différentes espèces de pathogènes fongiques, différents groupes d'espèces disposent d'un éventail diversifié de molécules permettant de mettre en place des stratégies d'infection distinctes, et d'acquérir des nutriments et compléter leur cycle de vie. Les agents pathogènes biotrophes ne peuvent acquérir leur subsistance que dans des tissus végétaux vivants. La colonisation des tissus végétaux par des agents pathogènes nécrotrophes s'accompagne de la mort des tissus colonisés, l'agent pathogène se nourrissant de la dégradation des tissus et des cellules végétales. Les agents pathogènes hémibiotrophes ont un cycle infectieux dans lequel une phase nécrotrophe va succéder à une phase biotrophe plus ou moins longue. Ce type d'agent pathogène ont d'abord besoin d'effecteurs pour supprimer les défenses des plantes et a ensuite besoin d'autres effecteurs qui tuent les cellules végétales.

La caractérisation des effecteurs basée sur le génome a fortement soutenu l'étude des déterminants pathogènes (Brown *et al.*, 2012; Syme *et al.*, 2018). Bien que le catalogue des effecteurs caractérisés chez *F. graminearum* reste limité, des efforts substantiels utilisant des approches génomiques ont fourni un certain nombre de nouveaux candidats. Basé sur le génome de référence PH-1, Brown *et al.* (2012) ont établi un sécrétome putatif de 574 protéines partageant des caractéristiques structurales de protéines sécrétées. Cela pourrait indiquer la présence d'un grand nombre d'effecteurs codés dans le génome de cette espèce qui peuvent potentiellement être délivrés dans différents compartiments cellulaires de la plante. Au cours du processus infectieux, les effecteurs peuvent être sécrétés dans l'espace extracellulaire des tissus hôtes (effecteurs apoplastiques ou extracellulaires) ou internalisés dans les cellules hôtes (effecteurs cytoplasmiques ou intracellulaires) (Toruño *et al.*, 2016). Les effecteurs extracellulaires sont sécrétés dans l'espace apoplastique où ils agissent en tant que facteur de virulence. Contrairement au système de sécrétion de type III bien connu chez les bactéries (Hueck, 1998), les structures impliquées dans la sécrétion des effecteurs dans le milieu apoplastique restent peu connues chez les champignons phytopathogènes filamenteux. Les effecteurs intracellulaires quant à eux sont sécrétés de la même façon que les effecteurs apoplastiques mais les modalités et sites d'entrée de ces protéines dans le cytoplasme de la cellule végétale font toujours débat (Kale *et al.*, 2010; Petre *et Kamoun*, 2014; Wawra *et al.*, 2012). Ces effecteurs intracellulaires se lient

sélectivement à un organite ou une protéine cible, régulent ensuite sa fonction biologique et donc peuvent perturber des processus physiologiques fondamentaux, les réponses immunitaires des plantes et ainsi faciliter la colonisation de l'hôte.

2.2.2.2. Rôle des effecteurs dans les interactions hôte-pathogène

a) Dégradation de la paroi cellulaire

En tant que barrière protégeant la cellule, la paroi cellulaire représente le premier obstacle que va devoir franchir un agent phytopathogène. Les micro-organismes ont ainsi développé un arsenal d'enzymes qui jouent un rôle essentiel dans la pénétration et la colonisation de la plante en participant à la dégradation de la paroi qui entoure les cellules végétales (Sinha *et al.*, 2013). Les effecteurs apoplastiques de type enzymes de dégradation des carbohydrates (CAZymes) sont produits pour la dégradation des matériaux polysaccharidiques de la paroi cellulaire. Avec cette dégradation, *F. graminearum* utilise les polysaccharides de la paroi cellulaire végétale comme source de nutriments lors de la progression de l'infection dans le tissu sous-jacent plus profond (Rauwane *et al.*, 2020). La diversité de ces enzymes est liée à la complexité structurale et à la dynamique de la paroi cellulaire, mais aussi au mode de vie et à l'adaptation de l'hôte par le pathogène (Levasseur *et al.*, 2013). Actuellement, les CAZymes présentes sous forme de différents modules déposés dans la base de données CAZy (www.cazy.org) sont classées par fonction : carbohydrolase (CE), activités auxiliaires (AA), modules de liaison aux glucides (CBM), glycoside hydrolase (GH), glycosyltransférase (GT), et polysaccharide lyase (PL). Des analyses du génome et du transcriptome des champignons phytopathogènes indiquent que différents profils de CAZymes sont produits au cours des différentes étapes de la transition entre les modes de vie biotrophes et nécrotrophes (McDowell, 2013). La production enzymatique de CAZymes s'est avérée être corrélée avec le degré de pathogénicité et la dégradation de la paroi cellulaire des plantes hôtes infectées (Ramzi *et al.*, 2019; Z. Zhao *et al.*, 2013). Compte tenu de l'importance de ces enzymes dans l'absorption des nutriments provenant de l'hôte végétal, elles sont considérées également comme un des principaux déterminants de la pathogénicité parmi les phytopathogènes. Par exemple, les phytopathogènes doivent développer des hyphes à travers les tissus et les cellules de l'hôte pour établir des maladies. King *et al.* (2017b) ont ainsi montré qu'un mutant de *Zymoseptoria tritici* en GT de type 2 « ZtGT2 » était dans l'incapacité d'étendre les hyphes de surface ce qui a entraîné une perte complète de virulence. Les CAZymes sont généralement sécrétées et peuvent être considérées comme favorisant la virulence car elles permettent une entrée réussie de l'agent pathogène. Ils entrent techniquement dans la définition d'effecteur (cf. « section 2.2.2 ») en raison de leurs capacités à modifier la structure des cellules de l'hôte.

b) Inhibition directe des protéines de défense de l'hôte

Un autre groupe d'effecteurs pertinents dans le processus phytopathogène sont les acteurs de la protéolyse, appelés protéases. Il a été démontré que les phytopathogènes sécrètent des inhibiteurs de protéase (IP) pour inhiber les protéases de défense des plantes et favoriser le développement de la maladie (Sabotič *et al.*, 2012). Ces effecteurs ciblent différents compartiments de l'hôte (Tian *et al.*, 2009) et sont classés soit par le type de protéase qu'ils inhibent, soit par leur mécanisme d'inhibition (Rawlings *et al.*, 2016). Un de ces effecteurs, Avr2 (inhibiteur de la cystéine protéase), sécrété par *Cladosporium fulvum* lors de l'infection, est nécessaire à la virulence de ce champignon sur la tomate (Rooney *et al.*, 2005; van Esse *et al.*, 2008). Avr2 inhibe des protéases apoplastiques de son hôte (Rcr3 et Pip1) pour soutenir la croissance de ce pathogène dans l'apoplaste. En outre, les protéases sécrétées jouent un rôle important dans la signalisation, l'approvisionnement nutritionnel et la dégradation des tissus de l'hôte. Avec les CAZymes, les protéases facilitent la pénétration et la dissémination efficace au sein de l'hôte en participant à la dégradation des barrières physiques de l'hôte (Figaj *et al.*, 2019).

2.2.3. Les trichothécènes : facteurs de virulence spécialisés

Les mycotoxines produites par le genre *Fusarium* représentent le groupe le plus diversifié de mycotoxines : plus de 140 métabolites identifiés à ce jour, dont les plus connus sont les trichothécènes, la zéaralénone, le buténolide, la fusaoctaxin A et les fusarines (Desjardins *et al.*, 2007; Nesic *et al.*, 2014; Sobrova *et al.*, 2010). Les trichothécènes sont les principales mycotoxines synthétisées par les membres du complexe d'espèces *F. graminearum* (FGSC). Sur la base de leur structure, les trichothécènes produits par *F. graminearum* peuvent être classés en deux groupes, type A et type B. Ceux de type A comprennent principalement les toxines T-2, NX-2 et le néosolaniol (Varga *et al.*, 2015), alors que les trichothécènes de type B comprennent le nivalénol (NIV), le déoxynivalénol (DON) et leurs dérivés acétylés le 3-acétyldéoxynivalénol (3-ADON) et le 15-acétyldéoxynivalénol (15-ADON) (Krnjaja *et al.*, 2018). La production de ces dérivés du DON (15-ADON ou 3-ADON) est principalement liée aux variations alléliques du gène TRI8 (Alexander *et al.*, 2011). Jusqu'à présent, 16 gènes ont été caractérisés dans la voie de biosynthèse du DON chez *F. graminearum* (Amarasinghe *et al.*, 2016). Ces gènes résident dans quatre loci différents sur différents chromosomes ; le core cluster TRI est composé de 12 gènes localisés sur le chromosome 2, les loci TRI1-TRI16 sur le chromosome 1, TRI101 sur le chromosome 4 et TRI15 sur le chromosome 3 (Figure 4).

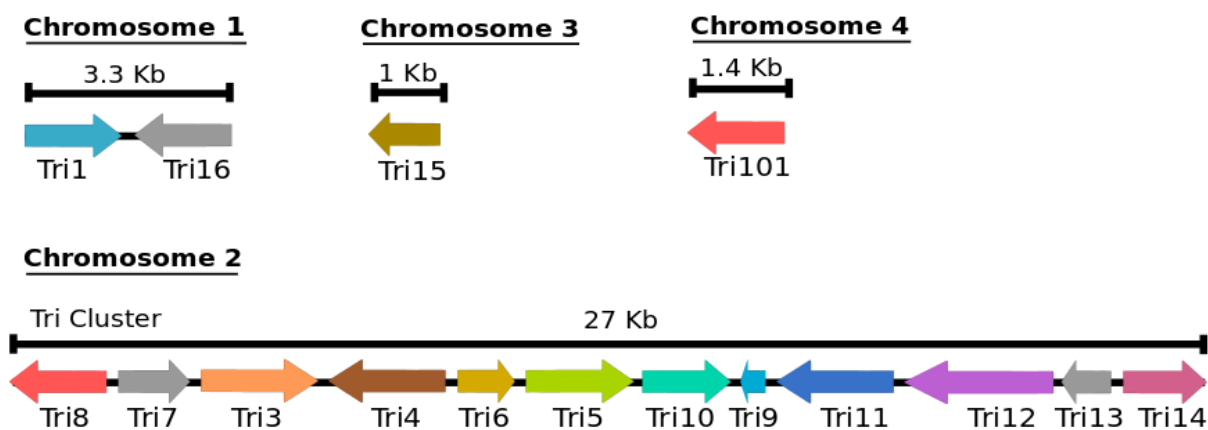


Figure 4. Le cluster Tri et quatre autres gènes – Tri1, Tri16, Tri15 et Tri101 – impliqués dans la biosynthèse des trichothécènes.

Parmi les différents gènes TRI, le gène TRI5 (trichodiène synthase) a reçu plus d'attention jusqu'à présent car il représente un bon indicateur de la synthèse de DON (Flynn *et al.*, 2019; Tretiakova *et al.*, 2022). Un mutant pour ce gène ne produit pas de DON mais reste capable de provoquer des lésions discrètes sur les épillets sans se déployer au rachis (Cuzick *et al.*, 2008; Jansen *et al.*, 2005). En plus de *tri5*, des mutants de délétion de *tri6* et *tri10* (codant deux protéines régulatrices du cluster *Tri*) ont été générés afin de comprendre la régulation de la synthèse de ces mycotoxines (Seong *et al.*, 2009). Ce travail a permis de montrer que le mutant *tri6* régule l'expression de plus de 200 gènes, dont des gènes impliqués aux fonctions des processus cellulaires essentiels, du métabolite secondaire et de la pathogenèse. Par conséquent, ce gène (*Tri6*) est un régulateur global, plutôt qu'un régulateur spécifique de la voie de biosynthèse des trichothécènes. Par ailleurs, il a été démontré que le gène *Tri12* (protéine de transport) était impliqué non seulement dans l'exportation de la toxine à l'extérieur de la cellule mais également à la tolérance à l'inhibition de la croissance dans des conditions induisant la biosynthèse des trichothécènes (Menke *et al.*, 2012). Varga *et al.* (2015) ont identifié plusieurs souches de *F. graminearum* d'Amérique du Nord qui ne produisaient aucune des trichothécènes connues sur le blé malgré l'apparition de symptômes de la maladie. Alors que d'autres études indiquent une forte corrélation entre l'agressivité de l'isolat et sa capacité à produire ces mycotoxines (Laurent *et al.*, 2017). Par conséquent, les trichothécènes seraient un facteur de virulence spécialisé plutôt qu'un facteur de pathogénicité.

2.2.4. Gènes de virulence et de pathogénicité vérifiés expérimentalement : PHI-base

Bien que de nombreux effecteurs protéiques dérivés de champignons soient impliqués dans les interactions hôte-pathogène, il est difficile d'établir la preuve directe de leur association (par perturbation ou suppression de gène) avec la pathogénicité ou la virulence à cause de la présence de gènes ayant des fonctions similaires. En effet, la redondance fonctionnelle fait référence à deux gènes

ou plus qui remplissent la même fonction et pour lesquels l'inactivation de l'un de ces gènes aurait peu ou pas d'effet sur le phénotype (Pérez-Pérez *et al.*, 2013; Tan *et al.*, 2015). En outre, du fait que des combinaisons d'effecteurs soient généralement employées dans le processus d'infection, le rôle d'un effecteur seul n'aurait pas nécessairement un effet fort sur l'infection. L'utilisation d'une liste de gènes dont il a été prouvé qu'ils affectent le résultat des interactions hôte-pathogène *via* des méthodes expérimentales, pourrait être un bon moyen de sélectionner des gènes candidats parmi les centaines de protéines sécrétées codées dans le génome de *F. graminearum* et d'autres phytopathogènes. PHI-base (*Pathogen-Host Interactions database* ; Urban *et al.*, 2020) est une base de données biologiques qui contient des informations sur des gènes, dont on a démontré l'implication dans le contrôle et le développement de l'interaction hôte-pathogène. Les gènes de pathogénicité ou de virulence dans PHI-base sont classés selon les phénotypes mutants, incluant ces catégories : (i) perte de pathogénicité, (ii) pathogénicité non affectée, (iii) virulence accrue et (iv) virulence réduite. Chez *F. graminearum*, PHI-base (consulté le 27/02/2022), donne des détails sur plus de 610 et 41 gènes associés au phénotype mutant « virulence réduite » et « perte de pathogénicité », respectivement, qui ont été validés expérimentalement. Parmi les voies et mécanismes responsables de la perte de pathogénicité, une proportion élevée comprend, soit la suppression, soit la perturbation des gènes responsables du mécanisme de transduction du signal, y compris les trois cascades de signalisation de la protéine kinase activée par un mitogène (MAP kinase). Ceci est principalement dû au fait que ces MAP kinases sont impliquées dans la régulation de plusieurs fonctions cellulaires. Par exemple, le GPMK1/MAP1 régule l'induction précoce des activités endoglucanases extracellulaires, xylanolytiques et protéolytiques (Jenczmionka *et Schäfer*, 2005). Le champignon utilise ces enzymes pour faciliter la dégradation de la paroi cellulaire de la plante et ainsi accéder aux sources de nutrition. Par conséquent, la perturbation du gène *Gpmk1* a fortement compromis la pathogénicité de *F. graminearum* (Jenczmionka *et Schäfer*, 2005 ; Ren *et al.*, 2022).

Tableau 1. Liste des déterminants de pathogénicité et de virulence identifiés chez *F. graminearum* (modifié à partir Rauwane *et al.*, 2020).

Catégorie/Classification		Fonction	Au moins une référence mentionnée
Enzymes dégradant la paroi cellulaire			
Pectinase	Enzyme dégradative	Décompose la pectine	Jenczmionka <i>et</i> Schäfer, 2005
Cellulase	Enzyme dégradative	Décompose la cellulose	Jenczmionka <i>et</i> Schäfer, 2005
Extracellular endoglucanase	Enzyme dégradative	Décompose le glucane	Jenczmionka <i>et</i> Schäfer, 2005
Proteolytic enzyme	Enzyme dégradative	Décompose les protéines	Jenczmionka <i>et</i> Schäfer, 2005
Xylanolytic enzyme	Enzyme dégradative	Décompose le xylane	Jenczmionka <i>et</i> Schäfer, 2005
Lipolytic enzyme	Enzyme dégradative	Décompose les lipides	Jenczmionka <i>et</i> Schäfer, 2005
Toxines			
Trichothecene NX-2	Trichothécène de type A (toxine)	Toxicité	Varga <i>et al.</i> , 2015
DON	Trichothécène de type B (toxine)	Toxicité	Goswami <i>et al.</i> , 2004
NIV	Trichothécène de type B (toxine)	Toxicité	Tralamazza <i>et al.</i> , 2016
Fusaoctaxin A	Toxine	Responsable de l'invasion de cellule à cellule du blé	Jia <i>et al.</i> , 2019
Gènes de biosynthèse des toxines			
Région génomique de cluster Tri	Gènes de biosynthèse du trichothécène	Impliqué dans la synthèse des mycotoxines trichothécènes	Kimura <i>et al.</i> , 2003
Autres déterminants de la pathogénicité et de la virulence			
Gpmk1 MAP kinase	MAP kinase	Impliqué dans l'accouplement, la conidiation et la pathogénicité	Jenczmionka <i>et</i> Schäfer, 2005
FGL1	Gène de la lipase	Augmente la virulence fongique contre le blé et le maïs	Voigt <i>et al.</i> , 2005
PKS2	Gène de la polykétide synthase	Responsable du métabolisme secondaire et de la virulence	Gaffoor <i>et al.</i> , 2005
FgABCC9	Gène transporteur de cassette de liaison à l'ATP	Impliqué dans la pathogénicité fongique du blé	Qi <i>et al.</i> , 2018
bZIP transcription factor	Facteur de transcription	Augmente la virulence de <i>F. graminearum</i> dans les plantes infectées	Wang <i>et al.</i> , 2011

2.2.5. Localisation subcellulaire pan-sécrétome dans la cellule végétale

La détermination de la localisation subcellulaire des effecteurs protéiques dans les cellules végétales peut fournir des informations précieuses sur les processus végétaux et les protéines ou acides nucléiques avec lesquels ils interagissent potentiellement. Cette connaissance facilite la priorisation des candidats pour l'analyse fonctionnelle (Petre *et* Kamoun, 2014). Une fois à l'intérieur des cellules hôtes, les effecteurs cytoplasmiques se transportent vers des compartiments distincts, y compris le noyau, le chloroplaste, le cytosol, la membrane plasmique ou le réticulum endoplasmique pour déréguler les fonctions végétales à leur avantage. Afin de cibler ces compartiments, les effecteurs ont des domaines ou des motifs (signaux d'adressage) dans leur séquence ; par exemple, des peptides transit N-terminaux dans le cas des chloroplastes et des mitochondries, ou des signaux de localisation nucléaire (NLS) dans le cas des noyaux (Bhushan *et al.*, 2006; Horton *et al.*, 2007; Sperschneider *et al.*, 2017). Toutefois, les peptides transit partagent rarement une séquence strictement identique et varient en longueur (Jeong *et al.*, 2021). Un défi majeur pour l'analyse fonctionnelle des effecteurs des champignons phytopathogènes dont *F. graminearum* est le grand nombre de protéines candidates codées dans leurs génomes (Brown *et al.*, 2012). Cela peut rendre la détermination de la localisation subcellulaire par des méthodes expérimentales purement prohibitive, laborieuse et coûteuse. Dans ce contexte, pour relever ces défis, l'analyse *in silico* via des outils bio-informatiques reste un bon choix alternatif pour aider les biologistes à identifier les cibles de l'agent pathogène et à réduire les listes de gènes candidats. Une compréhension complète des protéines sécrétées par *F. graminearum* fournira des cibles clés pour un meilleur contrôle de ce pathogène et pourrait améliorer la sélection de cultivars de blé résistants.

Contexte et objectifs de la Thèse

Ce travail de recherche a été développé dans le cadre d'un projet de Thèse en cotutelle établi entre le laboratoire de Biotechnologie Médicale (MedBiotech) de la Faculté de Médecine et de Pharmacie de l'Université Mohammed V de Rabat (Maroc) et l'équipe « Maladies des Céréales » (MDC) de l'Unité Mixte de Recherche Génétique, Diversité, Écophysiologie des Céréales (UMR INRAE- Université Clermont Auvergne 1095 GDEC, France). Ce projet vise à explorer les relations entre l'organisation du génome et la pathogénicité de deux micro-organismes étudiés dans les laboratoires de tutelle ; la bactérie pathogène opportuniste *A. baumannii* et le champignon phytopathogène *F. graminearum*. La stratégie proposée repose sur une approche commune de génomique comparative intra-espèce basée sur le séquençage de génomes complets d'une collection de 8 souches d'origine européenne de *F. graminearum* sélectionnées pour leur agressivité contrastée sur blé tendre et de 13 souches marocaines d'*A. baumannii* multirésistantes aux antibiotiques : 5 souches environnementales (isolées de l'environnement hospitalier) et 8 souches cliniques (isolées de patients), dont une souche résistante à la dernière ligne de traitement « la colistine ». Par cette démarche, mes travaux s'organisent autour de deux objectifs principaux et complémentaires :

Objectif 1 : Caractérisation de la diversité génomique structurale et fonctionnelle de deux micro-organismes

Chez *F. graminearum*, nous avons formulé l'hypothèse de l'existence d'un génome dispensable étendu reflétant sa capacité adaptative à des environnements très différents et à une large gamme d'hôtes. Dans ce cadre, ce travail a tout d'abord visé (1) à établir le premier pangénome de l'espèce, (2) à en caractériser les différentes composantes : core, dispensable et spécifique, et (3) à établir sa nature « ouverte » ou « fermée ». Ce travail a ensuite été poursuivi par la description de la fonction des groupes de gènes appartenant à ces différentes composantes.

Chez *A. baumannii*, nous sommes partis du postulat de l'existence, d'une part, d'une expansion sélective du contenu génomique spécifique/unique des souches environnementales par rapport aux souches cliniques qui pourrait contribuer à leur persistance et à leur adaptation à des spectres hôtes différents du milieu hospitalier et, d'autre part, de plusieurs clones circulant dans ce milieu hospitalier. Dans ce contexte, l'analyse comparative des génomes complets via la construction du pangénome, permettra de mieux comprendre (i) la diversité génétique des souches de deux origines, (ii) la structure de population d'*A. baumannii* à l'hôpital marocain, (iii) la relation entre ces deux caractères.

Objectif 2 : Étudier la relation entre la variabilité des gènes, leurs fonctions et la pathogénicité des souches

Chez *F. graminearum*, l'étude de cette relation a été abordée sous l'hypothèse forte que l'agressivité contrastée des différentes souches étudiées repose sur la présence/absence de déterminants génétiques variables. Cette variabilité a été évaluée principalement sur la base du contenu du pan-sécrétome (un ensemble complet des protéines sécrétées de toutes les souches) et sur l'impact de mutations LOF (potentiellement impliqués dans la perte de fonctions de protéines) présentes, d'une part, dans le pan-sécrétome et, d'autre part, dans les clusters de gènes de biosynthèse de métabolites secondaires. La démarche est structurée autour de plusieurs questions complémentaires permettant de caractériser les variations génétiques des protéines sécrétées notamment des effecteurs protéiques, leur distribution dans les différentes composantes du pan-sécrétome, leur localisation génomique et leurs cibles subcellulaires putatives chez la plante hôte.

Chez *A. baumannii*, cet objectif visait à comprendre et relier à grande échelle (i) le profil phénotypique de la résistance aux antibiotiques des souches, (ii) le contenu du pan-résistome (ensemble complet de gènes de résistance intrinsèques et acquis dans toutes les souches), (iii) la relation de ces caractères avec la ou les lignées clonales des souches. Dans ce cadre, deux questions principales ont été posées : (1) le profil phénotypique de résistance aux antibiotiques observé reflète-t-il le contenu du pan-résistome ? (2) Quel(s) déterminant(s) génétique(s) sont impliqué(s) dans le(s) mécanisme(s) de résistance à la dernière ligne de traitement « colistine » ?

Dans le premier chapitre de ce manuscrit, nous décrirons le workflow d'analyse bioinformatique utilisé et adapté pour les génomes de deux micro-organismes, en détaillant le principe et le but pour de chaque étape d'analyse avec les outils/packages choisis. Les résultats de deux micro-organismes, *F. graminearum* et *A. baumannii* seront détaillés dans **chapitre 2** et **chapitre 3**, respectivement. Le **chapitre 2** inclut deux articles publiés, le premier décrivant les caractéristiques du génome assemblé de la souche française la plus virulente (MDC_Fg1) et le deuxième comprend l'ensemble de nos résultats d'analyse comparative de génomes de *F. graminearum*. Le **chapitre 3** comprend deux articles, le premier décrivant l'assemblage du génome et les caractéristiques d'annotation de deux souches environnementales marocaines d'*A. baumannii* et le deuxième article, en préparation, comprend l'ensemble des résultats d'*A. baumannii*. Ensuite, une partie « **discussion et perspectives** » permettra de discuter conjointement l'ensemble des résultats obtenus au cours de ma Thèse avec des propositions de nouvelles perspectives d'études futures. Enfin, une section "**Annexes**" fournira un

état de l'art des technologies actuelles de séquençage NGS, d'assemblage de *nov*o et d'annotation du génome ainsi que les étapes pratiques nécessaires à la préparation de librairie "pour Illumina MiSeq". Cette section regroupe également l'ensemble des articles auxquels j'ai été amené à contribuer durant la période de ma Thèse.

Conception de l'étude :

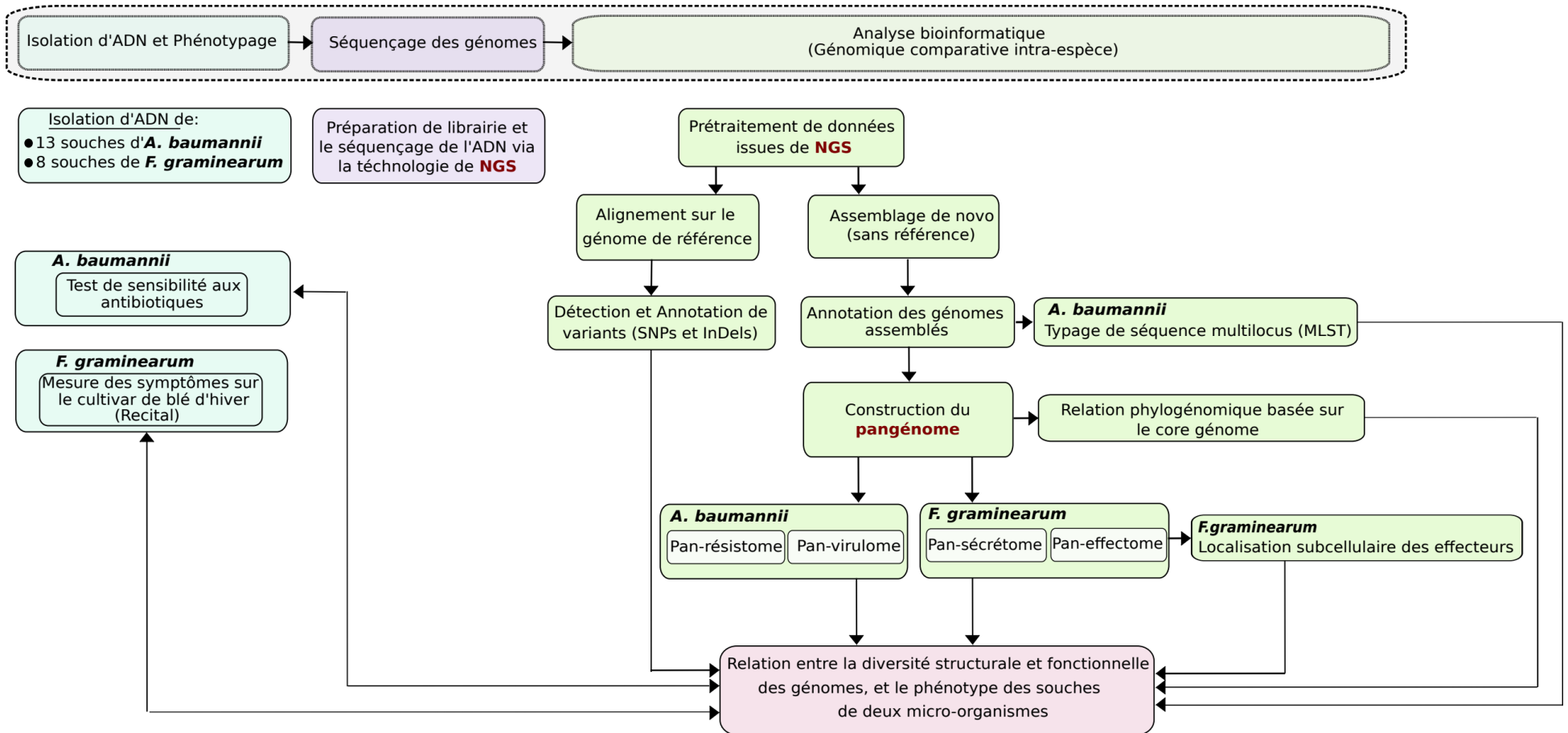


Figure 5. Workflow représentant les grandes étapes de l'étude.

Workflow d'analyse génomique

Chapitre 1

Chapitre 1 – Workflow d’analyse génomique utilisé pour les deux micro-organismes : la bactérie *A. baumannii* et le champignon *F. graminearum*

Pour extraire les informations dans les données fournies par les technologies de NGS, nous devons effectuer une analyse génomique qui repose sur un enchaînement d'étapes bioinformatiques (workflow). Après la conception expérimentale et le séquençage de l'ADN, le workflow d'analyse génomique a été commencé par le contrôle de qualité initial des données brutes jusqu'aux visualisations des résultats finaux prêts à être interprétés. Généralement, un workflow bioinformatique est établi pour une expérience spécifique en fonction de son objectif et de son principe. Au cours de ce travail de Thèse, nous avons adapté un workflow d'analyse pour les génomes des deux agents pathogènes ; la bactérie *A. baumannii* et le champignon *F. graminearum*. Celui-ci pourrait également s'appliquer à d'autres bactéries multirésistantes et champignons phytopathogènes avec quelques modifications spécifiques en fonction de l'espèce étudiée. Par exemple, le pathogène fongique *F. graminearum* est un génome d'organisme eucaryote, l'annotation génomique nécessite des évidences transcriptomiques et protéomiques de l'espèce étudiée et/ou des espèces apparentées en combinaison avec des méthodes *ab initio* dans le but de prédire les modèles de gènes.

Dans ce travail, le bio-informaticien doit non seulement avoir une expérience analytique, mais également comprendre le contexte biologique et la conception de l'expérience pour appliquer le workflow de manière appropriée. Parmi les aspects fondamentaux qui doivent être pris en compte lors du lancement d'un nouveau projet de génome, nous pouvons citer : (i) la taille du génome, (ii) le contenu en GC (iii) les niveaux de ploïdie et d'hétérozygotie. Ce chapitre vise à fournir des étapes pratiques pour faciliter l'achèvement du projet avant de se lancer dans un projet d'assemblage et d'annotation de génomes haploïdes. Il convient de noter que la plupart des étapes de ce workflow d'analyse ont été effectuées sur un cluster de calcul haute performance « HPC » de type SLURM (*Simple Linux Utility for Resource Management*).

I. Origines des souches d'*A. baumannii* et de *F. graminearum*

Treize nouvelles souches marocaines de bactéries d'*A. baumannii*, ainsi que huit nouvelles souches européennes de champignon *F. graminearum*, ont été inclus dans ces travaux de Thèse (Tableau 2). Les souches bactériennes ont été collectées et isolées au service de bactériologie de l'hôpital militaire Mohammed V de Rabat sous la direction du Pr. Mostafa Elouennass, dont huit souches cliniques récupérées sur des patients et cinq souches environnementales prélevées dans les chambres d'hôpital. Toutes ces souches ont été isolées à partir de différents sites d'isolement et ont été définies comme multirésistantes en raison de leur résistance à au moins trois classes

d'antibiotiques, dont une souche (M-C15) résistante à la colistine. Les huit souches européennes de *F. graminearum* étaient capables de produire du DON ainsi que d'induire la fusariose de l'épi mais néanmoins avec des niveaux d'agressivité contrastés sur le blé ; quatre ont été collectées en France, trois provenaient d'Italie et une souche a été collectée en Allemagne (**Tableau 2**) dans le cadre d'une collaboration internationale avec Dr. Joerg Bormann et Pr. Wilhelm Schäfer de l'Université de Hambourg en Allemagne ; Pr. Micheal Freitag de l'Université de l'Oregon aux Etats-Unis et la plateforme Gentyane de l'unité GDEC / INRAE, Clermont Ferrand.

Tableau 2. Origine, site et date d'isolement de treize et huit nouvelles souches d'*A. baumannii* et *F. graminearum*, respectivement, utilisées dans ces travaux Thèses.

Organisme	Souche		Site échantillonné	Date d'isolement	Origine	
<i>A. baumannii</i>	M-C03	Clinique	Bouche	2015	Maroc	
	M-C11		Marge anale	2015	Maroc	
	M-C12		Culture sanguine	2016	Maroc	
	M-C13		Aspiration bronchique	2015	Maroc	
	M-C14		Prélèvement distal protégé	2017	Maroc	
	M-C15		Urine	2016	Maroc	
	M-C38		Marge anale	2015	Maroc	
	M-C40		Marge anale	2015	Maroc	
	M-E03	Environnementale	Chariot patient	2015	Maroc	
	M-E08		Sol	2015	Maroc	
	M-E12		Sol	2015	Maroc	
	M-E21		Ventilateur médical	2015	Maroc	
	M-E27		Ventilateur médical	2015	Maroc	
	<i>F. graminearum</i>	MDC_Fg1		Épi de blé	2010	France
		MDC_Fg5		Épi de blé	2010	France
MDC_Fg8		Épi de blé	2018	Allemagne		
MDC_Fg13		Épi de blé	2012	France		
MDC_Fg202		Épi de blé	2018	Italie		
MDC_Fg593		Épi de blé	2018	Italie		
MDC_Fg851		Épi de blé	2018	Italie		
MDC_FgU1		Épi de blé	2010	France		

II. Analyse génomique des données issues de NGS

2.1. Prétraitement des *reads* issus de NGS

Pour chaque échantillon séquençé (via la technologie Illumina), nous avons obtenu deux fichiers bruts en format *.FastQ¹ (forward et reverse) produits du séquençage *paired-end* (Annexe B) constitué de milliers de *reads*² (en fonction de la profondeur de séquençage³). Dans le même *read*, toutes les bases nucléotidiques n'ont pas la même valeur de qualité, c'est-à-dire la probabilité que la base soit correcte n'est pas la même. Le contrôle qualité initial peut se focaliser sur plusieurs caractéristiques telles que (i) la qualité de séquence par base qui permet d'observer la distribution du score de qualité de chaque base au niveau de chaque position sur les *reads*, (ii) le contenu en bases des séquences qui permet de vérifier le pourcentage de chaque base (le pourcentage entre A-T et G-C), et (iii) le taux de bases inconnues,. Pour évaluer tous ces paramètres, FastQC v0.11 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>) a été exécuté sur des fichiers *.FastQ avant et après le filtrage des bases/*reads* de mauvaise qualité (Figure 6).

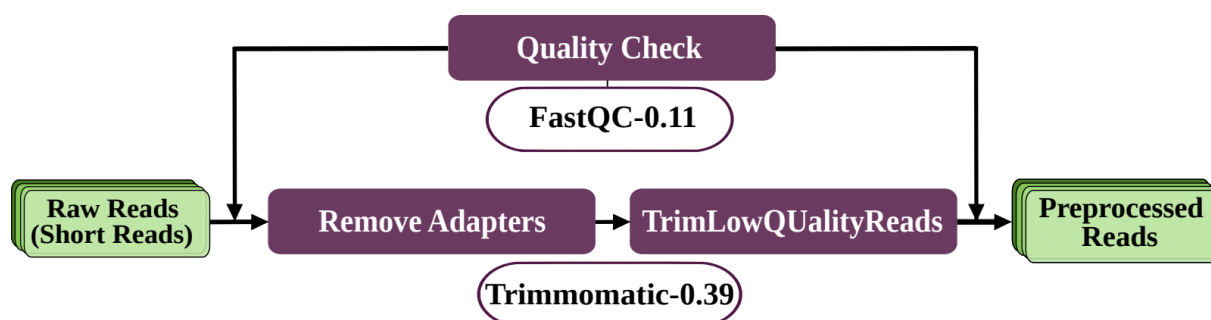


Figure 6 : Prétraitement des données brutes d'*A. baumannii* et *F. graminearum* issues de NGS

Comme l'ADN a été fragmenté en plus petits morceaux lors de la préparation de la librairie avec des adaptateurs ligaturés aux extrémités (Annexe B), il peut arriver que les séquences d'adaptateurs soit également partiellement séquençé. Nous avons utilisé Trimmomatic v0.39 (Bolger *et al.*, 2014) permettant la suppression de l'adaptateur et le filtrage des bases/*reads* de mauvaise qualité. Pour optimiser la détection des SNPs et des insertions/délétions (InDels), nous avons opté pour un filtrage de qualité plus strict basé sur l'approche de la fenêtre glissante pour supprimer les bases de *read* qui ont une qualité minimale de Q20 (probabilité d'avoir une base incorrecte sur 100 bases).

1 **FastQ**: format de fichier texte permettant de stocker à la fois des séquences nucléotidiques et les scores de qualité associées.

2 **Reads**: terme générique désignant les séquences nucléotidiques du fichier FastQ

3 **Profondeur de séquençage**: décrit le nombre de fois qu'un nucléotide donné dans le génome a été séquençé.

2.2. Détection et annotation des SNPs et InDels

Le principe de la détection des variants génétiques (SNPs et InDels) est d'utiliser un génome de référence de la souche en question, puis d'aligner les *reads* sur cette séquence de référence afin de détecter tous les variants génétiques. Ces variants sont ensuite filtrés par leur qualité et leur cohérence. Quatre étapes majeures sont réalisées pour détecter les SNPs et InDels (Figure 7) : (i) alignement de *reads* sur la séquence de référence, (ii) traitement post-alignement, (iii) détection et filtrage des variants, et (iv) annotation des variants détectés.

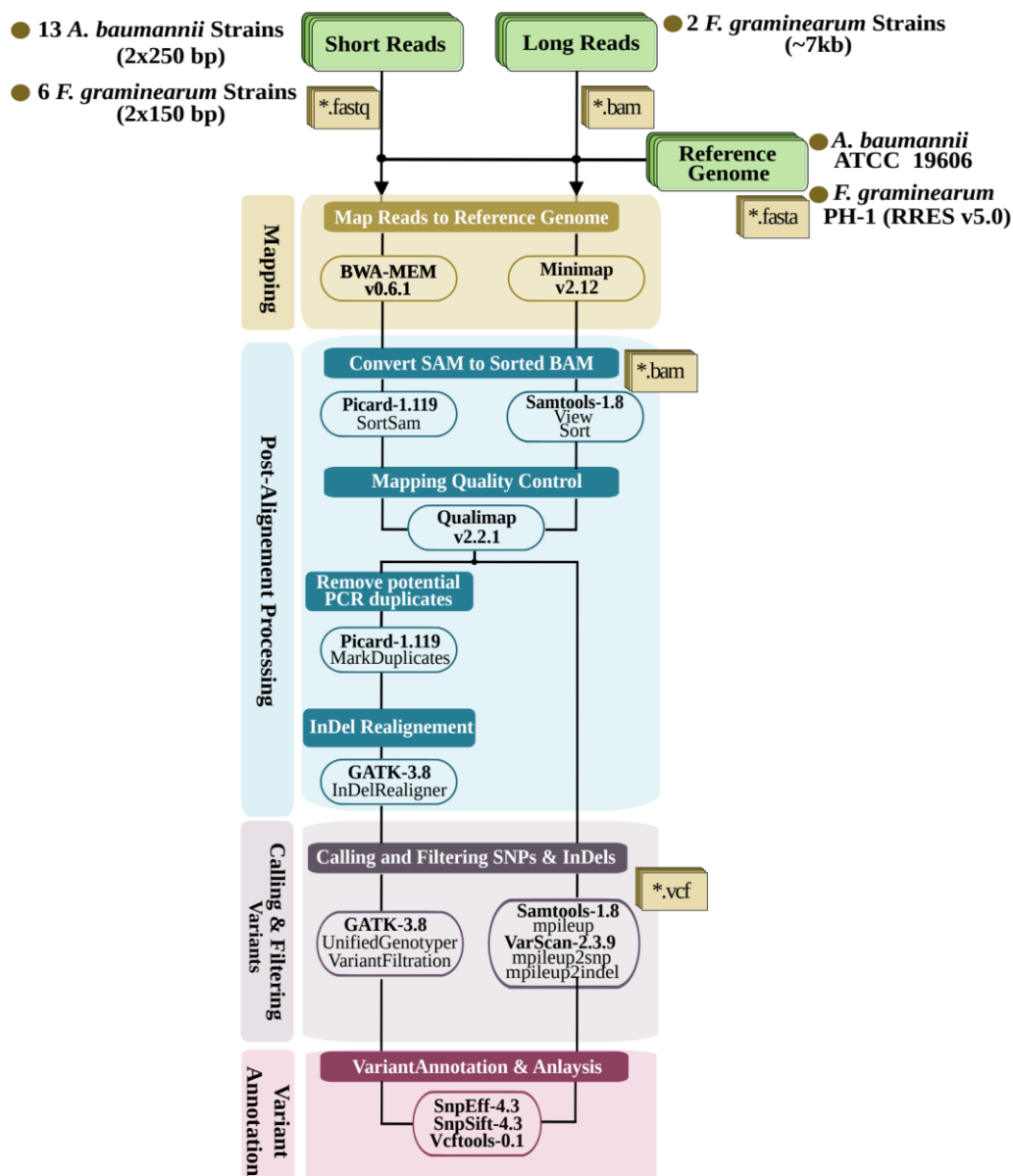


Figure 7 : Pipeline utilisé pour la détection de SNPs et InDels.

2.2.1. Alignement des *reads* sur la séquence de référence

L'objectif général lors d'alignement (ou mapping) de *reads* séquencés est de déterminer pour chaque *read* ou paire de *reads* le véritable emplacement (origine) par rapport à la référence (également appelé assemblage par référence). Cela nécessite des algorithmes sophistiqués à la fois rapides et efficaces. Pour les données de la plateforme Illumina, nous avons utilisé le programme BWA v0.6 (Li *et al.*, 2009) avec l'algorithme BWA-mem « *Maximal Exact Match* » qui a été conçu principalement pour des *reads* courts. En revanche, pour l'alignement de *reads* longs obtenus par PacBio Sequel, nous avons utilisé le programme minimap2 (Li, 2018), conçu pour aligner les *reads* de plusieurs kb de longueur. Il est à noter que le processus du mapping nécessite une étape initiale qui consiste à créer l'index de la séquence de référence, cette étape permet un accès plus rapide aux différentes parties de la référence lors de l'étape d'alignement.

2.2.2. Traitement post-alignement

Cette étape peut être considérée comme une préparation nécessaire aux étapes en aval. Tout d'abord, la conversion des fichiers d'*output* de BWA (format *.SAM⁴) en fichiers compatibles pour les analyses suivantes (format *.BAM⁵). Pour ce faire, le package Picard (<http://broadinstitute.github.io/picard>) a été utilisé pour convertir, trier (par coordonnées) et indexer les fichiers d'alignement produits. Ensuite, leur traitement a été effectué à différents niveaux. (i) Les *reads* dupliqués ont été identifiés via la fonction *MarkDuplicates* de Picard, car les *reads* techniquement amplifiés pourraient fortement biaiser les appels de variants avec les faux positifs en raison de l'amplification des erreurs de PCR (*Polymerase Chain Reaction*). (ii) Les positions des InDels ont été recalibrées en utilisant les fonctions *RealignerTargetCreator* et *IndelRealigner* du package GATK v3.8 (McKenna *et al.*, 2010). Ces étapes de marquage de *reads* dupliqués ne sont pas nécessaires pour les données PacBio car le séquençage a été effectué sur une seule molécule sans amplification (Annexe A). QualiMap2 (Okonechnikov *et al.*, 2016) a été ensuite exécuté pour générer des rapports de contrôle qualité d'alignement énumérant toutes les spécifications nécessaires des *reads* alignés, y compris (i) le pourcentage de *reads* alignés, (ii) le score de la qualité du mapping, (iii) le taux de couverture des *reads* sur la séquence de référence. Ce dernier nous permet de nous assurer que la couverture est suffisante pour détecter les SNP/InDels.

4 **SAM:** *Sequence Alignment Map*, est un format basé sur du texte à l'origine pour stocker les *reads* alignés sur une séquence de référence.

5 **BAM:** *Binary Alignment Map*, est la version binaire compressée d'un fichier *.SAM

2.2.3. Détection et filtrage des SNP et InDels

Les SNP et les InDels ont été détectés en utilisant la fonction *Unified-Genotyper* de GATK v3.8 (McKenna *et al.*, 2010) en mode haploïde. Ce dernier est conçu pour des *reads* courts comme ceux générés par Illumina. Ensuite, pour obtenir les SNPs et InDels de haute qualité, nous avons utilisé la fonction *VariantFiltration* de GATK pour effectuer un filtrage basé sur des seuils de qualité en suivant les recommandations des meilleures pratiques GATK (Van der Auwera *et al.*, 2013).

Pour identifier les SNPs et InDels à partir des fichiers *.BAM issus d'alignement des *reads* PacBio, nous avons d'abord généré des fichiers *.mpileup via le package SAMtools mpileup v1.8 (H. Li *et al.*, 2009). Une fois ces fichiers sont générés, VarScan v2.3.9 (Koboldt *et al.*, 2012) avec les deux fonctions, *mpileup2snp* et *mpileup2indel* pour identifier les SNPs et les InDels, respectivement. VarScan détecte et filtre les variants en fonction de plusieurs paramètres : la profondeur de variant, la qualité Phred minimale, etc. Enfin, les fichiers *.VCF⁶ est l'*output* de GATK et VarScan, contenant des informations de la détection de variants relatifs à une séquence de référence.

Il convient de noter que le calcul BAQ « *Base Alignment Quality* » ; un score de type Phred représentant la probabilité qu'une base de *read* soit mal aligné, est activé par défaut dans SAMtools mpileup, alors que les bases des *reads* issues de PacBio ne sont pas basés sur le score Phred, pour cela cette option a été désactivée avec l'option "-B".

2.2.4. Annotation fonctionnelle des SNP et InDels

Les SNPs et InDels détectés sont ensuite annotés. L'annotation fonctionnelle de variants vise à évaluer leur impact fonctionnel sur les gènes codant pour des protéines et des régions régulatrices. Nous avons utilisé SnpEff v 4.3 (Cingolani *et al.*, 2012) selon les catégories définies dans le manuel SnpEff (http://snpeff.sourceforge.net/SnpEff_manual.html). Ensuite, les effets des variants ont été analysés avec SnpSift v4.3 (Cingolani *et al.*, 2012) et VCFtools v0.1.16 (Danecek *et al.*, 2011). Les bases de données SnpEff ont été construites à l'aide des fichiers d'annotation *.GFF⁷ du génome de référence des deux micro-organismes (*A. baumannii* et *F. graminearum*).

2.3. Assemblage *de novo* et annotation des génomes

Générer un assemblage *de novo* et une annotation génomique de haute qualité est devenu un outil indispensable pour mieux comprendre la biologie de toute espèce. Dans cette section, nous décrivons la méthode choisie pour l'assemblage *de novo* selon la technologie de séquençage (Illumina

6 VCF: *Variant Call Format*, un format fichier texte pour stocker les résultats de variants génétiques

7 GFF: *General Feature Format*, est un fichier texte délimité par des tabulations pour décrire les caractéristiques génomiques.

et PacBio) et comment nous avons adapté et optimisé l'étape d'annotation des génomes assemblés des deux micro-organismes.

2.3.1. Assemblage *de novo* : des *reads* prétraités aux génomes assemblés

En principe, l'assemblage *de novo* est le processus d'alignement, d'orientation et de fusion de *reads* pour former un génome complet ou partiellement complet (sous forme de contigs⁸ ou scaffolds⁹) sans aucune séquence de référence.

Dans ces travaux, des assembleurs basés sur deux méthodes d'assemblages (Annexe C) ont été appliqués en fonction de la taille de *reads* générés par les deux technologies. La méthode du graphe de Bruijn (DBJ) qui fonctionne mieux sur les *reads* courts issus d'Illumina, alors que la méthode *Overlap-Layout-Consensus* (OLC) est mieux adaptée aux *reads* longs, y compris les *reads* issus de la plateforme PacBio. Nous avons utilisé trois assembleurs, ABySS v2.0 (Jackman *et al.*, 2017) et SPAdes v3.13 (Bankevich *et al.*, 2012 ; basés sur la méthode DBJ) pour assembler les *reads* Illumina ainsi que HGAP4 (Chin *et al.*, 2013 ; basée sur la méthode OLC) pour assembler les *reads* de PacBio. Il est intéressant de noter qu'avec le programme ABySS nous avons testé plusieurs tailles de k-mères jusqu'à ce que nous trouvions des valeurs optimales pour l'assemblage (41, 96 et 128). Alors que SPAdes, utilise systématiquement plusieurs tailles de k-mères de 21 à 127.

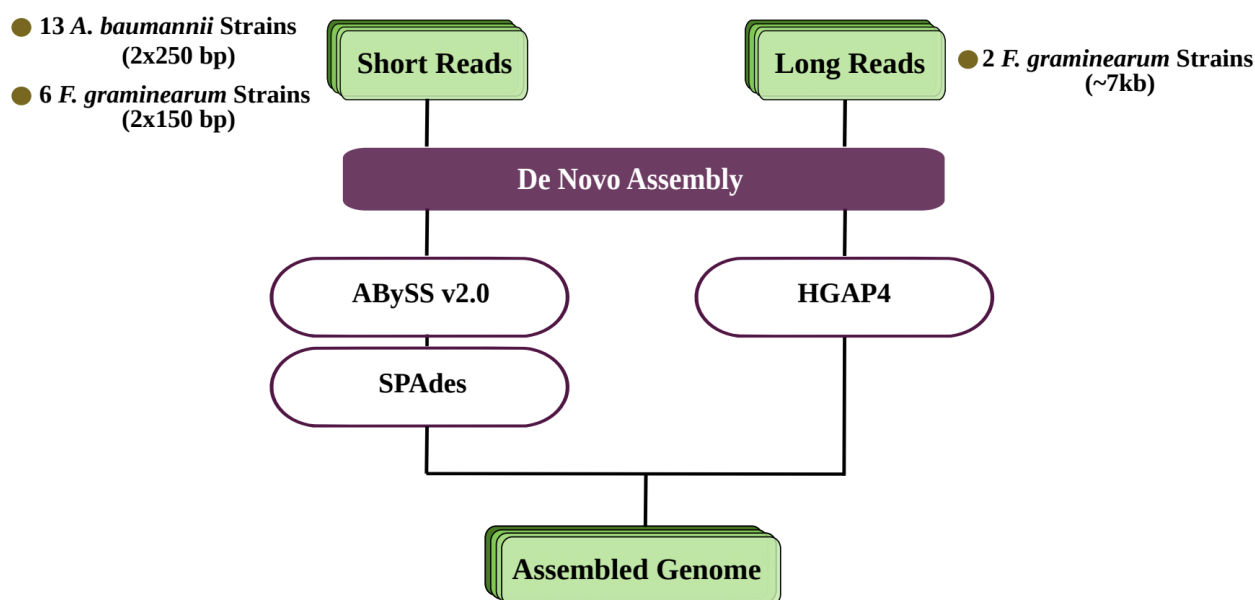


Figure 8 : Les outils utilisés pour l'assemblage *de novo* des *reads* courts (d'Illumina) et longs (de PacBio)

2.3.2. Évaluation de l'exhaustivité des génomes assemblés

⁸ **Contig**: est défini comme une longue séquence assemblée à partir d'un ensemble de fragments de séquence "*reads*".

⁹ **Scaffold**: est constitué de deux ou plusieurs contigs séparés par des *gaps*.

Une fois le processus d'assemblage *de novo* terminé, il est important de valider et de s'assurer que l'assemblage généré convient aux analyses en aval. Une attention particulière à cette étape est nécessaire, car il existe un risque d'assemblage incorrect avec des contigs fragmentés ou une taille du génome assemblé qui ne correspond pas à la taille attendue de l'espèce, etc. Des indices et des métriques statistiques ont été utilisés pour guider le processus dans la bonne direction et valider les génomes assemblés obtenus. Nous avons utilisé l'outil d'évaluation QUAST (*Quality Assessment Tool for Genome Assemblies* ; Gurevich *et al.*, 2013) pour calculer plusieurs de ces statistiques et fournir ainsi une base pour l'évaluation de l'assemblage, y compris (i) la taille du génome assemblé, (ii) le nombre total de contigs obtenus, (iii) N50 (Annexe C). Cependant, ces métriques ne suffisent pas, car elles dépendent uniquement de la taille des contigs sans évaluer leur exactitude. Pour cela, nous avons utilisé BUSCO v3 (*Benchmarking Universal Single-Copy Orthologs* ; Simão *et al.*, 2015), permettant de montrer la complétude d'assemblage en recherchant la présence d'un ensemble de gènes conservés. Pour *A. baumannii* et *F. graminearum*, respectivement, 452 et 3725 gènes orthologues conservés de gammaprotéobactéries et de la lignée Sordariomycètes, fournis par BUSCO ont été évalués pour leur présence afin d'estimer la complétude des génomes assemblés.

2.3.3. Annotation des génomes eucaryotes fongiques : *F. graminearum*

Les séquences génomiques assemblées et validées ont été ensuite annotées. Une annotation précise, en particulier des génomes eucaryotes, est une tâche difficile. Les deux principales méthodes sont la prédiction de gènes *ab initio* également appelée *de novo* (ou intrinsèques) et la méthode basée sur des évidences de séquences ou de motifs appelées extrinsèques (Annexe C). La découverte de gènes *ab initio* repose sur la seule séquence génomique assemblée qui est systématiquement recherchée pour les modèles intrinsèques des gènes codant pour les protéines. En revanche, la prédiction de gènes basée sur des évidences repose sur l'alignement précis des évidences d'expression existantes, telles que les marqueurs de séquences exprimées (EST ou *Expressed Sequence Tag*), les données de séquence d'ARN et les séquences de protéines de la même espèce et/ou des espèces proches. Dans ce travail, nous avons adapté le pipeline MAKER (Campbell *et al.*, 2014) pour les génomes de *F. graminearum* en combinant à la fois la prédiction de gènes *ab initio* et celle basée sur des évidences pour aboutir à un ensemble final de gènes modèles plus fiables. En outre, le pipeline a également associé des mesures de qualité telles que la distance d'édition d'annotation (AED), où un AED de 0 indique un alignement parfait avec les évidences fournies.

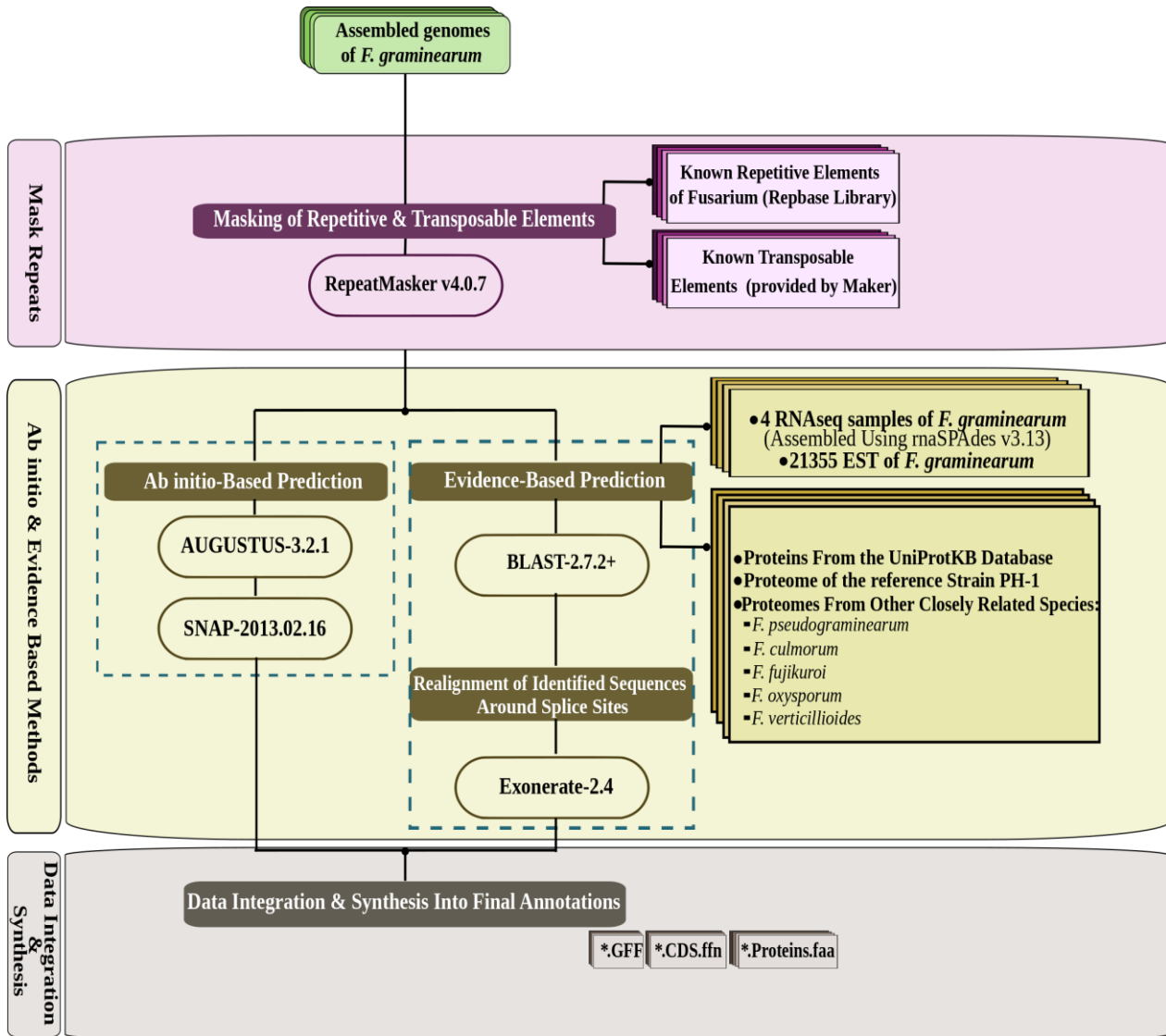


Figure 9 : Pipeline d'annotation du génome de *F. graminearum*.

Nous avons commencé à identifier les éléments répétitifs en utilisant RepeatMasker (Tempel, 2012) avec une librairie de tous les éléments répétés de *Fusarium* fournie par Repbase (Bao *et al.*, 2015). De même, la base de données "*te_proteins.fasta*" distribuée avec le pipeline MAKER a également été utilisée; cette base de données se compose de produits protéiques de gènes associés à des éléments transposables communément trouvés. Ensuite, Pour la prédiction de gènes basée sur des évidences, nous avons intégré dans le pipeline : (i) 21,355 ESTs de *F. graminearum* obtenues de la base de données NCBI (*National Center for Biotechnology Information*), (ii) quatre échantillons RNAseq de *F. graminearum* (*Run accessions*: SRR999640, SRR999641, SRR999642 et SRR999643) assemblés *de novo* par rnaSPAdes v3.13.0 (Bushmanova *et al.*, 2019), (iii) l'ensemble des protéines de la base de données UniProtKB (Boutet *et al.*, 2016) complétées par le protéome de la souche de référence PH-1 et des protéomes d'autres espèces de *Fusarium* (*F. pseudograminearum*, *F. culmorum*, *F.*

fujikuroi, *F. oxysporum* et *F. verticillioides*) obtenus à partir de la base de données Ensembl Fungi (mise à jour-40).

Les programmes de prédiction génétique *ab initio* prédisent des modèles de gènes basés sur des caractéristiques intrinsèques telles que les limites exon-intron, les codons d'initiation et stop, etc. Deux programmes, AUGUSTUS (Hoff *et* Stanke, 2019) et SNAP (<https://github.com/KorfLab/SNAP>) ont été optimisés pour l'annotation des génomes de *F. graminearum*. Tout d'abord, les données RNA-Seq susmentionnées, ont été réutilisées pour entraîner SNAP alors que AUGUSTUS contenait préalablement des modèles géniques de *F. graminearum* (et d'autres espèces).

Les résultats d'annotation obtenus (basées sur des évidences et *ab initio*) ont été intégrées et synthétisées par le pipeline MAKER. Les fichiers de sortie MAKER incluent des fichiers au format *.FASTA des séquences codantes (CDS) et des protéines pour chaque gène annoté ainsi que des fichiers d'annotation *.GFF. Ces fichiers fournissent également un certain nombre de mesures de qualité (y compris AED) pour chaque modèle de gène.

Il est intéressant de noter que parmi les huit génomes de *F. graminearum* de cette étude, nous avons utilisé douze autres génomes de la même espèce précédemment publiés, dont huit n'avaient que des données brutes (en format FastQ), et deux avec des génomes assemblés. Pour cela, nous avons appliqué la même méthode d'assemblage *de novo* (pour huit génomes) et d'annotation (pour les 10 génomes).

2.3.4. Annotation des génomes bactériens : *A. baumannii*

L'annotation des génomes assemblés d'*A. baumannii* a été réalisée via le pipeline d'annotation Prokka (Seemann, 2014). Ce pipeline regroupe plusieurs outils et est adapté pour les génomes bactériens. (i) Les régions génomiques codant pour les protéines ont été identifiées via Prodigal v2.6.3 (Hyatt *et al.*, 2010). (ii) Les régions codant pour les ARN de transfert (ARNt), les ARNr et les ARN non codant (ARNnc), ont été prédites avec Aragorn v1.2.38 (Laslett *et* Canback, 2004), Barrnap v0.9 (<https://github.com/tseemann/barrnap>) et Infernal v1.1.3 (Nawrocki, 2014), respectivement. (iii) La fonction de la protéine codée est prédite par la similitude avec une base de données spécialement configurée pour *A. baumannii* (Figure 10).

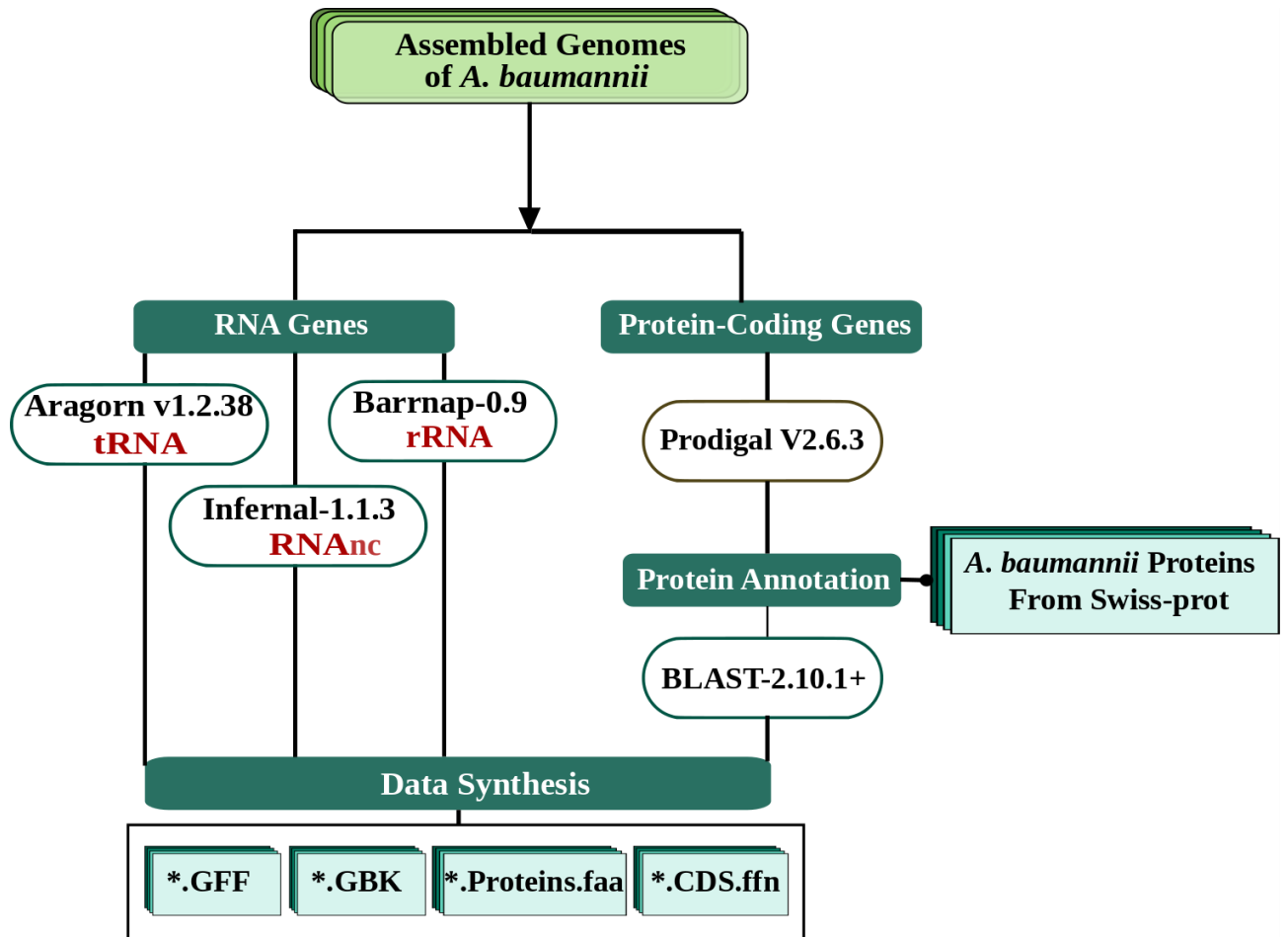


Figure 10 : Pipeline d'annotation du génome d'*A. baumannii*

Pour les fichiers *output* de Prokka : les fichiers *.GFF et *.GBK contiennent toutes les informations sur les entités annotées, le fichier *.faa contient les séquences protéiques et le fichier *.ffn contient les CDS.

2.3.5. Évaluation de l'exhaustivité de gènes annotés

Les gènes annotés ont été vérifiés pour la présence de gènes complets (gènes ayant à la fois des codons de démarrage et d'arrêt) et des modèles de gènes partiels (ne disposant pas de codon d'initiation ou de codon d'arrêt ou des deux). Ainsi, d'autres caractéristiques dont la taille moyenne des gènes, la taille des exons et le nombre d'exons par gène (pour *F. graminearum*) en utilisant le script "gag.py" (<https://github.com/genomeannotation/GAG>). Si les gènes partiels se retrouvent près des bords des contigs, ils peuvent être dus à la fragmentation de contigs. Pour cela, le script interne « *check_annotation.py* » nous permet de comparer la localisation des gènes annotés complets et partiels par rapport à leur emplacement le long de contigs entre tous les génomes.

2.4. Construction du pangénome et analyse phylogénomique

Quel que soit le but d'une étude pangénomique et/ou l'investigation d'une fonction moléculaire spécifique ou la diversité génétique d'un groupe spécifique d'organismes, la détermination d'un pangénome est souvent basée sur le *clustering* de séquences orthologues. En effet, la détection de gènes orthologues est souvent utilisée pour trouver des gènes fonctionnellement équivalents chez des individus de la même espèce, ce qui permet l'identification de familles de gènes appartenant au core génome et génome dispensable. La première étape de la plupart des algorithmes de *clustering* consiste à déterminer les alignements de séquences par paires de séquences protéiques ou de CDS dans un ensemble de données. La principale difficulté du *clustering* vient de la présence de paralogues¹⁰ qui peuvent rendre très difficile la reconnaissance de l'orthologue parmi les autres homologues. Ici, l'approche de *clustering* intra-spécifique a été réalisée sur 20 et 23 protéomes annotés de *F. graminearum* et d'*A. baumannii*, respectivement, en utilisant le pipeline ProteinOrtho (Lechner *et al.*, 2011). Les principaux objectifs des développeurs de ProteinOrtho étaient de réduire considérablement la quantité de mémoire requise pour l'analyse orthologique des protéines (comparativement aussi bonne que OrthoMCL (Fischer *et al.*, 2011)) et de traiter facilement un grand volume de données. Tout d'abord, ProteinOrtho identifie les (co-)orthologues entre les génomes via une comparaison initiale « *all-vs-all* » entre toutes les protéines de toutes les souches pour déterminer les similitudes des protéines. La deuxième étape est le *clustering* de gènes similaires en groupes (co-)orthologues significatifs. ProteinOrtho utilise l'approche de « *best hits* » bidirectionnels, également appelés « *best hits* » réciproques (Hirsh et Fraser, 2001 ; Jordan *et al.*, 2002) pour identifier les orthologues des paires de protéomes, cette méthode exige que pour une paire d'orthologues candidats "a", "b", que "a" soit le meilleur résultat pour "b" et que "b" soit le meilleur résultat pour "a". Ainsi, pour améliorer la précision de la prédiction, nous avons utilisé la « synténie » (*cf* « partie I » d'introduction générale) comme une caractéristique supplémentaire pour la discrimination de (co-)orthologues.

Il est intéressant de noter que le choix des seuils utilisés à l'étape de *clustering* des protéines pourrait affecter le nombre d'homologues détectés. Nous avons considéré de manière conservatrice les protéines ayant plus de 50 % et 70 %, et une couverture de séquence supérieure à 70 % et 80 % comme des homologues pour *F. graminearum* et *A. baumannii*, respectivement. La détermination de ces seuils était basée sur une comparaison de plusieurs valeurs d'identité (30% à 90%) et de couverture (60% à 90%) selon la conservation de la taille du pangénome construit comme rapporté par Plissonneau *et al.* (2018). D'autre part, pour extrapoler la taille du core et du pangénome et étudier leur nature (*open or closed*), nous avons ré-échantillonné toutes les combinaisons possibles de 1 à 20

¹⁰ **Paralogues** : Copies d'un gène présent dans le génome d'un organisme particulier issus d'évènements de duplication.

génomés pour *F. graminearum* et 1 à 23 pour *A. baumannii*, en utilisant « *proteinortho curves* » (Lechner *et al.*, 2011) avec un nombre de 100 itérations.

L'annotation des domaines fonctionnels du pangénome a été réalisée à l'aide d'InterProScan v5.13 (Jones *et al.*, 2014), avec les modèles prédictifs de la base de données Pfam et la classification de GO (*Gene Ontology*).

L'analyse des relations phylogénomique de chacune des deux espèces a été basée sur le core génome via un enchaînement d'étapes. Premièrement, les orthologues à copie unique identifiés dans les deux pangénomes ont été extraits et alignés à l'aide de MAFFT v7.471 (Katoh *et Standley*, 2013). Les régions mal alignées ont été supprimées en utilisant Gblocks v0.91b (Talavera *et Castresana*, 2007) avec les paramètres par défaut. L'analyse du maximum de vraisemblance (ML) a été réalisée avec RaxML v8.1.20 (Stamatakis, 2014) en utilisant le modèle de matrice de substitution d'acides aminés PROTGAMLMG avec 1,000 itérations de bootstrap. Ensuite les résultats des deux espèces ont été visualisés en utilisant FigTree v1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree>).

2.5. Prédiction du pan-résistome et du pan-virulome d'*A. baumannii*

Des bases de données avec un répertoire complet et précis de gènes de résistance aux antibiotiques et de virulence sont nécessaires pour identifier *in silico* le pan-virulome (fait référence à tous les gènes de virulence trouvés dans tous les génomes) et le pan-résistome de bactéries multirésistantes comme *A. baumannii*. Il existe plusieurs bases de données liées à la résistance aux antibiotiques qui sont spécialisées soit sur une seule espèce bactérienne comme *Escherichia coli* (Saha *et al.*, 2015) soit sur la résistance à une classe d'antibiotiques comme les β -lactamases (Naas *et al.*, 2017) ou multi-espèces (Doster *et al.*, 2020; Gibson *et al.*, 2015). Dans ce travail, le pan-résistome des treize souches d'*A. baumannii* a été prédit sur un ensemble de trois bases de données de référence. (i) BARRGD (BioProject : PRJNA313047), une base de données de référence contenant des séquences annotées pour plus de 5900 gènes de résistance aux antibiotiques. Ensuite, les résultats obtenus ont été validés et complétés avec deux autres bases de données largement utilisées, (ii) ResFinder v4.1 (Bortolaia *et al.*, 2020) et (iii) RGI (*Resistance Gene Identifies* ; (Jia *et al.*, 2017). Pour le pan-virulome, la base de données des facteurs de virulence des bactéries pathogènes (VFDB) contient plus de 1500 gènes connus pour être associés à la virulence (B. Liu *et al.*, 2019).

2.6. Prédiction du pan-sécrétome de *F. graminearum*

Dans cette section, nous décrivons les stratégies utilisées pour prédire le pan-sécrétome (fait référence à toutes les protéines sécrétées trouvées dans tous les génomes) de *F. graminearum* et leurs catégories fonctionnelles connues pour être associées potentiellement à la virulence ou pathogénicité

en utilisant des programmes et des bases de données spécialisés. Les protéines sécrétées partagent plusieurs caractéristiques structurales : (i) un peptide signal N-terminal, (ii) pas de domaine transmembranaire, (iii) pas de site d'ancrage glycosylphosphatidylinositol (GPI), (iv) aucun signal de localisation prédit pour cibler les mitochondries ou d'autres organites intracellulaires, etc. Il convient de noter également qu'il existe des protéines qui n'ont pas de séquence de peptide signal et malgré cela, elles peuvent être transloquées à leur emplacement respectif par la voie de sécrétion non conventionnelle « UPS » (Rabouille, 2017). L'analyse du pan-sécrétome de *F. graminearum* a été réalisée en utilisant le pipeline représenté dans la figure 11, qui associe différents outils complémentaires afin de produire des prédictions plus précises.

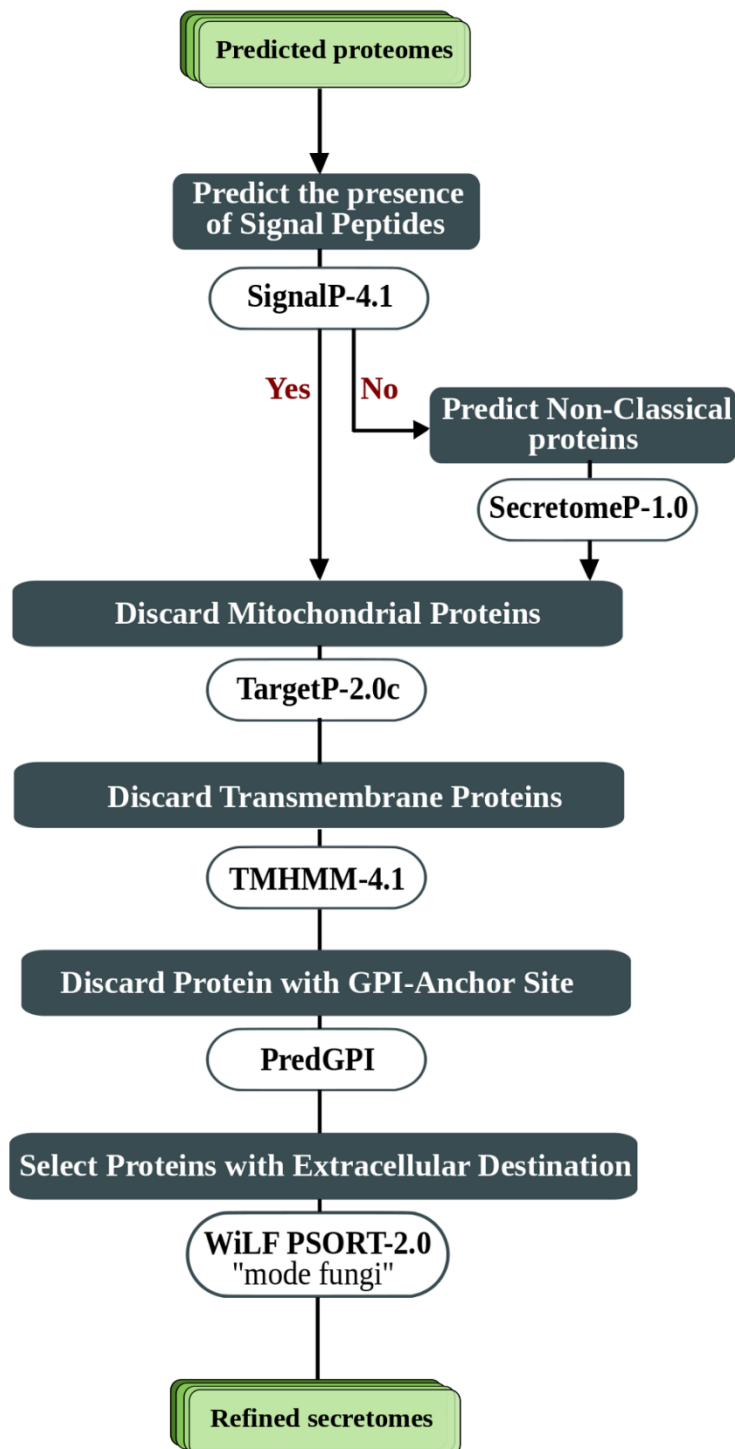


Figure 11 : Pipeline utilisé pour l'identification du sécrétome putatif.

La présence de peptide signal à l'extrémité N-terminale a été prédite à l'aide de SignalP 4.1 (Nielsen, 2017), permettant d'extraire des protéines susceptibles d'être sécrétées par une voie classique. Les séquences protéiques dépourvues du peptide signal ont été utilisées par SecretomeP v1.0 (Bendtsen *et al.*, 2004) pour la prédiction de l'UPS. Les résultats de SignalP et SecretomeP ont été analysés par TargetP v1.1b (Almagro Armenteros *et al.*, 2019) pour filtrer les protéines avec des signaux

d'adressage aux mitochondries ou à un autre organite intra-cellulaire. Ensuite, les protéines retenues ont été examinées pour la présence d'un domaine transmembranaire en utilisant TMHMM v2.0c (Krogh *et al.*, 2001), uniquement des protéines sans domaine transmembranaire et / ou ayant un seul domaine transmembranaire avec la présence d'un peptide signal à l'extrémité N-terminale a été retenu. Ainsi, toutes les protéines qui ont commencé avec la méthionine ont été utilisées comme *input* dans WolF-PSort v0.2 "mode fungi" (Horton *et al.*, 2007) pour sélectionner des protéines avec la destination extracellulaire ; ce programme rapporte les emplacements prédits de plusieurs compartiments ; seules les protéines ayant un score extracellulaire supérieur à 17 ont été retenues ; comme suggéré dans la méthode de Brown *et al.* (2012), ce qui indique une probabilité plus élevée que ces protéines soient effectivement sécrétées et « extracellulaires ». Enfin, seules les séquences protéiques sans GPI prédites par PredGPI (Pierleoni *et al.*, 2008) dans ces protéines extracellulaires ont été retenues comme candidates pour les protéines sécrétées.

Pour identifier les protéines sécrétées potentiellement associées à la virulence ou pathogénicité, la recherche BLAST « *Basic Local Alignment Search Tool Program* » pour protéine (BLASTP) a été effectuée par rapport à la PHI-base (cf. « section 2.2.4 » de l'introduction générale) qui comprend des gènes connus impliqués dans la virulence et la pathogénicité. Nous avons également prédit les enzymes protéolytiques et les IP présents dans le pan-sécrétome en utilisant la base de données MEROPS v12.1 (Rawlings *et al.*, 2018) ; une ressource d'informations organisée manuellement pour les enzymes protéolytiques et les inhibiteurs de protéines avec plus de 4000 séquences caractérisées et séquencées expérimentalement. Les recherches BLASTP sur les bases de données PHI et MEROPS ont été utilisées avec un seuil de e-value de 10^{-5} avec la meilleure correspondance et une identité supérieure à 40 %. Pour les CAZymes, nous avons effectué une recherche Hmmscan dans HMMER v3.1b2 (<http://hmmer.org/>) par rapport à la base de données de profils dbCAN HMM v6 (www.cazy.org ; détaillée dans la section 2.2.2.2 de l'introduction générale). Nous avons ensuite filtré les résultats à l'aide du script "hmmscan-parser.sh" fourni par dbCAN afin de sélectionner les correspondances significatives. En plus des effecteurs protéiques de type CAZymes, protéases et les IP, nous avons utilisé EffectorP v2.0 (Sperschneider *et al.*, 2018) ; une approche de prédiction d'apprentissage automatique spécialement développée pour prédire les candidats effecteurs à partir de sécrétome. Pour la détermination de la localisation du pan-effectome protéique (résultats d'effectorP et les trois bases de données PHI-base, MEROPS et dbCAN) à l'échelle de chromosome, nous avons d'abord extrait leurs correspondances en séquences CDS, puis GMAP v2020-06-30 (Wu *et al.*, 2016) a été utilisé pour les aligner sur les quatre chromosomes de *F. graminearum*. Les graphiques de densité génétique sur les chromosomes ont été générés en utilisant caryoploteR (Gel *et al.*, 2017).

2.7. Localisation subcellulaire des protéines sécrétées dans la cellule végétale

Des prédictions de localisation subcellulaire ont été faites sur le pan-sécrétome mature sous l'hypothèse que les protéines sécrétées doivent perdre leur séquence de peptide signal lors de la sécrétion et être transloquées dans l'hôte sans peptide signal. Le pan-sécrétome mature a été analysé pour leur localisation subcellulaire dans les cellules végétales en utilisant une combinaison de deux outils « avec des données de localisation vérifiées expérimentalement » : LOCALIZER v1.0 (Sperschneider *et al.*, 2017) en « mode effecteur » et WoLF PSORT v2.0 (Horton *et al.*, 2007) en « mode plante ». LOCALIZER facilite la prédiction de la translocation des protéines fongiques sécrétées dans les chloroplastes, les mitochondries et les noyaux végétaux en fonction de la présence de peptides transit et de NLS de plusieurs compartiments cellulaires (neuf sites). Il s'agit d'une méthode qui comprend des informations basées sur l'homologie des motifs fonctionnels et la recherche de signaux de localisation afin d'améliorer la précision de la prédiction de la localisation subcellulaire des protéines.

2.8. Résumé du workflow d'analyse génomique

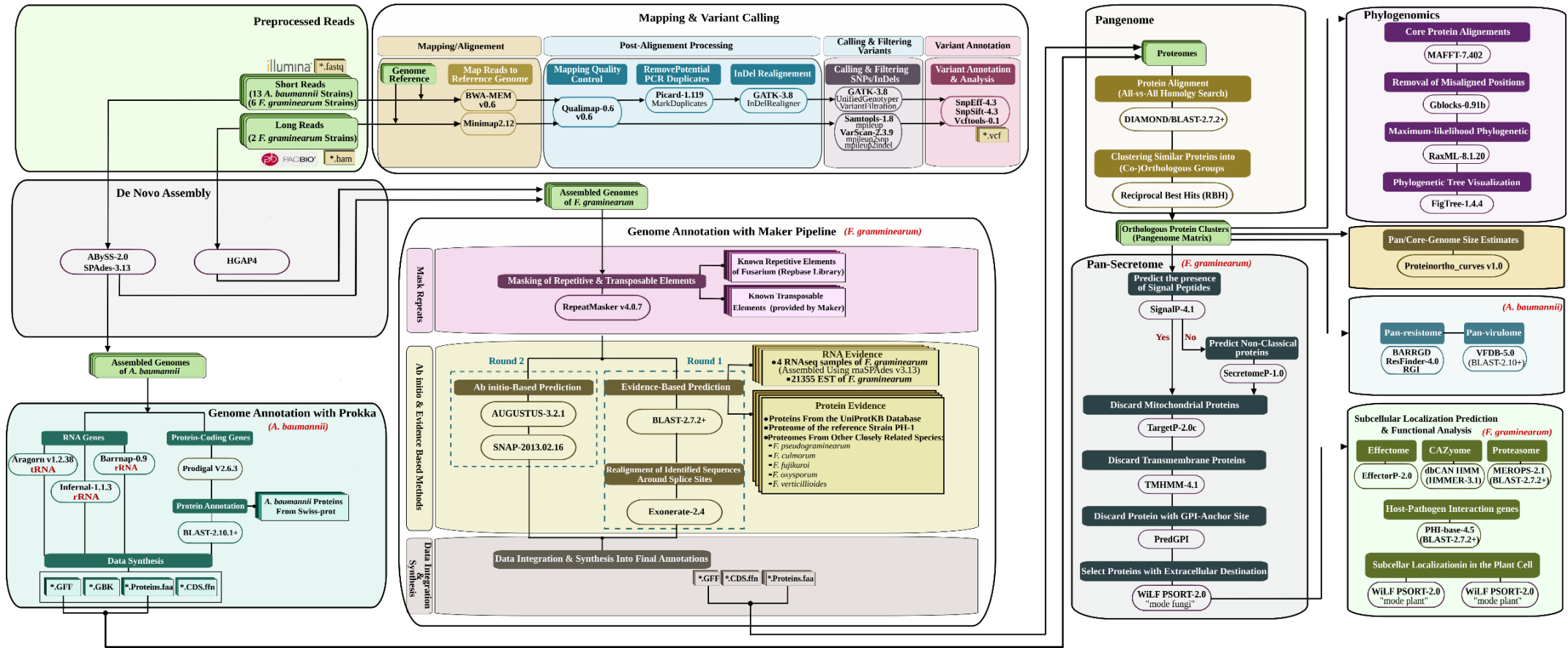


Figure 12 : Résumé du workflow d'analyse génomique de deux micro-organismes *A. baumannii* et *F. graminearum*.

2.9. Défis et orientations futures

Scores de confiance pour le *clustering* – Il n'existe pas une méthode de référence pour le *clustering* de séquences orthologues et les résultats de tout processus d'alignement de séquences dépendent fortement de l'algorithme et des paramètres choisis. Un seuil d'identité de séquence et de couverture trop strict entraînera une diminution du nombre de core gènes et une augmentation des gènes dispensables et vice versa, simplement parce que les séquences homologues sont séparées en différents groupes d'homologie. Cela pourrait impacter le niveau de conservation et de variabilité génétique et donc les analyses en aval. Comme indiqué à la « section 2.4 » (de ce chapitre), nous avons évalué la taille du pangénome en fonction de différents seuils d'identité et de couverture pour atteindre un équilibre acceptable entre les faux positifs (gènes qui ne sont pas orthologues mais classés comme tels) et les faux négatifs (vraies relations orthologues non détectées). Cette méthode nous permettra de générer plusieurs pangénomes de la même espèce afin de choisir la valeur optimale des seuils, ce qui n'est pas très pratique pour les grands génomes en termes de ressources de calcul, d'effort, et de temps. Cependant, une méthode plus sophistiquée qui pourrait générer systématiquement les scores de confiance sur la base de l'ensemble de gènes pour le *clustering* « sans passer par la construction de plusieurs pangénomes » est fortement recommandée.

Prédiction non conventionnelle des protéines sécrétées – La prédiction de l'UPS est une tâche beaucoup plus difficile que la prédiction des protéines sécrétées basée sur les peptides signaux, à la fois en raison du faible nombre de protéines UPS connues et de la connaissance limitée des voies impliquées, en particulier chez les eucaryotes. Cependant, l'idée que les protéines sécrétées partagent certaines propriétés quelle que soit la voie de sécrétion utilisée a permis de construire la méthode de prédiction la plus utilisée « SecretomeP » (Bendtsen *et al.*, 2004). SecretomeP a été entraîné sur des protéines de mammifères et de bactéries, cependant, ses performances sont moins bonnes que prévu pour d'autres types de données eucaryotes comme indiqué précédemment pour les protéines végétales (Lonsdale *et al.*, 2016 ; Nielsen *et al.*, 2019). Par conséquent, au lieu de s'appuyer sur des informations provenant d'autres lignées cellulaires qui peuvent être sujettes à des résultats faussement positifs, le développement futur d'outils de caractéristiques spécifiques aux protéines fongiques pourrait être une source d'améliorations significatives dans la prédiction des effecteurs protéiques.

Intégration des données avec le pangénome – Le stockage et la présentation des résultats dérivés d'analyses pangénomiques constituent également un défi. Les grands référentiels de données (ou bases de données internationales) devraient intégrer les informations relatives au pangénome de l'espèce et permettre un accès rapide à la communauté scientifique. L'intégration des données avec le

pangénome de l'espèce sera nécessaire pour relier le core génome et le génome dispensable avec des niveaux d'expression et des variants tels que les SNPs et les InDels. De telles informations fonctionnelles pourraient fournir un lien entre le pangénome et le trait observé.

Résultats – *F. graminearum*

Chapitre 2

Introduction

Comme indiqué dans la section d'introduction générale, la FHB causée principalement par l'agent pathogène *F. graminearum*, est une maladie dévastatrice des céréales à petits grains, en particulier le blé, entraînant une perte de rendement et une qualité réduite des céréales. Une analyse bibliographique des études réalisées précédemment à l'échelle génomique sur cette espèce révèle que les auteurs ont souvent concentré leurs efforts sur des analyses comparatives inter-espèces. Parmi ces projets, une analyse comparative de trois espèces du genre *Fusarium* à l'origine de la FHB : *F. graminearum*, *F. meridionale* et *F. asiaticum*, a révélé 12 % du répertoire total des gènes différenciant les trois espèces, et codent pour des protéines potentiellement associées à la FHB (Walkowiak *et al.*, 2016). En revanche, les analyses génomiques comparatives intra-espèces de *F. graminearum* sont largement limitées. Cependant, la conception et l'analyse du répertoire complet de gènes de *F. graminearum* via la construction du « pangénome » avec la définition de son ensemble de core génome et de son génome dispensable, devrait fournir de nouvelles informations précieuses dont la base génétique impliqué dans l'adaptation et les interactions plantes-pathogènes. Ce même concept du pangénome est attribué au pan-sécrétome, mais en mettant l'accent sur les gènes codant pour les protéines sécrétées, dont les effecteurs protéiques.

Au cours de ce travail, un séquençage du génome complet de huit nouvelles souches de *F. graminearum* sélectionnées pour leur agressivité contrastée sur blé. Cette agressivité a été évaluée sur des étapes précoces de l'interaction définies sur la base du travail de Thèse de Francis Fabre démontrant une forte discrimination de 3 souches fongiques 72 heures après le dépôt de spores dans les cavités florales (Fabre, *et al.*, 2019b). Nous avons ensuite utilisé et adapté un workflow d'analyse génomique comparative pour *F. graminearum* afin de : (i) caractériser la variabilité génomique de cette espèce via la construction de son premier pangénome, et (ii) étudier les liens possibles entre cette variabilité et la différence d'agressivité entre les souches, avec un accent particulier sur le « pan-sécrétome ». Ces analyses génèrent de nouveaux jeux de données permettant le développement d'un modèle descriptif pour le pangénome et le pan-sécrétome de *F. graminearum*, tout en constituant une ressource solide pour des investigations génomiques et génétiques ciblées.

number [SAMEA4778736](#) and BioProject accession number [PRJEB27611](#)). The PacBio reads are available in SRA under accession number [ERP109716](#). The version described in this paper is the first version, UIHA01000000.

ACKNOWLEDGMENTS

We thank Philippe Lecomte for preparing *Fusarium graminearum* DNA samples and the GENTYANE INRA platform for the Sequel sequencing (Pacific Biosciences). We also acknowledge Antoine Mahul and the Mesocentre Clermont-Auvergne for providing the computer calculation and storage facilities.

This work is supported by the Agence Nationale de la Recherche (NewMyco project, ANR-15-CE21-0010).

REFERENCES

1. Kazan K, Gardiner DM, Manners JM. 2012. On the trail of a cereal killer: recent advances in *Fusarium graminearum* pathogenomics and host resistance. *Mol Plant Pathol* 13:399–413. <https://doi.org/10.1111/j.1364-3703.2011.00762.x>.
2. Dean R, Van Kan JAL, Pretorius ZA, Hammond-Kosack KE, Di Pietro A, Spanu PD, Rudd JJ, Dickman M, Kahmann R, Ellis J, Foster GD. 2012. The top 10 fungal pathogens in molecular plant pathology. *Mol Plant Pathol* 13:414–430. <https://doi.org/10.1111/j.1364-3703.2011.00783.x>.
3. Cuomo CA, Guldener U, Xu J-R, Trail F, Turgeon BG, Di Pietro A, Walton JD, Ma L-J, Baker SE, Rep M, Adam G, Antoniw J, Baldwin T, Calvo S, Chang Y-L, DeCaprio D, Gale LR, Gnerre S, Goswami RS, Hammond-Kosack K, Harris LJ, Hilburn K, Kennell JC, Kroken S, Magnuson JK, Mannhaupt G, Mauceli E, Mewes H-W, Mitterbauer R, Muehlbauer G, Munsterkott M, Nelson D, O'Donnell K, Ouellet T, Qi W, Quesneville H, Roncero MIG, Seong K-Y, Tetko IV, Urban M, Waalwijk C, Ward TJ, Yao J, Birren BW, Kistler HC. 2007. The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. *Science* 317:1400–1402. <https://doi.org/10.1126/science.1143708>.
4. King R, Urban M, Hammond-Kosack MCU, Hassani-Pak K, Hammond-Kosack KE. 2015. The completed genome sequence of the pathogenic ascomycete fungus *Fusarium graminearum*. *BMC Genomics* 16:544. <https://doi.org/10.1186/s12864-015-1756-1>.
5. Gardiner DM, Stiller J, Kazan K. 2014. Genome sequence of *Fusarium graminearum* isolate CS3005. *Genome Announc* 2:e00227-14. <https://doi.org/10.1128/genomeA.00227-14>.
6. Zapparata A, Lio D, Somma S, Muñoz V, Malfatti L, Vannacci G, Moretti A, Baroncelli R, Sarrocco S. 2017. Genome sequence of *Fusarium graminearum* ITEM 124 (ATCC 56091), a mycotoxigenic plant pathogen. *Genome Announc* 5:12–13. <https://doi.org/10.1128/genomeA.01209-17>.
7. Laurent B, Moinard M, Spataro C, Ponts N, Barreau C, Foulongne-Oriol M. 2017. Landscape of genomic diversity and host adaptation in *Fusarium graminearum*. *BMC Genomics* 18:203. <https://doi.org/10.1186/s12864-017-3524-x>.
8. Chetouhi C, Bonhomme L, Lecomte P, Cambon F, Merlino M, Biron DG, Langin T. 2015. A proteomics survey on wheat susceptibility to *Fusarium* head blight during grain development. *Eur J Plant Pathol* 141:407–418. <https://doi.org/10.1007/s10658-014-0552-0>.
9. Chetouhi C, Bonhomme L, Lasserre-Zuber P, Cambon F, Pelletier S, Renou J-P, Langin T. 2016. Transcriptome dynamics of a susceptible wheat upon *Fusarium* head blight reveals that molecular responses to *Fusarium graminearum* infection fit over the grain development processes. *Funct Integr Genomics* 16:183–201. <https://doi.org/10.1007/s10142-016-0476-1>.
10. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>.



Article

Comparative Genomics of Eight *Fusarium graminearum* Strains with Contrasting Aggressiveness Reveals an Expanded Open Pangenome and Extended Effector Content Signatures

Tarek Alouane ¹, Hélène Rimbart ¹, Jörg Bormann ², Gisela A. González-Montiel ³, Sandra Loesgen ^{3,4}, Wilhelm Schäfer ², Michael Freitag ⁵, Thierry Langin ^{1,*} and Ludovic Bonhomme ^{1,*}

- ¹ Diversité, Ecophysiologie des Céréales, Université Clermont Auvergne, INRAE, UMR 1095 Génétique, 63000 Clermont-Ferrand, France; alouane.tarek@gmail.com (T.A.); helene.rimbart@inrae.fr (H.R.)
² Molekulare Phytopathologie, Institut für Pflanzenwissenschaften und Mikrobiologie, Universität Hamburg, 22609 Hamburg, Germany; bormannj@gmail.com (J.B.); wilhelm.schaefer@gmail.com (W.S.)
³ Department of Chemistry, Oregon State University, Corvallis, OR 97331, USA; gonzalgi@oregonstate.edu (G.A.G.-M.); sandra.loesgen@whitney.ufl.edu (S.L.)
⁴ Whitney Laboratory for Marine Bioscience, Department of Chemistry, University of Florida, St. Augustine, FL 32080, USA
⁵ Department of Biochemistry and Biophysics, Oregon State University, Corvallis, OR 97331, USA; Michael.Freitag@oregonstate.edu
* Correspondence: thierry.langin@inrae.fr (T.L.); ludovic.bonhomme@inrae.fr (L.B.)



Citation: Alouane, T.; Rimbart, H.; Bormann, J.; González-Montiel, G.A.; Loesgen, S.; Schäfer, W.; Freitag, M.; Langin, T.; Bonhomme, L. Comparative Genomics of Eight *Fusarium graminearum* Strains with Contrasting Aggressiveness Reveals an Expanded Open Pangenome and Extended Effector Content Signatures. *Int. J. Mol. Sci.* **2021**, *22*, 6257. <https://doi.org/10.3390/ijms22126257>

Academic Editor: Manuel Simões

Received: 7 May 2021

Accepted: 7 June 2021

Published: 10 June 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: *Fusarium graminearum*, the primary cause of Fusarium head blight (FHB) in small-grain cereals, demonstrates remarkably variable levels of aggressiveness in its host, producing different infection dynamics and contrasted symptom severity. While the secreted proteins, including effectors, are thought to be one of the essential components of aggressiveness, our knowledge of the intra-species genomic diversity of *F. graminearum* is still limited. In this work, we sequenced eight European *F. graminearum* strains of contrasting aggressiveness to characterize their respective genome structure, their gene content and to delineate their specificities. By combining the available sequences of 12 other *F. graminearum* strains, we outlined a reference pangenome that expands the repertoire of the known genes in the reference PH-1 genome by 32%, including nearly 21,000 non-redundant sequences and gathering a common base of 9250 conserved core-genes. More than 1000 genes with high non-synonymous mutation rates may be under diverse selection, especially regarding the trichothecene biosynthesis gene cluster. About 900 secreted protein clusters (SPCs) have been described. Mostly localized in the fast sub-genome of *F. graminearum* supposed to evolve rapidly to promote adaptation and rapid responses to the host's infection, these SPCs gather a range of putative proteinaceous effectors systematically found in the core secretome, with the chloroplast and the plant nucleus as the main predicted targets in the host cell. This work describes new knowledge on the intra-species diversity in *F. graminearum* and emphasizes putative determinants of aggressiveness, providing a wealth of new candidate genes potentially involved in the Fusarium head blight disease.

Keywords: *Fusarium graminearum*; Fusarium head blight; pangenome; secreted protein clusters; proteinaceous effectors; aggressiveness; intra-species genomic diversity

1. Introduction

Fusarium head blight (FHB), caused mainly by the Ascomycete fungus *Fusarium graminearum*, is a devastating disease of small grain cereals, and in particular wheat, which results in yield losses and reduced grain quality [1,2]. The main consequence of FHB is the accumulation of mycotoxins, such as trichothecenes deoxynivalenol (DON) and its derivatives, in the grain, which poses serious health risks, including immunological and teratogenic effects in humans [3–5]. *F. graminearum* has been defined as a member of the *F. graminearum* species complex (FGSC), comprising at least 16 phylogenetically distinct

species found worldwide [6,7]. *F. graminearum* is the predominant species associated with FHB in North and South America [8–12], in China [13,14] and in most European countries, e.g., France [15–17], Germany [18–20] and Italy [21–23].

Genome decoding is the first step to decipher the range of molecular mechanisms driving diversification, genetic variability and host–pathogen interactions. In *F. graminearum*, the genome size reaches 36–38 Mb, with about 14,000 identified gene models and a bipartite “two-speed genome” architecture [24–29]. Horizontal gene transfer events have already been suggested between *F. graminearum* and other fungi [30]. However, our knowledge of the complete gene set of *F. graminearum* still remains incomplete, because a single reference genome does not reflect the genetic variability within a species. Genetic variants, new haplotypes, change total genetic content and thus expand the number of unique gene models to the pangenome concept [31–34]. Hence, the pangenome includes (i) the core genome, i.e., genes conserved in all isolates or strains, and (ii) the dispensable or accessory genome, i.e., genes shared by only a subset of isolates or strains, or even specific to a single genome [35]. Such compendia offer unprecedented opportunities to discover new genes, explore genomic diversity and improve our knowledge of the evolutionary forces that shape the organization and dynamics of the genomes. In the *F. graminearum* species, designing and analyzing its full gene repertoire and defining its core gene set could provide new knowledge on the genetic determinants of plant–pathogen interactions, further discovering original ways of deciphering the molecular components of pathogenicity.

During the course of infection, the pathogen synthesizes a complex arsenal of secreted proteins, particularly proteinaceous effectors able to deregulate the main plant biological processes, such as the immune system, in order to facilitate the pathogen development in the plant host tissues [36,37]. These proteinaceous effectors are secreted and maintained either in the apoplast, or translocated into plant cells to target specific intracellular processes [38]. Known secreted effectors include protease inhibitors that override host proteases and active carbohydrate enzymes (CAZymes) that can weaken the plant cell walls [39–42]. Based on the first sequenced genome of *F. graminearum* (strain PH-1 [24]), Brown et al. [43] predicted a secretome of 574 proteins. The contrasting aggressiveness of *F. graminearum* strains on wheat and their ability to synthesize a quantitatively variable “effectome” [17] strongly suggest that the genetic determinant of pathogenicity, such as the effector content and sequence, are not similar. However, our knowledge of the profile and functional characteristics of the secretomes of strains harboring contrasting aggressiveness remains poorly documented and still requires to be connected with phenotype data.

Along with proteinaceous effectors, *F. graminearum* also has the ability to produce a set of secondary metabolites. Some of them are known to be involved in virulence, achieving effector-like impacts in the host plants [44–47]. Secondary metabolites are produced by proteins encoded by genes found in biosynthetic gene clusters (BGCs) [30,48], within which the expression of a few genes may be sufficient to the biosynthesis of a particular secondary metabolite [49–51]. However, until now, no study has evaluated the impact of nucleotide polymorphism of such key genes on the biosynthesis of *F. graminearum* secondary metabolites under different conditions, in planta or in vitro.

In this work, we analyzed the genome of *F. graminearum* strains displaying contrasting aggressiveness on bread wheat, their regular host in fields. First, we approximated the repertoire of the *F. graminearum* genes and analyzed their functional characteristics to qualify the intra-specific diversity. Then, we evaluated the contribution of this genomic diversity to the contrasting aggressiveness of the strains, with a particular emphasis on the putative secreted proteins, which are assumed to be involved in host–pathogen interaction. In addition, the impact of mutations in BGCs on the production of secondary metabolites and their relationship with metabolomic profiles was also assessed. These analyses generate new datasets allowing the development of a descriptive model of the complete genome of *F. graminearum*, providing access to a multitude of candidate secreted proteins potentially involved in the pathogenicity of *F. graminearum* and establishing a solid database for further targeted genomic and genetic research.

2. Results

2.1. Whole Genome Sequencing, Assembly and Annotation

We report the genome sequencing of eight European *F. graminearum* strains, including four French, three Italian and one German isolate (Supplementary Table S1). More than 23,849,853 reads were produced for each of the six strains sequenced by Illumina HiSeq and more than 854,603 subreads with an average length of 7 kb were generated using the PacBio Sequel for the two other strains. The size of the assembled genomes ranged from 36.28 to 37.13 Mb, with a GC content of approximately 48%, as expected for this species. The N50 value of these assemblies ranged between 0.51 and 2.28 Mb, illustrating a reliable contiguity quality. The detailed assembly statistics for the genome of all isolates are summarized in Table 1.

Table 1. Genome assembly and annotation statistics of the eight *F. graminearum* genomes of this study.

Strains	MDC_Fg1	MDC_Fg5	MDC_Fg8	MDC_Fg13	MDC_Fg202	MDC_Fg851	MDC_Fg593	MDC_FgU1
Assembly statistics								
Genome size (Mb)	36.80	36.56	36.28	36.75	37.13	36.75	36.45	37.01
k-mer for ABySS assembly	Na ^a	128	41	96	96	96	96	Na ^a
Total number of contigs	96	297	1.599	283	292	278	203	139
Total number of contigs > 1 Kb	95	131	1.327	176	185	155	140	139
Max contig size (Mb)	3.41	7.74	0.27	2.33	1.87	1.98	2.88	1.67
N50 (Mb)	1.65	2.28	0.51	0.79	0.77	0.65	1.02	0.65
G + C content (%)	47.97	48.16	48.29	48.32	48.28	48.32	48.33	48
BUSCO completeness (%)	98.63	99.41	98.90	99.46	99.38	99.41	99.41	97.85
Gene annotation								
Number of gene models	13,166	12,790	12,993	13,297	13,189	13,316	12,729	13,014
Total gene length (Mb)	21.69	22.80	22.63	21.93	23.14	22.02	21.66	22.85
Mean gene length (bp)	1648	1783	1742	1650	1755	1654	1702	1756
Coding region in genome (%)	50.7	52.6	52.8	51.6	52.6	51.8	51.3	51.6
Number of exons	36,189	38,400	38,475	36,415	39,181	36,670	35,721	38,824
Total exon length (Mb)	19.65	20.29	20.18	19.96	20.59	20.03	19.72	20.16
Mean exons per gene	2.7	3.0	3.0	2.7	3.0	2.8	2.8	3.0
Mean exon length	543	528	525	548	526	546	552	519
Number of introns	23,023	25,610	25,482	23,118	25,992	23,354	22,992	25,810
Total intron length (Mb)	2.09	2.57	2.50	2.03	2.60	2.04	2.0	2.74
Mean intron per gene	1.7	2.0	1.9	1.7	1.9	1.7	1.8	2.0
Mean intron length (bp)	91	100	98	88	100	87	87	106
Protein functions								
Genes with IPR/Pfam annotations	8536	8345	8409	8653	8504	8675	8634	8400
Genes with GO annotations	5501	5449	5469	5570	5531	5584	5558	5501
Secreted proteins	659	653	650	692	667	692	670	662

^a Assembled by HGAP4.

The completeness of the assembled genomes reached from 97.8% to 99.5% of 3725 core Sordariomycete genes (Supplementary Table S2), with a total number of gene models ranging from 12,729 to 13,316. The cumulative length of the total number of genes accounted for more than 50% of the total length, and the average gene length ranged from 1648 to 1783 bp. In each genome, more than 80% of all annotated genes displayed an Annotation Edit

Distance (AED) score between 0 and 0.25, showing high consistency between annotation and overlaps of the ESTs, the mRNA-Seq data and the protein homologies used as evidence. The functional annotation showed that more than 8300 proteins (approximately 65% of the predicted whole proteome) in each strain had known functions, of which more than 64% harbored GO terms (Table 1 and Supplementary Tables S3–S10). The most abundant Pfam domain was the transcription factor “Zn(2)-Cys(6) binuclear cluster” (PF00172), followed by Major Facilitator Superfamily (MFS) transporter (PF07690) with more than 257 and 226 proteins respectively, for each strain.

2.2. Pangenome Analysis and Phylogenomic Relationships of *F. graminearum* Strains

In order to characterize the intra-species genomic diversity of *F. graminearum* and approximate its complete repertoire of genes, we constructed a pangenome by clustering the protein sets encoded in twenty *F. graminearum* genomes, including the eight from this study and twelve other genomes publicly available (Supplementary Table S1). The clustering parameters were adjusted according to Plissonneau et al. [52] to obtain the most significant homology clusters (Supplementary Table S11). A total of 259,881 proteins were pooled into an open pangenome of 20,807 non-redundant orthologous proteins, of which 9247 (44%) were conserved in the 20 genomes and defined the core genome (Figure 1A, Supplementary Table S12). Nearly 75% (i.e., 6871) of the predicted proteins encoded by the core genome harbored known functions. The accessory genome accounted for 11,560 (56%) non-redundant predicted proteins, of which 7431 (64%) were shared in at least 2 genomes while 4129 (36%) were found in only 1 genome (singleton proteins). Only 34% (i.e., 3958) of the accessory genome had assigned functions. The number of singleton proteins across all genomes ranged from 32 for the INRA-171 strain to 760 for the PH-1 strain (Figure 1A). Among the eight strains of this study, the MDC_FgU1, MDC_Fg1 and MDC_Fg8 ones contained the largest number of singleton proteins, with 383, 318 and 239 respectively, while the MDC_Fg593 and MDC_Fg5 held the lowest number, with 43 and 67 singleton proteins, respectively. In addition to the non-redundant predicted proteins, only 150 redundant paralog proteins were found among the 20 genomes of *F. graminearum* (Supplementary Table S13). The gene accumulation curves show that the numbers of the core genes decreased continually with the addition of new strains (Figure 1B), as expected when sampling more diverging genomes of a species. In contrast, the increasing number of strains systematically yielded novel genes, demonstrating the open state of the pangenome of *F. graminearum*. The core genome was significantly enriched in fundamental processes, including the following GO terms: “biological regulation”, “regulation of cellular process”, “localization” and “transport” (Figure 1C). At the molecular function scale, the core genome was enriched with proteins mainly acting in “transcription regulator activity”, “small molecule sensor activity”, “phosphatase activity” and “phosphotransferase activity”. The accessory genome displayed significant enrichment in proteins involved in “metabolic process”, “binding”, “oxidoreductase activity”, “protein kinase activity” and other catalytic activity. In addition, the cellular component GO terms were enriched only in the core genome (Figure 1C).

The phylogenomic relationships were calculated using the protein sequence homologies of the core genome of the twenty *F. graminearum* genomes and three others of the genus *Fusarium* (*F. pseudograminearum*, *F. culmorum* and *F. fujikuroi*) used as out-groups. Based on 7640 single-copy orthologs, a close relationship between the European genomes of *F. graminearum* was observed, including 7 French, 3 Italian and the German one (Figure 2). The two French MDC_Fg1 and MDC_Fg5 genomes belonged to distinct clades compared to the European genomes. In addition, the non-European *F. graminearum* genomes scattered in the phylogenomic tree. For instance, among the three Chinese genomes, HN-Z6 and HN9-1 have been grouped with the American genome DAOM_233423, while the other American genome, PH-1, was closely related to the Australian strain, CS3005. This analysis also showed that, based on these data, the species *F. graminearum* was more closely related to *F. culmorum* than to *F. pseudograminearum*.

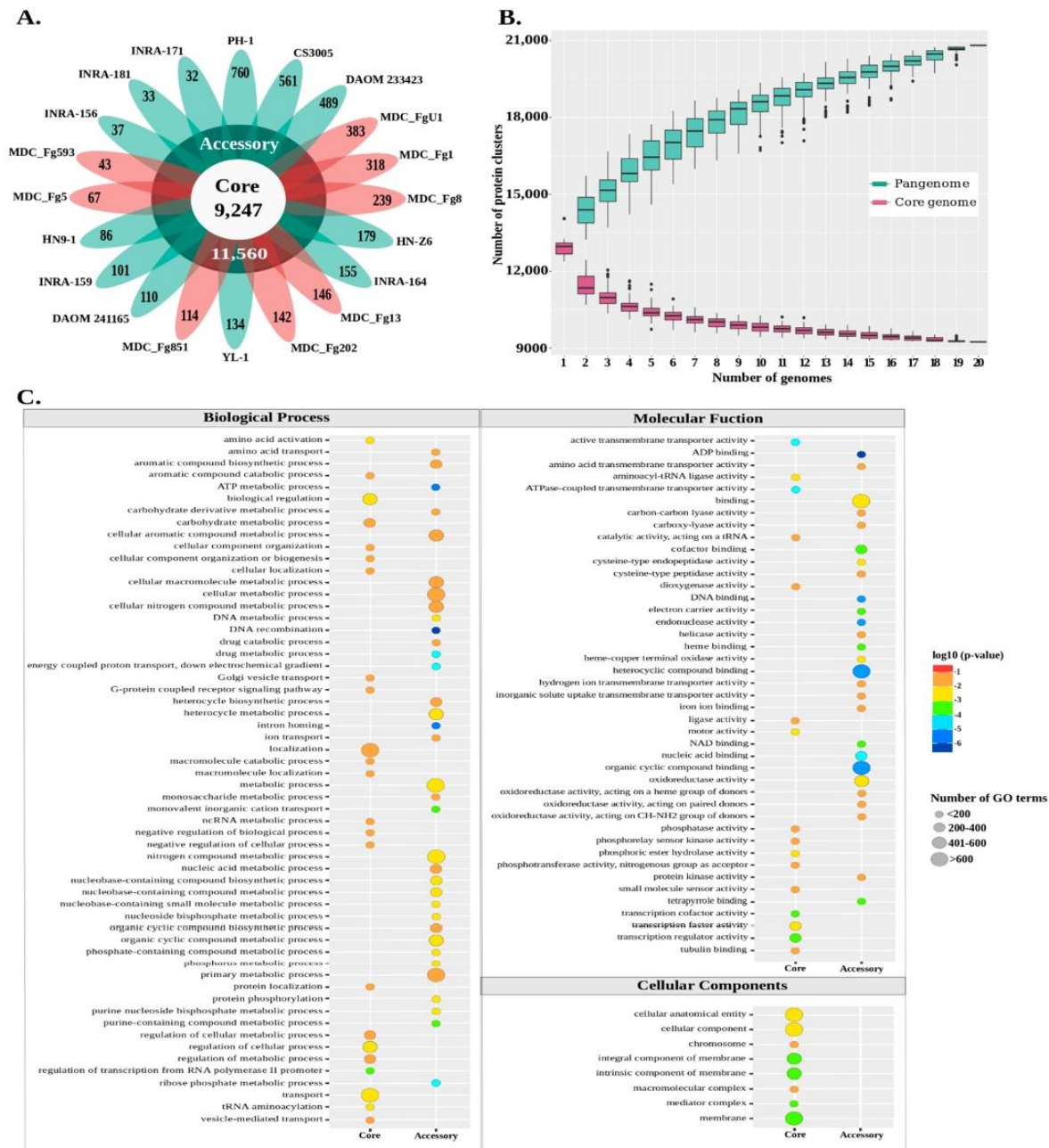


Figure 1. Structure and function of the *F. graminearum* pangenome. (A) The flower plot displays the number of core proteins (in the center) and accessory proteins (in the annulus), including the singletons (in the petals), encoded by the genome of 20 *F. graminearum* strains. (B) The *F. graminearum* pangenome accumulation curve is depicted with turquoise boxplots, indicating the number of accessory proteins discovered with the sequential addition of new genomes. The purple boxplots illustrate the number of core proteins discovered with the sequential addition of new genomes. (C) The density plot depicts the functional enrichment of the core and accessory *F. graminearum* genomes. Enriched GO terms belonging to biological processes, molecular functions and cellular components are presented. The color of the circle represents the statistical significance of the enriched GO terms. The size of the circles represents the number of occurrences of the corresponding GO term.

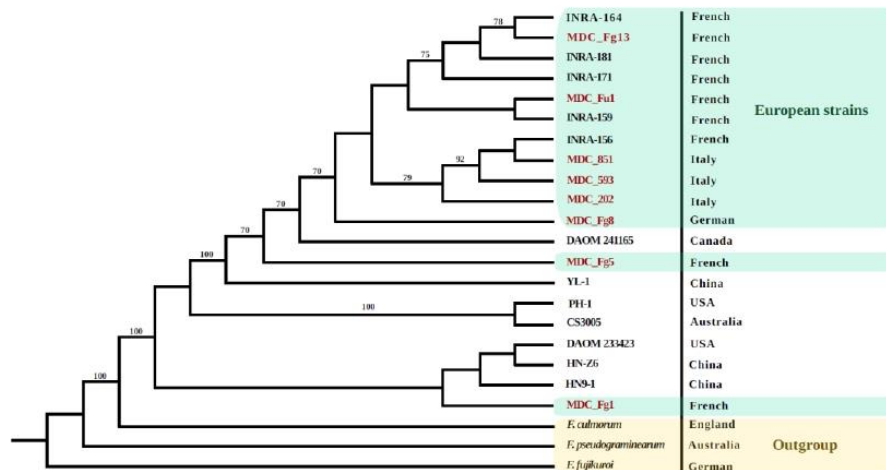


Figure 2. Phylogenetic relationship of 20 *F. graminearum* strains from different geographic origins. A Maximum likelihood bootstrapped cladogram was generated from 7640 single-copy protein sequences found in the 20 *F. graminearum* strains and from 3 other species of the *Fusarium* genus used as an outgroup (*F. pseudograminearum*, *F. culmorum* and *F. fujikuroi*). Numbers at nodes indicate a bootstrap value greater than 70%, performed with 1000 replications. The country of origin is shown on the right. Names of the 8 strains characterized in this study are shown in red.

2.3. Distribution and Analysis of Genetic Variants in the Eight *F. graminearum* Genomes

As a whole, 84% to 98% of the reads produced for the eight genomes were accurately mapped to the reference genome PH-1 (RRES v5.0), covering from 66.72- to 124.56-fold (Supplementary Table S14). Genetic variants were preferentially searched for SNPs and small InDels (less than 50 bp). A total of 353,055 highly confident variant sites differentiated the eight genomes, including 334,087 SNPs (95%) and 18,968 InDels (5%), with an average transition to transversion mutation ratio of 2.57. The number of variants per genome ranged from 108,634 to 145,405 for SNPs and from 4948 to 6568 for InDels (Supplementary Table S14), which is consistent with previous reports [16,53]. The distribution of the expected effects of genetic variants in genomic regions has been studied. The annotation of the 353,055 variant sites predicted nearly 527,000 functional effects. Of these, 61.2% were distributed in the intergenic regions (including the downstream and upstream variants) and 38.8% in the intragenic regions (Figure 3A). Among intragenic variant effects, nearly 87.5% were found in the exonic regions, with more than half (55.8%) having a synonymous effect and 42.1% a non-synonymous one (41.1% were missense effects, 0.9% produced a gain or a loss of stop codon and 0.1% produced a loss of start codon). The remaining 2.1% included both in-frame deletions (deletion of one or several codons) and frameshifts. The number of mutations per gene in each strain showed that MDC_Fg1 systematically displayed the lowest number of genes, with at least one synonymous mutation, followed by the MDC_Fg5 strain (Figure 3B). A similar ranking was observed for the number of genes containing 1 to 2 non-synonymous mutations.

Considering the 8 genomes, a total of 1128 genes accumulate more than 6 non-synonymous mutations and display a ratio of non-synonymous mutation on the total number of synonymous + non-synonymous mutations higher than 0.6 (Supplementary Table S15). About 46% of these genes ($n = 522$) displayed known functions, including transcription factors, cytochrome P450, protein kinases and enzymes involved in the secondary metabolism. Their distribution along the four chromosomes of the PH-1 reference genome showed that they are preferentially located in the proximal regions of the telomeres and in the central regions of the chromosomes. Chromosome 2 had the highest number of these genes despite being shorter than chromosomes 1 and 4 (Figure 3C).

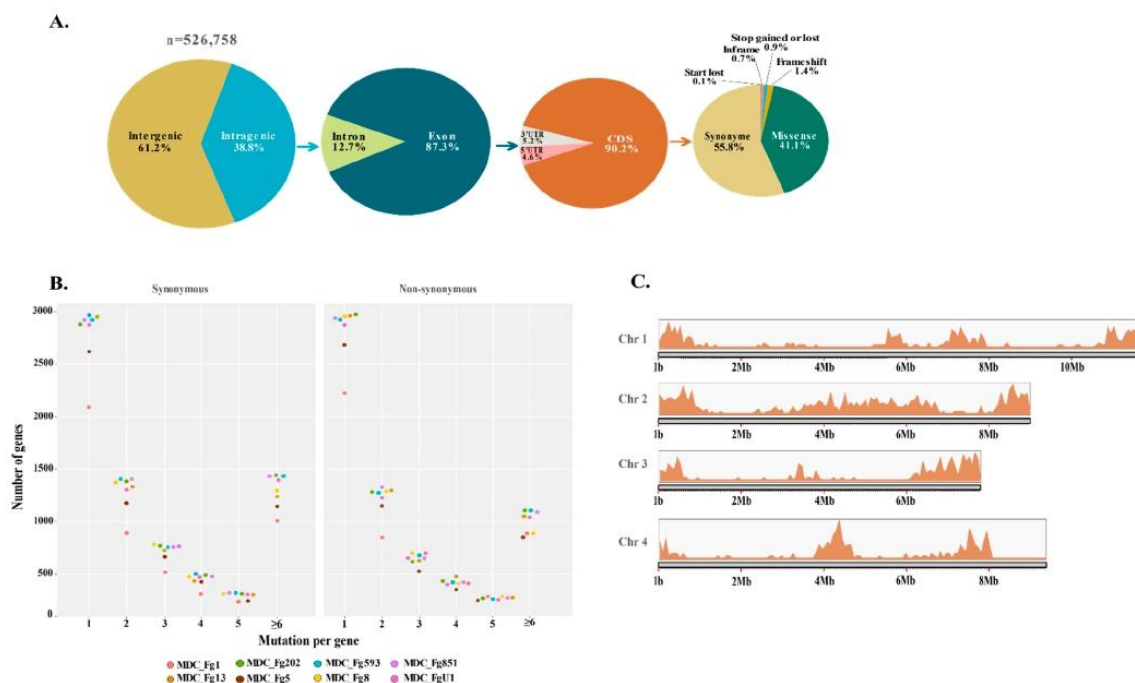


Figure 3. Genome-wide genetic variant distributions. (A) Pie charts represent the percentage distribution of the effects of mutations in different genomic regions (inter- and intra-genic), including different types of identified mutations in coding sequences. All the effects expected from the mutations identified in the eight *F. graminearum* genomes, relative to the PH-1 reference sequence, are combined. (B) Count distribution of synonymous and non-synonymous mutations per gene is depicted for the eight *F. graminearum* strains. The colored dots correspond to each of the eight genomes. (C) Density and localization of genes with a high rate of non-synonymous mutations are depicted along the four *F. graminearum* chromosomes. Only genes with at least six non-synonymous mutations and exhibiting a non-synonymous/(synonymous + non-synonymous) ratio greater than 0.55 were included. The size of the chromosomes is in Mb, in 2 Mb increments.

Among the range of identified mutations, we noticed 1283 genes displaying loss-of-function (LOF) mutations (Supplementary Table S16). The number of LOF genes for each genome ranged from 396 (in the MDC_Fg1 strain) to 582 (in the MDC_FgU1) (Supplementary Figure S1) and only 6% ($n = 79$) were shared in the 8 genomes, while more than 30 genes were specifically lost in each genome.

2.4. Accumulation of Mutations in Secondary Metabolite Gene Clusters and Relationships with Metabolomic Profiles

We examined genetic variants in the 67 known secondary metabolite gene clusters characterized in *F. graminearum* [30]. The trichothecene biosynthetic gene cluster (BGC), encoding enzymes and proteins directly involved in the virulence on wheat tissues [54], displayed the highest number of variants compared to all the other BGCs. The trichothecene BGCs are defined by 15 *TRI* genes: 12 are grouped in chromosome 2, while 3 are distributed on 3 other chromosomes (Supplementary Table S17). A total of 1316 and 1165 different mutations was found in the intra- and inter-genic regions of trichothecene BGCs, respectively (Supplementary Table S18). The two Italian genomes MDC_Fg202 and MDC_Fg593 displayed a particularly high number of genetic variants, gathering 96% of the total intra-genic mutations. These corresponded to 306 non-synonymous and 6 frameshift mutations in MDC_Fg593 and 224 non-synonymous and 7 frameshifts in MDC_Fg202. The *TRI12* (coding for the MFS transporter involved in the efflux of trichothecene) and the *TRI8* (coding for the C-3 esterase [55]) genes accumulate more non-synonymous mutations than

other *TRI* genes (68 in *TRI12* and 66 in *TRI8* found in MDC_Fg593, and 55 in *TRI12* and 72 in *TRI8* found in MDC_Fg202). The other six genomes harbored a smaller number of non-synonymous mutations in the *TRI* genes, with only two in MDC_Fg1 and MDC_Fg5, thirteen in MDC_Fg8 and MDC_Fg851 and sixteen in MDC_Fg13 and MDC_FgU1. In addition, a total of 210 LOF mutations were identified in 107 genes belonging to 47 secondary metabolite BGCs, including the trichothecene, the fusarin, the fusaristatin, the butanolide and the zearalenone BGCs (Supplementary Table S19).

Given the high rate of genetic variants in these secondary metabolite BGCs, we profiled the metabolite content of each of the eight *F. graminearum* strains, using a shot-gun metabolomics approach. Metabolite profiling shows substantial differences in metabolite abundance, in the eight strains, measured by ion abundances (Figure 4). MDC_FgU1 produced the most 12,13-epoxy-9,10-trichoene-2-ol. The two Italian strains (MDC_Fg202 and MDC_Fg593), that accumulate a high number of genetic variants in the trichothecene BGC, displayed the highest abundance of the trichothecene intermediates (Figure 4 and Supplementary Figure S2). These two strains also displayed the highest amount of fusarins. Regarding the range of LOF mutations, no obvious links were evidenced between enzymes with truncated sequences and the expected metabolite they are supposed to produce. For instance, intermediate abundances of fusarin C were detected in the MDC_Fg8 strain when it was observed to harbor a LOF mutation in the *fus1* gene (FGSG_07798), one of the pivotal enzymes required for the synthesis of this metabolite. Another discrepancy has been noted for zearalenone found at the highest abundance in the MDC_Fg851 strain, where this strain includes a LOF mutation in a gene responsible for its production. A non-ribosomal peptide synthetase (NRPS7, FGSG_08209) belonging to the gene cluster that mediates the biosynthesis of fusaristatin A, and a transporter (FGSG_08084) belonging to the butanolide BGC, were found with a LOF mutation in MDC_Fg5 and MDC_FgU1 respectively, while high abundance of these metabolites was found in these two strains.

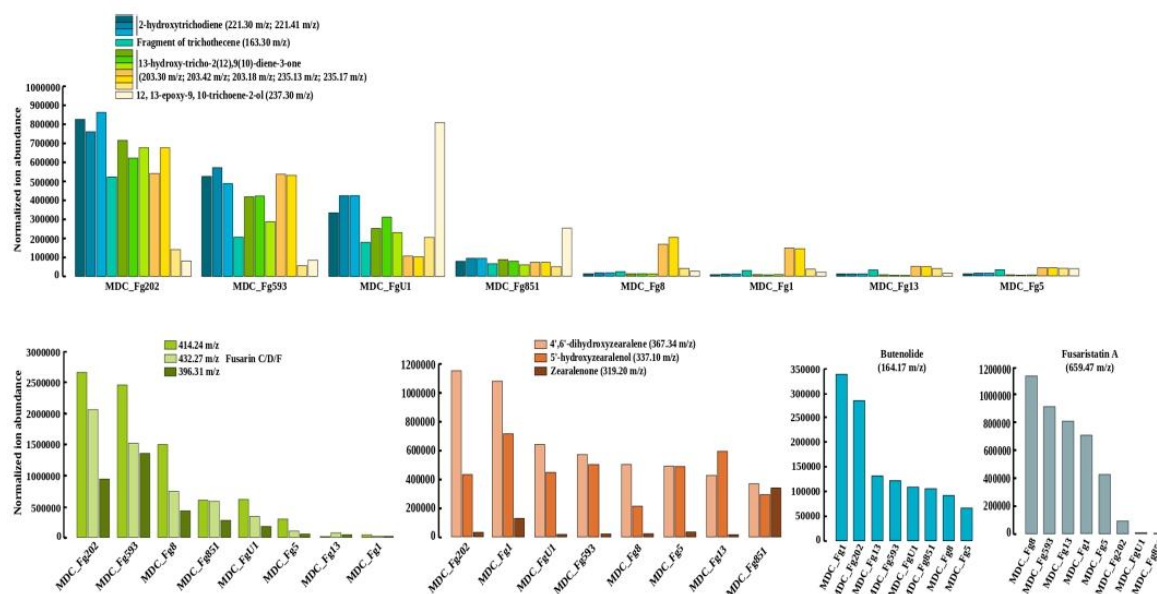


Figure 4. Abundances of selected metabolites in the eight *F. graminearum* strains. Ion abundances were derived from extracted ion chromatography (XIC) profiles obtained from mass spectrometry analysis (LC/MS) of the eight strains grown on PDA medium. Different detected ions corresponding to trichothecene biosynthesis (see Supplementary Figure S2 for details, mass-to-charge ratio (m/z): 163.30, 203.30, 203.42, 203.18, 235.13, 235.17, 221.41, 237.30), fusarin isoforms (retention time 19.8 min and m/z values of 396.31 for M-2xH₂O+H, 414.24 for M-H₂O+H, 432.27 for M+H), zearalenone and its intermediates (retention times 21.1, 23.5 and 34.3 min with m/z values of 319.20, 337.10, 367.34, respectively), Butanolide (retention time 2.9 min with m/z value of 164.17) and Fusaristatin A (retention time 27.9 min and m/z value of 659.47) are reported.

2.5. Characterization of the Aggressiveness of the Eight *F. graminearum* Strains on Wheat Spikes

Our previous work has shown that a rating of symptoms during the first 72 h on the FHB-susceptible winter wheat Recital clearly differentiated three different strains of *F. graminearum* [17]. The aggressiveness of the eight *F. graminearum* strains was thus evaluated according to the same procedure by rating inoculated spikelets at 72 hpi, specifically on this wheat cultivar (Figure 5). The two MDC_Fg1 and MDC_Fg593 strains systematically induced the most intense symptoms compared to the MDC_Fg202, MDC_Fg13, MDC_Fg5 and MDC_Fg851 which produced intermediate symptoms on the wheat spikelets. The German MDC_Fg8 and the French MDC_FgU1 were the least aggressive strains, especially the latter one that induced very weak symptoms at 72 hpi.

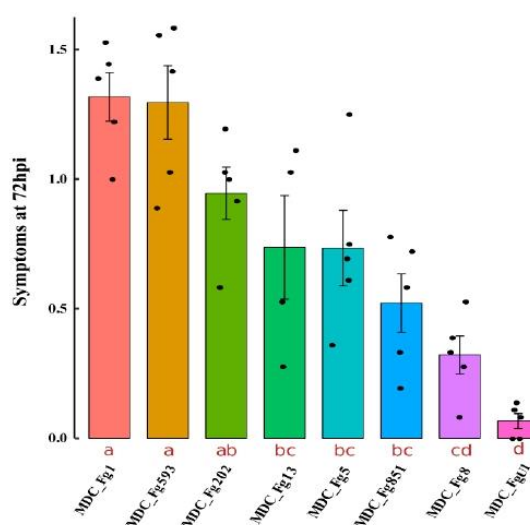


Figure 5. Symptom severity produced by the eight *F. graminearum* strains in bread wheat spikes during the FHB disease progress. The eight strains of *F. graminearum* were inoculated on the Recital wheat cultivar and symptoms were monitored at 72 hpi according the scoring described by Fabre et al. [56]. The Y-axis represents the symptom severity scale ranging from 0 to a maximum of 4. The error bars are the means $\pm$ SE of five biological replicates, and each replicate was characterized by the average value (represented as black dots) computed from six spikelets of three inoculated spikes. The different letters indicate significant differences at $p \leq 0.05$ (Duncan's multiple range test for post-hoc ANOVA mean separation).

2.6. Comparative Analysis of Secreted Protein Contents in *F. graminearum* Strains of Contrasting Aggressiveness

In each strain, about 5% of the theoretical proteins harbored secretory characteristics. The pan-secretome included 893 non-redundant secreted protein clusters (SPCs), of which 476 (53%) belonged to the core secretome (SPCs found in the 8 strains), 292 (33%) to the accessory secretome (SPCs found in 2 to 7 strains) and 125 (14%) were specific to 1 strain (Figure 6A, Supplementary Table S20). More than three-quarters of the pan-secretome belonged to the fast sub-genome, including 75%, 78% and 83% of the core, accessory and singleton secreted proteins, respectively.

Nearly 60% of the SPCs ($n = 519$) displayed known functions (Figure 6C), half of which ($n = 261$) corresponded to putative secreted CAZymes (Table S20), including 110 glycoside hydrolases (GH). CAZymes also included 59 enzymes belonging to auxiliary activities (AA) and 51 carbohydrate esterases (CE). In addition, 40% ($n = 207$) were putative secreted effectors and 142 SPCs corresponded to putative secreted proteases, including 82 serine peptidases and 32 metallo-peptidases. Overall, nearly 15% of all SPCs belonged to at least two functional categories of CAZymes, proteases and secreted effectors (Figure 6C), and

nearly 60% of the CAZymes, proteases and putative effectors are part of the core secretome, and are thus expected to be conserved in the eight strains (Figure 6B).

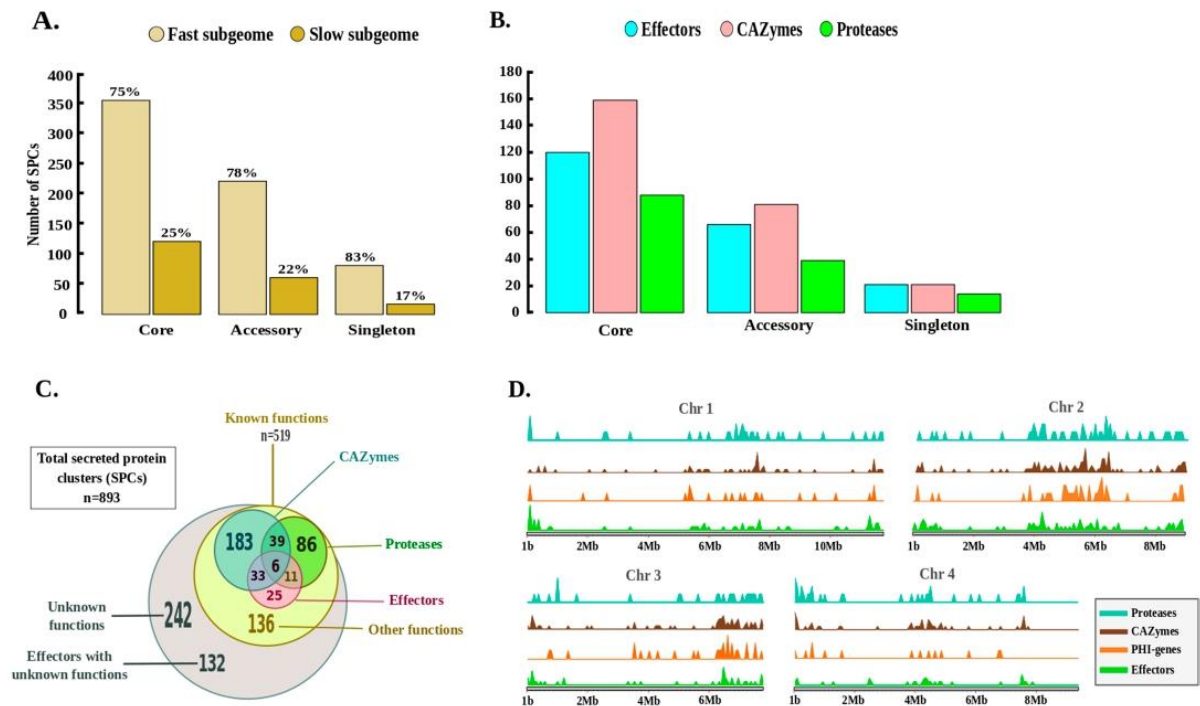


Figure 6. Characterization of the *F. graminearum* pan-secretome. (A) Categorization of putative secreted protein clusters (SPCs) of the eight *F. graminearum* strains into core, accessory and singleton secreted proteins. In each of the three categories, the percentage of SPCs assigned to the fast and slow sub-genome is reported. (B) The bar graph depicts the number of SPCs encoding CAZymes, proteases and putative proteinaceous effectors (according to EffectorP v2.0) in each of the three pan-secretome categories. (C) The Venn diagram depicts the number of SPCs displaying one or more labels (among CAZyme, protease or effector), as well as those with no predicted functions. (D) Density and localization of the genes encoding secreted CAZymes, secreted proteases, secreted effectors and genes potentially involved in the host–pathogen interaction (according to PHI-base) along the four *F. graminearum* chromosomes. The size of the chromosomes is in Mb, in 2 Mb increments.

To go further into the possible links between secreted protein roles and fungal aggressiveness, the pathogen–host interaction (PHI) database was also queried. A total of 108 known PHI genes were retrieved in the 8 genomes, among which two-thirds belonged to the core secretome (Supplementary Table S19). Overall, 50 proteins were assigned to the “unaffected pathogenicity” group, 38 to the “reduced virulence” group, 4 to the “avirulence plant determinant” and 1 from the “loss of pathogenicity” group, while 15 were part of the “mixed outcome” category (i.e., found in more than one group). The unique secreted protein belonging to the “loss of pathogenicity” group corresponds to a superoxide dismutase [57] detected in four strains of strong or intermediate aggressiveness (MDC_Fg1, FG001_06131; MDC_Fg593, FG593_06654; MDC_Fg13, FG013_08273; MDC_Fg851, FG851_03677). Likewise, among the four proteins belonging to the “avirulence plant determinant” group, three were found in all strains, while one containing a LysM domain was observed only in the aggressive strain MDC_Fg593 (FG593_11638) (Supplementary Table S20). The density and localization of genes encoding these SPCs, in particular those encoding CAZymes, proteases, effectors and PHI genes, along the four chromosomes of PH-1 evidenced the presence of highly enriched and compartmentalized genomic regions, with conserved dis-

tributions in each chromosome (Figure 6D). The proximal regions of the telomeres and the central regions of the four chromosomes especially harbored these SPCs, and a particular accumulation was found in chromosome 2 compared to the other three chromosomes.

LOF mutations in SPCs were further investigated. A total of 162 secreted proteins belonging to 77 SPCs harbored LOF mutations (Supplementary Table S20). The least aggressive strain MDC_FgU1 displayed the highest number of truncated secreted proteins (LOF mutations in 36 proteins) compared to the other seven strains. Among these, 13 were expected to remain intact in all the other genomes, including a phosphoesterase (FU001_11254), three CAZymes of the PL4, AA5 and GH145 families (FU001_02749, FU001_08346 and FU001_07816, respectively), two serine peptidases (FU001_06394, FU001_06650) and two secreted proteins belonging to both CE10 CAZyme and peptidase groups (FU001_08354 and FU001_08976).

2.7. Predicted Subcellular Localization of Secreted Proteins in the Plant Cell

Prediction of the subcellular localization of all SPCs on mature proteins (without the signal peptide sequence) demonstrated the targeting of three major subcellular compartments (Figure 7, Supplementary Table S20). The chloroplast target was predicted for 277 (31%) SPCs, of which 91 were putative effectors. Only 26 of these displayed known functional domains, including fourteen CAZymes (seven CE, three CBM, two GH, one AA and one PL), five proteases (two serine peptidases, two metallo-peptidases and one peptidase inhibitor), two Killer toxin (kp4) and one Ecp2 effector. The plant nucleus localization was identified for 188 (21%) SPCs, including 39 putative effectors, 10 of which displayed significant homology with CAZymes (six GH, two PL, one AA and one CBM), five with proteases (two serine peptidases, two metallo-peptidases and one cysteine-peptidase), one coding for a fungal lipase, one cerato-platanin and one for Killer toxin (kp4). The cytosol target was predicted for 167 SPCs (18.7%), including 28 putative effectors, of which half display known functions. A multiple localization signature (SPCs expected to target more than one cell compartment) was evidenced in 140 SPCs (15.7%), including 19 putative effectors. The 13.6% remaining SPCs were expected to target other plant compartments, such as the mitochondria (4.4%), the apoplast (3.47%), the cytoskeleton (2.57%), the endoplasmic reticulum (0.9%), the plasma membrane (0.9%), the peroxisomes (0.7%), the Golgi apparatus (0.33%) or the vacuolar membranes (0.33%).

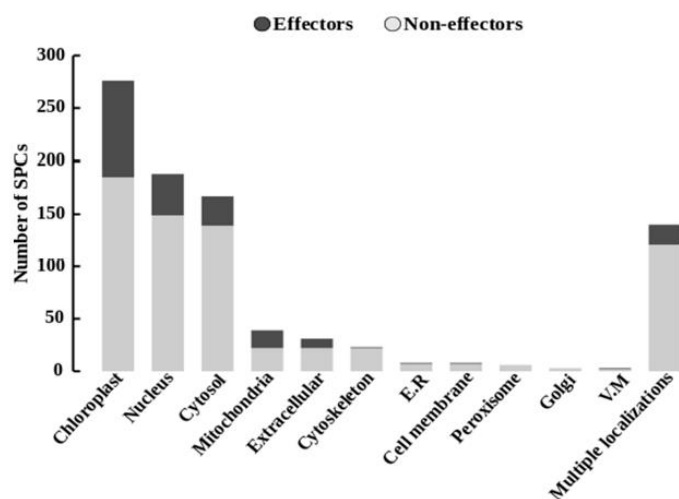


Figure 7. Expected subcellular targeting of SPCs in the plant cell. Bar graphs represent the predicted subcellular localization of putative secreted proteins (light gray), including the secreted putative effectors (dark gray). The pipeline used for the prediction of the secretion characteristics is described in Supplementary Figure S3.

3. Discussion

3.1. Structure and Functional Diversity at the Intra-Species Scale of *F. graminearum*

Previous comparative genomic analyses of *F. graminearum* mainly focused on the genome diversity within the *F. graminearum* species complex (FGSC), particularly on shared and absent genes as compared to a single reference genome [58–61]. In the present study, an intra-specific comparative genomic analysis was conducted to study the variability within the genomes of *F. graminearum* species. Walkowiak et al. [59] identified a list of 15,297 genes from 10 FGSC genomes, including 12% accessory genes (genes absent in at least one genome). Our study extends the pangenome of *F. graminearum* to approximate its full repertoire of genes based on twelve already reported genomes of different geographical origins and complemented with eight newly sequenced genomes of European strains. Our results showed that *F. graminearum* contains a significantly larger gene repertoire than that previously known from the reference genome PH-1 [25,26], with 20,807 non-redundant proteins. Less than half (45%) characterized all the genomes. The size of this core genome appears to reach a constant estimate and it is unlikely that the addition of other new genomes could significantly extensively reduce its size. The phylogeny, based on the shared proteins, shows that the European genomes are mostly grouped and closely related, as compared to non-European genomes, suggesting that they belong to a related population. Only the two French genomes, MDC_Fg1 and MDC_Fg5, belonged to separate clades. This could be due to the high rate of singleton SNPs which was only found in these two genomes, as well as a lower number of mutations per gene as compared to the other genomes of our study. Besides these core proteins, the high proportion of accessory predicted proteins suggests that the pangenome of *F. graminearum* is of the open type, indicating that the inclusion of other annotated genomes is likely to increase the size of the gene repertoire. This indicates that the strains harbor a high level of unique genomic diversity, which may reflect the ability of *F. graminearum* to survive in different environments [62]. Reaching 55% of the pangenome, the number of accessory and specific genes emphasizes the occurrence of potentially redundant functions. Given that our pangenome analysis and previous analyses of the PH-1 reference genome revealed little evidence of paralogy [24,60], it is unlikely that functional redundancy has evolved due to gene duplications. The action of the repetition-induced mutation system (RIP), which operates in *F. graminearum* [63], is known to protect the fungal genome against the activity of transposable elements and also limits the genome evolution by duplication of genes [64,65]. This could also argue for the low number of repeated sequences compared to other phytopathogenic species [66,67].

Considering the eight genomes we sequenced, a substantial fraction of the proteins ($n = 1283$) potentially lost their function because of LOF mutations such as frameshifts, gains of stop codon, losses of start codon, or splice acceptors and donors. Although these numerous truncated proteins are unlikely to perform a biological function identical to the intact corresponding protein, no direct link has been evidenced with strain development, pathogenicity or virulence through experimental tests, suggesting that functional redundancy compensation processes could occur, as already reported in *Botrytis cinerea* [68]. Besides functional redundancy and genomic truncations, the analysis of non-synonymous mutation rates indicated that 1129 other genes may be under diverse selection. As other phytopathogens, *F. graminearum* is known to have a “two-speed” genome with comparable sizes and number of genes [28,29]. More than 90% of the genes showing an excess rate of non-synonymous mutations were over-represented in the fast sub-genome, emphasizing that this sub-genome evolves faster than its slow counterpart. In addition, these genes are preferably grouped together and located in the proximal regions of the telomeres of the four chromosomes, or in the most central regions of the three chromosomes (1, 2 and 4). This is consistent with previous work showing links between the SNP density and high recombination frequency occurring in these chromosomal areas [16,24]. Our results support the idea that interstitial regions with high mutation rates result from ancestral chromosome fusions that merged ancestral telomeres [24,58,69].

Functional analysis showed that trichothecene BGC particularly held high rates of non-synonymous mutations, especially in the *Tri8* and *Tri12* genes. *Tri8* is a C3-esterase able to influence the synthesis of different forms of trichothecenes (both 15-acetyl-DON and T-2 toxin), while the *Tri12* gene is supposed to be involved in trichothecene production and efflux [55,70]. As a whole, the links between genetic variations of *TRI* genes and the aggressiveness of *F. Graminearum* strains are not clear, emphasizing that trichothecenes are only part of the molecular determinants of pathogenicity. At the strain level, only the two Italian genomes, MDC_Fg202 and MDC_Fg593, systematically accumulated a high rate of genetic variation in these two aforementioned genes, as well as a remarkable number of variants found in the intra- and inter-genic regions of trichothecene BGC, while they are among the most aggressive strains. These results partly corroborate the in vitro metabolomic profiling observed, in which only MDC_Fg202 and MDC_Fg593 exhibited an increased abundance of the trichothecene precursors, while little difference can be observed in the trichothecene products, suggesting that the accumulation of mutations identified in these *TRI* genes could play an important role in the accumulation of trichothecene precursors. Many previous studies have already shown that genetic variants can affect metabolite production in *A. fumigatus*, *A. nidulans* and *B. cinerea* [71–74]. However, in our study, no direct link was observed between other metabolites and severe effects of genetic variants, such as LOF mutations affecting the genes responsible for the production of these metabolites. This suggests that the accumulation of numerous mutations in BGCs could have a wider impact on metabolite production than LOF mutations in a single gene of the pathway. This also suggests the existence of a possible functional redundancy/compensation in the genes required for metabolite biosynthesis.

3.2. Links between the Genetic Variations in *F. graminearum* Secretomes and the Strain Aggressiveness in Wheat Tissues

In this analysis, we studied the refined secretomes of eight *F. graminearum* strains with high contrasting aggressiveness on a winter wheat genotype. Brown et al. [43] already predicted a secretome in the reference PH1 strain, indicating that 4.2% of the total genes encode classical secreted proteins. Here, approximately 5% of the total gene number harbored secretion characteristics (classical and non-classical secreted proteins) in each of the eight genomes. This proportion seems to be a regular content which should be observed in any refined secretome of *F. graminearum*. Overall, 893 SPCs defined the pan-secretome of the 8 strains and are considered to be putative secreted proteins. The functional analysis revealed 374 (42%) SPCs of unknown function, of which more than a third ($n = 132$) were predicted as putative secreted effectors. This exemplifies that unknown functions of secreted proteins and putative fungal effectors are usual characteristics in *F. graminearum*, as in other filamentous plant pathogens [41,75–77]. The prediction of their subcellular localization in the host cell demonstrated that the chloroplast could be a primary target of *F. graminearum* effectors, followed by the plant nucleus. Although functional characterizations of these proteins might be necessary to identify their intracellular localization and their role in the infection process, this clearly corroborates previous dual-proteomics on the FHB disease in bread wheat that identified a close relationship between chloroplast adjustments and the accumulation of *F. graminearum* proteins harboring the features of a chloroplast transit peptide [17,56,78]. Among SPCs with known functions, more than 70% are closely linked to plant/pathogen interaction and pathogenicity, including CAZymes, proteases and PHI genes. This finding emphasizes that *F. graminearum* aggressiveness depends on an arsenal of secreted proteins. Since they were mostly located in regions known as chromosomal recombination hotspots [24,43], this could therefore facilitate the occurrence of genetic variations to adapt with changes in the host responses. More than two-thirds of all SPCs belonged to the fast sub-genome, where genes also rapidly evolve to facilitate swift adaptation and infection responses [28,29]. *F. graminearum* colonizes the surface of wheat via runner hyphae (RH) and develops infection cushions (IC) which are essential for the penetration of the fungus into the plant cuticle [19,79,80]. Compared to mycelium grown in complete medium, both RH and IC have already shown that 982 genes were

systematically upregulated during initial infection of wheat floral tissues [19]. Of these genes, 194 were included in our pan-secretome, of which 74% belonged to the core secretome and 88% were located in the fast sub-genome, emphasizing their pivotal role in the infection process. Moreover, in another previous study, we identified 321 proteins of the MDC_Fg1 strain that were significantly regulated in planta during a 96 hpi period [56]. Among these, nearly three-quarters (73%) belonged to the slow sub-genome, suggesting that this slow sub-genome includes the necessary drivers of the disease progression through the accumulation of proteins with critical roles in pathogenicity.

Given the large differences in aggressiveness in the eight strains, we sought the possible links with the content of their secretomes, including with known secreted proteins involved in pathogenicity. For instance, 23 secreted proteins proved to be specific to the aggressive MDC_Fg1 strain, including 7 proteins predicted as proteinaceous effectors, with only 1 (FG001_05220) assigned to a known function encoding α/β -hydrolase and displaying a putative nuclear localization signal. Although a number of other specific secreted proteins have been identified in the most aggressive strains, their connections with aggressiveness require further functional analyses to be validated. In contrast, most of the putative secreted proteins linked to pathogenicity, including the PHI-genes, CAZymes, proteases and effectors candidates, were found in the core secretome, thus systematically shared by the eight strains. This corroborates previous results demonstrating that the three French, MDC_Fg1, MDC_Fg13 and MDC_FgU1, strains synthesize in planta very similar proteins at 72 hpi, and that the differences in aggressiveness could rather be due to differences in protein accumulation [17]. Consequently, no obvious qualitative and quantitative relationship can be defined between the SPCs linked to pathogenicity and the strain aggressiveness, and this high level of conservation in the eight secretomes suggests that *F. graminearum* secreted proteins can be involved in both the ability to survive as a pathogen and as a saprophyte [43]. In spite of a few qualitative differences in the SPCs among the eight strains, we nevertheless detected a number of truncated secreted proteins by LOF mutations, specifically in the least aggressive MDC_FgU1 strain. Among these SPCs, proteins with crucial roles in the hydrolysis of polysaccharide components of the cell wall of plants and in the degradation of host defense proteins such as GH145, PL4, AA5, serine peptidases and carboxylesterase type B were specifically truncated in MDC_FgU1, while they were intact in the other strains. This further suggests that different *F. graminearum* strains are characterized by a core SPC signature, with, however, marginal differences that may partly lead to discrepancies in aggressiveness.

4. Materials and Methods

4.1. *Fusarium graminearum* Isolates and Strain Aggressiveness Rating

Eight European *F. graminearum* isolates were reported in this study, and all isolates were able to produce the metabolite deoxynivalenol as well as induce FHB on bread wheat. Of these isolates, four were collected from France (MDC_Fg1, MDC_Fg5, MDC_Fg13 and MDC_FgU1) in field trials between 2010 and 2012 (Supplementary Table S1). MDC_Fg1, MDC_Fg13 and MDC_FgU1 have already been characterized in a previous study dealing with fungal proteomics during FHB [17]. Three other isolates were originated from Sardinia (Italy: MDC_Fg202, MDC_Fg593 and MDC_Fg851), and they were collected in 2018. The MDC_Fg8 isolate was collected from Germany and was obtained from Dr. T. Miedaner [81] and extensively characterized within a number of reports [54,79–84]. For each isolate, the preparation of the *F. graminearum* inocula was performed as described by Fabre et al. [56] to obtain a concentration of 10,000 spores per mL.

Aggressiveness of the eight *F. graminearum* strains was evaluated through symptom monitoring on the winter wheat cultivar Recital, as described by Fabre et al. [17,56]. After sowing seeds, plantlets underwent an 8-week vernalization at 4 °C, then they were transferred in a growth cabinet with optimal conditions to allow synchronized flowering (16 h daylight at 20 °C, 8 h darkness at 18 °C and 80% relative humidity). A total of 40 homogeneous plants were prepared in individual 4 L pots and divided into 5 complete

randomized blocks, surrounded by 30 additional plants to control any edge effect. In each block, one plant was used to monitor the aggressiveness of each of the eight strains, which corresponds to five different replicates. At the mid-anthesis stage, the eight strains were individually inoculated into randomly selected plants in each block by depositing 10 µL of inoculum in the floral cavity of six contiguous spikelets located in the middle zone of three synchronized spikes per plant. These represented a total of 18 spikelets per plant (90 measurements per strain) that were individually scored for the symptom dynamics at 72 h after the inoculation (hpi). This duration was chosen on the basis of our previous studies [17,56,78] demonstrating a turning point in the FHB progress at this time, enabling to discriminate strains of contrasted aggressiveness. Along with the forty infected plants, five additional ones were inoculated with water and were used as a control, and no symptoms were observed on these plants.

4.2. Genome Sequencing and Assembly

The fungal strains were grown on potato dextrose broth and the genomic DNA was prepared from the mycelium using a NucleoBond kit according to the manufacturer's instructions (Macherey-Nagel, Düren, Germany). *F. graminearum* MDC_Fg5, MDC_Fg8, MDC_Fg13, MDC_202, MDC_Fg593 and MDC_Fg851 were sequenced using the Illumina HiSeq platform. MDC_Fg1 and MDC_FgU1 were sequenced by the PacBio SMRT Sequel platform. A 150 bp paired-end library was prepared following the manufacturer's instructions. After quality control, the percentage of reads with a Q20 and Q30 score ranged from 95 to 98.5 and 90 to 97, respectively. The reads were preprocessed with Trimmomatic v.36 [85] using default settings. Over 87% of preprocessed reads were used for de novo assembly with ABySS v2.0.0 [86] using three k-mer sizes (41, 96 and 128).

The SMRT Bell library was prepared using a DNA template Prep 1.0 kit (Pacific Biosciences, CA, USA) according to the manufacturer's protocol. A size selection of 12 kb was realized by using the BluePippin system (Sage Scientific, Beverly, MA, USA) according to the protocol "20 kb Template Preparation Using BluePippin Size-Selection System". The raw PacBio Sequel subreads were filtered and assembled de novo using the Hierarchical Genome Assembly Process (HGAP4) implemented in SMRT Link v5.0 [87] using default settings, which keeps only subreads with read quality (rq) of ≥ 0.7 . The completeness of the assembled genomes was assessed by the presence of 3725 single-copy conserved genes from the Sordariomyceta lineage using Benchmarking Universal Single Copy Orthologs Version 3 (BUSCO v3) [88].

4.3. Gene Prediction and Genome Annotation

The assembled genomes were annotated using the MAKER2 Annotation Pipeline [89], with a combination of evidence-based methods (transcriptome data and homology with known proteomes) and ab initio gene prediction. In brief, four samples of transcriptomic RNAseq reads of methylation-depressed *F. graminearum* obtained from NCBI (Run accessions: SRR999640, SRR999641, SRR999642 and SRR999643) were assembled by rnaSPAdes v3.13.0 [90] and used as evidence, with 21,355 expressed sequence tags (EST) of *F. graminearum* PH-1 (downloaded from NCBI).

Protein evidence from UniProtKB (update-2018_09) supplemented with the proteomes of *F. graminearum* PH-1, *F. pseudograminearum*, *F. culmorum*, *F. fujikuroi*, *F. oxysporum* and *F. verticillioides* retrieved from Ensembl Fungi (update-40) were also used for homology research. Two different ab initio gene annotation programs have been trained for use with MAKER2. SNAP v2013-02-16 [91] was used with transcriptome sequences and Augustus v3.2.1 [92] was used with *F. graminearum* gene models. The repetitive elements of the whole genomes were masked using the library of all *Fusarium* repeated elements of the Repbase database (update-01.02.2017) via RepeatMasker v4.0.7 [93]. The same annotation method was used for ten other *F. graminearum* genomes with only published raw data, including five French, three Chinese, one American and one Canadian (Supplementary Table S1).

4.4. Pangenome Construction and GO Enrichment Analysis

The eight genomes of our study and twelve other genomes of *F. graminearum*, including five French (INRA-156, INRA-159, INRA-164, INRA-171 and INRA-181), three Chinese (YL-1, HN9-1 and HN-Z6), two American (PH-1 and DAOM_233423), one Canadian (DAOM_241165) and one Australian (CS3005) (Supplementary Table S1) were used for the pangenome construction. Reciprocal best BLAST hits [94,95] were performed to identify the set of orthologous genes using the Proteinortho v5.16b [96]. After comparing several values of identity (30% to 90%) and coverage (60% to 90%), based on conservation of the pangenome size [52], proteins with an identity score above 50% and a sequence coverage above 70% (e-value below 1×10^{-5}) were deemed as significant hits. Core genome and pangenome size curves were generated using proteinortho_curves [96] with 100 iterations. To analyze the GO enrichment, functional annotation was performed using InterProScan v5.13-52.0 [97], including the Pfam database families and the gene ontology (GO) classification. The enrichment of GO terms was evaluated using the Python package GOATOOLS v0.8.2 [98] and the go-basic.obo database (version 2020-02-06). The *p*-values used to assess the significant enrichment of the GO terms were calculated based on a Fisher's exact test with propagation of the parent term and correction for the false discovery rate (FDR, $p < 0.05$).

4.5. Phylogenomic Analysis

To construct a phylogeny of the twenty genomes of *F. graminearum* described below, completed by three other genomes of the genus *Fusarium* (*F. pseudograminearum*, *F. culmorum* and *F. fujikuroi*) used as outgroups, we used protein sequences found in exactly one copy in all the representatives of the twenty-three genomes based on a cluster analysis. The identified single-copy orthologs were aligned using MAFFT v7.402 [99] and poorly aligned regions were removed by using Gblocks v0.91b [100] with default settings. Maximum likelihood (ML) analysis was performed with RaxML v8.1.20 [101] using the “-f a” option and the PROTGAMMLG amino acid substitution matrix model with 1000 bootstrap replicates. The phylogenetic tree was visualized using FigTree v1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree> (accessed on 7 July 2020)).

4.6. Analysis of Secreted Proteins and Prediction of Their Subcellular Localization in the Plant Cell

The eight annotated proteomes of *F. graminearum* of this study were used for the prediction of secretomes downstream by performing the pipeline described in Supplementary Figure S3. The secretory signal peptide was predicted using SignalP v4.1 [102]. Protein sequences lacking the signal peptide were used by SecretomeP v1.0 [103] for the prediction of non-classical secreted proteins. After that, the signal sequences approved by SignalP and SecretomeP were analyzed by TargetP v1.1b [104] to further remove mitochondrial proteins. The sets of proteins were examined for the presence of a transmembrane domain using TMHMM v2.0c [105], and only proteins without a transmembrane domain or one transmembrane domain in the N-terminal signal peptide were kept. WoLFPSORT v0.2 [106] was used to select the proteins with an extracellular destination with a score greater than 17, as mentioned by Brown et al. [43]. Lastly, only sequences without GPI (glycosylphosphatidylinositol) anchors predicted by PredGPI [107] in these extracellular proteins were retained as candidates for secreted proteins. The subcellular localization of secreted proteins in plant cells were predicted using a combination of three tools, LOCALIZER [108] running in “effector mode” and WoLFPSORT [106] running in “plant mode” after removing the fungal secretion signal, while nuclear localization signals (NLS) were confirmed using the cNLS Mapper [109].

4.7. Pathogenicity-Related Genes' Prediction

The refined secretomes were screened for Carbohydrate Active enZymes (CAZymes, Appendix A) using Hmmscan in HMMER v3.1b2 (<http://hmmer.org/> (accessed on 16 July 2020)) and the dbCAN HMM profile database v6 [110]. The outputs were further

processed by the hmmscan-parser script provided by dbCAN to select significant matches. Putative peptidases and peptidase inhibitors were predicted using the MEROPS database (update-12.1) [111]. To identify proteins potentially associated with pathogenicity, the BLASTP search was performed against the Pathogen–Host Interaction database v4.5 (PHI-base) [112,113]. BLASTP searches against the MEROPS database and PHI-base were used with an e-value threshold of 10^{-5} with the best hits and an identity greater than 40%. Besides the PHI genes, CAZymes and secreted peptidases, we also searched for proteinaceous effectors by characterizing the secreted proteins using the pipeline described above (Supplementary Table S3) and by adding a search with EffectorP v2.0 to distinguish the secreted proteinaceous effectors in the predicted fungal secretomes [114].

To determine the location in the chromosome and the distribution of genes coding for putative secreted proteins linked to pathogenicity (CAZymes, peptidases, PHI-genes and effectors), we aligned these coding genes with the PH-1 reference genome RRES v5.0 using GMAP v2020-06-30 [115] using the default settings. The gene density plots on the four chromosomes were generated using karyoploteR [116].

4.8. Variant Calling and Analyses

The filtered reads were mapped on the PH-1 reference genome RRES v5.0 using BWA-MEM v0.7.17-r1188 (with option -M) [117] for the data generated by Illumina HiSeq and Minimap v2.12-r847 [118] for those released by PacBio SMRT Sequel. The mapping quality was assessed by Qualimap v2.2.1 [119].

The six BAM files (results from Illumina HiSeq data mapping) were sorted, and duplicated reads were removed by Picard SortSam and MarkDuplicates v1.119, respectively (<http://broadinstitute.github.io/picard/> (accessed on 7 September 2020)). Positions of insertions/deletions (InDels) were recalibrated using the RealignerTargetCreator and the IndelRealigner modules of the Genome Analysis Toolkit (GATK, version 3.8-1-0-gf15c1c3ef) [120] with standard parameters. The single nucleotide polymorphisms (SNPs) and short InDels were named using the GATK Unified Genotyper in haploid mode. To obtain high-quality variants, we used the GATK VariantFiltration to perform hard filtering based on quality cut-offs following the GATK best practice recommendations. For SNPs, we filtered out: $QD < 2.0 \mid FS > 60.0 \mid QUAL < 250 \mid MQ < 40.0 \mid SOR > 3.0 \mid MappingQualityRankSum < -12.5 \mid ReadPosRankSum < -8.0$. For InDels, we filtered out: $QD < 2.0 \mid FS > 200.0 \mid SOR > 10.0 \mid QUAL < 250 \mid ReadPosRankSum < -20$. Only insertions and deletions shorter than or equal to 50 bp were kept.

To identify variants from the two other Bam files (PacBio Sequel subread mapping results), we first generated mpileup files using SAMtools, v1.8, with -Q0 -e 5 -C50 -B options [121]. The result was used as input to VarScan, version 2.3.9 [122], with the “mpileup2snp” and “mpileup2indel” functions to identify SNPs and short indels respectively, with a minimum average coverage of 10, a minimum variant allele frequency of 0.2, minimum frequency to call homozygote of 0.85 and a *p*-value threshold of 0.05.

The final call set of the eight genomes was annotated and their impact was predicted using SnpEff v 4.3t [123] with the -ud 500 and -lof options. Variant effects were parsed with SnpSift v4.3t [124] and vcfutils v0.1.16 [125]. The SnpEff databases were built locally using annotations of the reference genome PH-1 RRES v5.0 obtained in GTF format from Ensembl Fungi (update-40). The SnpEff database was used to annotate SNPs and InDels with putative functional effects according to the categories defined in the SnpEff manual (http://snpeff.sourceforge.net/SnpEff_manual.html (accessed on 27 November 2020)). Genes were accepted with an excess of mutational mutations if they contained at least more than six non-synonymous mutations and showed a ratio of non-synonymous mutations to the total number of synonymous and non-synonymous mutations greater than 0.6, according to Laurent et al. [16].

4.9. Secondary Metabolite Analysis

Strains were grown on PDA (Potato Dextrose Agar, Difco, 1.5% agar) plates for eight days at 25 °C. For metabolomics analysis, four plates of each strain were frozen, then thawed and extracted with 50 mL ethyl acetate (EtOAc). The mixture was shaken at 200 rpm and ambient temperature for 48 h, then filtered through filter paper. The ethyl acetate phase was separated from the water and dried in vacuo. Each extract was prepared to 40 mg/mL in 1:1 ACN:methanol (MeOH) for LC/MS analysis. Each sample was analyzed in duplicate on an Agilent 1100 series LC/MS platform. Positive mode ionization was used to detect metabolites. Datasets were exported from Agilent's Chemstation software as .netCDF files and imported into MZmine v2.3 [126]. Peak picking was performed with established protocols [48]. Mass detection was centroid with 1.7×10^4 minimum height. Chromatogram building was limited to peaks greater than 0.1 min with 0.1 mass-to-charge ratio (m/z) tolerance and 1.8×10^4 minimum height. Chromatogram deconvolution utilized local minimum search, with 92–99% threshold, 40 min search range, 1% relative height, 1.9×10^4 minimum abs. height, 1 minimum ratio of top/edge and peak duration from 0.10 to 10.00 min. All treatments were then aligned with a tolerance of 0.05 m/z and relative retention time tolerance of 3.0 (%). Peak finder gap filling was performed with 25% intensity tolerance and 0.1 m/z tolerance. Duplicated peaks were combined with a tolerance of 0.5 m/z and 0.3 min. The normalized data were processed using MetaboAnalyst software v4.0 [127]. Data were log-transformed and scaled to unit variance for a multivariate modeling approach to check for variability based on overall metabolite profiles. Data were quantitated to scalar graphs for *Fusarium* compounds by their extracted ion detection relative to each of the eight *F. graminearum* isolates.

5. Conclusions

This study expanded our understanding of the intra-species genomic diversity of *F. graminearum* and provided new data on its genetic content and its relationship with the FHB disease. First, this work showed that a single reference genome sequence depicting less than 68% of a pangenome does not fully reflect the gene content of any strain belonging to this species due to the high degree of observed genomic variation. This emphasizes that switching from a reference genome to a reference pangenome, gathering the dispensable or accessory genomic regions we described in this study, could arguably refine a wide range of genomics or transcriptomics analyses, and further improve variant calling and the accurate quantification of gene expression. Lastly, although further analyses are required to fill the gap between comparative genomics and strain pathogenicity, all of these data substantially improve our knowledge about the genomic variations of *F. graminearum* strains displaying contrasting aggressiveness. They provide pivotal clues to further decipher the molecular determinants of the FHB disease and bring to the forefront a wealth of candidate secreted proteins potentially involved in the pathogenicity of *F. graminearum*, including a number of new candidate effectors that will guide future research.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/ijms22126257/s1>.

Author Contributions: Conceptualization, L.B. and T.L.; methodology, T.A., H.R., G.A.G.-M. and S.L.; resources, J.B., W.S. and M.F.; formal analysis, T.A. and L.B.; data curation, T.A. and G.A.G.-M.; writing—original draft, T.A. and L.B.; writing—review and editing, T.A., L.B., T.L., S.L., J.B., W.S. and M.F.; supervision, L.B. and T.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the Agence Nationale de la Recherche of the French government (NewMyco project, ANR-15-CE21-0010), and the authors also acknowledge the support received from the Agence Nationale de la Recherche of the French government through the program “Investissements d’Avenir” (16-IDEX-0001 CAP 20-25). Work on *F. graminearum* in the Freitag Lab is supported by a grant from the US NSF (MCB1818006). Work in the Loesgen Lab has been supported by US NSF (CH1808717).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The Whole Genome Shotgun (WGS) projects; raw data, assembled and annotated genomes of the eight strains of this study were deposited in ENA/GenBank/DDBJ under the accession numbers listed in Table S1 (BioProject accession no. PRJEB27611).

Acknowledgments: The authors thank the Mesocentre Clermont-Auvergne for providing the computer calculation and storage facilities. We acknowledge M.H. Lebrun (UMR1290 BIOGER) for his help in the MDC_Fg13 and MDC_Fg5 genome sequencing.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

GH, glycoside hydrolases; CE, carbohydrate esterases; AA, auxiliary activities; GT: glycosyltransferases, CBM, carbohydrate-binding modules; PL, polysaccharide lyases; S, serine; M, metallo; C, cysteine; T: threonine; A, aspartate, N, asparagine; G, glutamate; I, inhibitors.

Appendix A

We compared the global distribution of CAZymes and proteolytic enzymes in *F. graminearum* with 15 other fungal genomes of the Sordariomycetes class and one genome from the Dothideomycetes class (used as an outgroup) with different lifestyles. We observed that *F. graminearum* contained a larger number of genes coding for CAZymes compared to other pathogenic fungi, such as *Verticillium dahliae* (necrotrophic), *Gaeumannomyces graminis* (necrotrophic), *Verticillium alfalfae* (necrotrophic) and *Magnaporthe oryzae* (hemibiotrophic) (Supplementary Figure S4a). In the eight genomes of *F. graminearum* sequenced in this study, an average of 654 proteins distributed in 139 CAZyme families were systemically predicted, of which 66 (48%) were conserved in all the analyzed fungal species. In addition, four CAZyme families were found in the *F. graminearum* genomes only, GH14 (β -amylase), GH73 (lysozyme), GT61 (β -1,2-xylosyl-transferase) and CBM47 (Fucose-binding activity).

Regarding proteases, our results revealed that the *F. graminearum* had the fourth highest number of predicted representatives, just after *Nectria haematococca* (Necrotrophic), *Fusarium oxysporum* (Necrotrophic) and *Magnaporthe oryzae* (Hemibiotrophic) (Supplementary Figure S4b), with an average of 742 proteases belonging to 122 families and 38 protease inhibitors from 16 families. Serine peptidases were observed as the largest category, followed by metalloproteases and cysteine peptidases. A similar trend was observed in the other analyzed fungal species. Furthermore, the comparative analysis identified only one protease family specifically found in the genomes of *F. graminearum* (M26, IgA1-specific metallopeptidase).

References

1. Goswami, R.S.; Kistler, H.C. Heading for Disaster: *Fusarium graminearum* on Cereal Crops. *Mol. Plant Pathol.* **2004**, *5*, 515–525. [CrossRef] [PubMed]
2. Xu, X.; Nicholson, P. Community Ecology of Fungal Pathogens Causing Wheat Head Blight. *Annu. Rev. Phytopathol.* **2009**, *47*, 83–103. [CrossRef] [PubMed]
3. Pestka, J.J.; Smolinski, A.T. Deoxynivalenol: Toxicology and Potential Effects on Humans. *J. Toxicol. Environ. Health B Crit. Rev.* **2005**, *8*, 39–69. [CrossRef]
4. Desjardins, A.E. *Fusarium Mycotoxins: Chemistry, Genetics, and Biology*; APS Press: St. Paul, MN, USA, 2006.
5. Yu, M.; Chen, L.; Peng, Z.; Nüssler, A.K.; Wu, Q.; Liu, L.; Yang, W. Mechanism of Deoxynivalenol Effects on the Reproductive System and Fetus Malformation: Current Status and Future Challenges. *Toxicol. Vitro* **2017**, *41*, 150–158. [CrossRef]
6. O'Donnell, K.; Ward, T.J.; Geiser, D.M.; Corby Kistler, H.; Aoki, T. 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.* **2004**, *41*, 600–623. [CrossRef] [PubMed]
7. Starkey, D.E.; Ward, T.J.; Aoki, T.; Gale, L.R.; Kistler, H.C.; Geiser, D.M.; Suga, H.; Tóth, B.; Varga, J.; O'Donnell, K. Global Molecular Surveillance Reveals Novel *Fusarium* Head Blight Species and Trichothecene Toxin Diversity. *Fungal Genet. Biol.* **2007**, *44*, 1191–1204. [CrossRef]

8. Scoz, L.B.; Astolfi, P.; Reartes, D.S.; Iii, D.G.S.; Moraes, M.G.; Ponte, E.M.D. Trichothecene Mycotoxin Genotypes of *Fusarium graminearum* Sensu Stricto and *Fusarium meridionale* in Wheat from Southern Brazil. *Plant Pathol.* **2009**, *58*, 344–351. [\[CrossRef\]](#)
9. Reynoso, M.M.; Ramirez, M.L.; Torres, A.M.; Chulze, S.N. Trichothecene Genotypes and Chemotypes in *Fusarium graminearum* Strains Isolated from Wheat in Argentina. *Int. J. Food Microbiol.* **2011**, *145*, 444–448. [\[CrossRef\]](#)
10. Castañares, E.; Albuquerque, D.R.; Dinolfo, M.I.; Pinto, V.F.; Patriarca, A.; Stenglein, S.A. Trichothecene Genotypes and Production Profiles of *Fusarium graminearum* Isolates Obtained from Barley Cultivated in Argentina. *Int. J. Food Microbiol.* **2014**, *179*, 57–63. [\[CrossRef\]](#)
11. Ward, T.J.; Clear, R.M.; Rooney, A.P.; O'Donnell, K.; Gaba, D.; Patrick, S.; Starkey, D.E.; Gilbert, J.; Geiser, D.M.; Nowicki, T.W. An Adaptive Evolutionary Shift in *Fusarium* Head Blight Pathogen Populations Is Driving the Rapid Spread of More Toxicogenic *Fusarium graminearum* in North America. *Fungal Genet. Biol.* **2008**, *45*, 473–484. [\[CrossRef\]](#)
12. Crippin, T.; Renaud, J.B.; Sumarah, M.W.; Miller, J.D. Comparing Genotype and Chemotype of *Fusarium graminearum* from Cereals in Ontario, Canada. *PLoS ONE* **2019**, *14*, e0216735. [\[CrossRef\]](#)
13. Zhang, X.-W.; Jia, L.-J.; Zhang, Y.; Jiang, G.; Li, X.; Zhang, D.; Tang, W.-H. In Planta Stage-Specific Fungal Gene Profiling Elucidates the Molecular Strategies of *Fusarium graminearum* Growing inside Wheat Coleoptiles. *Plant Cell* **2012**, *24*, 5159–5176. [\[CrossRef\]](#)
14. Zhu, Z.; Chen, L.; Zhang, W.; Yang, L.; Zhu, W.; Li, J.; Liu, Y.; Tong, H.; Fu, L.; Liu, J.; et al. Genome-Wide Association Analysis of *Fusarium* Head Blight Resistance in Chinese Elite Wheat Lines. *Front. Plant Sci.* **2020**, *11*. [\[CrossRef\]](#)
15. Boutigny, A.-L.; Ward, T.J.; Ballois, N.; Iancu, G.; Ios, R. Diversity of the *Fusarium graminearum* Species Complex on French Cereals. *Eur. J. Plant Pathol.* **2014**, *138*, 133–148. [\[CrossRef\]](#)
16. Laurent, B.; Moinard, M.; Spataro, C.; Ponts, N.; Barreau, C.; Foulongne-Oriol, M. Landscape of Genomic Diversity and Host Adaptation in *Fusarium graminearum*. *BMC Genom.* **2017**, *18*, 203. [\[CrossRef\]](#)
17. Fabre, F.; Bormann, J.; Urbach, S.; Roche, S.; Langin, T.; Bonhomme, L. Unbalanced Roles of Fungal Aggressiveness and Host Cultivars in the Establishment of the *Fusarium* Head Blight in Bread Wheat. *Front. Microbiol.* **2019**, *10*. [\[CrossRef\]](#)
18. Talas, F.; Parzies, H.K.; Miedaner, T. 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.* **2011**, *131*, 39–48. [\[CrossRef\]](#)
19. Mentges, M.; Glasenapp, A.; Boenisch, M.; Malz, S.; Henrissat, B.; Frandsen, R.J.N.; Güldener, U.; Münsterkötter, M.; Bormann, J.; Lebrun, M.-H.; et al. Infection Cushions of *Fusarium graminearum* Are Fungal Arsenals for Wheat Infection. *Mol. Plant Pathol.* **2020**, *21*, 1070–1087. [\[CrossRef\]](#)
20. Birr, T.; Hasler, M.; Verreet, J.-A.; Klink, H. Composition and Predominance of *Fusarium* Species Causing *Fusarium* Head Blight in Winter Wheat Grain Depending on Cultivar Susceptibility and Meteorological Factors. *Microorganisms* **2020**, *8*, 617. [\[CrossRef\]](#)
21. Prodi, A.; Tonti, S.; Nipoti, P.; Pancaldi, D.; Pisi, A. Identification of deoxynivalenol and nivalenol producing chemotypes of *Fusarium graminearum* isolates from durum wheat in a restricted area of northern Italy. *J. Plant Pathol.* **2009**, *91*, 727–731. [\[CrossRef\]](#)
22. Prodi, A.; Purahong, W.; Tonti, S.; Salomoni, D.; Nipoti, P.; Covarelli, L.; Pisi, A. Difference in Chemotype Composition of *Fusarium graminearum* Populations Isolated from Durum Wheat in Adjacent Areas Separated by the Apennines in Northern-Central Italy. *Plant Pathol. J.* **2011**, *27*, 354–359. [\[CrossRef\]](#)
23. Bentivenga, G.; Spina, A.; Ammar, K.; Allegra, M.; Cacciola, S.O. Screening of Durum Wheat (*Triticum Turgidum* L. Subsp. Durum (Desf.) Husn.) Italian Cultivars for Susceptibility to *Fusarium* Head Blight Incited by *Fusarium graminearum*. *Plants* **2021**, *10*, 68. [\[CrossRef\]](#)
24. Cuomo, C.A.; Güldener, U.; Xu, J.-R.; Trail, F.; Turgeon, B.G.; Di Pietro, A.; Walton, J.D.; Ma, L.-J.; Baker, S.E.; Rep, M.; et al. The *Fusarium graminearum* Genome Reveals a Link between Localized Polymorphism and Pathogen Specialization. *Science* **2007**, *317*, 1400–1402. [\[CrossRef\]](#)
25. King, R.; Urban, M.; Hammond-Kosack, M.C.U.; Hassani-Pak, K.; Hammond-Kosack, K.E. The Completed Genome Sequence of the Pathogenic Ascomycete Fungus *Fusarium graminearum*. *BMC Genom.* **2015**, *16*, 544. [\[CrossRef\]](#)
26. King, R.; Urban, M.; Hammond-Kosack, K.E. Annotation of *Fusarium graminearum* (PH-1) Version 5.0. *Genome Announc.* **2017**, *5*. [\[CrossRef\]](#)
27. Alouane, T.; Rimbart, H.; Fabre, F.; Cambon, F.; Langin, T.; Bonhomme, L. Genome Sequence of *Fusarium graminearum* Strain MDC_Fg1, Isolated from Bread Wheat Grown in France. *Microbiol. Resour. Announc.* **2018**, *7*. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Wang, Q.; Jiang, C.; Wang, C.; Chen, C.; Xu, J.-R.; Liu, H. Characterization of the Two-Speed Subgenomes of *Fusarium graminearum* Reveals the Fast-Speed Subgenome Specialized for Adaption and Infection. *Front. Plant Sci.* **2017**, *8*. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Dong, S.; Raffaele, S.; Kamoun, S. The Two-Speed Genomes of Filamentous Pathogens: Waltz with Plants. *Curr. Opin. Genet. Dev.* **2015**, *35*, 57–65. [\[CrossRef\]](#)
30. Sieber, C.M.K.; Lee, W.; Wong, P.; Münsterkötter, M.; Mewes, H.-W.; Schmeitzl, C.; Varga, E.; Berthiller, F.; Adam, G.; Güldener, U. The *Fusarium graminearum* Genome Reveals More Secondary Metabolite Gene Clusters and Hints of Horizontal Gene Transfer. *PLoS ONE* **2014**, *9*, e110311. [\[CrossRef\]](#)
31. Ballouz, S.; Dobin, A.; Gillis, J.A. Is It Time to Change the Reference Genome? *Genome Biol.* **2019**, *20*, 159. [\[CrossRef\]](#)
32. Gage, J.L.; Vaillancourt, B.; Hamilton, J.P.; Manrique-Carpintero, N.C.; Gustafson, T.J.; Barry, K.; Lipzen, A.; Tracy, W.F.; Mikel, M.A.; Kaeppler, S.M.; et al. Multiple Maize Reference Genomes Impact the Identification of Variants by Genome-Wide Association Study in a Diverse Inbred Panel. *Plant. Genome* **2019**, *12*, 180069. [\[CrossRef\]](#)
33. Bayer, P.E.; Golicz, A.A.; Scheben, A.; Batley, J.; Edwards, D. Plant Pan-Genomes Are the New Reference. *Nat. Plants* **2020**, *6*, 914–920. [\[CrossRef\]](#)

34. Della Coletta, R.; Qiu, Y.; Ou, S.; Hufford, M.B.; Hirsch, C.N. How the Pan-Genome Is Changing Crop Genomics and Improvement. *Genome Biol.* **2021**, *22*, 3. [\[CrossRef\]](#)
35. Tettelin, H.; Maignani, V.; Cieslewicz, M.J.; Donati, C.; Medini, D.; Ward, N.L.; Angiuoli, S.V.; Crabtree, J.; Jones, A.L.; Durkin, A.S.; et al. Genome Analysis of Multiple Pathogenic Isolates of *Streptococcus agalactiae*: Implications for the Microbial “Pan-Genome”. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 13950–13955. [\[CrossRef\]](#)
36. Cook, D.E.; Mesarich, C.H.; Thomma, B.P.H.J. Understanding Plant Immunity as a Surveillance System to Detect Invasion. *Annu. Rev. Phytopathol.* **2015**, *53*, 541–563. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Rovenich, H.; Boshoven, J.C.; Thomma, B.P.H.J. Filamentous Pathogen Effector Functions: Of Pathogens, Hosts and Microbiomes. *Curr. Opin. Plant Biol.* **2014**, *20*, 96–103. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Kamoun, S. A Catalogue of the Effector Secretome of Plant Pathogenic Oomycetes. *Annu. Rev. Phytopathol.* **2006**, *44*, 41–60. [\[CrossRef\]](#) [\[PubMed\]](#)
39. Misas-Villamil, J.C.; van der Hooft, R.A.L. Enzyme-Inhibitor Interactions at the Plant-Pathogen Interface. *Curr. Opin. Plant Biol.* **2008**, *11*, 380–388. [\[CrossRef\]](#) [\[PubMed\]](#)
40. Jashni, M.K.; Dols, I.H.M.; Iida, Y.; Boeren, S.; Beenen, H.G.; Mehrabi, R.; Collemare, J.; de Wit, P.J.G.M. Synergistic Action of a Metalloprotease and a Serine Protease from *Fusarium oxysporum* f. sp. *Lycopersici* Cleaves Chitin-Binding Tomato Chitinases, Reduces Their Antifungal Activity, and Enhances Fungal Virulence. *Mol. Plant Microbe Interact.* **2015**, *28*, 996–1008. [\[CrossRef\]](#)
41. Lo Presti, L.; Lanver, D.; Schweizer, G.; Tanaka, S.; Liang, L.; Tollot, M.; Zuccaro, A.; Reissmann, S.; Kahmann, R. Fungal Effectors and Plant Susceptibility. *Annu. Rev. Plant Biol.* **2015**, *66*, 513–545. [\[CrossRef\]](#)
42. Lyu, X.; Shen, C.; Fu, Y.; Xie, J.; Jiang, D.; Li, G.; Cheng, J. Comparative Genomic and Transcriptional Analyses of the Carbohydrate-Active Enzymes and Secretomes of Phytopathogenic Fungi Reveal Their Significant Roles during Infection and Development. *Sci. Rep.* **2015**, *5*, 15565. [\[CrossRef\]](#)
43. Brown, N.A.; Antoniw, J.; Hammond-Kosack, K.E. The Predicted Secretome of the Plant Pathogenic Fungus *Fusarium graminearum*: A Refined Comparative Analysis. *PLoS ONE* **2012**, *7*, e33731. [\[CrossRef\]](#)
44. Varga, E.; Wiesenberger, G.; Hametner, C.; Ward, T.J.; Dong, Y.; Schöffbeck, D.; McCormick, S.; Broz, K.; Stückler, R.; Schuhmacher, R.; et al. New Tricks of an Old Enemy: Isolates of *Fusarium graminearum* Produce a Type A Trichothecene Mycotoxin. *Environ. Microbiol.* **2015**, *17*, 2588–2600. [\[CrossRef\]](#)
45. Bahadoor, A.; Brauer, E.K.; Bosnich, W.; Schneiderman, D.; Johnston, A.; Aubin, Y.; Blackwell, B.; Melanson, J.E.; Harris, L.J. Gramillin A and B: Cyclic Lipopeptides Identified as the Nonribosomal Biosynthetic Products of *Fusarium graminearum*. *J. Am. Chem. Soc.* **2018**, *140*, 16783–16791. [\[CrossRef\]](#) [\[PubMed\]](#)
46. Jia, L.-J.; Tang, H.-Y.; Wang, W.-Q.; Yuan, T.-L.; Wei, W.-Q.; Pang, B.; Gong, X.-M.; Wang, S.-F.; Li, Y.-J.; Zhang, D.; et al. A Linear Nonribosomal Octapeptide from *Fusarium graminearum* Facilitates Cell-to-Cell Invasion of Wheat. *Nat. Commun.* **2019**, *10*, 922. [\[CrossRef\]](#)
47. Villafana, R.T.; Ramdass, A.C.; Rampersad, S.N. Selection of *Fusarium* Trichothecene Toxin Genes for Molecular Detection Depends on TRI Gene Cluster Organization and Gene Function. *Toxins* **2019**, *11*, 36. [\[CrossRef\]](#) [\[PubMed\]](#)
48. Adpressa, D.A.; Connolly, L.R.; Konkel, Z.M.; Neuhaus, G.F.; Chang, X.L.; Pierce, B.R.; Smith, K.M.; Freitag, M.; Loesgen, S. A Metabolomics-Guided Approach to Discover *Fusarium graminearum* Metabolites after Removal of a Repressive Histone Modification. *Fungal Genet. Biol.* **2019**, *132*, 103256. [\[CrossRef\]](#)
49. Niehaus, E.-M.; Kleigrew, K.; Wiemann, P.; Stüttgen, L.; Sieber, C.M.K.; Connolly, L.R.; Freitag, M.; Güldener, U.; Tudzynski, B.; Humpf, H.-U. Genetic Manipulation of the *Fusarium fujikuroi* Fusarin Gene Cluster Yields Insight into the Complex Regulation and Fusarin Biosynthetic Pathway. *Chem. Biol.* **2013**, *20*, 1055–1066. [\[CrossRef\]](#) [\[PubMed\]](#)
50. Gaffoor, I.; Brown, D.W.; Plattner, R.; Proctor, R.H.; Qi, W.; Trail, F. Functional Analysis of the Polyketide Synthase Genes in the Filamentous Fungus *Gibberella Zeae* (Anamorph *Fusarium graminearum*). *Eukaryot. Cell* **2005**, *4*, 1926–1933. [\[CrossRef\]](#) [\[PubMed\]](#)
51. Lysoe, E.; Klemsdal, S.S.; Bone, K.R.; Frandsen, R.J.N.; Johansen, T.; Thrane, U.; Giese, H. The PKS4 Gene of *Fusarium graminearum* Is Essential for Zearalenone Production. *Appl. Environ. Microbiol.* **2006**, *72*, 3924–3932. [\[CrossRef\]](#)
52. Plissonneau, C.; Hartmann, F.E.; Croll, D. Pangenome Analyses of the Wheat Pathogen *Zymoseptoria tritici* Reveal the Structural Basis of a Highly Plastic Eukaryotic Genome. *BMC Biol.* **2018**, *16*, 5. [\[CrossRef\]](#)
53. Walkowiak, S.; Bonner, C.T.; Wang, L.; Blackwell, B.; Rowland, O.; Subramaniam, R. Intraspecific Interaction of *Fusarium graminearum* Contributes to Reduced Toxin Production and Virulence. *Mol. Plant Microbe Interact.* **2015**, *28*, 1256–1267. [\[CrossRef\]](#)
54. Jansen, C.; von Wettstein, D.; Schäfer, W.; Kogel, K.-H.; Felk, A.; Maier, F.J. Infection Patterns in Barley and Wheat Spikes Inoculated with Wild-Type and Trichodiene Synthase Gene Disrupted *Fusarium graminearum*. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 16892–16897. [\[CrossRef\]](#)
55. McCormick, S.P.; Alexander, N.J. *Fusarium Tri8* Encodes a Trichothecene C-3 Esterase. *Appl. Environ. Microbiol.* **2002**, *68*, 2959–2964. [\[CrossRef\]](#)
56. Fabre, F.; Vignassa, M.; Urbach, S.; Langin, T.; Bonhomme, L. Time-Resolved Dissection of the Molecular Crosstalk Driving *Fusarium* Head Blight in Wheat Provides New Insights into Host Susceptibility Determinism. *Plant Cell Environ.* **2019**, *42*, 2291–2308. [\[CrossRef\]](#)
57. Narasipura, S.D.; Chaturvedi, V.; Chaturvedi, S. Characterization of *Cryptococcus Neoformans* Variety Gattii SOD2 Reveals Distinct Roles of the Two Superoxide Dismutases in Fungal Biology and Virulence. *Mol. Microbiol.* **2005**, *55*, 1782–1800. [\[CrossRef\]](#)

58. Ma, L.-J.; van der Does, H.C.; Borkovich, K.A.; Coleman, J.J.; Daboussi, M.-J.; Di Pietro, A.; Dufresne, M.; Freitag, M.; Grabherr, M.; Henrissat, B.; et al. Comparative Genomics Reveals Mobile Pathogenicity Chromosomes in *Fusarium*. *Nature* **2010**, *464*, 367–373. [\[CrossRef\]](#)
59. Walkowiak, S.; Rowland, O.; Rodrigue, N.; Subramaniam, R. Whole Genome Sequencing and Comparative Genomics of Closely Related *Fusarium* Head Blight Fungi: *Fusarium graminearum*, *F. meridionale* and *F. asiaticum*. *BMC Genom.* **2016**, *17*, 1014. [\[CrossRef\]](#)
60. King, R.; Brown, N.A.; Urban, M.; Hammond-Kosack, K.E. Inter-Genome Comparison of the Quorn Fungus *Fusarium venenatum* and the Closely Related Plant Infecting Pathogen *Fusarium graminearum*. *BMC Genom.* **2018**, *19*, 269. [\[CrossRef\]](#)
61. Tralamazza, S.M.; Rocha, L.O.; Oggenfuss, U.; Corrêa, B.; Croll, D. Complex Evolutionary Origins of Specialized Metabolite Gene Cluster Diversity among the Plant Pathogenic Fungi of the *Fusarium graminearum* Species Complex. *Genome Biol. Evol.* **2019**, *11*, 3106–3122. [\[CrossRef\]](#)
62. Leplat, J.; Friberg, H.; Abid, M.; Steinberg, C. Survival of *Fusarium graminearum*, the Causal Agent of Fusarium Head Blight. A Review. *Agron. Sustain. Dev.* **2013**, *33*, 97–111. [\[CrossRef\]](#)
63. Pomraning, K.R.; Connolly, L.R.; Whalen, J.P.; Smith, K.M.; Freitag, M. Repeat-Induced Point Mutation, DNA Methylation and Heterochromatin in *Gibberella Zeae* (Anamorph: *Fusarium graminearum*). In *Fusarium Genomics, Molecular and Cellular Biology*; Brown, D., Proctor, R.H., Eds.; Horizon Scientific Press: Norwich, UK, 2013; pp. 93–109.
64. Galagan, J.E.; Selker, E.U. RIP: The Evolutionary Cost of Genome Defense. *Trends Genet.* **2004**, *20*, 417–423. [\[CrossRef\]](#) [\[PubMed\]](#)
65. Rouxel, T.; Grandaubert, J.; Hane, J.K.; Hoede, C.; van de Wouw, A.P.; Couloux, A.; Dominguez, V.; Anthouard, V.; Bally, P.; Bourras, S.; et al. Effector Diversification within Compartments of the *Leptosphaeria maculans* Genome Affected by Repeat-Induced Point Mutations. *Nat. Commun.* **2011**, *2*, 202. [\[CrossRef\]](#) [\[PubMed\]](#)
66. Amyotte, S.G.; Tan, X.; Pennerman, K.; Jimenez-Gasco, M. del M.; Klosterman, S.J.; Ma, L.-J.; Dobinson, K.F.; Veronese, P. Transposable Elements in Phytopathogenic *Verticillium* Spp.: Insights into Genome Evolution and Inter- and Intra-Specific Diversification. *BMC Genom.* **2012**, *13*, 314. [\[CrossRef\]](#)
67. Mat Razali, N.; Cheah, B.H.; Nadarajah, K. Transposable Elements Adaptive Role in Genome Plasticity, Pathogenicity and Evolution in Fungal Phytopathogens. *Int. J. Mol. Sci.* **2019**, *20*, 3597. [\[CrossRef\]](#)
68. Aguilera, G.; Lengelle, J.; Chiapello, H.; Giraud, T.; Viaud, M.; Fournier, E.; Rodolphe, F.; Marthey, S.; Ducasse, A.; Gendrault, A.; et al. Genes under Positive Selection in a Model Plant Pathogenic Fungus, *Botrytis*. *Infect. Genet. Evol.* **2012**, *12*, 987–996. [\[CrossRef\]](#) [\[PubMed\]](#)
69. Connolly, L.R.; Smith, K.M.; Freitag, M. The *Fusarium graminearum* Histone H3 K27 Methyltransferase KMT6 Regulates Development and Expression of Secondary Metabolite Gene Clusters. *PLoS Genet.* **2013**, *9*, e1003916. [\[CrossRef\]](#) [\[PubMed\]](#)
70. Menke, J.; Dong, Y.; Kistler, H.C. *Fusarium graminearum* Tri12p Influences Virulence to Wheat and Trichothecene Accumulation. *Mol. Plant Microbe Interact.* **2012**, *25*, 1408–1418. [\[CrossRef\]](#)
71. Lind, A.L.; Wisecaver, J.H.; Lameiras, C.; Wiemann, P.; Palmer, J.M.; Keller, N.P.; Rodrigues, F.; Goldman, G.H.; Rokas, A. Drivers of Genetic Diversity in Secondary Metabolic Gene Clusters within a Fungal Species. *PLoS Biol.* **2017**, *15*, e2003583. [\[CrossRef\]](#)
72. Throckmorton, K.; Lim, F.Y.; Kontoyiannis, D.P.; Zheng, W.; Keller, N.P. Redundant Synthesis of a Conidial Polyketide by Two Distinct Secondary Metabolite Clusters in *Aspergillus fumigatus*. *Environ. Microbiol.* **2016**, *18*, 246–259. [\[CrossRef\]](#) [\[PubMed\]](#)
73. Drott, M.T.; Bastos, R.W.; Rokas, A.; Ries, L.N.A.; Gabaldón, T.; Goldman, G.H.; Keller, N.P.; Greco, C. Diversity of Secondary Metabolism in *Aspergillus nidulans* Clinical Isolates. *mSphere* **2020**, *5*. [\[CrossRef\]](#)
74. Schumacher, J.; Gautier, A.; Morgant, G.; Studt, L.; Ducrot, P.-H.; Pêcheur, P.L.; Azeddine, S.; Fillinger, S.; Leroux, P.; Tudzynski, B.; et al. A Functional Bikaverin Biosynthesis Gene Cluster in Rare Strains of *Botrytis cinerea* Is Positively Controlled by VELVET. *PLoS ONE* **2013**, *8*, e53729. [\[CrossRef\]](#)
75. Rafiqi, M.; Ellis, J.G.; Ludowici, V.A.; Hardham, A.R.; Dodds, P.N. Challenges and Progress towards Understanding the Role of Effectors in Plant-Fungal Interactions. *Curr. Opin. Plant Biol.* **2012**, *15*, 477–482. [\[CrossRef\]](#) [\[PubMed\]](#)
76. Van Schie, C.C.N.; Takken, F.L.W. Susceptibility Genes 101: How to Be a Good Host. *Annu. Rev. Phytopathol.* **2014**, *52*, 551–581. [\[CrossRef\]](#) [\[PubMed\]](#)
77. Fabre, F.; Rocher, F.; Alouane, T.; Langin, T.; Bonhomme, L. Searching for FHB Resistances in Bread Wheat: Susceptibility at the Crossroad. *Front. Plant Sci.* **2020**, *11*. [\[CrossRef\]](#) [\[PubMed\]](#)
78. Fabre, F.; Urbach, S.; Roche, S.; Langin, T.; Bonhomme, L. Proteomics-Based Data Integration of Wheat Cultivars Facing *Fusarium graminearum* Strains Revealed a Core-Responsive Pattern Controlling Fusarium Head Blight. *Front. Plant Sci.* **2021**, *12*. [\[CrossRef\]](#)
79. Boenisch, M.J.; Schäfer, W. *Fusarium graminearum* Forms Mycotoxin Producing Infection Structures on Wheat. *BMC Plant Biol.* **2011**, *11*, 110. [\[CrossRef\]](#)
80. Blümke, A.; Falter, C.; Herrfurth, C.; Sode, B.; Bode, R.; Schäfer, W.; Feussner, I.; Voigt, C.A. Secreted Fungal Effector Lipase Releases Free Fatty Acids to Inhibit Innate Immunity-Related Callose Formation during Wheat Head Infection. *Plant Physiol.* **2014**, *165*, 346–358. [\[CrossRef\]](#)
81. Miedaner, T.; Reinbrecht, C.; Schilling, A.G. Association among Aggressiveness, Fungal Colonization, and Mycotoxin Production of 26 Isolates of *Fusarium graminearum* in Winter Rye Head Blight/Beziehung Zwischen Aggressivität, Myzelwachstum Und Mykotoxinproduktion von 26 *Fusarium graminearum*-Isolaten Bei Der Ährenfusariose Des Winterroggens. *Z. Pflanzenkrankh. Pflanzenschutz J. Plant. Dis. Prot.* **2000**, *107*, 124–134.
82. Voigt, C.A.; Schäfer, W.; Salomon, S. A Secreted Lipase of *Fusarium graminearum* Is a Virulence Factor Required for Infection of Cereals. *Plant J.* **2005**, *42*, 364–375. [\[CrossRef\]](#)

83. Bönninghausen, J.; Gebhard, D.; Kröger, C.; Hader, B.; Tumforde, T.; Lieberei, R.; Bergemann, J.; Schäfer, W.; Bormann, J. Disruption of the GABA Shunt Affects Mitochondrial Respiration and Virulence in the Cereal Pathogen *Fusarium Graminearum*. *Mol. Microbiol.* **2015**, *98*, 1115–1132. [CrossRef]
84. Bönninghausen, J.; Schauer, N.; Schäfer, W.; Bormann, J. Metabolic Profiling of Wheat Rachis Node Infection by *Fusarium graminearum*—Decoding Deoxynivalenol-Dependent Susceptibility. *New Phytol.* **2019**, *221*, 459–469. [CrossRef] [PubMed]
85. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef] [PubMed]
86. Simpson, J.T.; Wong, K.; Jackman, S.D.; Schein, J.E.; Jones, S.J.M.; Birol, I. ABySS: A Parallel Assembler for Short Read Sequence Data. *Genome Res.* **2009**, *19*, 1117–1123. [CrossRef] [PubMed]
87. Chin, C.-S.; Alexander, D.H.; Marks, P.; Klammer, A.A.; Drake, J.; Heiner, C.; Clum, A.; Copeland, A.; Huddleston, J.; Eichler, E.E.; et al. Nonhybrid, Finished Microbial Genome Assemblies from Long-Read SMRT Sequencing Data. *Nat. Methods* **2013**, *10*, 563–569. [CrossRef]
88. Waterhouse, R.M.; Seppey, M.; Simão, F.A.; Manni, M.; Ioannidis, P.; Klioutchnikov, G.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. *Mol. Biol. Evol.* **2018**, *35*, 543–548. [CrossRef]
89. Holt, C.; Yandell, M. MAKER2: An Annotation Pipeline and Genome-Database Management Tool for Second-Generation Genome Projects. *BMC Bioinform.* **2011**, *12*, 491. [CrossRef]
90. Bushmanova, E.; Antipov, D.; Lapidus, A.; Pribelski, A.D. RnaSPAdes: A de Novo Transcriptome Assembler and Its Application to RNA-Seq Data. *GigaScience* **2019**, *8*. [CrossRef]
91. Korf, I. Gene Finding in Novel Genomes. *BMC Bioinform.* **2004**, *5*, 59. [CrossRef] [PubMed]
92. Stanke, M.; Keller, O.; Gunduz, I.; Hayes, A.; Waack, S.; Morgenstern, B. AUGUSTUS: Ab Initio Prediction of Alternative Transcripts. *Nucleic Acids Res.* **2006**, *34*, W435–W439. [CrossRef]
93. Smit, A.; Hubley, R.; Green, P. RepeatMasker Open-4.0. Available online: <http://www.repeatmasker.org> (accessed on 8 October 2020).
94. Hirsh, A.E.; Fraser, H.B. Protein Dispensability and Rate of Evolution. *Nature* **2001**, *411*, 1046–1049. [CrossRef]
95. Jordan, I.K.; Rogozin, I.B.; Wolf, Y.I.; Koonin, E.V. Essential Genes Are More Evolutionarily Conserved than Are Nonessential Genes in Bacteria. *Genome Res.* **2002**, *12*, 962–968. [CrossRef]
96. Lechner, M.; Findeiß, S.; Steiner, L.; Marz, M.; Stadler, P.F.; Prohaska, S.J. Proteinortho: Detection of (Co-)Orthologs in Large-Scale Analysis. *BMC Bioinform.* **2011**, *12*, 124. [CrossRef]
97. Jones, P.; Binns, D.; Chang, H.-Y.; Fraser, M.; Li, W.; McAnulla, C.; McWilliam, H.; Maslen, J.; Mitchell, A.; Nuka, G.; et al. InterProScan 5: Genome-Scale Protein Function Classification. *Bioinformatics* **2014**, *30*, 1236–1240. [CrossRef] [PubMed]
98. Klopfenstein, D.V.; Zhang, L.; Pedersen, B.S.; Ramírez, F.; Warwick Vesztrocy, A.; Naldi, A.; Mungall, C.J.; Yunes, J.M.; Botvinnik, O.; Weigel, M.; et al. GOATOOLS: A Python Library for Gene Ontology Analyses. *Sci. Rep.* **2018**, *8*, 10872. [CrossRef] [PubMed]
99. Katoh, K.; Standley, D.M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [CrossRef]
100. Castresana, J. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Mol. Biol. Evol.* **2000**, *17*, 540–552. [CrossRef]
101. Stamatakis, A. RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [CrossRef]
102. Petersen, T.N.; Brunak, S.; von Heijne, G.; Nielsen, H. SignalP 4.0: Discriminating Signal Peptides from Transmembrane Regions. *Nat. Methods* **2011**, *8*, 785–786. [CrossRef] [PubMed]
103. Bendtsen, J.D.; Jensen, L.J.; Blom, N.; von Heijne, G.; Brunak, S. Feature-Based Prediction of Non-Classical and Leaderless Protein Secretion. *Protein Eng. Des. Sel.* **2004**, *17*, 349–356. [CrossRef]
104. Emanuelsson, O.; Nielsen, H.; Brunak, S.; von Heijne, G. Predicting Subcellular Localization of Proteins Based on Their N-Terminal Amino Acid Sequence. *J. Mol. Biol.* **2000**, *300*, 1005–1016. [CrossRef]
105. Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E.L. Predicting Transmembrane Protein Topology with a Hidden Markov Model: Application to Complete Genomes. *J. Mol. Biol.* **2001**, *305*, 567–580. [CrossRef] [PubMed]
106. Horton, P.; Park, K.-J.; Obayashi, T.; Fujita, N.; Harada, H.; Adams-Collier, C.J.; Nakai, K. WoLF PSORT: Protein Localization Predictor. *Nucleic Acids Res.* **2007**, *35*, W585–W587. [CrossRef] [PubMed]
107. Pierleoni, A.; Martelli, P.L.; Casadio, R. PredGPI: A GPI-Anchor Predictor. *BMC Bioinform.* **2008**, *9*, 392. [CrossRef] [PubMed]
108. Sperschneider, J.; Catanzariti, A.-M.; DeBoer, K.; Petre, B.; Gardiner, D.M.; Singh, K.B.; Dodds, P.N.; Taylor, J.M. LOCALIZER: Subcellular Localization Prediction of Both Plant and Effector Proteins in the Plant Cell. *Sci. Rep.* **2017**, *7*, 44598. [CrossRef] [PubMed]
109. Kosugi, S.; Hasebe, M.; Tomita, M.; Yanagawa, H. Systematic Identification of Cell Cycle-Dependent Yeast Nucleocytoplasmic Shuttling Proteins by Prediction of Composite Motifs. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 10171–10176. [CrossRef]
110. Yin, Y.; Mao, X.; Yang, J.; Chen, X.; Mao, F.; Xu, Y. DbCAN: A Web Resource for Automated Carbohydrate-Active Enzyme Annotation. *Nucleic Acids Res.* **2012**, *40*, W445–W451. [CrossRef]
111. Rawlings, N.D.; Barrett, A.J.; Thomas, P.D.; Huang, X.; Bateman, A.; Finn, R.D. The MEROPS Database of Proteolytic Enzymes, Their Substrates and Inhibitors in 2017 and a Comparison with Peptidases in the PANTHER Database. *Nucleic Acids Res.* **2018**, *46*, D624–D632. [CrossRef]

112. Winnenburg, R.; Baldwin, T.K.; Urban, M.; Rawlings, C.; Köhler, J.; Hammond-Kosack, K.E. PHI-Base: A New Database for Pathogen Host Interactions. *Nucleic Acids Res.* **2006**, *34*, D459–D464. [\[CrossRef\]](#)
113. Urban, M.; Cuzick, A.; Seager, J.; Wood, V.; Rutherford, K.; Venkatesh, S.Y.; De Silva, N.; Martinez, M.C.; Pedro, H.; Yates, A.D.; et al. PHI-Base: The Pathogen–Host Interactions Database. *Nucleic Acids Res.* **2020**, *48*, D613–D620. [\[CrossRef\]](#)
114. Sperschneider, J.; Dodds, P.N.; Gardiner, D.M.; Singh, K.B.; Taylor, J.M. Improved Prediction of Fungal Effector Proteins from Secretomes with EffectorP 2.0. *Mol. Plant Pathol.* **2018**, *19*, 2094–2110. [\[CrossRef\]](#) [\[PubMed\]](#)
115. Wu, T.D.; Reeder, J.; Lawrence, M.; Becker, G.; Brauer, M.J. GMAP and GSNAP for Genomic Sequence Alignment: Enhancements to Speed, Accuracy, and Functionality. *Methods Mol. Biol.* **2016**, *1418*, 283–334. [\[CrossRef\]](#) [\[PubMed\]](#)
116. Gel, B.; Serra, E. KaryoploteR: An R/Bioconductor Package to Plot Customizable Genomes Displaying Arbitrary Data. *Bioinformatics* **2017**, *33*, 3088–3090. [\[CrossRef\]](#) [\[PubMed\]](#)
117. Li, H. Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM. *arXiv* **2013**, arXiv:1303.3997.
118. Li, H. Minimap2: Pairwise Alignment for Nucleotide Sequences. *Bioinformatics* **2018**, *34*, 3094–3100. [\[CrossRef\]](#)
119. Okonechnikov, K.; Conesa, A.; García-Alcalde, F. Qualimap 2: Advanced Multi-Sample Quality Control for High-Throughput Sequencing Data. *Bioinformatics* **2016**, *32*, 292–294. [\[CrossRef\]](#)
120. der Auwera, G.A.V.; Carneiro, M.O.; Hartl, C.; Poplin, R.; del Angel, G.; Levy-Moonshine, A.; Jordan, T.; Shakir, K.; Roazen, D.; Thibault, J.; et al. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Curr. Protoc. Bioinform.* **2013**, *43*, 11.10.1–11.10.33. [\[CrossRef\]](#)
121. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [\[CrossRef\]](#)
122. Koboldt, D.C.; Zhang, Q.; Larson, D.E.; Shen, D.; McLellan, M.D.; Lin, L.; Miller, C.A.; Mardis, E.R.; Ding, L.; Wilson, R.K. VarScan 2: Somatic Mutation and Copy Number Alteration Discovery in Cancer by Exome Sequencing. *Genome Res.* **2012**, *22*, 568–576. [\[CrossRef\]](#)
123. Cingolani, P.; Platts, A.; Wang, L.L.; Coon, M.; Nguyen, T.; Wang, L.; Land, S.J.; Lu, X.; Ruden, D.M. A Program for Annotating and Predicting the Effects of Single Nucleotide Polymorphisms, SnpEff. *Fly* **2012**, *6*, 80–92. [\[CrossRef\]](#)
124. Cingolani, P.; Patel, V.M.; Coon, M.; Nguyen, T.; Land, S.J.; Ruden, D.M.; Lu, X. Using *Drosophila Melanogaster* as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. *Front. Genet.* **2012**, *3*, 35. [\[CrossRef\]](#) [\[PubMed\]](#)
125. Danecek, P.; Auton, A.; Abecasis, G.; Albers, C.A.; Banks, E.; DePristo, M.A.; Handsaker, R.E.; Lunter, G.; Marth, G.T.; Sherry, S.T.; et al. The Variant Call Format and VCFtools. *Bioinformatics* **2011**, *27*, 2156–2158. [\[CrossRef\]](#) [\[PubMed\]](#)
126. Pluskal, T.; Castillo, S.; Villar-Briones, A.; Orešič, M. MZmine 2: Modular Framework for Processing, Visualizing, and Analyzing Mass Spectrometry-Based Molecular Profile Data. *BMC Bioinform.* **2010**, *11*, 395. [\[CrossRef\]](#) [\[PubMed\]](#)
127. Chong, J.; Wishart, D.S.; Xia, J. Using MetaboAnalyst 4.0 for Comprehensive and Integrative Metabolomics Data Analysis. *Curr. Protoc. Bioinform.* **2019**, *68*, e86. [\[CrossRef\]](#)

Conclusion

L'objectif de ce travail était d'étudier les liens entre la diversité structurale et fonctionnelle des génomes de *F. graminearum* et leur agressivité contrastée. Cette analyse a été réalisée via l'approche de séquençage et d'analyse comparative intra-espèce de huit nouveaux génomes européens sélectionnés pour leur niveau d'agressivité contrasté. Un résultat original obtenu lors de ces travaux réside dans la construction du premier pangénome de *F. graminearum*, offrant de nouvelles opportunités sans précédent pour découvrir de nouveaux gènes, explorer la diversité génétique et améliorer notre connaissance des forces évolutives qui façonnent l'organisation et la dynamique des génomes de l'espèce *F. graminearum*. Outre le pangénome, une proportion élevée (plus de 70%) du pan-sécrétome, notamment des gènes codant pour les effecteurs protéiques appartenaient au core sécrétome, suggérant que l'agressivité des souches de *F. graminearum* est caractérisée par une signature partagée entre les souches. Nos résultats ont également fourni des informations précieuses sur l'organisation de ces effecteurs protéiques et sur l'ensemble des gènes susceptibles d'une sélection diversifiée le long des quatre chromosomes. Cependant, leurs localisations préférentielles dans des compartiments en évolution rapide pourraient jouer un rôle clé dans la capacité de ce phytopathogène de répondre à ses forts besoins évolutifs. Ainsi, leurs localisations subcellulaires dans le chloroplaste et le noyau de la plante comme cible privilégiés devraient constituer un point stratégique pour le succès de l'infection et une source d'information originale pour établir de nouvelles hypothèses sur les déterminants moléculaires de l'interaction.

Résultats – *A. baumannii*

Chapitre 3

Introduction

Outre l'importance sanitaire et économique via la contamination des céréales, l'une des crises sanitaires mondiales les plus imminentes est la prévalence croissante et la dissémination mondiale de bactéries pathogènes hautement résistantes, capables de persister en milieu hospitalier malgré les procédures de contrôle des infections. L'OMS a classé *A. baumannii* en premier ordre (priorité 1 : critique) des agents pathogènes nosocomiaux en raison de leur remarquable capacité à acquérir des déterminants de résistance à tous les types d'antibiotiques (WHO). La connaissance du catalogue complet des gènes, des variants génétiques et des lignées clonales associées, est une des premières étapes dans la recherche de nouvelles stratégies pour contrôler ce micro-organisme. La plupart des études portant sur la caractérisation du pangénome d'*A. baumannii* se sont focalisées particulièrement sur des souches cliniques (isolées de patients), cependant, la définition du core génome et du génome dispensable de cette espèce en combinaison avec une collection de souches environnementales reste peu documentée. Ainsi, bien que les clones internationaux soient considérés comme la source de la plupart des infections, l'identité de lignées clonales et leur contenu en virulome et résistome n'ont pas été autant étudiés que dans les pays développés.

Dans ces travaux de recherche, treize nouvelles souches marocaines multirésistantes d'*A. baumannii* ont été sélectionnées : 8 souches isolées de patients et cinq isolées de l'environnement hospitalier, dont une souche résistante à la colistine (dernière ligne de traitement). Les tests de sensibilité aux antibiotiques ont été effectués comme décrit par Uwingabiye *et al.* (2017) au sein de l'hôpital militaire de Rabat. Nous avons ensuite réalisé un séquençage de génomes entiers couplé à des analyses génomiques comparatives intra-espèces afin d'étudier leur diversité structurale et fonctionnelle avec un accent particulier sur le résistome, tout en établissant un lien avec leur relation clonale. Ces analyses fournissent des informations importantes sur l'évolution d'*A. baumannii* en tant qu'agent pathogène nosocomial.



First Whole-Genome Sequences of Two Multidrug-Resistant *Acinetobacter baumannii* Strains Isolated from a Moroccan Hospital Floor

Tarek Alouane,^a Jean Uwingabiye,^b Abdelhay Lemnouer,^b Lamiaa Lahlou,^a Meriem Laamarti,^a Souad Kartti,^a Oussama Benhrif,^a Hamza El Misbahi,^a Mohammed Frikh,^b Yassine Benlahlou,^b Fatna Bssaibis,^b Soukaina El Abbassi,^b Samia Kabbage,^b Adil Maleb,^b Mostafa Elouennass,^b Azeddine Ibrahimia^a

Biotechnology Lab (MedBiotech), Rabat Medical and Pharmacy School, Mohammed V University, Rabat, Morocco^a; Department of Bacteriology, Mohammed V Military Teaching Hospital, Research Unit, Epidemiology and Bacterial resistance, Rabat Medical and Pharmacy School, Mohammed V University, Rabat, Morocco^b

ABSTRACT This report describes the whole-genome shotgun sequences of two multidrug-resistant *Acinetobacter baumannii* strains, ABE8_07 and ABE12_M, isolated from a Moroccan hospital floor. These two genome sequences will initiate the study and characterization of the *Acinetobacter baumannii* genome in Morocco.

Acinetobacter baumannii is a Gram-negative coccobacillus known for its remarkable persistence in hospital environments worldwide; it causes nosocomial infection, and reports have shown that it is potentially resistant to several antibiotics, thus limiting current therapeutic options (1, 2). Previous Moroccan studies have shown that, during the period of 2003 to 2016, this microorganism represented 6.94% of all bacterial infections and 9.6% of all Gram-negative bacilli (3) and that its antibiotic resistance rates increased from 23 to 76% for imipenem, 63 to 86% for ceftazidime, 41 to 52% for amikacin, and 68 to 87% for ciprofloxacin (3, 4).

In this study, we present the draft genome sequences of two multidrug-resistant *A. baumannii* strains, isolated from a Moroccan hospital floor, that demonstrate resistance to various antibiotics, including ticarcillin, piperacillin, ticarcillin-clavulanic acid, piperacillin-tazobactam, ceftazidime, cefepime, imipenem, gentamicin, tobramycin, amikacin, netilmicin, ciprofloxacin, tetracycline, and cotrimoxazole (but not colistin).

Genomic DNA was extracted using a DNA IQ system kit (Promega) and quantified using a NanoVue Plus spectrophotometer (Biochrom). The library of genomic DNA was prepared using a Nextera XT DNA library preparation kit (Illumina), with dual indexing adapters, and sequenced using an Illumina MiSeq sequencer with a 2 × 251-bp paired-end configuration. The sequencing generated 2,843,500 reads for ABE8_07 and 601,494 reads for ABE12_M. The sequences were trimmed and *de novo* assembled using CLC Genomics Workbench version 9.5 (Qiagen). Strain ABE8_07 comprised 187 contigs with a total length of 3,898,083 bp, an *N*₅₀ of 57,519 bp, and an average contig length of 20,845 bp, while strain ABE12_M comprised 196 contigs with a total size of 3,902,530 bp, an *N*₅₀ of 54,846 bp, and an average contig length of 19,911 bp. The G+C contents for ABE8_07 and ABE12_M were 38.97% and 38.95%, respectively.

The genomes were annotated using Prokka version 1.7 (5) and the Rapid Annotations using Subsystems Technology (RAST) server (6). The annotation of strain ABE8_07 detected 3,657 protein-coding genes, 63 tRNA genes, and 6 rRNAs, representing 456 subsystems. The annotation of strain ABE12_M detected 3,669 protein-coding genes, 61 tRNA genes, and 3 rRNA genes, representing 453 subsystems.

Received 11 March 2017 Accepted 15 March 2017 Published 4 May 2017

Citation Alouane T, Uwingabiye J, Lemnouer A, Lahlou L, Laamarti M, Kartti S, Benhrif O, El Misbahi H, Frikh M, Benlahlou Y, Bssaibis F, El Abbassi S, Kabbage S, Maleb A, Elouennass M, Ibrahim A. 2017. First whole-genome sequences of two multidrug-resistant *Acinetobacter baumannii* strains isolated from a Moroccan hospital floor. Genome Announc 5:e00298-17. <https://doi.org/10.1128/genomeA.00298-17>.

Copyright © 2017 Alouane et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Address correspondence to Tarek Alouane, tarek.alouane@um5s.net.ma.

Analysis of antimicrobial resistance genes by ResFinder version 2.1 (7) indicated the presence of multiple genes encoding for resistance to β -lactams (*bla*_{ADC-25s}, *bla*_{OXA-23s}, *bla*_{OXA-66s}, and *bla*_{TEM-1D}), aminoglycosides [*aph*(3')-VIa, *aph* [3']-Ic, *aadA1*, *aacA4*, *strA*, *strB*, and *armA*], fluoroquinolone-aminoglycosides [*aac*(6')Ib-cr], macrolide-lincosamide-streptogramin [*msr*(E) and *mph*(E)], sulfonamide (*sul1*), phenicol (*catB8*), and tetracycline [*tet*(B)] in both the genomes, except for *aadB*, which was present in ABE12_M only.

Accession number(s). This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession numbers [FPKN00000000](#) and [FPEF00000000](#) for strains ABE8_07 and ABE12_M, respectively. The versions described in this article are the first versions, FPKN01000000 and FPEF01000000.

ACKNOWLEDGMENTS

This work was carried out under national funding from the Moroccan Ministry of Higher Education and Scientific Research (PPR program) to A.I. This work was also supported by a grant from the NIH for H3Africa BioNet to A.I.

REFERENCES

1. Roca I, Espinal P, Vila-Farrés X, Vila J. 2012. The *Acinetobacter baumannii* oxymoron: commensal hospital dweller turned pan-drug-resistant menace. *Front Microbiol* 3:148. <https://doi.org/10.3389/fmicb.2012.00148>.
2. Weber BS, Harding CM, Feldman MF. 2015. Pathogenic *Acinetobacter*: from the cell surface to infinity and beyond. *J Bacteriol* 198:880–887. <https://doi.org/10.1128/JB.00906-15>.
3. Uwingabiye J, Frikh M, Lemnouer A, Bssaibis F, Belefquih B, Maleb A, Dahraoui S, Belyamani L, Bait A, Haimeur C, Louzi L, Ibrahimi A, Elouennass M. 2016. *Acinetobacter* infections prevalence and frequency of the antibiotics resistance: comparative study of intensive care units versus other hospital units. *Pan Afr Med J* 23:191. <https://doi.org/10.11604/pamj.2016.23.191.7915>.
4. Elouennass M, Bajou T, Lemnouer AH, Foissaud V, Hervé V, Baaj AJ. 2003. *Acinetobacter baumannii*: étude de la sensibilité des souches isolées à l'hôpital militaire d'instruction Mohammed V, Rabat, Maroc. *Med Mal Infect* 33:361–364. [https://doi.org/10.1016/S0399-077X\(03\)00205-1](https://doi.org/10.1016/S0399-077X(03)00205-1).
5. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>.
6. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST server: rapid annotations using subsystems technology. *BMC Genomics* 9:75. <https://doi.org/10.1186/1471-2164-9-75>.
7. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen MV. 2012. Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother* 67:2640–2644. <https://doi.org/10.1093/jac/dks261>.

Comparative genomics of multidrug-resistant *Acinetobacter baumannii* strains isolated from patients and environment in the intensive care unit of a Moroccan hospital

Tarek Alouane¹, Jean Uwingabiy², Meriem Laamarti¹, Mostafa Elouennass², and Azeddine Ibrahimi^{1,*}

¹ Medical Biotechnology Laboratory (MedBiotech), Bioinova Research Center, Rabat Medical and Pharmacy School, Mohammed Vth University, Rabat 10100, Morocco; alouane.tarej@gmail.com (T.A); M_laamarti@yahoo.com (M.L); a.ibrahimi@um5r.ac.ma (A.I);

² Department of Clinical Bacteriology, Mohammed V Military Teaching Hospital, Research Team of Epidemiology and Bacterial Resistance, Faculty of Medicine and Pharmacy, Mohammed V University, Rabat, Morocco elouennassm@yahoo.fr (M.E).

* Correspondence: a.ibrahimi@um5s.net.ma (A.I)

Abstract: *Acinetobacter baumannii* (*A. baumannii*) is an opportunistic pathogen commonly associated with nosocomial infections. The spread of multidrug-resistant (MDR) strains with their ability to persist and survive in hospital environments poses an urgent threat to public health. In this study, we present multiscale comparative analyzes of the first complete genomes of MDR *A. baumannii* Moroccan strains, divided into 5 strains isolated from the hospital environment and 8 clinical strains isolated from patients, including a strain resistant to colistin. Our results revealed three sequence types (ST2, ST85, ST315) circulating in the Moroccan hospital, belonging to the three distinct phylogenomic clusters. Moreover, a closely related relationship between clinical and environmental strains was observed, being both shared more than 80% of the total identified virulence genes. While a remarkable variability of resistance genes found was mainly linked to the clonal lineages of the strains. In addition, we also identified a mutation (p.Arg263Ser) in the PmrB gene, potentially involved in resistance to colistin. On the other hand, we have shown that the choice of the reference genome of *A. baumannii* has a direct impact on the analysis of genetic variants, including the identification of SNPs according to the clonal identity of strains.

Overall, our analyses described new knowledge for a better understanding of the genomic diversity of clinical and environmental strains of MDR *A. baumannii* and their clonal relationships.

Keywords: *Acinetobacter baumannii*; Environment; Clinical; comparative genomics; carbapenem-resistant; colistin-resistant; Clonal Lineage

1. Introduction

A. baumannii is an opportunistic Gram-negative bacterium considered to be one of the most prevalent pathogens associated with nosocomial infections [1–2]. Its ability to persist in biotic and abiotic environments and its high level of intrinsic and acquired resistance to antibiotics have made *A. baumannii* an effective pathogen in the hospital setting [3–4]. Carbapenem-resistant *A. baumannii* is currently classified as the top pathogen on the World Health Organization's (WHO) list of critical priorities for the development of new antibiotics [5]. Carbapenem resistance is increasingly reported among strains of *A. baumannii* [6–11]. However, current treatment options are extremely limited and administration of colistin (polymyxin E) has been reintroduced as a last-line antibiotic to treat Gram-negative infections like *A. baumannii* [12–15].

A. baumannii remains a heterogeneous species with over 2000 sequence types (STs) defined via multilocus sequence type typing "MLST" (<https://pubmlst.org/abaumannii>; accessed 15/03/2022). Although this species is composed of several well-differentiated clones and many of them display a multidrug-resistant (MDR) phenotype, their dissemination has been studied mainly from clinical strains (isolated from patients) [16–18]. Consequently, our knowledge of the clonal relationships of environmental strains in hospital settings remains poorly documented and has yet to be linked to their virulence and antibiotic resistance profile. Additionally, although this species is known for relatively few virulence factors compared to other gram-negative pathogens [19], genome-wide comparative analysis of clinical and environmental strains can elucidate clues regarding potential virulence factors and their role in adaptation to the hospital environment.

In this study, we performed a sequencing and a comparative genomic analysis of 13 new Moroccan strains of *A. baumannii* resistant to carbapenems including one resistant to colistin and divided into eight clinical strains isolated from different patients and five environmental strains from the same hospital environment. We first predicted the gene repertoire of *A. baumannii* via pangenome construction to reveal genetic characteristics between strains from two isolation sources (clinical and environmental) with their phylogenomic relationship completed with 10 genomes belonged to the three international clones (IC-I to IC-III). We then identified and studied the distribution of genetic determinants associated with virulence and antimicrobial resistance between the strains studied with the epidemic potential clonal lineages. Additionally, we assessed for the first time how the choice of reference genomes can impact mutation detection in *A. baumannii*.

2. Materials and Methods

2.1. *A. baumannii* strains and Antibiotic susceptibility

In this study, we report thirteen newly sequenced strains of *A. baumannii* collected between March 2015 and January 2017 (Table S1) at Mohammed V University Military Hospital in Morocco. Eight of these strains were isolated from infected patients (clinical strains) while five were isolated from hospital settings (environmental strains). Environmental isolates were collected from floors, medical ventilators, and patient carts, while clinical isolates were recovered from the mouth, anal margin, urine, bronchial aspirate, and Protected Distal Sampling (PDS). The sampling and antimicrobial susceptibility tests were performed as described by Uwingabiye et al. [20]. These tests were carried out with β -lactam (imipenem, cefepime, ceftazidime, piperacillin, ticarcillin, ticarcillin-Clavulanic acid, piperacillin-Tazobactam), aminoglycosides (gentamicin, tobramycin, netilmicin), Fluoroquinolone (ciprofloxacin), sulfonamide (cotrimoxazole), tetracycline, rifamycin (Rifampicin), and polymyxin E (colistin).

2.2. Library Preparation and Sequencing

DNA extractions from overnight cultures were performed using DNA IQ™ System (Promega Corporation, Madison, WI, USA) according to the manufacturer's instructions. The quality and quantity of extracted genomic DNA were determined by NanoVue™ Plus Spectrophotometer (GE Health Care, Buckinghamshire, United Kingdom) and the Quantus™ Fluorometer (Promega, Germany), respectively. A total of 1 nanogram of genomic DNA was used as input for library preparation with Nextera XT DNA Library Preparation kit (Illumina, San Diego, CA), according to manufacturer's instructions, and sequenced using Illumina MiSeq platform, with a 2 × 251-bp paired-end configuration following the manufacturer's instructions.

2.3. Genome Assembly, Gene Prediction, and Annotation

Prior to genome assembly, a preprocessing analysis was performed. The quality of the raw reads from sequencing was evaluated using the FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>), and preprocessed with Trimmomatic v.36 [21] using default settings. The genomes were assembled using SPAdes v3.13 [22] with k values of 21, 33, 55, 77, 99, and 127 and their completeness was assessed by the presence of 366 single-copy conserved genes of the Gammaproteobacteria lineage using Benchmarking Universal Single-Copy Orthologs Version 5.1.2 (BUSCO v5.1.2) [23]. The assembled genomes were structurally annotated using Prokka pipeline v1.14.6 [24], while functional annotation was performed using InterProScan v5.13-52 [25], including the Pfam database families.

2.4. Pangenome Construction and Phylogenomic Analysis

The genomes of thirteen strains of this study and ten other genomes of the same species representing the three IC of *A. baumannii*: IC-I (A1, AYE, AB0057, AB307-0294), IC-II (ACICU, MDR-TJ, 1656-2, BJAB0868, TCDC-AB0715) and IC-III (OIFC1378), were used to calculate the

pangenome with two different approaches: Agenome-based approach using the ppsPCP pipeline [26] to find the complete complement of genomic sequences of the 23 genomes and a gene-based approach that reciprocal the best BLAST hits [27] to identify the set of orthologous genes using the Proteinortho v5.16b [28], with an identity score above 70% and a sequence coverage above 80% (e-value below 1×10^{-5}). Core genome and pangenome size curves were generated using proteinortho curves [28] with 100 iterations.

The construction of the phylogenomic tree of the 23 genomes was based on the core genome; single-copy orthologs. Alignment of these orthologous proteins was performed using MAFFT v7.471 [29] and misaligned regions were removed using Gblocks v0.91b [30] with default parameters. Maximum likelihood (ML) analysis was performed with RaxML v8.1.20 [31] using the PROTGAMLMG amino acid substitution matrix model with 1000 bootstrap replicates. The phylogenomic tree was visualized using FigTree v1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree>).

2.5. Resistome and virulome prediction

The resistome of each of the thirteen *A. baumannii* strains was characterized from predicted proteomes provided by Prokka, using the Bacterial Antimicrobial Resistance Reference Gene Database (BARRGD) (BioProject number PRJNA313047, accessed December 2020). To confirm the BARRGD results, the ResFinder v4.1 database [32] and the Resistance Gene Identifier (RGI) software [33] of CARD (The Comprehensive Antibiotic Resistance Database) were used. Resistance genes detected were only considered to be present if they had an identity $\geq 80\%$ and a coverage $\geq 60\%$. For virulome prediction, we interrogated the proteomes of thirteen strains against the Virulence Factors of Pathogenic Bacteria Database (VFDB) harbored known genes related to virulence factors [34].

2.6. MLST, Capsule (K) and O-Antigen (OC) Biosynthetic Gene Clusters Analysis

The ST or multilocus sequence typing (MLST) of each genome were determined using MLST 2.0 (Larsen et al., 2012) available on the Center for Genomic Epidemiology website (<http://www.genomicepidemiology.org>) according to the *Acinetobacter baumannii*#2 referred to as the Pasteur scheme (Diancourt et al., 2010). The results were confirmed using PubMLST (<http://pubmlst.org/abaumannii/>; consulted in 15/11/2020) according to the Institute Pasteur scheme. Capsular polysaccharide K locus (KL) and outer nucleus locus (OCL) types were typed using the Kaptive tool [35] and organized databases of reference sequences annotated for KL and OCL of *A. baumannii* [36].

2.6. Variant Calling and Analyses

To characterize the genetic variants of thirteen genomes in this study, five reference genomes of *A. baumannii* commonly used and representing IC-I (A1, GenBank: CP010781), IC-II (ACICU, GenBank: CP031380; MDR-TJ, GenBank: CP003500), and the non-IC lineages (with American Type Culture Collection; ATCC 19606, Genbank: AP022836; ATCC 17978, Genbank:

CP053098) were selected as comparators. The filtered reads of each strain were mapped to the five reference genomes of *A. baumannii* independently, using BWA-MEM v0.7.17-r1188 with -M option) [37]. The mapping quality (BAM files) was assessed by Qualimap v2.2.1 [38]. The BAM files were sorted, and duplicated reads were removed by Picard SortSam and MarkDuplicates v2.25.4, respectively (<http://broadinstitute.github.io/picard/>). Single Nucleotide Polymorphisms (SNPs) and short insertions and deletions (InDels) were called using GATK v3.8-1 [39] with the Unified Genotyper in haploid mode. To achieve high-quality variants, we used the GATK VariantFiltration to perform strict filtering based on quality thresholds, following GATK best practice recommendations. For SNPs, we filtered: QD <2.0 || DP <10 || FS > 60.0 || MQ <40.0 || SOR > 3.0 || MQRankSum <-12.5 || ReadPosRankSum <-8.0. For InDels, we have filtered: QD <2.0 || DP <10 || FS > 200.0 || MQ <40.0 || SOR > 10.0 || MQRankSum <-12.5 || ReadPosRankSum <-20.0. The genomes were annotated using SnpEff v4.3 [40]. Intersections between genetic variants were performed via Bcftools isec v1.10 [41]. The SnpEff databases were built locally using annotations of each of five reference genomes obtained in GBK format from NCBI and used to annotate SNPs and InDels with putative functional effects according to the categories defined in the SnpEff manual (http://snpeff.sourceforge.net/SnpEff_manual.html).

3. Results

3.1. Genomic features of thirteen *A. baumannii* strains

We report the genome sequencing of thirteen Moroccan carbapenem-resistant *A. baumannii*, including a strain resistant to colistin. The statistics for *de novo* assembly and annotation of sequenced genomes are summarized in Table 1. The size of the assembled genomes ranged from 3.91 to 4.06 Mb, with a GC content of approximately 39%, as expected for this species. The N50 value of these assemblies ranged between 44.7 and 131 Kb. The completeness of the assembled genomes reached 97.2% to 98.4% of the 366 conserved genes of gammaproteobacteria (Table S2), with a total number of coding genes ranging from 3668 to 3838. The cumulative length of the total number of genes was over 86% of the total length, and the average gene length ranged from 898 bp to 906 bp. Functional annotation showed that 80% of the entire proteome predicted in each strain was assigned to known functions, of which over 65% carried GO terms (Table 1 and Table S3-S15).

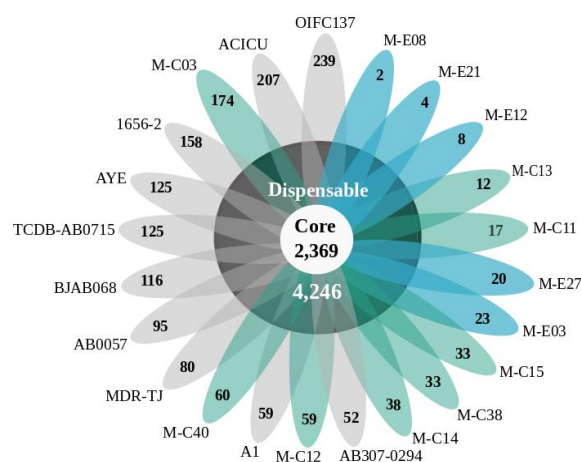
Table 1. Genome assembly and annotation statistics for the thirteen *A. baumannii* strains from this study.

Strains	M-C03	M-C11	M-C12	M-C13	M-C14	M-C15	M-C38	M-C40	M-E03	M-E08	M-E12	M-E21	M-E27
Assembly stats													
Genome size (Mb)	3.91	3.92	3.95	3.96	3.94	4.05	4.03	4.03	4.01	3.93	3.91	3.93	4.06
Number of contigs	193	96	182	100	163	136	217	174	166	106	118	147	158
Number contigs > 1Kb	180	85	155	87	137	82	197	141	150	95	105	132	133
Max contig size (Kb)	116.6	257.6	176.8	238.6	167	267.4	174.7	257.2	142.9	256.2	195.3	178.	257.2
N50 (Kb)	44.7	131.1	57.6	106.6	59.6	131.1	50.1	84.5	54.8	130.7	79.07	60.5	59.3
G + C content (%)	39.1	39.1	38.9	38.9	38.9	39.1	39.1	39	39.1	39	39	39	39
Completeness (%)	97.2	98.4	98	98	98	98.4	98.4	98.4	98.4	98.4	98.4	98.4	98.4
Gene annotation													
Number of genes	3 800	3 774	3 826	3 833	3 805	3 924	3 896	3 911	3 873	3 814	3 796	3 820	3 939
Number of CDS	3 684	3 668	3 714	3 726	3 697	3 818	3 788	3 802	3 766	3 708	3 690	3 714	3 838
Misc_RNA	49	38	36	39	37	39	36	37	37	38	38	38	37
rRNA	3	4	6	4	6	5	7	7	6	4	4	4	7
tRNA	63	63	69	63	64	61	64	64	63	63	63	63	64
Total gene length (Mb)	34.12	34.2	34.49	34.65	34.42	35.35	35	35	35.07	34.5	34.28	34.54	35.43
Mean gene length (bp)	898	906	901	904	905	901	898	895	906	904	903	904	899
Coding region in genome (%)	87.3	87.6	87.3	87.6	87.2	87.3	86.7	86.9	87.4	87.8	87.6	87.9	87.3
Protein functions													
Genes with IPR/Pfam	2 978	3 016	3 022	3 017	3 010	3 052	3 059	3 043	3 053	3 013	2 995	3 014	3 076
Genes with GO	1 948	1 957	1 959	1 945	1 952	1 956	1 993	1 972	1 988	1 943	1 929	1 942	2 004

3.2. Genomic diversity of clinical and environmental *A. baumannii* strains

In order to assess the intra-species genetic variability of clinical and environmental strains of *A. baumannii*, we first constructed a pangenome from thirteen strains from this study and another ten genomes that represent the three ICs and are publicly available (see Materials and Methods section 2.4), using two different approaches: (i) A gene-based approach was used by clustering the predicted protein sets. A total of 79,157 proteins were clustered into an open pangenome of 6,616 non-redundant orthologous proteins (Figure 1, Table S16), of which 2,369 (35.8%) were conserved in all genomes and defined as core genome.

A.



B.

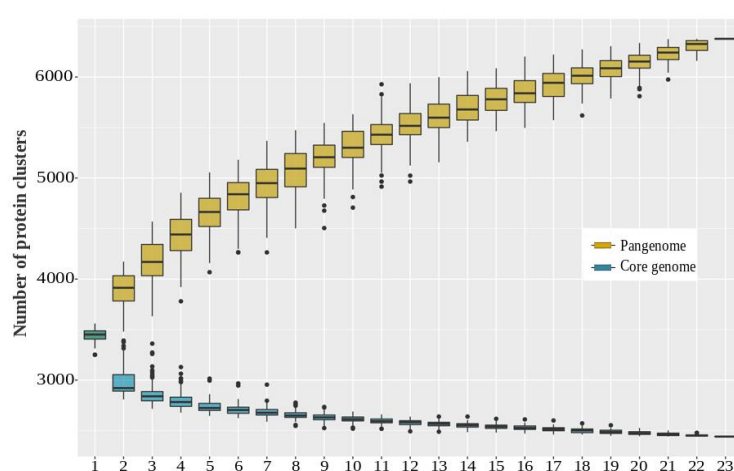


Figure 1. (A) The flower plot displays the number of core proteins (in the center) and accessory proteins (in the annulus), including the singletons (in the petals), encoded by the genome of 23 *A. baumannii* strains. **(B)** The *A. baumannii* pangenome accumulation curve is depicted with dark yellow boxplots, indicating the number of accessory proteins discovered with the sequential addition of new genomes. The blue boxplots illustrate the number of core proteins discovered with the sequential addition of new genomes.

More than 80% (i.e., 1929) of the predicted proteins encoded by the core genome assigned to known functions. The dispensable genome represented 4,246 (64.2%) predicted non-redundant proteins, of which the average number of dispensable proteins per genome was 2508 (59%). While 1739 (41%) were found in a single genome (singletons proteins), highlighting the high degree of diversity in the gene repertoire as expected of this species. The number of singleton proteins in all genomes ranged from 2 for the environmental strain M-E08 to 239 for the public strain OIFC137. Thus, the greatest number of singletons proteins of our strains was found in the clinical strain M-C03 with 174. Except for this strain (M-C03), all other Moroccan strains regardless of clinical and environmental hold the smallest number of proteins singletons, 2 to 60, indicating the strong resemblance between the clinical and environmental strains used in this study. (ii) Genome-based approach from the 23 assembled genomes showed that the final size of the calculated pangenome is 6.75 MB, about 2.77 MB (41%) larger than the size of the known genome of the species *A. baumannii*. This extra size encoded a total of, 2963 protein-coding genes.

3.3. Phylogenomic relationships and diversity of MLST, K and OC loci.

To define how the different strains in this study are related to each other, we first analyzed their STs. The thirteen Moroccan MDR *A. baumannii* strains showed three allelic profiles corresponding to ST2, ST85, and ST31. Eight strains (5 clinical and 3 environmental) belonged to ST2, corresponding to IC-II. Four strains (2 clinical and 2 environmental) belonged to ST85, while the remaining clinical strains belonged to ST315. In combination with MLST, KL and OCL typing is considered to be a promising epidemiological marker (Table S3). Four distinct gene clusters associated with KL synthesis and three with OCL were revealed. It is interesting to note that all the strains of *A. baumannii* ST2 analyzed were assigned as OCL1, while the strains ST85 and ST315 were found to have, respectively, the type OCL6 and OCL5. In contrast, two types of KL; KL3 and KL13 were found in strains belonging to ST2, while strains ST85 and ST315 belonged to KL124 and KL3, respectively.

Phylogenomic relationships were calculated using the orthologous protein sequences. Based on the 2,369 singles copy orthologous proteins (core genome), a close relationship between genomes of the same STs was observed (Figure 2). Of which 13 genomes belonged to ST1 (IC-I), four to ST2 (IC-II), and one genome at ST3 (IC-III).

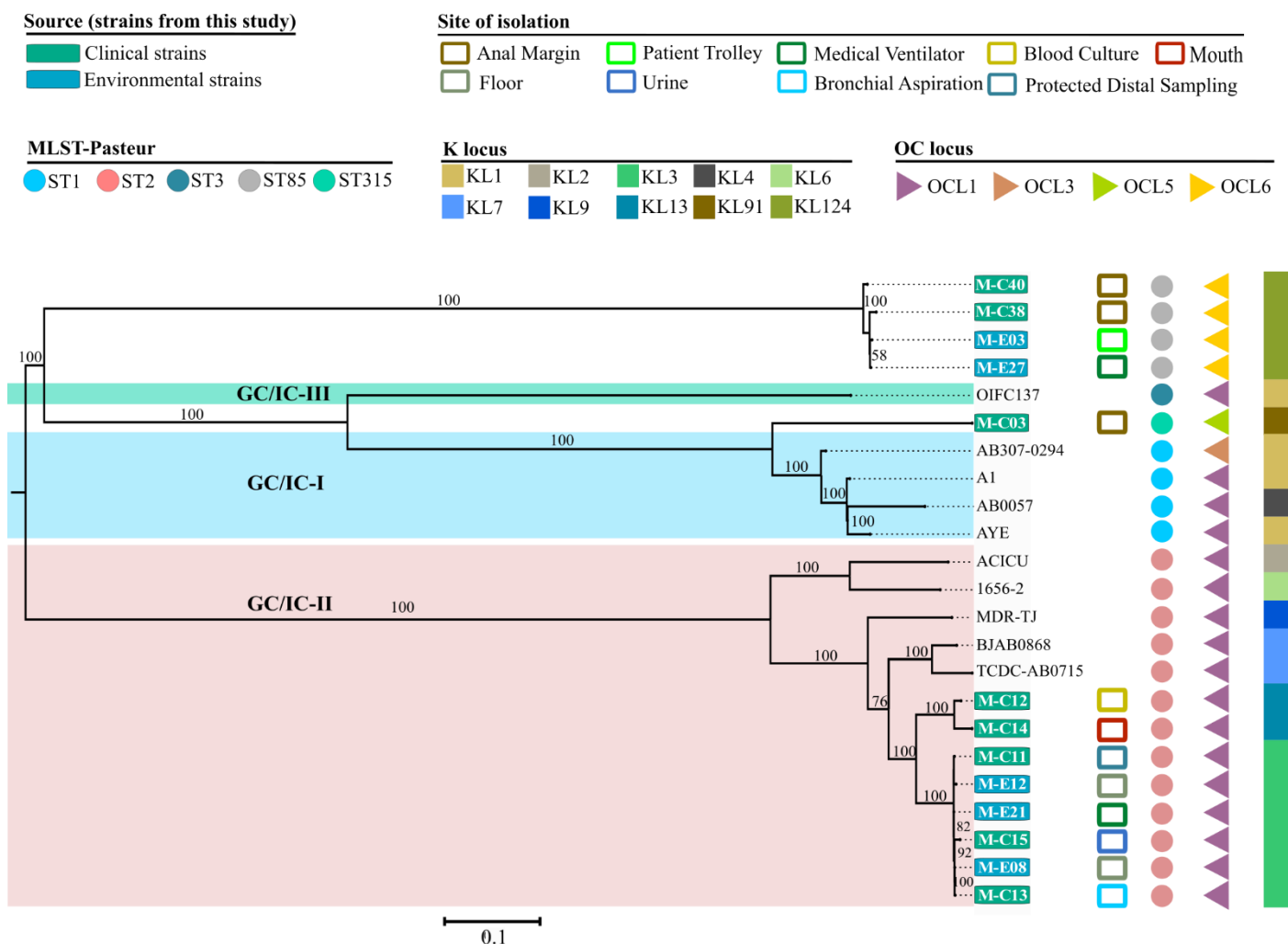


Figure 2. Phylogenomic relationship of 23 strains of *A. baumannii* based on the core genome and their associations to STs, KL and OCL. A maximum likelihood bootstrap was generated from 2396 single copy protein sequences found in the 23 *A. baumannii* strains from different clone lineages. The numbers at the nodes indicate a percentage value made with 1000 replications. The colors of STs, KL, OCL, and different isolation sites are indicated in the legend.

In addition, the four genomes assigned as ST85 from this study were grouped and belonged to the same clade cluster. Likewise, the clinical strain MC-03 assigned as ST315 was distributed in a single clade. We then elucidated the associations between the phylogenetic clades and the types of KL and OCL identified. Except for the genome AB307-0294 (which belonged to ST1), the strains belonging to ST1, ST2, and ST3 were assigned to OCL1. However, the cluster harboring the four strains of *A. baumannii* ST85 was assigned to OCL6. The analysis at the sub-cluster scale reveals an association between the types of KL identified and these sub-clusters. For example, the clone ST2 included a sub-cluster of six closely related strains, including three clinical (M-C11, M-C15, M-C13) and three environments (M-E12, M-E21, M-E08) which have been specifically assigned to the KL3. Likewise, a sub-cluster of the same clone ST2, comprised two clinical strains (M-C12, M-C14) that belonged to KL13.

3.4. Distribution of genetic variants in the three STs and the impact of the choice of the reference genome.

In order to study the genetic variants, in particular the SNPs between the three STs, five reference genomes were used, two of which belonged to IC-II (ACICU and MDR-TJ), one to IC-I (A1) and two lines non-IC (ATCC 19696 and ATCC 17978). The analysis showed a high difference in the rate of SNPs according to the ST of the strain (FIG. 3).

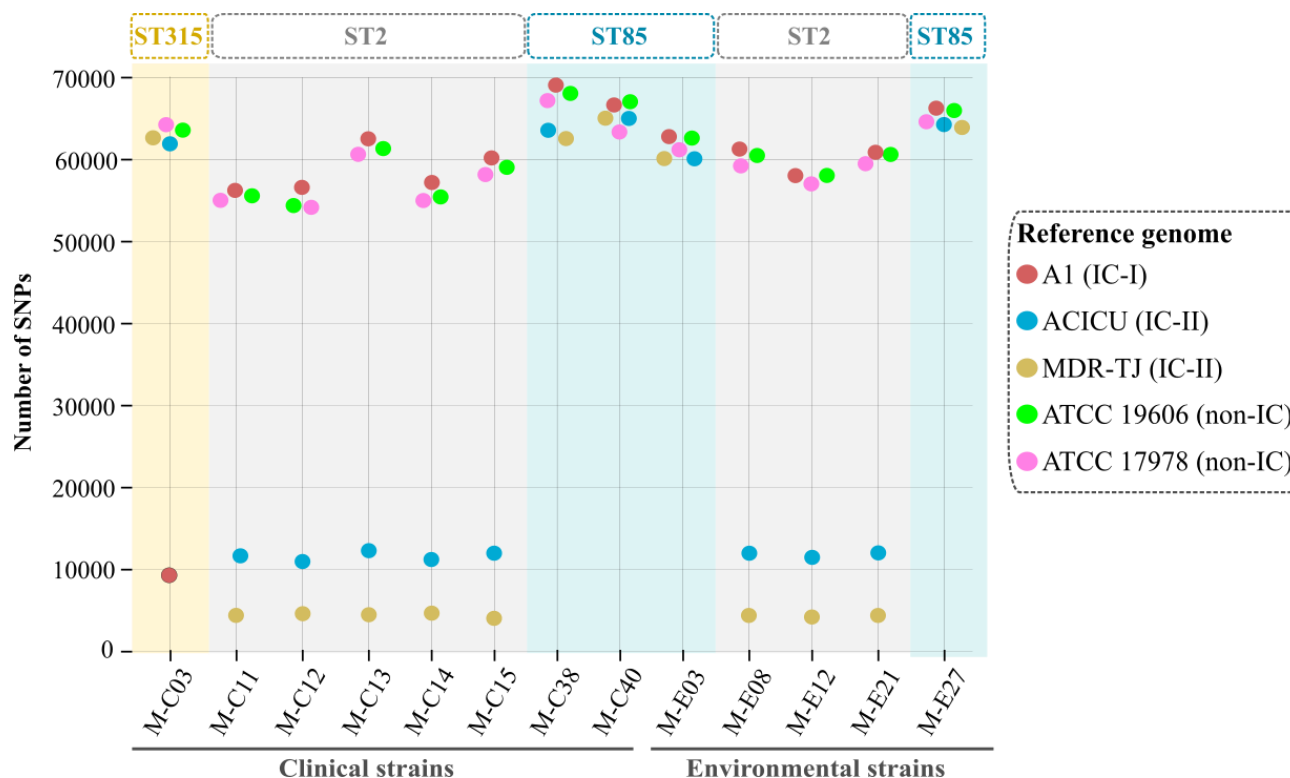


Figure 3. Distribution of the number of SNPs in thirteen strains according to the choice of the reference genome and their relationship to the ST of the strain. The colored dots correspond to each of the five reference genomes of *A. baumannii* used.

The lowest values were obtained in the eight strains of *A. baumannii* ST2 compared to two reference genomes of the same type sequence, MDR-TJ and ACICU, with approximately 5,000 and 12,000 genetic variants, respectively. While the mapping of these two reference genomes (MDR-TJ and ACICU) with respect to strains belonging to ST85 and ST315, revealed a high rate of variants with more than 60,000 SNPs found in each strain. On the other hand, strain ST315 contained a low rate of variant with the reference sequence A1 of ST1, indicating a close similarity between these two STs.

A similar number of variants was observed in the two non-IC references, ATCC 17978 and ATCC 19606. This reflects the approximate similarity in size of the central SNPs of the thirteen genomes with these two reference sequences (ATCC 17978 and ATCC 19606) with 14 734 and 15,333 SNPs, respectively (Figure 4A). In contrast, the intersections of their results showed only 0.67% ($n=3531$) of the shared SNPs (Fig. 4B). Similarly, for the other two references from the same ST, ACICU and MDR-TJ, showed only 0.44% ($n=2330$) of shared SNPs.

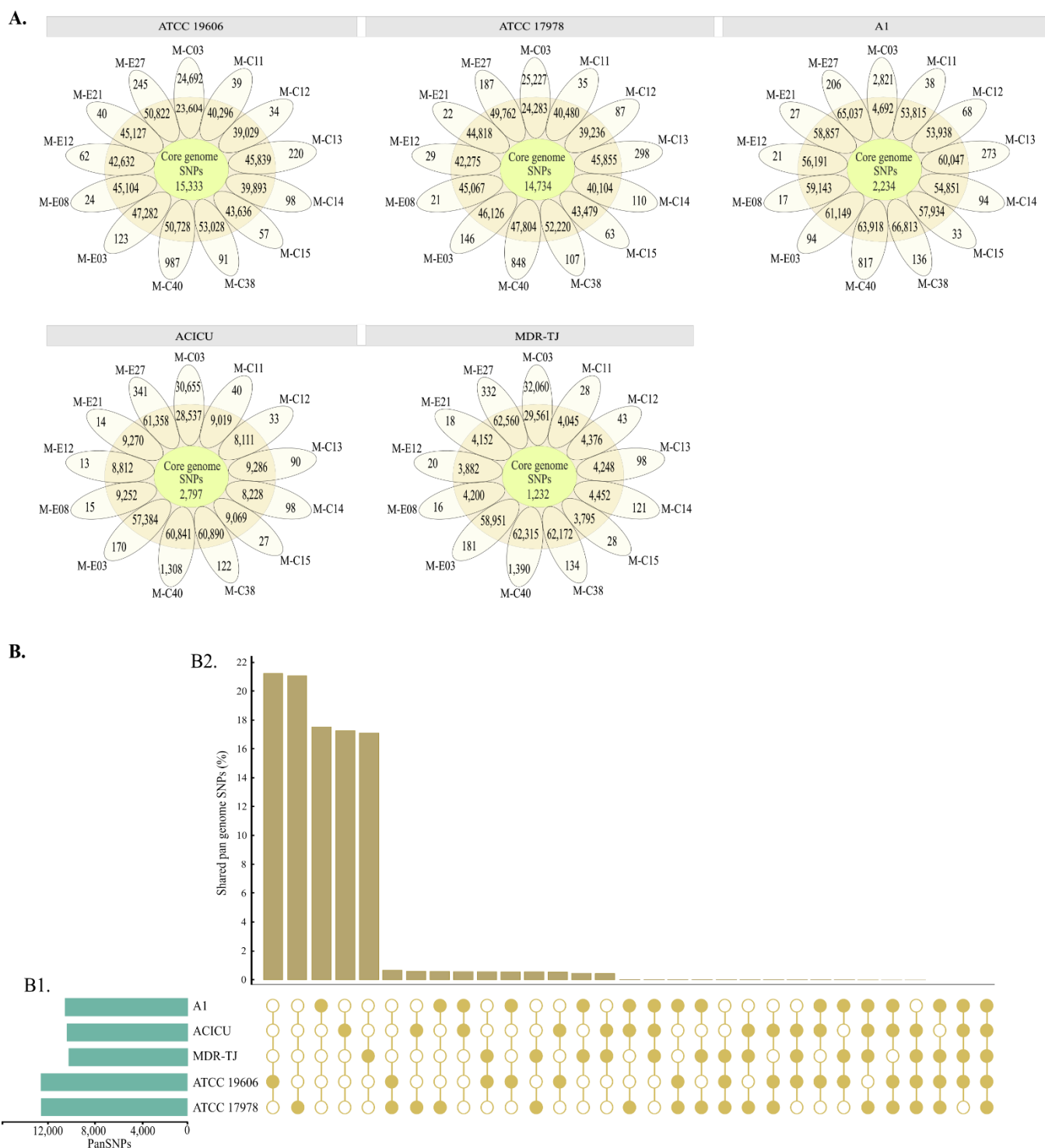


Figure 4. (A) The flower plot displays the number of core genome SNPs (in the center) and SNPs shared between 2 to 12 genomes (in the ring), and the singletons SNPs (in the petals) of 13 genomes of this study compared to each of five sequences of references. **(B1)** The size of the identified pangenome SNPs of the 13 genomes compared to the reference genomes. **(B2)** The rate of shared and singleton SNPs between the five reference genomes.

3.5. Profile Antibiotic Resistance and Related Genes

The thirteen *A. baumannii* strains were tested for their sensitivity to a variety of antibiotics. All strains showing a multi-antibiotic resistant phenotype with a broader resistance profile to different classes of antibiotics, including resistance to the colistin by M-C15 strain. Resistance to a wide range of antibiotics, including beta-lactams (in association to the

carbapenems tested), aminoglycosides, fluoroquinolones, sulfonamides, tetracycline, and rifamycin was detected (Figure 5). This phenotypic resistance is correlated by a diversified content of the identified antibiotic resistance genes and also associated with strain ST. For example, two genes encoding enzymes hydrolyzing carbapenems, NDM-1 and OXA-94, were detected in *A. buamannii* strains ST85, while the strains belonged to ST2 and ST315 harbored OXA-66 and OXA-69, respectively. An additional OXA-like carbapenemase gene (OXA-23) was found in ten strains belonging to ST2, ST315 and ST85. Moreover, among the eleven genes coding for resistance to aminoglycosides, eight were detected in the ST2 strains, including four shared with the ST315, while the remaining genes were found in the ST85 strains. Resistance to fluoroquinolones also appeared in the thirteen strains via mutations in *gyrA* (S81L) and *parC* (S84L; V104I; D105E), except that the strain ST315 had an absence of one of the three mutations of the *parC* gene (S84L). In addition, the two predicted genes *sul1* and *sul2* conferred resistance to sulphonamide, the first of which was detected only in ST2 strains, while the second was detected in ST85 and ST315 strains. Moreover, 12 genes associated with efflux pumps were identified, including Four genes (*adeA*, *adeC*, *adeG*, and *adeH*) were found to be missing exclusively in the ST85 strains, while two additional genes (*adeA* and *adeN*) were found to be absent only in the ST315 strain M-C03. Details on the predicted resistance genes of all isolates are summarized in Figure 5.

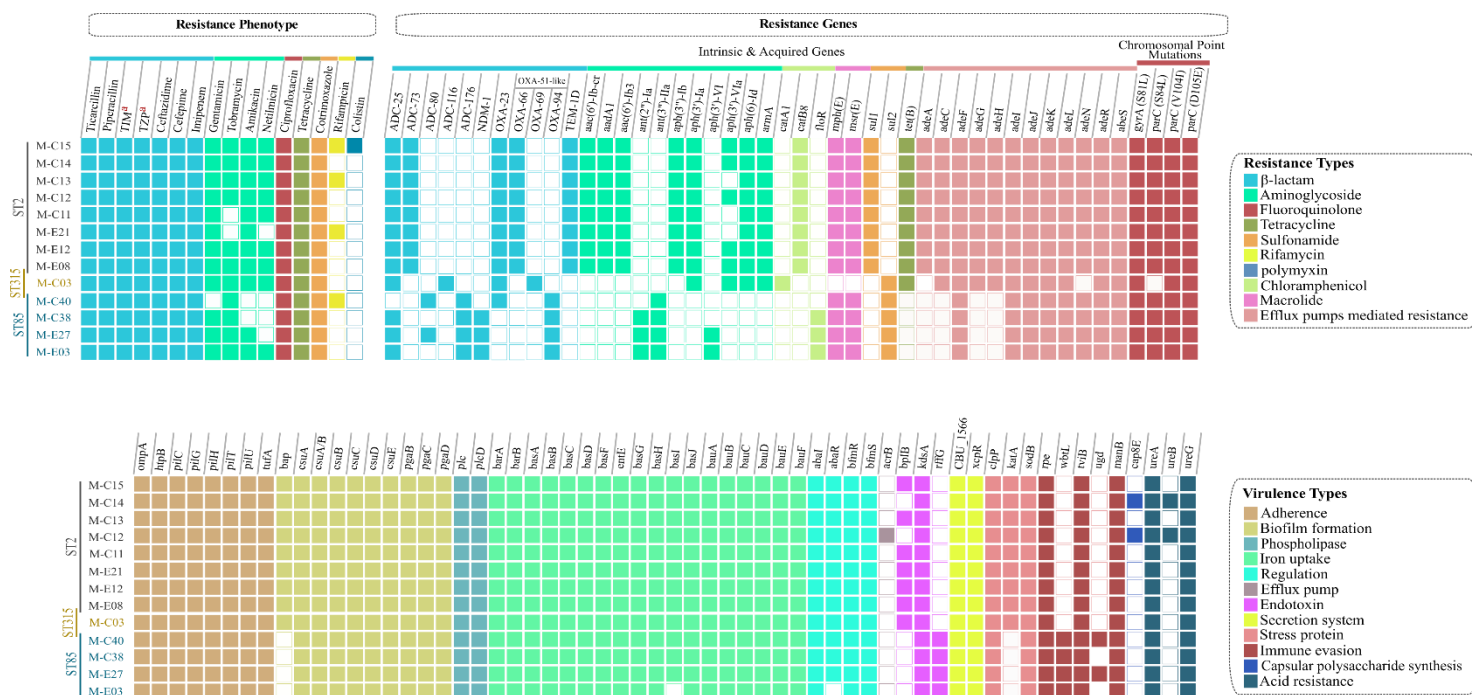


Figure 5. (A) Comparison of phenotypic antibiotic resistance profiles and resistomes. **(B)** Predicted virulence genes. The ST assignment for each of the strains is shown on the left. Color code represents each antibiotic class and type of virulence.

The analysis of the virulome revealed the presence of 63 virulence genes, of which a high proportion (83%) belonged to the core virulome. Minor genetic variability was observed. Several genes related with biofilm formation (*bap*), iron absorption (*basI*) and regulation

(abaR), and stress protein (katA) were found to be lacking from the genomes of *A. baumannii* ST85, whereas other genes were present in the genomes of all strains.

3.6. Mutation in histidine kinase PmrB associated with resistance to colistin in *A. baumannii* M-C15

To investigate the genetic cause of colistin resistance for the M-C15 strain, we targeted the genes for PmrAB, LpxACD, and other genes previously described as being associated with colistin resistance. Compared to the twelve colistin-susceptible strains in this study, only one specific mutation of the M-C15 genome in the pmrB gene was reported (Figure 6). Indeed, genetic analysis of these genes showed four non-synonymous mutations (A138T, R263S, N440H, A444V), including the single mutation from arginine to serine at amino acid 263 (R263S), which corresponds to the cytosine to adenine (C787A) substitution in the kinase of PmrB was only identified in the colistin-resistant strain (Figure 6), as previously suggested, this mutation was involved in colistin resistance.

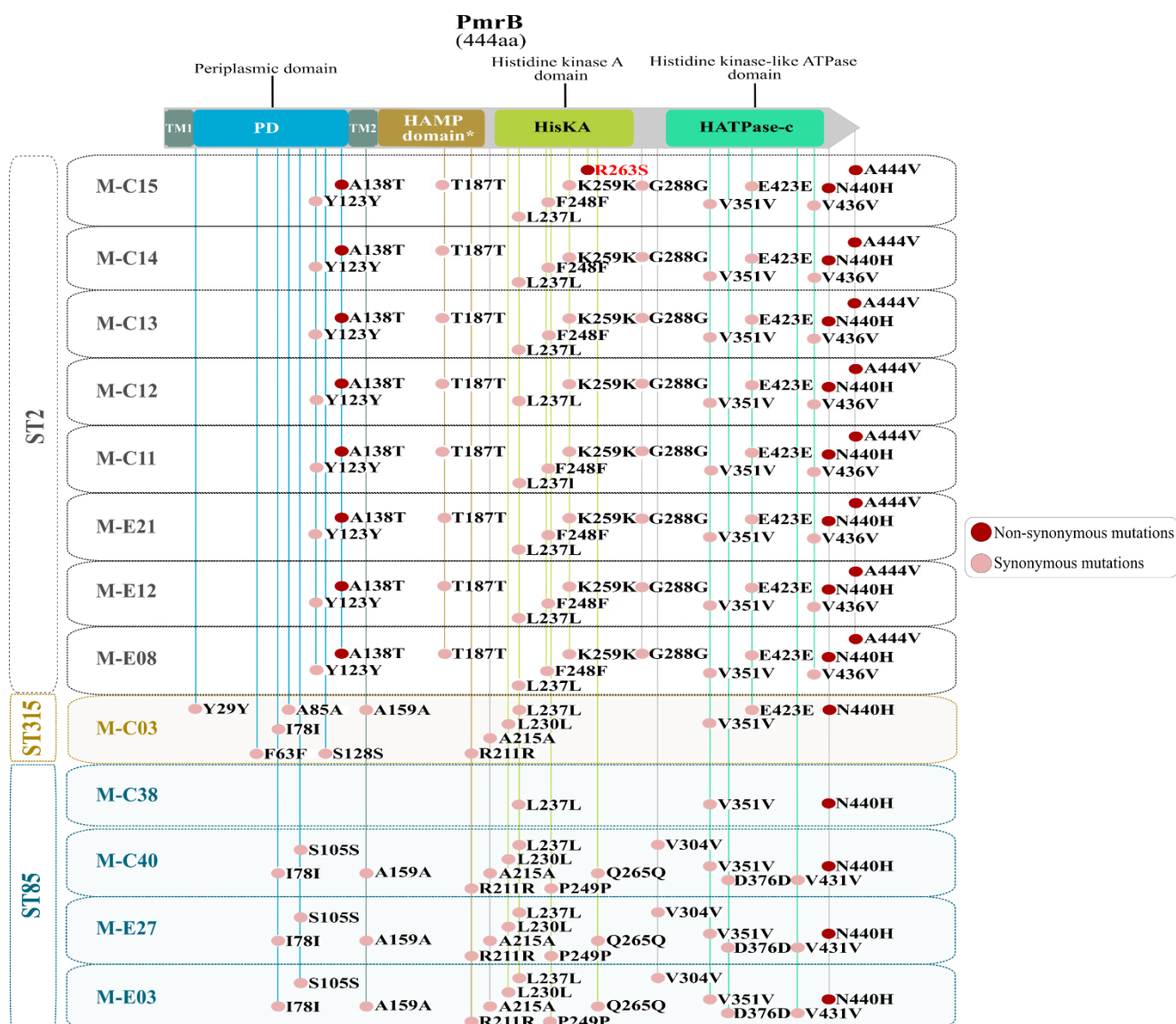


Figure 6. Distribution of amino acid substitutions along pmrB of the colistin resistant strain (M-C15) and twelve other colistin sensitive strains. Amino acid alterations are indicated by different colors depending on the domains of the pmrB gene. the assignment of ST is indicated. The specific mutation of M-C15 and stained red.

4. Discussion

A. baumannii has emerged as an important nosocomial pathogen in recent years, exhibiting the ability to survive on biotic and abiotic surfaces, as well as resistance to almost all classes of antibiotics [14,20,42]. Yakkala et al. [43] revealed the existence of strain-specific genetic content of *A. baumannii* isolated from agricultural soil, favoring its adaptation to this particular environment compared to strains of the same species of various origins. Moreover, the molecular basis behind such strain-specific adaptability to the hospital environment remains largely unknown. This work aims to carry out a comparative genomic analysis based on clinical (isolated from patients) and environmental (isolated from the hospital environment) MDR *A. baumannii* Moroccan strains. We have (i) elucidated the genetic variability between strains from the two sources of isolation and (ii) assessed the contribution of this genetic variability to virulence and antibiotic resistance in relation to their clonal identity and resistance phenotypic profile.

In agreement with previous reports showing an open pangenome of this species [44–45], our results showed an increase in the size of the gene repertoire after the inclusion of new genomes. In addition, to the gene-scale panegnome, this work provided new insights into the size of the pangenome of *A. baumannii* at the assembled genome-scale. We were able to estimate a size of 6.75 Mb, which is about 50% larger than the size of the known reference genomes of this species [46–51]. The comparative analysis demonstrated a high similarity in genetic content between the genomes in this study. Although no significant differences were revealed in specific genes between clinical and environmental strains, a minority of specific genes were found in environmental strains such as luxR, a transcriptional regulator in Gram-negative bacteria [52] and the outer membrane protein (OmpW) [53].

Pan-virulome analysis also revealed strong conservation of genes potentially involved in virulence (more than 80% of the total pan-virulome) between clinical and environmental strains, including the genetic machinery involved in biofilm formation, reflecting the adaptation of *A. baumannii* to the pathogenic lifestyle regardless of their source of isolation. Although the virulome was conserved, some virulence genes were found to be specific to STs. For example, rffG and wbtL associated with endotoxin and immune evasion, respectively, were only found in ST85, indicating some genetic specificity of virulence might be related to the clonal lineage of the strain.

Our results reported the presence of three clonal lineages or STs of *A. baumannii* circulating in the Moroccan hospital. These three STs were found in both clinical and environmental strains. The most frequent was ST2 (or IC-II), considered a predominant international clone in the world [54–57]. Follow-up of ST85, known for its rapid emergence and worldwide diffusion in

strains of different origins: Tunisia, Syria, Turkey, and Spain [58–61]. Finally, ST315 was also previously reported in strains from Germany [62] and Belgium [63]. The KL and OCL loci identified were closely related to the classification of the three reported STs as well as their nucleus-based phylogenomic relationships. This indicates that clinical strains isolated from patients were closely related to those isolated from the hospital environment. Corroborating the result of our previous study [20] based on pulsed-field electrophoresis (PFGE) of 83 clinical and environmental strains. However, our data suggest that there have been local and international transmission events.

The phenotypic profile of resistance to several classes of antibiotics perfectly reflects the profile of resistance genes predicted in the 13 genomes "pan-resistome". This pan-resistome, presented a diversified content according to the clonal identity of the strain, indicating that the ability to exchange certain genetic elements was not uniform between inter-clonal strains and that genomes belonging to the same ST "intra-clonal" were more likely to exchange these elements between them [64,65]. In other words, certain resistance genes clearly associated with a specific ST. For example, three allelic variants of OXA-51-like genes known for their resistance to carbapenems (β -lactam); OXA-66, OXA-69, OXA-94, were found in strains belonged to ST2, ST315 and ST85, respectively [59,64,66]. In addition to carbapenem resistance genes, in 2009, Adams et al. [3] compared the sequences of the PmrAB two-component system in strains of *A. baumannii* sensitive and resistant to the last line of treatment "colistin" found mutations in resistant strains pmrA and pmrB, which led to the hypothesis that these two genes regulate the sensitivity to colistin in *A. baumannii*. Since then, mutations have been reported in pmrA and/or pmrB and implicated in colistin in many strains of *A. baumannii* [67–72]. Interestingly, our results revealed one in mutations from arginine to serine at position 263 of histidine kinase (R263S) of the pmrB gene. This mutation has been reported recently [73] and suggested to being involved in colistin resistance.

5. Conclusions

This work has expanded our knowledge of the intra-species genomic diversity of MDR *A. baumannii*. Comparative genomic analysis showed no obvious difference in dispensable genome as well as virulome between clinical and environmental strains isolated from the same hospital setting. Genetic variability, particularly in the resistome, was mainly associated with the clonal identity of the strains studied regardless of their sources of isolation (environmental or clinical). This work further considered as the first study reporting the presence of a strain resistant to colistin in Morocco caused by a mutation of the pmrB gene. Furthermore, we also demonstrated that the choice of the reference genome of *A. baumannii* could bias the downstream analyses according to the STs of the strains. This underlines that the transition from a reference genome to a "pan-chromosome" reference, bringing together the dispensable and conserved genomic regions of the species, could refine a wide range of genomic analyses.

Author Contributions

Conceptualization, T.A and A.I.; methodology, T.A., J.U and M.L; Formal analysis, T.A and A.I; data curation, T.A.; writing—original draft, T.A; supervision, A.I. and M.E. All authors have read and agreed to the published version of the manuscript.

Acknowledgments:

This work was carried out under national funding from the Moroccan Ministry of Higher Education and Scientific Research (PPR program) to Azeddine Ibrahimi. This work was also supported by a grant from the Institute of Cancer Research of the Lalla Salma Foundation.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lee, C.-R.; Lee, J.H.; Park, M.; Park, K.S.; Bae, I.K.; Kim, Y.B.; Cha, C.-J.; Jeong, B.C.; Lee, S.H. Biology of *Acinetobacter baumannii*: Pathogenesis, Antibiotic Resistance Mechanisms, and Prospective Treatment Options. *Front Cell Infect Microbiol* **2017**, *7*, 55, doi:10.3389/fcimb.2017.00055.
2. MacNair, C.R.; Stokes, J.M.; Carfrae, L.A.; Fiebig-Comyn, A.A.; Coombes, B.K.; Mulvey, M.R.; Brown, E.D. Overcoming Mcr-1 Mediated Colistin Resistance with Colistin in Combination with Other Antibiotics. *Nat Commun* **2018**, *9*, 458, doi:10.1038/s41467-018-02875-z.
3. Adams, M.D.; Nickel, G.C.; Bajaksouzian, S.; Lavender, H.; Murthy, A.R.; Jacobs, M.R.; Bonomo, R.A. Resistance to Colistin in *Acinetobacter baumannii* Associated with Mutations in the PmrAB Two-Component System. *Antimicrob Agents Chemother* **2009**, *53*, 3628–3634, doi:10.1128/AAC.00284-09.
4. Vázquez-López, R.; Solano-Gálvez, S.G.; Juárez Vignon-Whaley, J.J.; Abello Vaamonde, J.A.; Padró Alonzo, L.A.; Rivera Reséndiz, A.; Muleiro Álvarez, M.; Vega López, E.N.; Franyuti-Kelly, G.; Álvarez-Hernández, D.A.; et al. *Acinetobacter baumannii* Resistance: A Real Challenge for Clinicians. *Antibiotics (Basel)* **2020**, *9*, doi:10.3390/antibiotics9040205.
5. World Health Organization. Global Priority List of Antibiotic-Resistant Bacteria to Guide Research, Discovery, and Development of New Antibiotics. **2017**. Available online: <https://apps.who.int/iris/bitstream/handle/10665/311820/WHO-EMP-IAU-2017.12-eng.pdf> (accessed on 14 November 2021).
6. Aurilio, C.; Sansone, P.; Barbarisi, M.; Pota, V.; Giaccari, L.G.; Coppolino, F.; Barbarisi, A.; Passavanti, M.B.; Pace, M.C. Mechanisms of Action of Carbapenem Resistance. *Antibiotics* **2022**, *11*, 421, doi:10.3390/antibiotics11030421.

7. Poirel, L.; Nordmann, P. Carbapenem Resistance in *Acinetobacter baumannii*: Mechanisms and Epidemiology. *Clin Microbiol Infect* **2006**, *12*, 826–836, doi:10.1111/j.1469-0691.2006.01456.x.
8. Kempf, M.; Rolain, J.-M.; Diatta, G.; Azza, S.; Samb, B.; Mediannikov, O.; Sow, A.G.; Diene, S.M.; Fenollar, F.; Raoult, D. Carbapenem Resistance and *Acinetobacter baumannii* in Senegal: The Paradigm of a Common Phenomenon in Natural Reservoirs. *PLOS ONE* **2012**, *7*, e39495, doi:10.1371/journal.pone.0039495.
9. Alcántar-Curiel, M.D.; Rosales-Reyes, R.; Jarillo-Quijada, M.D.; Gayosso-Vázquez, C.; Fernández-Vázquez, J.L.; Toledano-Tableros, J.E.; Giono-Cerezo, S.; Garza-Villafuerte, P.; López-Huerta, A.; Vences-Vences, D.; et al. Carbapenem-Resistant *Acinetobacter baumannii* in Three Tertiary Care Hospitals in Mexico: Virulence Profiles, Innate Immune Response and Clonal Dissemination. *Front. Microbiol.* **2019**, *10*, doi:10.3389/fmicb.2019.02116.
10. Said, D.; Willrich, N.; Ayobami, O.; Noll, I.; Eckmanns, T.; Markwart, R. The Epidemiology of Carbapenem Resistance in *Acinetobacter baumannii* Complex in Germany (2014–2018): An Analysis of Data from the National Antimicrobial Resistance Surveillance System. *Antimicrobial Resistance & Infection Control* **2021**, *10*, 45, doi:10.1186/s13756-021-00909-8.
11. Li, S.; Duan, X.; Peng, Y.; Rui, Y. Molecular Characteristics of Carbapenem-Resistant *Acinetobacter Spp.* from Clinical Infection Samples and Fecal Survey Samples in Southern China. *BMC Infectious Diseases* **2019**, *19*, 900, doi:10.1186/s12879-019-4423-3.
12. Henry, R.; Vithanage, N.; Harrison, P.; Seemann, T.; Coutts, S.; Moffatt, J.H.; Nation, R.L.; Li, J.; Harper, M.; Adler, B.; et al. Colistin-Resistant, Lipopolysaccharide-Deficient *Acinetobacter baumannii* Responds to Lipopolysaccharide Loss through Increased Expression of Genes Involved in the Synthesis and Transport of Lipoproteins, Phospholipids, and Poly- β -1,6-N-Acetylglucosamine. *Antimicrob Agents Chemother* **2012**, *56*, 59–69, doi:10.1128/AAC.05191-11.
13. Qureshi, Z.A.; Hittle, L.E.; O'Hara, J.A.; Rivera, J.I.; Syed, A.; Shields, R.K.; Pasculle, A.W.; Ernst, R.K.; Doi, Y. Colistin-Resistant *Acinetobacter baumannii*: Beyond Carbapenem Resistance. *Clinical Infectious Diseases* **2015**, *60*, 1295–1303, doi:10.1093/cid/civ048.
14. Papathanakos, G.; Andrianopoulos, I.; Papathanasiou, A.; Priavali, E.; Koulenti, D.; Koulouras, V. Colistin-Resistant *Acinetobacter baumannii* Bacteremia: A Serious Threat for Critically Ill Patients. *Microorganisms* **2020**, *8*, 287, doi:10.3390/microorganisms8020287.

15. Sato, Y.; Ubagai, T.; Tansho-Nagakawa, S.; Yoshino, Y.; Ono, Y. Effects of Colistin and Tigecycline on Multidrug-Resistant *Acinetobacter baumannii* Biofilms: Advantages and Disadvantages of Their Combination. *Sci Rep* **2021**, *11*, 11700, doi:10.1038/s41598-021-90732-3.
16. Royer, S.; Faria, A.L.S.; Seki, L.M.; Chagas, T.P.G.; Campos, P.A. de; Batistão, D.W. da F.; Asensi, M.D.; Gontijo Filho, P.P.; Ribas, R.M. Spread of Multidrug-Resistant *Acinetobacter baumannii* and *Pseudomonas Aeruginosa* Clones in Patients with Ventilator-Associated Pneumonia in an Adult Intensive Care Unit at a University Hospital. *Braz J Infect Dis* **2015**, *19*, 350–357, doi:10.1016/j.bjid.2015.03.009.
17. Bonnin, R.A.; Cuzon, G.; Poirel, L.; Nordmann, P. Multidrug-Resistant *Acinetobacter baumannii* Clone, France. *Emerg Infect Dis* **2013**, *19*, 822–823, doi:10.3201/eid1905.121618.
18. Matsui, M.; Suzuki, M.; Suzuki, M.; Yatsuyanagi, J.; Watahiki, M.; Hiraki, Y.; Kawano, F.; Tsutsui, A.; Shibayama, K.; Suzuki, S. Distribution and Molecular Characterization of *Acinetobacter baumannii* International Clone II Lineage in Japan. *Antimicrob Agents Chemother* **2018**, *62*, e02190-17, doi:10.1128/AAC.02190-17.
19. Harding, C.M.; Hennon, S.W.; Feldman, M.F. Uncovering the Mechanisms of *Acinetobacter baumannii* Virulence. *Nat Rev Microbiol* **2018**, *16*, 91–102, doi:10.1038/nrmicro.2017.148.
20. Uwingabiye, J.; Lemnouer, A.; Roca, I.; Alouane, T.; Frikh, M.; Belefquih, B.; Bssaibis, F.; Maleb, A.; Benlahlou, Y.; Kassouati, J.; et al. Clonal Diversity and Detection of Carbapenem Resistance Encoding Genes among Multidrug-Resistant *Acinetobacter baumannii* Isolates Recovered from Patients and Environment in Two Intensive Care Units in a Moroccan Hospital. *Antimicrobial Resistance & Infection Control* **2017**, *6*, 99, doi:10.1186/s13756-017-0262-4.
21. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* **2014**, *30*, 2114–2120, doi:10.1093/bioinformatics/btu170.
22. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Prjibelski, A.D.; et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J Comput Biol* **2012**, *19*, 455–477, doi:10.1089/cmb.2012.0021.
23. Seppely, M.; Manni, M.; Zdobnov, E.M. BUSCO: Assessing Genome Assembly and Annotation Completeness. *Methods Mol Biol* **2019**, *1962*, 227–245, doi:10.1007/978-1-4939-9173-0_14.
24. Seemann, T. Prokka: Rapid Prokaryotic Genome Annotation. *Bioinformatics* **2014**, *30*, 2068–2069, doi:10.1093/bioinformatics/btu153.

25. Jones, P.; Binns, D.; Chang, H.-Y.; Fraser, M.; Li, W.; McAnulla, C.; McWilliam, H.; Maslen, J.; Mitchell, A.; Nuka, G.; et al. InterProScan 5: Genome-Scale Protein Function Classification. *Bioinformatics* **2014**, *30*, 1236–1240, doi:10.1093/bioinformatics/btu031.
26. Tahir Ul Qamar, M.; Zhu, X.; Xing, F.; Chen, L.-L. PpsPCP: A Plant Presence/Absence Variants Scanner and Pan-Genome Construction Pipeline. *Bioinformatics* **2019**, *35*, 4156–4158, doi:10.1093/bioinformatics/btz168.
27. Jordan, I.K.; Rogozin, I.B.; Wolf, Y.I.; Koonin, E.V. Essential Genes Are More Evolutionarily Conserved than Are Nonessential Genes in Bacteria. *Genome Res* **2002**, *12*, 962–968, doi:10.1101/gr.87702.
28. Lechner, M.; Findeiß, S.; Steiner, L.; Marz, M.; Stadler, P.F.; Prohaska, S.J. Proteinortho: Detection of (Co-)Orthologs in Large-Scale Analysis. *BMC Bioinformatics* **2011**, *12*, 124, doi:10.1186/1471-2105-12-124.
29. Katoh, K.; Standley, D.M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol* **2013**, *30*, 772–780, doi:10.1093/molbev/mst010.
30. Castresana, J. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Mol Biol Evol* **2000**, *17*, 540–552, doi:10.1093/oxfordjournals.molbev.a026334.
31. Stamatakis, A. RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313, doi:10.1093/bioinformatics/btu033.
32. Bortolaia, V.; Kaas, R.S.; Ruppe, E.; Roberts, M.C.; Schwarz, S.; Cattoir, V.; Philippon, A.; Allesoe, R.L.; Rebelo, A.R.; Florensa, A.F.; et al. ResFinder 4.0 for Predictions of Phenotypes from Genotypes. *J Antimicrob Chemother* **2020**, *75*, 3491–3500, doi:10.1093/jac/dkaa345.
33. McArthur, A.G.; Waglechner, N.; Nizam, F.; Yan, A.; Azad, M.A.; Baylay, A.J.; Bhullar, K.; Canova, M.J.; De Pascale, G.; Ejim, L.; et al. The Comprehensive Antibiotic Resistance Database. *Antimicrob Agents Chemother* **2013**, *57*, 3348–3357, doi:10.1128/AAC.00419-13.
34. Liu, B.; Zheng, D.; Jin, Q.; Chen, L.; Yang, J. VFDB 2019: A Comparative Pathogenomic Platform with an Interactive Web Interface. *Nucleic Acids Research* **2019**, *47*, D687–D692, doi:10.1093/nar/gky1080.
35. Wick, R.R.; Heinz, E.; Holt, K.E.; Wyres, K.L. Kaptive Web: User-Friendly Capsule and Lipopolysaccharide Serotype Prediction for Klebsiella Genomes. *J Clin Microbiol* **2018**, *56*, doi:10.1128/JCM.00197-18.

36. Wyres, K.L.; Cahill, S.M.; Holt, K.E.; Hall, R.M.; Kenyon, J.J. Identification of *Acinetobacter baumannii* Loci for Capsular Polysaccharide (KL) and Lipooligosaccharide Outer Core (OCL) Synthesis in Genome Assemblies Using Curated Reference Databases Compatible with Kaptive. *Microb Genom* **2020**, *6*, doi:10.1099/mgen.0.000339.
37. Li, H. Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM. *arXiv:1303.3997 [q-bio]* **2013**.
38. Okonechnikov, K.; Conesa, A.; García-Alcalde, F. Qualimap 2: Advanced Multi-Sample Quality Control for High-Throughput Sequencing Data. *Bioinformatics* **2016**, *32*, 292–294, doi:10.1093/bioinformatics/btv566.
39. Van der Auwera, G.A.; Carneiro, M.O.; Hartl, C.; Poplin, R.; Del Angel, G.; Levy-Moonshine, A.; Jordan, T.; Shakir, K.; Roazen, D.; Thibault, J.; et al. From FastQ Data to High Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Curr Protoc Bioinformatics* **2013**, *43*, 11.10.1–11.10.33, doi:10.1002/0471250953.bi1110s43.
40. Cingolani, P.; Platts, A.; Wang, L.L.; Coon, M.; Nguyen, T.; Wang, L.; Land, S.J.; Lu, X.; Ruden, D.M. A Program for Annotating and Predicting the Effects of Single Nucleotide Polymorphisms, SnpEff. *Fly (Austin)* **2012**, *6*, 80–92, doi:10.4161/fly.19695.
41. Danecek, P.; McCarthy, S.A. BCFtools/Csq: Haplotype-Aware Variant Consequences. *Bioinformatics* **2017**, *33*, 2037–2039, doi:10.1093/bioinformatics/btx100.
42. Da Silva, G.J.; Domingues, S. Interplay between Colistin Resistance, Virulence and Fitness in *Acinetobacter baumannii*. *Antibiotics (Basel)* **2017**, *6*, 28, doi:10.3390/antibiotics6040028.
43. Yakkala, H.; Samantarri, D.; Gribskov, M.; Siddavattam, D. Comparative genome analysis reveals niche specific genome expansion in *Acinetobacter baumannii* strains. *PLOS ONE* **2019**, *7*, 000690, doi: 10.1099/mgen.0.000690.
44. Mangas, E.L.; Rubio, A.; Álvarez-Marín, R.; Labrador-Herrera, G.; Pachón, J.; Pachón-Ibáñez, M.E.; Divina, F.; Pérez-Pulido, A.J.Y. 2019 Pangenome of *Acinetobacter baumannii* Uncovers Two Groups of Genomes, One of Them with Genes Involved in CRISPR/Cas Defence Systems Associated with the Absence of Plasmids and Exclusive Genes for Biofilm Formation. *Microbial Genomics* **5**, e000309, doi:10.1099/mgen.0.000309.
45. Hassan, A.; Naz, A.; Obaid, A.; Paracha, R.Z.; Naz, K.; Awan, F.M.; Muhmmad, S.A.; Janjua, H.A.; Ahmad, J.; Ali, A. Pangenome and Immuno-Proteomics Analysis of *Acinetobacter baumannii* Strains Revealed the Core Peptide Vaccine Targets. *BMC Genomics* **2016**, *17*, 732, doi:10.1186/s12864-016-2951-4.

46. Lucaßen, K.; Gerson, S.; Xanthopoulou, K.; Wille, J.; Wille, T.; Seifert, H.; Higgins, P.G. Comparison of the *A. baumannii* Reference Strains ATCC 17978 and ATCC 19606 in Antimicrobial Resistance Mediated by the AdeABC Efflux Pump. *Antimicrob Agents Chemother* **2021**, AAC0057021, doi:10.1128/AAC.00570-21.
47. Tsubouchi, T.; Suzuki, M.; Niki, M.; Oinuma, K.-I.; Niki, M.; Kakeya, H.; Kaneko, Y. Complete Genome Sequence of *Acinetobacter baumannii* ATCC 19606T, a Model Strain of Pathogenic Bacteria Causing Nosocomial Infection. *Microbiol Resour Announc* **2020**, 9, doi:10.1128/MRA.00289-20.
48. Hamidian, M.; Hawkey, J.; Wick, R.; Holt, K.E.; Hall, R.M. Evolution of a Clade of *Acinetobacter baumannii* Global Clone 1, Lineage 1 via Acquisition of Carbapenem- and Aminoglycoside-Resistance Genes and Dispersion of ISAbal. *Microb Genom* **2019**, 5, e000242, doi:10.1099/mgen.0.000242.
49. Holt, K.E.; Hamidian, M.; Kenyon, J.J.; Wynn, M.T.; Hawkey, J.; Pickard, D.; Hall, R.M. Genome Sequence of *Acinetobacter baumannii* Strain A1, an Early Example of Antibiotic-Resistant Global Clone 1. *Genome Announc* **2015**, 3, doi:10.1128/genomeA.00032-15.
50. Iacono, M.; Villa, L.; Fortini, D.; Bordoni, R.; Imperi, F.; Bonnal, R.J.P.; Sicheritz-Ponten, T.; De Bellis, G.; Visca, P.; Cassone, A.; et al. Whole-Genome Pyrosequencing of an Epidemic Multidrug-Resistant *Acinetobacter baumannii* Strain Belonging to the European Clone II Group. *Antimicrob Agents Chemother* **2008**, 52, 2616–2625, doi:10.1128/AAC.01643-07.
51. Alouane, T.; Uwingabiye, J.; Lemnouer, A.; Lahlou, L.; Laamarti, M.; Kartti, S.; Benhrif, O.; El Misbahi, H.; Frikh, M.; Benlahlou, Y.; et al. First Whole-Genome Sequences of Two Multidrug-Resistant *Acinetobacter baumannii* Strains Isolated from a Moroccan Hospital Floor. *Genome Announc* **2017**, 5, e00298-17, doi:10.1128/genomeA.00298-17.
52. Hudaiberdiev, S.; Choudhary, K.S.; Vera Alvarez, R.; Gelencsér, Z.; Ligeti, B.; Lamba, D.; Pongor, S. Census of Solo LuxR Genes in Prokaryotic Genomes. *Front Cell Infect Microbiol* **2015**, 5, 20, doi:10.3389/fcimb.2015.00020.
53. Catel-Ferreira, M.; Marti, S.; Guillon, L.; Jara, L.; Coadou, G.; Molle, V.; Bouffartigues, E.; Bou, G.; Shalk, I.; Jouenne, T.; et al. The Outer Membrane Porin OmpW of *Acinetobacter baumannii* Is Involved in Iron Uptake and Colistin Binding. *FEBS Lett* **2016**, 590, 224–231, doi:10.1002/1873-3468.12050.
54. Principe, L.; Piazza, A.; Giani, T.; Bracco, S.; Caltagirone, M.S.; Arena, F.; Nucleo, E.; Tammaro, F.; Rossolini, G.M.; Pagani, L.; et al. Epidemic Diffusion of OXA-23-Producing *Acinetobacter baumannii* Isolates in Italy: Results of the First Cross-Sectional Countrywide Survey. *J Clin Microbiol* **2014**, 52, 3004–3010, doi:10.1128/JCM.00291-14.

55. Jeannot, K.; Diancourt, L.; Vaux, S.; Thouverez, M.; Ribeiro, A.; Coignard, B.; Courvalin, P.; Brisse, S. Molecular Epidemiology of Carbapenem Non-Susceptible *Acinetobacter baumannii* in France. *PLOS ONE* **2014**, *9*, e115452, doi:10.1371/journal.pone.0115452.
56. Nowak, J.; Zander, E.; Stefanik, D.; Higgins, P.G.; Roca, I.; Vila, J.; McConnell, M.J.; Cisneros, J.M.; Seifert, H. High Incidence of Pandrug-Resistant *Acinetobacter baumannii* Isolates Collected from Patients with Ventilator-Associated Pneumonia in Greece, Italy and Spain as Part of the MagicBullet Clinical Trial. *J Antimicrob Chemother* **2017**, *72*, 3277–3282, doi:10.1093/jac/dkx322.
57. Pournaras, S.; Dafopoulou, K.; Del Franco, M.; Zarkotou, O.; Dimitroulia, E.; Protonotariou, E.; Poulou, A.; Zarrilli, R.; Tsakris, A.; Greek Study Group on Acinetobacter Antimicrobial Resistance Predominance of International Clone 2 OXA-23-Producing-*Acinetobacter baumannii* Clinical Isolates in Greece, 2015: Results of a Nationwide Study. *Int J Antimicrob Agents* **2017**, *49*, 749–753, doi:10.1016/j.ijantimicag.2017.01.028.
58. Jaidane, N.; Naas, T.; Oueslati, S.; Bernabeu, S.; Boujaafar, N.; Bouallegue, O.; Bonnin, R.A. Whole-Genome Sequencing of NDM-1-Producing ST85 *Acinetobacter baumannii* Isolates from Tunisia. *Int J Antimicrob Agents* **2018**, *52*, 916–921, doi:10.1016/j.ijantimicag.2018.05.017.
59. Salloum, T.; Tannous, E.; Alousi, S.; Arabaghian, H.; Rafei, R.; Hamze, M.; Tokajian, S. Genomic Mapping of ST85 BlaNDM-1 and BlaOXA-94 Producing *Acinetobacter baumannii* Isolates from Syrian Civil War Victims. *Int J Infect Dis* **2018**, *74*, 100–108, doi:10.1016/j.ijid.2018.07.017.
60. Heydari, F.; Mammina, C.; Koksai, F. 2015 NDM-1-Producing *Acinetobacter baumannii* ST85 Now in Turkey, Including One Isolate from a Syrian Refugee. *Journal of Medical Microbiology* **64**, 1027–1029, doi:10.1099/jmm.0.000132.
61. Montaña, S.; Vilacoba, E.; Fernandez, J.S.; Traglia, G.M.; Sucari, A.; Pennini, M.; Iriarte, A.; Centron, D.; Melano, R.G.; Ramírez, M.S. Genomic Analysis of Two *Acinetobacter baumannii* Strains Belonging to Two Different Sequence Types (ST172 and ST25). *Journal of Global Antimicrobial Resistance* **2020**, *23*, 154–161, doi:10.1016/j.jgar.2020.09.006.
62. Eigenbrod, T.; Reuter, S.; Gross, A.; Kocer, K.; Günther, F.; Zimmermann, S.; Heeg, K.; Mutters, N.T.; Nurjadi, D. Molecular Characterization of Carbapenem-Resistant *Acinetobacter baumannii* Using WGS Revealed Missed Transmission Events in Germany from 2012–15. *J Antimicrob Chemother* **2019**, *74*, 3473–3480, doi:10.1093/jac/dkz360.
63. De Vos, D.; Pirnay, J.-P.; Bilocq, F.; Jennes, S.; Verbeken, G.; Rose, T.; Keersebilck, E.; Bosmans, P.; Pieters, T.; Hing, M.; et al. Molecular Epidemiology and Clinical Impact of *Acinetobacter calcoaceticus-baumannii* Complex in a Belgian Burn Wound Center. *PLoS One* **2016**, *11*, e0156237, doi:10.1371/journal.pone.0156237.

64. Leal, N.C.; Campos, T.L.; Rezende, A.M.; Docena, C.; Mendes-Marques, C.L.; de Sá Cavalcanti, F.L.; Wallau, G.L.; Rocha, I.V.; Cavalcanti, C.L.B.; Veras, D.L.; et al. Comparative Genomics of *Acinetobacter baumannii* Clinical Strains From Brazil Reveals Polyclonal Dissemination and Selective Exchange of Mobile Genetic Elements Associated With Resistance Genes. *Front Microbiol* **2020**, *11*, doi:10.3389/fmicb.2020.01176.
65. Graña-Miraglia, L.; Evans, B.A.; López-Jácome, L.E.; Hernández-Durán, M.; Colín-Castro, C.A.; Volkow-Fernández, P.; Cevallos, M.A.; Franco-Cendejas, R.; Castillo-Ramírez, S. Origin of OXA-23 Variant OXA-239 from a Recently Emerged Lineage of *Acinetobacter baumannii* International Clone V. *mSphere* **2020**, *5*, e00801-19, doi:10.1128/mSphere.00801-19.
66. Karah, N.; Sundsfjord, A.; Towner, K.; Samuelsen, Ø. Insights into the Global Molecular Epidemiology of Carbapenem Non-Susceptible Clones of *Acinetobacter baumannii*. *Drug Resist Updat* **2012**, *15*, 237–247, doi:10.1016/j.drug.2012.06.001.
67. Ghahraman, M.R.K.; Hosseini-Nave, H.; Azizi, O.; Shakibaie, M.R.; Mollaie, H.R.; Shakibaie, S. Molecular Characterization of LpxACD and PmrA/B Two-Component Regulatory System in the Colistin Resistance *Acinetobacter baumannii* Clinical Isolates. *Gene Reports* **2020**, *21*, 100952, doi:10.1016/j.genrep.2020.100952.
68. Sun, B.; Liu, H.; Jiang, Y.; Shao, L.; Yang, S.; Chen, D. New Mutations Involved in Colistin Resistance in *Acinetobacter baumannii*. *mSphere* **2020**, *5*, doi:10.1128/mSphere.00895-19.
69. Gerson, S.; Betts, J.W.; Lucaßen, K.; Nodari, C.S.; Wille, J.; Josten, M.; Göttig, S.; Nowak, J.; Stefanik, D.; Roca, I.; et al. Investigation of Novel PmrB and EptA Mutations in Isogenic *Acinetobacter baumannii* Isolates Associated with Colistin Resistance and Increased Virulence In Vivo. *Antimicrob Agents Chemother* **2019**, *63*, doi:10.1128/AAC.01586-18.
70. Mustapha, M.M.; Li, B.; Pacey, M.P.; Mettus, R.T.; McElheny, C.L.; Marshall, C.W.; Ernst, R.K.; Cooper, V.S.; Doi, Y. Phylogenomics of Colistin-Susceptible and Resistant XDR *Acinetobacter baumannii*. *Journal of Antimicrobial Chemotherapy* **2018**, *73*, 2952–2959, doi:10.1093/jac/dky290.
71. Dahdouh, E.; Gómez-Gil, R.; Sanz, S.; González-Zorn, B.; Daoud, Z.; Mingorance, J.; Suárez, M. A Novel Mutation in PmrB Mediates Colistin Resistance during Therapy of *Acinetobacter baumannii*. *Int J Antimicrob Agents* **2017**, *49*, 727–733, doi:10.1016/j.ijantimicag.2017.01.031.
72. Hong, Y.K.; Lee, J.Y.; Wi, Y.M.; Ko, K.S. High rate of colistin dependence in *Acinetobacter baumannii*. *J. Antimicrob. Chemother.* **2016**, *71*, 2346–2348

73. Jovicic, B.; Novovic, K.; Dekic, S.; Hrenovic, J. Colistin Resistance in Environmental Isolates of *Acinetobacter baumannii*. *Microb Drug Resist* **2021**, *27*, 328–336, doi:10.1089/mdr.2020.0188.

Conclusion

L'objectif principal de ce travail est d'étudier la diversité génétique des souches marocaines multirésistantes d'*A. baumannii* isolées de patients et du milieu hospitalier, avec un accent particulier sur le répertoire des gènes de résistance et leurs relations clonales. Il s'agit de la première étude incluant des génomes complets de souches marocaines de ce pathogène et rapportant au moins trois lignées clonales circulantes dans ce milieu hospitalier, dont une lignée appartenait à l'un des trois clones internationaux « IC- II ». Malgré la grande diversité génétique de cette espèce et sa nature de pangénome ouvert, aucune différence significative n'a été trouvée dans le contenu total du génome dispensable, du pan-résistome et du pan-virulome spécifique aux souches environnementales. Cela a été également confirmé via l'analyse phylogénomique basée sur le génome montrant une relation étroitement apparentée entre les deux les souches de deux origines. Cependant, le réservoir environnemental pourrait être une source importante d'épidémies d'*A. baumannii* multirésistant résultant exclusivement par la contamination et la transmission en milieu hospitalier. Ce travail est également considéré comme la première étude rapporte la présence d'une souche résistante à la dernière ligne de traitement (colistine) d'une souche marocaine via la présence d'une mutation dans le domaine histidine kinase du gène *pmrB*.

Discussion générale et perspectives

Discussion générale et perspectives

Les travaux de cette Thèse ont porté sur l'étude (i) de la diversité structurale et fonctionnelle du génome de deux microorganismes pathogènes (*F. graminearum* et *A. baumannii*), ainsi que sur (ii) la contribution de cette diversité génomique à la pathogénicité des souches, avec un accent particulier sur les gènes de résistance de la bactérie opportuniste *A. baumannii*, et les protéines sécrétées du champignon phytopathogène *F. graminearum*. L'utilisation des technologies NGS couplée à des analyses bioinformatiques a permis de reconstruire les génomes de huit souches européennes de *F. graminearum* sélectionnées pour leur agressivité contrastée. Dans cette analyse de génomique comparative, j'ai également utilisé les informations issues de 12 génomes supplémentaires de la même espèce disponibles dans des bases de données accessibles au public. Pour *A. baumannii*, 13 nouvelles souches marocaines multirésistantes ont été utilisées et réparties en 8 souches environnementales (isolées de l'environnement hospitalier) et 5 souches cliniques (isolées de patients), dont une souche résistante à la colistine. Nous avons complété ce panel de souches d'*A. baumannii* avec 10 génomes accessibles au public, sélectionnés pour leur appartenance aux trois clones internationaux.

I – Construction du premier pangénome de *Fusarium graminearum*

En dépit de son importance sanitaire et économique, aucune étude n'avait été menée pour caractériser le pangénome de *F. graminearum*. En effet jusqu'à présent, seul le génome de la souche américaine "PH-1", composé de quatre chromosomes d'une longueur de 38 Mb codant 14,145 gènes au total, faisait référence pour l'espèce (King *et al.*, 2015; King *et al.*, 2017a). Dans le cadre de cette Thèse, nous avons défini le premier pangénome, et élargit le répertoire de gènes connus dans le génome de référence de 47 % (Alouane *et al.*, 2021). Toutefois, en dépit de l'utilisation de 20 souches de *F. graminearum*, ce pangénome reste de type ouvert, suggérant qu'un nombre peut être élevé de gènes dispensables restent encore à être identifiés via l'assemblage des génomes de souches d'origine variées. Le nombre élevé de gènes accessoires (près de 12000) questionne sur leur rôle dans la biologie du champignon. Ces gènes dispensables pourraient refléter la capacité de *F. graminearum* de se développer dans des environnements très différents et/ou de s'adapter à une large gamme d'hôtes potentiels. Cette observation est étayée par le fait qu'environ 90% des gènes faisant l'objet d'une sélection diversifiante sont surreprésentés dans les compartiments du génome à évolution rapide, comme les

régions proximales des télomères et les régions les plus centrales des quatre chromosomes de *F. graminearum*, connues pour disposer d'une fréquence élevée de recombinaison (Cuomo *et al.*, 2007; Laurent *et al.*, 2017). Cette localisation préférentielle a été formulée dans le modèle génomique à deux vitesses d'évolution observé dans *F. graminearum* comme de nombreux autres phytopathogènes filamenteux (Dong *et al.*, 2015; Q. Wang *et al.*, 2017).

L'expansion du pangénome au sein d'une espèce est souvent attribuable à des événements de transfert horizontal de gènes, de duplications d'un segment d'ADN chromosomique (duplication segmentaire) ou de gènes (Brockhurst *et al.*, 2019; Tettelin *et Medini*, 2020). Ce travail a révélé très peu de séquences paralogues ($\approx 1\%$ du pangénome total) correspondant aux protéines paralogues, corroborant le résultat du système de mutation fongique induit par répétition "RIP" qui opère dans *F. graminearum* (Cuomo *et al.*, 2007 ; Pomraning *et al.*, 2013), connu pour protéger le génome fongique contre l'activité des éléments transposables. Malgré cela, des événements de transfert horizontal ont été suggérés entre *F. graminearum* et d'autres espèces de *Fusarium* et en dehors du phylum *Fusarium* (Sieber *et al.*, 2014). Deux hypothèses non exclusives peuvent être formulées pour expliquer l'existence de ces gènes dispensables : la perte progressive de gènes à partir d'un génome ancestral et/ou le gain de gènes par transfert horizontal.

Outre l'échange génétique, les variants ponctuels dans les gènes préexistants pourraient également être une source majeure de cette expansion du pangénome. Nos travaux ont révélé que plus de 40% des variants intra-géniques étaient associés à un effet non synonyme, dont la perte du codon stop et du décalage de cadre de lecture modifiant les séquences protéiques. Plus de 90% des gènes porteurs de mutations de type LOF (potentiellement impliquées dans la perte de fonction des protéines) appartenaient au génome dispensable, suggérant que le génome dispensable ne relèverait pas seulement d'un gain successif de gènes mais pourraient aussi être expliqué par l'élimination sélective de certains gènes. Outre ces variants ponctuels, les génomes peuvent différer par les variants structuraux (SV), où les différences peuvent être beaucoup plus importantes, conduisant ainsi à l'expansion du pangénome. Les SVs sont considérés comme un moteur important d'évolution et d'adaptation à de nouveaux environnements (Gorkovskiy *et Verstrep*, 2021). Il serait donc intéressant de compléter le catalogue des variants génétiques identifiés dans le cadre de ce travail par l'identification et la caractérisation des SVs présents dans les génomes de *F. graminearum*. Enfin, le mode de vie de cette espèce pourrait jouer un rôle majeur dans cette diversité, comme déjà rapporté chez certaines

espèces sympatriques qui peuvent avoir un pangénome plus important que des espèces isolées ou très spécialisées (Rouli *et al.*, 2015; Snipen *et al.*, 2009).

La nature variable du pangénome observée dans *F. graminearum*, indique qu'un génome de référence issu d'une seule souche est insuffisante pour représenter toute la diversité génomique de cette espèce. Bien que le génome de référence soit censé être un standard, ce que cela signifie en pratique n'est pas clair. Le contenu génomique de référence PH-1 n'est pas une moyenne de l'ensemble de la population de cette espèce, mais contient plutôt des séquences spécifiques d'une souche américaine. L'utilisation du pangénome d'espèce peut servir de génome de référence universel pour une meilleure exploration des variations génomiques sous-jacentes et devrait améliorer les analyses des données d'expression et l'efficacité de l'identification de variants génétiques comme cela a déjà été rapporté pour certaines espèces (Burgess, 2021; Gage *et al.*, 2019; Sherman *et al.*, 2019; Valiente-Mullor *et al.*, 2021).

L'analyse phylogénomique des souches utilisées dans cette analyse, basée sur le core génome, montre clairement un apparentement de la plupart des souches européennes par rapport aux souches d'autres origines. Ceci suggère que notre étude, si elle a permis de construire le premier pangénome de l'espèce *F. graminearum*, n'a probablement pas permis d'en capter toute l'importance. Fort des résultats et des connaissances acquises, de nouvelles perspectives se dessinent. Il serait d'ailleurs pertinent de sélectionner d'autres échantillons susceptibles de capturer une plus grande part de la diversité de cette espèce afin de définir un pangénome plus complet. Ce panel de souches devrait inclure non seulement des souches d'origines différentes, mais aussi de souches collectées sur un large éventail d'hôtes. Outre la sélection d'échantillons, l'adoption du pangénome de l'espèce comme un pangénome de référence universelle reflète non seulement l'inclusion de nouvelles séquences génomiques, mais également l'introduction d'une nouvelle structure et d'un nouveau format de données. En d'autres termes, par rapport à un génome de référence linéaire, les emplacements (ordre et orientation) de ces nouvelles séquences sont difficiles à intégrer lors de la création d'un système de coordonnées de référence de pangénome. Nous avons proposé un pangénome de référence basée sur les génomes assemblés à la fin de la période doctorale via l'approche « *map-to-pan* » (Aggarwal *et al.*, 2022) en combinant le génome référence PH-1 avec des séquences représentatives non-références nouvellement identifiées, mais les emplacements ou les coordonnées génomiques de ces séquences restent inconnues. Cela doit être poursuivi avec d'autres approches telles « *graph-based pangenome* » (H. Li *et al.*, 2020), tout en préservant les coordonnées de référence linéaires.

II – Réévaluation à la hausse du pangénome d'*Acinetobacter baumannii* :

Il est bien établi que l'infection et la colonisation par des souches d'*A. baumannii* posent des problèmes sanitaires majeurs en milieu nosocomial en raison de sa capacité d'adaptation aux environnements biotiques et abiotiques. Cette capacité d'adaptation suggère qu'*A. baumannii* dispose d'un répertoire de gènes lui permettant de s'adapter rapidement à des environnements très différents (Yakkala *et al.*, 2019). La quasi-totalité des études génomiques précédentes se sont principalement concentrées sur les souches cliniques (isolées de patients). Dans ces travaux, une analyse génomique comparative intra-espèce a été réalisée via la construction du pangénome en combinant les génomes de souches marocaines, dont 5 souches isolées en milieu hospitalier, 8 souches cliniques isolées de patients, et 10 autres génomes appartenant aux trois clones internationaux (IC-I, IC-II et IC-III), couramment associées aux épidémies hospitalières (Hamidian *et al.*, 2019; Martins *et al.*, 2016; Matsui *et al.*, 2018; Skerniškytė *et al.*, 2019).

En accord avec les rapports précédents (Chan *et al.*, 2015; Hassan *et al.*, 2016; Urhan *et al.*, s. d.), nos résultats ont révélé un pangénome ouvert (près de 6400 gènes) élargissant le répertoire des gènes connus pour cette espèce d'environ 68 %. Malgré cette nature ouverte du pangénome d'*A. baumannii*, une faible proportion de gènes (53 gènes) s'est avérée spécifique aux génomes des souches environnementales par rapport aux souches cliniques. Parmi eux, seuls 13 gènes assignés à des fonctions connues, comme la protéine de membrane externe W « OmpW », une porine jouant un rôle central dans l'absorption de substances nutritionnelles telles que le fer et contribue à l'adaptation au stress chez plusieurs espèces (Ye *et al.*, 2018; Zhang *et al.*, 2019). Leurs liens avec l'adaptation et la survie d'*A. baumannii* en milieu hospitalier nécessitent des analyses fonctionnelles complémentaires pour être validées. Ainsi, la constitution génétique nécessaire à la formation du biofilm, une caractéristique essentielle à la colonisation de l'hôte, a été observé dans tous les génomes, et donc aussi bien dans les souches isolées des patients que dans le milieu hospitalier. En revanche, une expansion sélective du contenu génomique unique et spécifique des génomes appartenant à la même lignée clonale a été observée indépendamment à leurs origines. Ceci a été confirmé par une analyse phylogénomique basée sur le core génome, montrant une relation étroite entre les souches environnementales et les souches cliniques présentant la même identité clonale, suggérant une transmission probable due à une contamination

provenant de l'environnement hospitalier, avec une propagation beaucoup plus importante de ST2 par rapport aux deux autres clones. Cette ST est considérée également comme un clone majeur d'*A.baumannii* en raison de sa distribution dans plusieurs pays sur tous les continents (Karah *et al.*, 2012). ST85 fait partie également d'une émergence rapide et s'étend à de nombreuses régions mondiales comme l'Espagne, la Turquie et la Syrie (Heydari *et al.*, 2015; Jaidane *et al.*, 2018; Mann *et al.*, 2022; Salloum *et al.*, 2018), tandis que ST315 a été signalé dans des souches d'Allemagne et avait des antécédents d'importation de Tunisie (Eigenbrod *et al.*, 2019). Cela suggère également une diffusion internationale de ces trois clones retrouvés dans l'hôpital marocain, probablement facilitée par le transfert de patients entre différents hôpitaux, voire différents pays. Sur la base de cette diversité clonale une nouvelle question s'est posée : les séquences de référence d'*A. baumannii* représentent-elles la diversité génétique de ces trois ST ? Ce travail de Thèse a évalué pour la première fois l'impact du choix des génomes de référence d'*A. baumannii*. Pour répondre à cette question, nous avons effectué une série d'analyses en utilisant cinq génomes de références les plus utilisés. Nous avons montré que le fait de ne s'appuyer que sur un seul génome de référence d'*A. baumannii* dans une étude incluant plusieurs souches de différentes STs peut générer des biais. Ceci est élucidé par la preuve que : (i) la divergence génétique (basée sur la proportion de SNPs) de nos souches par rapport aux cinq génomes de référence, varie considérablement en fonction de l'identité clonale de référence utilisée. Les génomes appartenant à un clone différent du génome de référence ont montré un taux élevé de divergence génétique et vice versa. (ii) La construction du pangénome en utilisant l'approche basée sur les génomes assemblés a défini une taille du "pan-chromosome" de 6,75 Mb, soit environ 50% de plus par rapport à la taille du génome connu d'*A. baumannii*. Ces résultats corroborent un rapport récent (Valiente-Mullor *et al.*, 2021) montrant que le choix du génome de référence dans chacune des cinq bactéries étudiées (*Klebsiella pneumoniae*, *Legionella pneumophila*, *Neisseria gonorrhoeae*, *Pseudomonas aeruginosa* et *Serratia marcescens*) s'est avéré avoir des biais sur presque toutes les analyses nécessitant un génome de référence telles que l'inférence génétique basée sur les SNP. Cependant, le passage d'un génome de référence issu d'une seule souche à un pangénome de référence, en particulier le « pan-chromosome », pourrait sans aucun doute affiner les analyses ultérieures et réduirait le temps et les efforts nécessaires pour mapper les génomes sur plusieurs séquences.

Ces travaux ont permis de générer des connaissances nouvelles et originales sur la diversité génétique d'*A. baumannii*. Ces résultats ouvrent la voie à l'analyse des caractéristiques génétiques des souches cliniques et environnementales isolées d'un milieu hospitalier au Maroc. Ceci ouvre de nouvelles perspectives sur sa

diversité génomique et l'évolution du génome pour de futures études complémentaires : (i) Recherche de la présence de nouveaux clones de différents niches écologiques dans un contexte d'infection humaine. (ii) L'expansion du pangénome et l'évolution des clones de souches à partir de deux sources d'isolement doivent encore être liées à leur capacité à acquérir des éléments génétiques mobiles (intégrons, plasmides, transposons, séquences d'insertion, etc.) ce qui pourrait contribuer au succès et à la persistance en milieu hospitalier « comparaison intra- et inter-spécifique ». (iii) Enfin, sélectionner des génomes représentatifs de l'ensemble de la diversité génétique (de la population mondiale) de cette espèce afin de suivre la caractérisation du pangénome de référence « pan-chromosome ». La sélection des génomes devrait être basée principalement sur l'ensemble des lignées clonales d'*A. Baumannii* rapportées (plus de 2000 STs différentes sont actuellement définis via MLST « <https://pubmlst.org/abaumannii> »)

III – Quelles sont les relations entre organisation structurale et organisation fonctionnelle des génomes de ces deux micro-organismes ?

3.1. Chez *F. graminearum*

Bien que toutes les souches de *F. graminearum* ne soient pas identiques dans leur capacité à induire des symptômes (Fabre et al., 2019a ; Fabre et al., 2019b ; Goswami et Kistler, 2004 ; Laurent et al., 2017), les gènes, fonctions et mécanismes déterminant le niveau d'agressivité restent encore largement méconnus. La production de mycotoxines est parfois utilisée comme mesure de l'agressivité bien que la relation entre les deux n'est pas toujours directe et reste débattue dans la littérature (Fabre et al., 2019a ; Laurent et al., 2017 ; Mesterházy, 2002 ; Pierron et al., 2016). L'agressivité des souches de cette espèce issues de ce travail a été phénotypée en se basant sur les résultats de Fabre et al. (2019b) en phénotypant les épillets inoculés à 72 hpi, spécifiquement sur le cultivar de blé d'hiver Récital (sensible à la fusariose). Une agressivité contrastée entre les huit souches a été observée, permettant de classer ces souches par catégorie de symptômes produits (forts, intermédiaires et légers). Compte tenu de cette grande différence d'agressivité, nous avons recherché des liens possibles entre cette agressivité et le contenu du pan-sécrétome (composé de protéines sécrétées putatives communes et variables entre toutes les souches). Une précédente étude (Brown et al., 2012) basée sur le génome de référence PH-1 a montré que 4,2% du total des gènes codent pour des protéines sécrétées par la voie classique (présence de peptide signal). Nos résultats basés sur une approche classique et non-classique

ou UPS (ne nécessitant aucun peptide signal pour être sécrété) ont montré qu'environ 5% du nombre total de gènes présentaient des caractéristiques de sécrétion dans chacun des huit génomes de cette étude. Cette proportion semble être constante dans tout sécrétome de *F. graminearum*. Par rapport au sécrétome dispensable, plus de deux tiers de protéines sécrétées présentant des caractéristiques d'effecteurs, de CAZymes, de protéases ou présentant une forte similitude avec les gènes répertoriés dans "PHI-base" a été trouvée dans le core sécrétome, et donc systématiquement partagées entre les souches agressives et moins agressives, suggérant que la pathogénicité des différentes souches de cette espèce relève d'une signature protéique hautement conservée appartenant au core sécrétome. Les gènes codant pour ces effecteurs protéiques étaient préférentiellement localisés dans les régions appartenant au sous-génome rapide de *F. graminearum*, soulignant le rôle central du core sécrétome et du sous-génome rapide dans les processus initiaux d'infection. Ces résultats appuient les travaux d'une étude précédente démontrant que des souches de *F. graminearum* d'agressivité contrastée expriment un ensemble de protéines qualitativement comparable mais quantitativement discriminant lors de leur interaction avec la plante (Fabre *et al.*, 2019b).

Bien qu'aucune relation qualitative et quantitative évidente ne puisse être définie entre le pan-sécrétome et l'agressivité contrastée des souches, nous avons étendu notre analyse en recherchant les mutations LOF présentes dans l'ensemble de ces protéines sécrétées. 11 protéines perdent potentiellement leur fonction, retrouvées spécifiquement dans la souche MDC_FgU1 (la moins agressive), alors qu'elles étaient intactes dans les autres souches. Parmi eux, ceux codés pour l'hydrolyse des composants polysaccharidiques de la paroi cellulaire végétale et dans la dégradation des protéines de défense de l'hôte. Leurs implications ou contribution à la virulence des souches nécessitent des analyses fonctionnelles complémentaires pour être validées. En revanche, la présence de certaines protéines tronquées dans des souches plus agressives comme MDC_Fg1, suggère l'existence d'une éventuelle redondance/compensation fonctionnelle pour maintenir la fonction des gènes (Aguileta *et al.*, 2012). Ce même processus (de redondance/compensation fonctionnelle) semble être largement présent dans des familles de gènes impliqués dans la production de métabolites secondaires. Par exemple, le cluster de gènes de biosynthèse de la fusarine se compose de neuf gènes (fus1-fus9), dont seulement quatre (fus1, fus2, fus8 et fus9) sont nécessaires et suffisants pour assurer la biosynthèse de ce métabolite (Niehaus *et al.*, 2013; Yörük *et al.*, 2015). Bien qu'une mutation LOF ait été trouvée dans le gène fus1, des abondances de ce métabolite ont été observées via l'approche de profilage métabolomique (*in vitro*).

L'un des résultats majeurs de ces travaux de Thèse est la surreprésentation des motifs d'adressage aux chloroplastes et aux noyaux des cellules végétales, qui confirme les résultats de Fabre *et al.* (2019a) proposant le chloroplaste comme cible privilégiée des effecteurs protéiques de ce pathogène. De nombreuses études antérieures avaient déjà identifié le noyau et le chloroplaste comme étant une cible privilégiée des effecteurs fongiques ou bactériens (Canonne *et Rivas*, 2012; Kretschmer *et al.*, 2019; Petre *et al.*, 2016; Q. Xu *et al.*, 2019; F. Yang *et al.*, 2021). En raison de leurs implications dans un grand nombre de processus physiologiques végétaux, le contrôle de ces organites par un pathogène pourrait constituer un point stratégique pour le succès de l'infection. Ces résultats ouvrent également de nouvelles perspectives sur la biologie de l'interaction *blé-F. graminearum*. La première consiste à valider expérimentalement les localisations subcellulaires des effecteurs protéiques dans les deux organites, le chloroplaste et le noyau de la cellule végétale, comme étant les cibles privilégiées de la plante. La seconde consiste à étudier plus spécifiquement le rôle du sécrétome singleton (protéines sécrétées spécifiques d'une souche). Cette étape analytique pourrait s'appuyer d'une part sur les effecteurs protéiques présents uniquement dans les souches les plus agressives et, d'autre part, sur les effecteurs potentiellement perdus leurs fonctions (via des mutations LOF) dans les souches les moins agressives. Ces analyses permettraient de compléter les résultats de ce travail de Thèse en apportant une résolution beaucoup plus fine du pan-sécrétome de *F. graminearum*.

3.1. Chez *A. baumannii*

Bien que les génomes d'*A. baumannii* analysés dans ce travail de Thèse soient isolés de différentes sources et appartenant aux différentes lignées clonales, une conservation élevée du contenu de pan-virulome (l'ensemble des gènes de virulence identifiés dans tous les génomes analysés) avec plus de 80% faisant partie du core virulome (conservé dans tous les génomes), suggérant la nature pathogène de la plupart des souches d'*A. baumannii*. Corroborant les résultats d'une étude précédente, montrant que les fonctions associées à la virulence se trouvent dans le core génome (Yakkala *et al.*, 2019).

Outre la conservation du pan-virulome, un contenu diversifié du pan-résistome (l'ensemble des gènes de résistance aux antibiotiques identifiés dans tous les génomes) a été identifié. Toutes les souches de ce travail sont des souches multirésistantes, le profil phénotypique de résistance à plusieurs classes d'antibiotiques s'est également reflété dans le contenu des gènes de résistance. Ces déterminants étaient principalement plus variables en fonction de l'identité clonale de la souche, quelles que soient leurs origines (cliniques ou environnementaux). En d'autres termes, les souches isolées de patients et de l'environnement hospitalier

partageaient la plupart des gènes de résistance, en particulier dans les génomes appartenant aux mêmes STs, suggérant des événements d'échanges génétiques préférentiels entre les souches de même clone. En revanche, une variabilité génétique intra-clonale a été également observée, notamment pour la lignée ST85. Par exemple, NDM-1 (une enzyme qui rend les bactéries résistantes à une large gamme de bêta-lactamines) est souvent associé à la lignée ST85 (Heydari *et al.*, 2015; Jaidane *et al.*, 2018; Salloum *et al.*, 2018). Ce gène était absent dans l'une des quatre souches (M-C40) de cette lignée, ce qui pourrait refléter la nature génétiquement hétérogène de cette espèce. Cette variabilité intra-clonale, il a été précédemment décrit dans des souches provenant des hôpitaux de différentes zones géographiques, qui présentaient le même clone, mais avec une acquisition de gènes résistances différents (Leal *et al.*, 2020), indiquant une dissémination d'éléments mobiles éventuellement variable en raison des conditions diverses existant dans les différentes zones géographiques. La variabilité génétique entre les STs comprend non seulement les déterminants de la résistance acquise, mais également les déterminants de résistance intrinsèques (situés sur le chromosome). Ces derniers pourraient également être proposés comme marqueurs des lignées clonales d'*A. baumannii*. L'exemple le plus connu est la carbapénémase OXA-51 (Rafei *et al.*, 2015). Conformément aux rapports précédents, nous avons révélé trois variants d'OXA-51 « OXA-66, OXA-69 et OXA-94 » appartenant respectivement aux ST2, ST315 et ST85, respectivement (Fernández-Cuenca *et al.*, 2020; Leal *et al.*, 2020; Salloum *et al.*, 2018). Le variant OXA-69 a également été signalé antérieurement dans des souches appartenant à ST1 (correspond à IC-I ; Rafei *et al.*, 2015), ce qui pourrait expliquer en partie la parenté observée entre ces deux lignées "ST315 et ST1" via l'analyse phylogénomique basée sur le core génome. Cette proximité pourrait également refléter le faible niveau de variants génétiques identifiés entre notre souche ST315 et le génome de référence de ST1.

Des mutations dans les gènes constitutifs peuvent également conduire à des niveaux élevés de résistance aux principales classes d'antibiotiques (Adams *et al.*, 2009). La souche clinique M-C15 est considérée comme la première souche marocaine d'*A. baumannii* résistante à la dernière ligne de traitement (colistine). Selon les mécanismes responsables de la résistance à la colistine chez *A. baumannii* (cf. "section 2.1.3.4" de l'introduction générale), l'analyse génomique comparative a révélé une mutation (R263S) dans le domaine histidine kinase du gène *pmrB* dans la souche M-C15 comparé aux autres souches sensibles à la colistine. Comme suggéré récemment, cette mutation a été révélée dans quatre souches résistantes à la colistine isolées d'un environnement hospitalier (eaux usées hospitalières) en Croatie (Jovcic *et al.*, 2021). Les mutations de *pmrB* sont souvent liées à une expression accrue de *pmrC*, qui à son tour est associée à une augmentation de

la CMI de la colistine (Adams *et al.*, 2009). En raison de la CMI élevée de colistine (≥ 256 mg/L) observée dans M-C15, nous pensons que cette souche peut présenter un niveau d'expression accru de *PmrC* par rapport aux souches sensibles à la colistine. Une analyse transcriptionnelle plus poussée est désormais nécessaire non seulement pour vérifier leur niveau d'expression, mais également pour valider l'association de *pmrB* avec la résistance à la colistine. L'émergence de souches résistantes à tous les antibiotiques comme M-C15, soulignant la nécessité d'une surveillance épidémiologique génomique continue au niveau national et à travers les réseaux internationaux pour détecter rapidement ces organismes et prévenir leur propagation.

Conclusion générale

Conclusion générale

Cette Thèse présente l'analyse de génomes complets de souches européennes de *Fusarium graminearum* et de souches marocaines d'*Acinetobacter baumannii* avec comme objectif la caractérisation de la variabilité intraspécifique de l'organisation structurale et fonctionnelle, et la contribution potentielle de cette variabilité à la virulence et/ou au changement phénotypique (l'agressivité sur le blé pour *F. graminearum* et la résistance aux antibiotiques pour *A. baumannii*).

Ce travail a permis de revisiter l'image que l'on avait de l'organisation du génome de *Fusarium graminearum*, avec deux résultats importants :

(1) la mise en évidence pour la première fois d'une très forte variabilité intraspécifique de son organisation génomique, avec en particulier, une variation importante du nombre de gènes présents dans les génomes des différentes souches. Ces résultats ont permis de définir le premier pangénome « ouvert » de cette espèce. Cette forte variabilité est probablement à mettre en lien avec la capacité adaptative de *F. graminearum* (environnements, hôtes).

(2) l'existence d'un « core sécrétome/effectome », partagé par l'ensemble des souches analysées, sans lien direct avec les différences d'agressivité, évaluées pour ces différentes souches sur une même plante hôte, le blé. L'analyse de ce core-effectome a permis de confirmer la présence en grands nombre de candidats effecteurs ciblant des compartiments sub-cellulaires tels que le chloroplaste et le noyau. Ces résultats ouvrent la voie à des approches plus fonctionnelles ayant comme objectif la validation du rôle de ces candidats effecteurs dans la virulence de ces souches.

Chez *A. baumannii*, les résultats notables issus de ce travail résident dans (1) l'identification, parmi les souches marocaines, de trois clones génétiquement distincts, précédemment signalés dans de nombreuses régions du monde, notamment en Europe. Ce résultat suggère non seulement une diffusion internationale de ces clones, mais également, des transmissions locales en raison de la relation étroite observée entre les souches isolées chez des patients et celles isolées dans le milieu hospitalier. En outre, (2) les variations génétiques entre ces trois lignées clonales d'*A. baumannii* impliquant en particulier le contenu génétique de résistance aux antibiotiques. Ce dernier reflétait les différents phénotypes de résistance aux antibiotiques, dont la résistance à la colistine (dernière ligne de traitement). Ces résultats indiquent que le renforcement des mesures de

contrôle et de prévention des infections nosocomiales serait actuellement le moyen le plus fiable pour limiter la propagation de ces clones multirésistants.

Les résultats obtenus pour les deux micro-organismes pathogènes montrent clairement les limites de ce que l'on appelle « les génomes de référence », qui ne permettent pas de capter toute l'information génomique des deux espèces. Ils soulignent ainsi l'importance de la caractérisation de pangénomes et de leurs composantes (core, accessoire et unique) pour définir de nouvelles références génomiques plus pertinentes pour améliorer notre compréhension de la biologie des espèces.

Références bibliographiques

- Abdi, S. N., Ghotaslou, R., Ganbarov, K., Mobed, A., Tanomand, A., Yousefi, M., Asgharzadeh, M., and Kafil, H. S. (2020). *Acinetobacter baumannii* Efflux Pumps and Antibiotic Resistance. *Infection and Drug Resistance*, 13, 423-434. <https://doi.org/10.2147/IDR.S228089>
- Adams, M. D., Goglin, K., Molyneaux, N., Hujer, K. M., Lavender, H., Jamison, J. J., MacDonald, I. J., Martin, K. M., Russo, T., Campagnari, A. A., Hujer, A. M., Bonomo, R. A., and Gill, S. R. (2008). Comparative Genome Sequence Analysis of Multidrug-Resistant *Acinetobacter baumannii*. *Journal of Bacteriology*, 190(24), 8053-8064. <https://doi.org/10.1128/JB.00834-08>
- Adams, M. D., Nickel, G. C., Bajaksouzian, S., Lavender, H., Murthy, A. R., Jacobs, M. R., and Bonomo, R. A. (2009). Resistance to Colistin in *Acinetobacter baumannii* Associated with Mutations in the PmrAB Two-Component System. *Antimicrobial Agents and Chemotherapy*, 53(9), 3628-3634. <https://doi.org/10.1128/AAC.00284-09>
- Aggarwal, S. K., Singh, A., Choudhary, M., Kumar, A., Rakshit, S., Kumar, P., Bohra, A., and Varshney, R. K. (2022). Pangenomics in Microbial and Crop Research: Progress, Applications, and Perspectives. *Genes*, 13(4), 598. <https://doi.org/10.3390/genes13040598>
- Aguileta, G., Lengelle, J., Chiapello, H., Giraud, T., Viaud, M., Fournier, E., Rodolphe, F., Marthey, S., Ducasse, A., Gendrault, A., Poulain, J., Wincker, P., and Gout, L. (2012). Genes under positive selection in a model plant pathogenic fungus, *Botrytis*. *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, 12(5), 987-996. <https://doi.org/10.1016/j.meegid.2012.02.012>
- Al Atrouni, A., Joly-Guillou, M.-L., Hamze, M., and Kempf, M. (2016). Reservoirs of Non-*baumannii* *Acinetobacter* Species. *Frontiers in Microbiology*, 7. <https://www.frontiersin.org/article/10.3389/fmicb.2016.00049>

- Alcaraz, L. D., Moreno-Hagelsieb, G., Eguiarte, L. E., Souza, V., Herrera-Estrella, L., and Olmedo, G. (2010). Understanding the evolutionary relationships and major traits of *Bacillus* through comparative genomics. *BMC Genomics*, 11, 332. <https://doi.org/10.1186/1471-2164-11-332>
- Alexander, N. J., McCormick, S. P., Waalwijk, C., van der Lee, T., and Proctor, R. H. (2011). The genetic basis for 3-ADON and 15-ADON trichothecene chemotypes in *Fusarium*. *Fungal Genetics and Biology: FG & B*, 48(5), 485-495. <https://doi.org/10.1016/j.fgb.2011.01.003>
- Alfaro, M., Oguiza, J. A., Ramírez, L., and Pisabarro, A. G. (2014). Comparative analysis of secretomes in basidiomycete fungi. *Journal of Proteomics*, 102, 28-43. <https://doi.org/10.1016/j.jprot.2014.03.001>
- Aliramezani, A., Soleimani, M., Fard, R. M. N., and Nojoomi, F. (2019). Virulence determinants and biofilm formation of *Acinetobacter baumannii* isolated from hospitalized patients. *Germs*, 9(3), 148-153. <https://doi.org/10.18683/germs.2019.1171>
- Almagro Armenteros, J. J., Salvatore, M., Emanuelsson, O., Winther, O., von Heijne, G., Elofsson, A., and Nielsen, H. (2019). Detecting sequence signals in targeting peptides using deep learning. *Life Science Alliance*, 2(5), e201900429. <https://doi.org/10.26508/lsa.201900429>
- Alouane, T., Rimbert, H., Bormann, J., González-Montiel, G. A., Loesgen, S., Schäfer, W., Freitag, M., Langin, T., and Bonhomme, L. (2021). Comparative Genomics of Eight *Fusarium graminearum* Strains with Contrasting Aggressiveness Reveals an Expanded Open Pangenome and Extended Effector Content Signatures. *International Journal of Molecular Sciences*, 22(12), 6257. <https://doi.org/10.3390/ijms22126257>
- Amarasinghe, C. C., and Fernando, W. G. D. (2016). Comparative Analysis of Deoxynivalenol Biosynthesis Related Gene Expression among Different Chemotypes of *Fusarium graminearum* in Spring Wheat. *Frontiers in Microbiology*, 7, 1229. <https://doi.org/10.3389/fmicb.2016.01229>
- Antunes, L. C. S., Visca, P., and Towner, K. J. (2014). *Acinetobacter baumannii*: Evolution of a global pathogen. *Pathogens and Disease*, 71(3), 292-301. <https://doi.org/10.1111/2049-632X.12125>
- Antunes, N. T., Lamoureaux, T. L., Toth, M., Stewart, N. K., Frase, H., and Vakulenko, S. B. (2014). Class D β -Lactamases: Are They All Carbapenemases? *Antimicrobial Agents and Chemotherapy*, 58(4), 2119-2125. <https://doi.org/10.1128/AAC.02522-13>

- Arroyo, L. A., Herrera, C. M., Fernandez, L., Hankins, J. V., Trent, M. S., and Hancock, R. E. W. (2011). The pmrCAB operon mediates polymyxin resistance in *Acinetobacter baumannii* ATCC 17978 and clinical isolates through phosphoethanolamine modification of lipid A. *Antimicrobial Agents and Chemotherapy*, 55(8), 3743-3751. <https://doi.org/10.1128/AAC.00256-11>
- Atasoy, A. R., Ciftci, I. H., and Petek, M. (2015). Modifying enzymes related aminoglycoside : Analyses of resistant *Acinetobacter* isolates. *International Journal of Clinical and Experimental Medicine*, 8(2), 2874-2880.
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., and Pevzner, P. A. (2012). SPAdes : A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology*, 19(5), 455-477. <https://doi.org/10.1089/cmb.2012.0021>
- Bao, W., Kojima, K. K., and Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA*, 6, 11. <https://doi.org/10.1186/s13100-015-0041-9>
- Barajas, H. R., Romero, M. F., Martínez-Sánchez, S., and Alcaraz, L. D. (2019). Global genomic similarity and core genome sequence diversity of the *Streptococcus* genus as a toolkit to identify closely related bacterial species in complex environments. *PeerJ*, 6, e6233. <https://doi.org/10.7717/peerj.6233>
- Bartual, S. G., Seifert, H., Hippler, C., Luzon, M. A. D., Wisplinghoff, H., and Rodríguez-Valera, F. (2005). Development of a Multilocus Sequence Typing Scheme for Characterization of Clinical Isolates of *Acinetobacter baumannii*. *Journal of Clinical Microbiology*, 43(9), 4382-4390. <https://doi.org/10.1128/JCM.43.9.4382-4390.2005>
- Baumann, P. (1968). Isolation of *Acinetobacter* from Soil and Water. *Journal of Bacteriology*, 96(1), 39-42.
- Bendtsen, J. D., Jensen, L. J., Blom, N., Von Heijne, G., and Brunak, S. (2004). Feature-based prediction of non-classical and leaderless protein secretion. *Protein Engineering, Design & Selection: PEDS*, 17(4), 349-356. <https://doi.org/10.1093/protein/gzh037>
- Bergogne-Bérézin, E., and Towner, K. J. (1996). *Acinetobacter* spp. as nosocomial pathogens : Microbiological, clinical, and epidemiological features. *Clinical Microbiology Reviews*, 9(2), 148-165.

- Bhushan, S., Kuhn, C., Berglund, A.-K., Roth, C., and Glaser, E. (2006). The role of the N-terminal domain of chloroplast targeting peptides in organellar protein import and miss-sorting. *FEBS Letters*, 580(16), 3966-3972. <https://doi.org/10.1016/j.febslet.2006.06.018>
- Blackwell, G. A., and Hall, R. M. (2017). The tet39 Determinant and the msrE-mphE Genes in *Acinetobacter* Plasmids Are Each Part of Discrete Modules Flanked by Inversely Oriented pdif (XerC-XerD) Sites. *Antimicrobial Agents and Chemotherapy*, 61(8), e00780-17. <https://doi.org/10.1128/AAC.00780-17>
- Blanco, P., Hernando-Amado, S., Reales-Calderon, J. A., Corona, F., Lira, F., Alcalde-Rico, M., Bernardini, A., Sanchez, M. B., and Martinez, J. L. (2016). Bacterial Multidrug Efflux Pumps : Much More Than Antibiotic Resistance Determinants. *Microorganisms*, 4(1), 14. <https://doi.org/10.3390/microorganisms4010014>
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic : A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114-2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Bonnin, R. A., Nordmann, P., Potron, A., Lecuyer, H., Zahar, J.-R., and Poirel, L. (2011). Carbapenem-Hydrolyzing GES-Type Extended-Spectrum β -Lactamase in *Acinetobacter baumannii*. *Antimicrobial Agents and Chemotherapy*, 55(1), 349-354. <https://doi.org/10.1128/AAC.00773-10>
- Bortolaia, V., Kaas, R. S., Ruppe, E., Roberts, M. C., Schwarz, S., Cattoir, V., Philippon, A., Allesoe, R. L., Rebelo, A. R., Florensa, A. F., Fagelhauer, L., Chakraborty, T., Neumann, B., Werner, G., Bender, J. K., Stingl, K., Nguyen, M., Coppens, J., Xavier, B. B., ... Aarestrup, F. M. (2020). ResFinder 4.0 for predictions of phenotypes from genotypes. *The Journal of Antimicrobial Chemotherapy*, 75(12), 3491-3500. <https://doi.org/10.1093/jac/dkaa345>
- Bottalico, A., and Perrone, G. (2002). Toxigenic *Fusarium* species and Mycotoxins Associated with Head Blight in Small-Grain Cereals in Europe. *European Journal of Plant Pathology*, 108(7), 611-624. <https://doi.org/10.1023/A:1020635214971>
- Boutet, E., Lieberherr, D., Tognolli, M., Schneider, M., Bansal, P., Bridge, A. J., Poux, S., Bougueleret, L., and Xenarios, I. (2016). UniProtKB/Swiss-Prot, the Manually Annotated Section of the UniProt KnowledgeBase : How to Use the Entry View. *Methods in Molecular Biology (Clifton, N.J.)*, 1374, 23-54. https://doi.org/10.1007/978-1-4939-3167-5_2

- Brockhurst, M. A., Harrison, E., Hall, J. P. J., Richards, T., McNally, A., and MacLean, C. (2019). The Ecology and Evolution of Pangenomes. *Current Biology*, 29(20), R1094-R1103. <https://doi.org/10.1016/j.cub.2019.08.012>
- 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(4), e33731. <https://doi.org/10.1371/journal.pone.0033731>
- Brown, S., Young, H. K., and Amyes, S. G. B. (2005). Characterisation of OXA-51, a novel class D carbapenemase found in genetically unrelated clinical strains of *Acinetobacter baumannii* from Argentina. *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases*, 11(1), 15-23. <https://doi.org/10.1111/j.1469-0691.2004.01016.x>
- Burgess, D. (2021). Quantifying the effect of reference genome choice. *Nature Reviews Genetics*, 22(8), 481-481. <https://doi.org/10.1038/s41576-021-00392-2>
- Bushmanova, E., Antipov, D., Lapidus, A., and Prjibelski, A. D. (2019). rnaSPAdes : A de novo transcriptome assembler and its application to RNA-Seq data. *GigaScience*, 8(9), giz100. <https://doi.org/10.1093/gigascience/giz100>
- Campbell, M. S., Holt, C., Moore, B., and Yandell, M. (2014). Genome Annotation and Curation Using MAKER and MAKER-P. *Current Protocols in Bioinformatics*, 48, 4.11.1-39. <https://doi.org/10.1002/0471250953.bi0411s48>
- Canonne, J., and Rivas, S. (2012). Bacterial effectors target the plant cell nucleus to subvert host transcription. *Plant Signaling and Behavior*, 7(2), 217-221. <https://doi.org/10.4161/psb.18885>
- Cerezales, M., Xanthopoulou, K., Wille, J., Krut, O., Seifert, H., Gallego, L., and Higgins, P. G. (2020). Mobile Genetic Elements Harboring Antibiotic Resistance Determinants in *Acinetobacter baumannii* Isolates From Bolivia. *Frontiers in Microbiology*, 11, 919. <https://doi.org/10.3389/fmicb.2020.00919>
- Chan, A. P., Sutton, G., DePew, J., Krishnakumar, R., Choi, Y., Huang, X.-Z., Beck, E., Harkins, D. M., Kim, M., Lesho, E. P., Nikolich, M. P., and Fouts, D. E. (2015). A novel method of consensus pan-

- chromosome assembly and large-scale comparative analysis reveal the highly flexible pan-genome of *Acinetobacter baumannii*. *Genome Biology*, 16(1), 143. <https://doi.org/10.1186/s13059-015-0701-6>
- Chin, C.-S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E. E., Turner, S. W., and Korlach, J. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nature Methods*, 10(6), 563-569. <https://doi.org/10.1038/nmeth.2474>
- Choi, C. H., Lee, E. Y., Lee, Y. C., Park, T. I., Kim, H. J., Hyun, S. H., Kim, S. A., Lee, S.-K., and Lee, J. C. (2005). Outer membrane protein 38 of *Acinetobacter baumannii* localizes to the mitochondria and induces apoptosis of epithelial cells. *Cellular Microbiology*, 7(8), 1127-1138. <https://doi.org/10.1111/j.1462-5822.2005.00538.x>
- Choi, M.-J., and Ko, K. S. (2014). Mutant prevention concentrations of colistin for *Acinetobacter baumannii*, *Pseudomonas aeruginosa* and *Klebsiella pneumoniae* clinical isolates. *Journal of Antimicrobial Chemotherapy*, 69(1), 275-277. <https://doi.org/10.1093/jac/dkt315>
- Cingolani, P., Patel, V. M., Coon, M., Nguyen, T., Land, S. J., Ruden, D. M., and Lu, X. (2012). Using *Drosophila melanogaster* as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. *Frontiers in Genetics*, 3, 35. <https://doi.org/10.3389/fgene.2012.00035>
- Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., and Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly*, 6(2), 80-92. <https://doi.org/10.4161/fly.19695>
- Coelho, J., Woodford, N., Afzal-Shah, M., and Livermore, D. (2006). Occurrence of OXA-58-Like Carbapenemases in *Acinetobacter spp.* Collected over 10 Years in Three Continents. *Antimicrobial Agents and Chemotherapy*, 50(2), 756-758. <https://doi.org/10.1128/AAC.50.2.756-758.2006>
- Cuomo, C. A., Guldener, U., Xu, J.-R., Trail, F., Turgeon, B. G., Di Pietro, A., Walton, J. D., Ma, L.-J., Baker, S. E., Rep, M., Adam, G., Antoniw, J., Baldwin, T., Calvo, S., Chang, Y.-L., Decaprio, D., Gale, L. R., Gnerre, S., Goswami, R. S., ... Kistler, H. C. (2007). The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. *Science (New York, N.Y.)*, 317(5843), 1400-1402. <https://doi.org/10.1126/science.1143708>

- Cuzick, A., Urban, M., and Hammond-Kosack, K. (2008). *Fusarium graminearum* gene deletion mutants map1 and tri5 reveal similarities and differences in the pathogenicity requirements to cause disease on Arabidopsis and wheat floral tissue. *The New Phytologist*, 177(4), 990-1000. <https://doi.org/10.1111/j.1469-8137.2007.02333.x>
- Damier-Piolle, L., Magnet, S., Brémont, S., Lambert, T., and Courvalin, P. (2008). AdeIJK, a resistance-nodulation-cell division pump effluxing multiple antibiotics in *Acinetobacter baumannii*. *Antimicrobial Agents and Chemotherapy*, 52(2), 557-562. <https://doi.org/10.1128/AAC.00732-07>
- Desjardins, A. E., and Proctor, R. H. (2007). Molecular biology of *Fusarium* mycotoxins. *International Journal of Food Microbiology*, 119(1-2), 47-50. <https://doi.org/10.1016/j.ijfoodmicro.2007.07.024>
- Diancourt, L., Passet, V., Nemec, A., Dijkshoorn, L., and Brisse, S. (2010). The population structure of *Acinetobacter baumannii* : Expanding multiresistant clones from an ancestral susceptible genetic pool. *PloS One*, 5(4), e10034. <https://doi.org/10.1371/journal.pone.0010034>
- Ding, L., Xu, H., Yi, H., Yang, L., Kong, Z., Zhang, L., Xue, S., Jia, H., and Ma, Z. (2011). Resistance to Hemi-Biotrophic *F. graminearum* Infection Is Associated with Coordinated and Ordered Expression of Diverse Defense Signaling Pathways. *PLOS ONE*, 6(4), e19008. <https://doi.org/10.1371/journal.pone.0019008>
- Dong, S., Raffaele, S., and Kamoun, S. (2015). The two-speed genomes of filamentous pathogens : Waltz with plants. *Current Opinion in Genetics & Development*, 35, 57-65. <https://doi.org/10.1016/j.gde.2015.09.001>
- Doster, E., Lakin, S. M., Dean, C. J., Wolfe, C., Young, J. G., Boucher, C., Belk, K. E., Noyes, N. R., and Morley, P. S. (2020). MEGARes 2.0 : A database for classification of antimicrobial drug, biocide and metal resistance determinants in metagenomic sequence data. *Nucleic Acids Research*, 48(D1), D561-D569. <https://doi.org/10.1093/nar/gkz1010>
- Dunbar, H. E., Wilson, A. C. C., Ferguson, N. R., and Moran, N. A. (2007). Aphid thermal tolerance is governed by a point mutation in bacterial symbionts. *PLoS Biology*, 5(5), e96. <https://doi.org/10.1371/journal.pbio.0050096>

- Eigenbrod, T., Reuter, S., Gross, A., Kocer, K., Günther, F., Zimmermann, S., Heeg, K., Mutters, N. T., and Nurjadi, D. (2019). Molecular characterization of carbapenem-resistant *Acinetobacter baumannii* using WGS revealed missed transmission events in Germany from 2012-15. *The Journal of Antimicrobial Chemotherapy*, 74(12), 3473-3480. <https://doi.org/10.1093/jac/dkz360>
- Eliopoulos, G. M., Maragakis, L. L., and Perl, T. M. (2008). *Acinetobacter baumannii* : Epidemiology, Antimicrobial Resistance, and Treatment Options. *Clinical Infectious Diseases*, 46(8), 1254-1263. <https://doi.org/10.1086/529198>
- Evans, B. A., and Amyes, S. G. B. (2014). OXA β -Lactamases. *Clinical Microbiology Reviews*, 27(2), 241-263. **Erreur ! Référence de lien hypertexte non valide.**
- Fabre, F., Vignassa, M., Urbach, S., Langin, T., and Bonhomme, L. (2019a). Time-resolved dissection of the molecular crosstalk driving Fusarium head blight in wheat provides new insights into host susceptibility determinism. *Plant, Cell & Environment*, 42(7), 2291-2308. <https://doi.org/10.1111/pce.13549>
- Fabre, F., Bormann, J., Urbach, S., Roche, S., Langin, T., and Bonhomme, L. (2019b). Unbalanced Roles of Fungal Aggressiveness and Host Cultivars in the Establishment of the Fusarium Head Blight in Bread Wheat. *Frontiers in Microbiology*, 10, 2857. <https://doi.org/10.3389/fmicb.2019.02857>
- Falagas, M. E., and Kasiakou, S. K. (2006). Toxicity of polymyxins : A systematic review of the evidence from old and recent studies. *Critical Care*, 10(1), R27. <https://doi.org/10.1186/cc3995>
- Fedrigo, N. H., Xavier, D. E., Cerdeira, L., Fuga, B., Marini, P. V. B., Shinohara, D. R., Carrara-Marroni, F. E., Lincopan, N., and Tognim, M. C. B. (2022). Genomic insights of *Acinetobacter baumannii* ST374 reveal wide and increasing resistome and virulome. *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, 97, 105148. <https://doi.org/10.1016/j.meegid.2021.105148>
- Fernandes, N., Case, R. J., Longford, S. R., Seyedsayamdost, M. R., Steinberg, P. D., Kjelleberg, S., and Thomas, T. (2011). Genomes and Virulence Factors of Novel Bacterial Pathogens Causing Bleaching Disease in the Marine Red Alga *Delisea pulchra*. *PLOS ONE*, 6(12), e27387. <https://doi.org/10.1371/journal.pone.0027387>

- Fernández-Cuenca, F., Pérez-Palacios, P., Galán-Sánchez, F., López-Cerero, L., López-Hernández, I., López Rojas, R., Arca-Suárez, J., Díaz-de Alba, P., Rodríguez Iglesias, M., and Pascual, A. (2020). First identification of blaNDM-1 carbapenemase in blaOXA-94-producing *Acinetobacter baumannii* ST85 in Spain. *Enfermedades Infecciosas Y Microbiología Clínica (English Ed.)*, 38(1), 11-15. <https://doi.org/10.1016/j.eimc.2019.03.008>
- Fiester, S. E., Arivett, B. A., Schmidt, R. E., Beckett, A. C., Ticak, T., Carrier, M. V., Ghosh, R., Ohneck, E. J., Metz, M. L., Sellin Jeffries, M. K., and Actis, L. A. (2016). Iron-Regulated Phospholipase C Activity Contributes to the Cytolytic Activity and Virulence of *Acinetobacter baumannii*. *PloS One*, 11(11), e0167068. <https://doi.org/10.1371/journal.pone.0167068>
- Figaj, D., Ambroziak, P., Przepiora, T., and Skorko-Glonek, J. (2019). The Role of Proteases in the Virulence of Plant Pathogenic Bacteria. *International Journal of Molecular Sciences*, 20(3), 672. <https://doi.org/10.3390/ijms20030672>
- Fischer, S., Brunk, B. P., Chen, F., Gao, X., Harb, O. S., Iodice, J. B., Shanmugam, D., Roos, D. S., and Stoeckert, C. J. (2011). Using OrthoMCL to assign proteins to OrthoMCL-DB groups or to cluster proteomes into new ortholog groups. *Current protocols in bioinformatics / editorial board, Andreas D. Baxevanis ... [et al.]*, CHAPTER, Unit-6.1219. <https://doi.org/10.1002/0471250953.bi0612s35>
- Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A., and Merrick, J. M. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science (New York, N.Y.)*, 269(5223), 496-512. <https://doi.org/10.1126/science.7542800>
- Flynn, C. M., Broz, K., Jonkers, W., Schmidt-Dannert, C., and Kistler, H. C. (2019). Expression of the *Fusarium graminearum* terpenome and involvement of the endoplasmic reticulum-derived toxosome. *Fungal Genetics and Biology*, 124, 78-87. <https://doi.org/10.1016/j.fgb.2019.01.006>
- Fournier, P.-E., Vallenet, D., Barbe, V., Audic, S., Ogata, H., Poirel, L., Richet, H., Robert, C., Mangenot, S., Abergel, C., Nordmann, P., Weissenbach, J., Raoult, D., and Claverie, J.-M. (2006). Comparative Genomics of Multidrug Resistance in *Acinetobacter baumannii*. *PLOS Genetics*, 2(1), e7. <https://doi.org/10.1371/journal.pgen.0020007>

- Firdich, E., and Whitfield, C. (2005). Lipopolysaccharide inner core oligosaccharide structure and outer membrane stability in human pathogens belonging to the Enterobacteriaceae. *Journal of Endotoxin Research*, 11(3), 133-144. <https://doi.org/10.1179/096805105X46592>
- Gaffoor, I., Brown, D. W., Plattner, R., Proctor, R. H., Qi, W., and Trail, F. (2005). Functional Analysis of the Polyketide Synthase Genes in the Filamentous Fungus *Gibberella zeae* (Anamorph *Fusarium graminearum*). *Eukaryotic Cell*, 4(11), 1926-1933. <https://doi.org/10.1128/EC.4.11.1926-1933.2005>
- Gage, J. L., Vaillancourt, B., Hamilton, J. P., Manrique-Carpintero, N. C., Gustafson, T. J., Barry, K., Lipzen, A., Tracy, W. F., Mikel, M. A., Kaeppler, S. M., Buell, C. R., and de Leon, N. (2019). Multiple Maize Reference Genomes Impact the Identification of Variants by Genome-Wide Association Study in a Diverse Inbred Panel. *The Plant Genome*, 12(2), 180069. <https://doi.org/10.3835/plant-genome2018.09.0069>
- Gallagher, L. A., Ramage, E., Weiss, E. J., Radey, M., Hayden, H. S., Held, K. G., Huse, H. K., Zurawski, D. V., Brittnacher, M. J., and Manoil, C. (2015). Resources for Genetic and Genomic Analysis of Emerging Pathogen *Acinetobacter baumannii*. *Journal of Bacteriology*, 197(12), 2027-2035. <https://doi.org/10.1128/JB.00131-15>
- Gel, B., and Serra, E. (2017). karyoploteR : An R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics (Oxford, England)*, 33(19), 3088-3090. <https://doi.org/10.1093/bioinformatics/btx346>
- Gerson, S., Betts, J. W., Lucaßen, K., Nodari, C. S., Wille, J., Josten, M., Göttig, S., Nowak, J., Stefanik, D., Roca, I., Vila, J., Cisneros, J. M., La Ragione, R. M., Seifert, H., and Higgins, P. G. (2019). Investigation of Novel *pmrB* and *eptA* Mutations in Isogenic *Acinetobacter baumannii* Isolates Associated with Colistin Resistance and Increased Virulence In Vivo. *Antimicrobial Agents and Chemotherapy*, 63(3), e01586-18. <https://doi.org/10.1128/aac.01586-18>
- Giannouli, M., Antunes, L. C., Marchetti, V., Triassi, M., Visca, P., and Zarrilli, R. (2013). Virulence-related traits of epidemic *Acinetobacter baumannii* strains belonging to the international clonal lineages I-III and to the emerging genotypes ST25 and ST78. *BMC Infectious Diseases*, 13, 282. <https://doi.org/10.1186/1471-2334-13-282>

- Gibson, M. K., Forsberg, K. J., and Dantas, G. (2015). Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. *The ISME Journal*, 9(1), 207-216. <https://doi.org/10.1038/ismej.2014.106>
- Godeux, A.-S., Svedholm, E., Lupo, A., Haenni, M., Venner, S., Laaberki, M.-H., and Charpentier, X. (2020). Scarless Removal of Large Resistance Island AbaR Results in Antibiotic Susceptibility and Increased Natural Transformability in *Acinetobacter baumannii*. *Antimicrobial Agents and Chemotherapy*, 64(10), e00951-20. <https://doi.org/10.1128/AAC.00951-20>
- Goh, H. M. S., Beatson, S. A., Totsika, M., Moriel, D. G., Phan, M.-D., Szubert, J., Runnegar, N., Sidjabat, H. E., Paterson, D. L., Nimmo, G. R., Lipman, J., and Schembri, M. A. (2013). Molecular Analysis of the *Acinetobacter baumannii* Biofilm-Associated Protein. *Applied and Environmental Microbiology*, 79(21), 6535-6543. <https://doi.org/10.1128/AEM.01402-13>
- Golicz, A. A., Bayer, P. E., Barker, G. C., Edger, P. P., Kim, H., Martinez, P. A., Chan, C. K. K., Severn-Ellis, A., McCombie, W. R., Parkin, I. A. P., Paterson, A. H., Pires, J. C., Sharpe, A. G., Tang, H., Teakle, G. R., Town, C. D., Batley, J., and Edwards, D. (2016). The pangenome of an agronomically important crop plant *Brassica oleracea*. *Nature Communications*, 7, 13390. <https://doi.org/10.1038/ncomms13390>
- Golicz, A. A., Bayer, P. E., Bhalla, P. L., Batley, J., and Edwards, D. (2020). Pangenomics Comes of Age : From Bacteria to Plant and Animal Applications. *Trends in Genetics*, 36(2), 132-145. <https://doi.org/10.1016/j.tig.2019.11.006>
- Gordon, S. P., Contreras-Moreira, B., Woods, D. P., Des Marais, D. L., Burgess, D., Shu, S., Stritt, C., Roulin, A. C., Schackwitz, W., Tyler, L., Martin, J., Lipzen, A., Dochy, N., Phillips, J., Barry, K., Geuten, K., Budak, H., Juenger, T. E., Amasino, R., ... Vogel, J. P. (2017). Extensive gene content variation in the *Brachypodium distachyon* pan-genome correlates with population structure. *Nature Communications*, 8(1), 2184. <https://doi.org/10.1038/s41467-017-02292-8>
- Gorkovskiy, A., and Verstrepen, K. J. (2021). The Role of Structural Variation in Adaptation and Evolution of Yeast and Other Fungi. *Genes*, 12(5), 699. <https://doi.org/10.3390/genes12050699>

- Goswami, R. S., and Kistler, H. C. (2004). Heading for disaster : *Fusarium graminearum* on cereal crops. *Molecular Plant Pathology*, 5(6), 515-525. <https://doi.org/10.1111/j.1364-3703.2004.00252.x>
- Guimarães, L. C., Florczak-Wyspianska, J., de Jesus, L. B., Viana, M. V. C., Silva, A., Ramos, R. T. J., Soares, S. de C., and Soares, S. de C. (2015). Inside the Pan-genome—Methods and Software Overview. *Current Genomics*, 16(4), 245-252. <https://doi.org/10.2174/1389202916666150423002311>
- Gupta, R., Lee, S. E., Agrawal, G. K., Rakwal, R., Park, S., Wang, Y., and Kim, S. T. (2015). Understanding the plant-pathogen interactions in the context of proteomics-generated apoplastic proteins inventory. *Frontiers in Plant Science*, 6. <https://www.frontiersin.org/article/10.3389/fpls.2015.00352>
- Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. *Bioinformatics (Oxford, England)*, 29(8), 1072-1075. <https://doi.org/10.1093/bioinformatics/btt086>
- Guyon, K., Balagué, C., Roby, D., and Raffaele, S. (2014). Secretome analysis reveals effector candidates associated with broad host range necrotrophy in the fungal plant pathogen *Sclerotinia sclerotiorum*. *BMC Genomics*, 15(1), 336. <https://doi.org/10.1186/1471-2164-15-336>
- Hägglblom, P., and Nordkvist, E. (2015). Deoxynivalenol, zearalenone, and *Fusarium graminearum* contamination of cereal straw; field distribution; and sampling of big bales. *Mycotoxin Research*, 31(2), 101-107. <https://doi.org/10.1007/s12550-015-0220-z>
- Hakemi Vala, M., Hallajzadeh, M., Hashemi, A., Goudarzi, H., Tarhani, M., Sattarzadeh Tabrizi, M., and Bazmi, F. (2014). Detection of Ambler class A, B and D β -lactamases among *Pseudomonas aeruginosa* and *Acinetobacter baumannii* clinical isolates from burn patients. *Annals of Burns and Fire Disasters*, 27(1), 8-13.
- Hall, B. G., and Barlow, M. (2005). Revised Ambler classification of {beta}-lactamases. *The Journal of Antimicrobial Chemotherapy*, 55(6), 1050-1051. <https://doi.org/10.1093/jac/dki130>
- Hameed, F., Khan, M. A., Muhammad, H., Sarwar, T., Bilal, H., and Rehman, T. U. (2019). Plasmid-mediated mcr-1 gene in *Acinetobacter baumannii* and *Pseudomonas aeruginosa* : First report from Pakistan. *Revista Da Sociedade Brasileira De Medicina Tropical*, 52, e20190237. <https://doi.org/10.1590/0037-8682-0237-2019>

- Hamidian, M., Blasco, L., Tillman, L. N., To, J., Tomas, M., and Myers, G. S. A. (2020). Analysis of Complete Genome Sequence of *Acinetobacter baumannii* Strain ATCC 19606 Reveals Novel Mobile Genetic Elements and Novel Prophage. *Microorganisms*, 8(12), 1851. <https://doi.org/10.3390/microorganisms8121851>
- Hamidian, M., Hawkey, J., Wick, R., Holt, K. E., and Hall, R. M. (2019). Evolution of a clade of *Acinetobacter baumannii* global clone 1, lineage 1 via acquisition of carbapenem- and aminoglycoside-resistance genes and dispersion of ISAbal. *Microbial Genomics*, 5(1), e000242. <https://doi.org/10.1099/mgen.0.000242>
- Hamouda, A., and Amyes, S. G. B. (2004). Novel gyrA and parC point mutations in two strains of *Acinetobacter baumannii* resistant to ciprofloxacin. *Journal of Antimicrobial Chemotherapy*, 54(3), 695-696. <https://doi.org/10.1093/jac/dkh368>
- Hassan, A., Naz, A., Obaid, A., Paracha, R. Z., Naz, K., Awan, F. M., Muhmmad, S. A., Janjua, H. A., Ahmad, J., and Ali, A. (2016). Pangenome and immuno-proteomics analysis of *Acinetobacter baumannii* strains revealed the core peptide vaccine targets. *BMC Genomics*, 17, 732. <https://doi.org/10.1186/s12864-016-2951-4>
- Heydari, F., Mammina, C., and Koksai, F. (2015). NDM-1-producing *Acinetobacter baumannii* ST85 now in Turkey, including one isolate from a Syrian refugee. *Journal of Medical Microbiology*, 64(9), 1027-1029. <https://doi.org/10.1099/jmm.0.000132>
- Hirsch, C. N., Foerster, J. M., Johnson, J. M., Sekhon, R. S., Muttoni, G., Vaillancourt, B., Peñagaricano, F., Lindquist, E., Pedraza, M. A., Barry, K., de Leon, N., Kaeppler, S. M., and Buell, C. R. (2014). Insights into the maize pan-genome and pan-transcriptome. *The Plant Cell*, 26(1), 121-135. <https://doi.org/10.1105/tpc.113.119982>
- Hoff, K. J., and Stanke, M. (2019). Predicting Genes in Single Genomes with AUGUSTUS. *Current Protocols in Bioinformatics*, 65(1), e57. <https://doi.org/10.1002/cpbi.57>
- Hogg, J. S., Hu, F. Z., Janto, B., Boissy, R., Hayes, J., Keefe, R., Post, J. C., and Ehrlich, G. D. (2007). Characterization and modeling of the *Haemophilus influenzae* core and supragenomes based on the

- complete genomic sequences of Rd and 12 clinical nontypeable strains. *Genome Biology*, 8(6), R103. <https://doi.org/10.1186/gb-2007-8-6-r103>
- Hood, M. I., Becker, K. W., Roux, C. M., Dunman, P. M., and Skaar, E. P. (2013). Genetic Determinants of Intrinsic Colistin Tolerance in *Acinetobacter baumannii*. *Infection and Immunity*, 81(2), 542-551. <https://doi.org/10.1128/IAI.00704-12>
- Horton, P., Park, K.-J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J., and Nakai, K. (2007). WoLF PSORT : Protein localization predictor. *Nucleic Acids Research*, 35(Web Server issue), W585-587. <https://doi.org/10.1093/nar/gkm259>
- Howard, A., O'Donoghue, M., Feeney, A., and Sleator, R. D. (2012). *Acinetobacter baumannii*. *Virulence*, 3(3), 243-250. <https://doi.org/10.4161/viru.19700>
- Huang, C.-H., Lee, C.-L., Lin, A. C.-M., Chen, W.-Y., Teng, P.-C., Lee, S.-H., Hsieh, Y.-J., and Jang, T.-N. (2011). Different strains of *Acinetobacter baumannii* spreading in an intensive care unit. *Journal of Acute Medicine*, 1(1), 5-10. <https://doi.org/10.1016/j.jacme.2011.07.003>
- Hueck, C. J. (1998). Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiology and Molecular Biology Reviews: MMBR*, 62(2), 379-433. <https://doi.org/10.1128/MMBR.62.2.379-433.1998>
- Jackman, S. D., Vandervalk, B. P., Mohamadi, H., Chu, J., Yeo, S., Hammond, S. A., Jahesh, G., Khan, H., Coombe, L., Warren, R. L., and Birol, I. (2017). ABySS 2.0 : Resource-efficient assembly of large genomes using a Bloom filter. *Genome Research*, 27(5), 768-777. <https://doi.org/10.1101/gr.214346.116>
- Jacobs, A. C., Hood, I., Boyd, K. L., Olson, P. D., Morrison, J. M., Carson, S., Sayood, K., Iwen, P. C., Skaar, E. P., and Dunman, P. M. (2010). Inactivation of Phospholipase D Diminishes *Acinetobacter baumannii* Pathogenesis. *Infection and Immunity*, 78(5), 1952-1962. <https://doi.org/10.1128/IAI.00889-09>
- Jaidane, N., Naas, T., Oueslati, S., Bernabeu, S., Boujaafar, N., Bouallegue, O., and Bonnin, R. A. (2018). Whole-genome sequencing of NDM-1-producing ST85 *Acinetobacter baumannii* isolates from

- Tunisia. *International Journal of Antimicrobial Agents*, 52(6), 916-921. <https://doi.org/10.1016/j.ijantimicag.2018.05.017>
- Janik, E., Niemcewicz, M., Ceremuga, M., Stela, M., Saluk-Bijak, J., Siadkowski, A., and Bijak, M. (2020). Molecular Aspects of Mycotoxins—A Serious Problem for Human Health. *International Journal of Molecular Sciences*, 21(21), 8187. <https://doi.org/10.3390/ijms21218187>
- Jansen, C., von Wettstein, D., Schäfer, W., Kogel, K.-H., Felk, A., and Maier, F. J. (2005). Infection patterns in barley and wheat spikes inoculated with wild-type and trichodiene synthase gene disrupted *Fusarium graminearum*. *Proceedings of the National Academy of Sciences of the United States of America*, 102(46), 16892-16897. <https://doi.org/10.1073/pnas.0508467102>
- Jenczmionka, N. J., and Schäfer, W. (2005). The Gpmk1 MAP kinase of *Fusarium graminearum* regulates the induction of specific secreted enzymes. *Current Genetics*, 47(1), 29-36. <https://doi.org/10.1007/s00294-004-0547-z>
- Jeong, J., Hwang, I., and Lee, D. W. (2021). Functional Organization of Sequence Motifs in Diverse Transit Peptides of Chloroplast Proteins. *Frontiers in Physiology*, 12. <https://www.frontiersin.org/article/10.3389/fphys.2021.795156>
- Jia, B., Raphenya, A. R., Alcock, B., Waglechner, N., Guo, P., Tsang, K. K., Lago, B. A., Dave, B. M., Pereira, S., Sharma, A. N., Doshi, S., Courtot, M., Lo, R., Williams, L. E., Frye, J. G., Elsayegh, T., Sardar, D., Westman, E. L., Pawlowski, A. C., ... McArthur, A. G. (2017). CARD 2017 : Expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Research*, 45(D1), D566-D573. <https://doi.org/10.1093/nar/gkw1004>
- Jia, L.-J., Tang, H.-Y., Wang, W.-Q., Yuan, T.-L., Wei, W.-Q., Pang, B., Gong, X.-M., Wang, S.-F., Li, Y.-J., Zhang, D., Liu, W., and Tang, W.-H. (2019). A linear nonribosomal octapeptide from *Fusarium graminearum* facilitates cell-to-cell invasion of wheat. *Nature Communications*, 10(1), 922. <https://doi.org/10.1038/s41467-019-08726-9>
- Joly-Guillou, M.-L. (2005). Clinical impact and pathogenicity of *Acinetobacter*. *Clinical Microbiology and Infection*, 11(11), 868-873. <https://doi.org/10.1111/j.1469-0691.2005.01227.x>

- Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R., and Hunter, S. (2014). InterProScan 5 : Genome-scale protein function classification. *Bioinformatics (Oxford, England)*, 30(9), 1236-1240. <https://doi.org/10.1093/bioinformatics/btu031>
- Jouybari, M. A., Ahanjan, M., Mirzaei, B., and Goli, H. R. (2021). Role of aminoglycoside-modifying enzymes and 16S rRNA methylase (ArmA) in resistance of *Acinetobacter baumannii* clinical isolates against aminoglycosides. *Revista Da Sociedade Brasileira De Medicina Tropical*, 54, e05992020. <https://doi.org/10.1590/0037-8682-0599-2020>
- Jovicic, B., Novovic, K., Dekic, S., and Hrenovic, J. (2021). Colistin Resistance in Environmental Isolates of *Acinetobacter baumannii*. *Microbial Drug Resistance (Larchmont, N.Y.)*, 27(3), 328-336. <https://doi.org/10.1089/mdr.2020.0188>
- Juhas, M., van der Meer, J. R., Gaillard, M., Harding, R. M., Hood, D. W., and Crook, D. W. (2009). Genomic islands : Tools of bacterial horizontal gene transfer and evolution. *Fems Microbiology Reviews*, 33(2), 376-393. <https://doi.org/10.1111/j.1574-6976.2008.00136.x>
- Kaase, M., Nordmann, P., Wichelhaus, T. A., Gatermann, S. G., Bonnin, R. A., and Poirel, L. (2011). NDM-2 carbapenemase in *Acinetobacter baumannii* from Egypt. *The Journal of Antimicrobial Chemotherapy*, 66(6), 1260-1262. <https://doi.org/10.1093/jac/dkr135>
- Kale, S. D., Gu, B., Capelluto, D. G. S., Dou, D., Feldman, E., Rumore, A., Arredondo, F. D., Hanlon, R., Fudal, I., Rouxel, T., Lawrence, C. B., Shan, W., and Tyler, B. M. (2010). External lipid PI3P mediates entry of eukaryotic pathogen effectors into plant and animal host cells. *Cell*, 142(2), 284-295. <https://doi.org/10.1016/j.cell.2010.06.008>
- Karah, N., Sundsfjord, A., Towner, K., and Samuelsen, Ø. (2012). Insights into the global molecular epidemiology of carbapenem non-susceptible clones of *Acinetobacter baumannii*. *Drug Resistance Updates: Reviews and Commentaries in Antimicrobial and Anticancer Chemotherapy*, 15(4), 237-247. <https://doi.org/10.1016/j.drug.2012.06.001>
- Karlsson, I., Persson, P., and Friberg, H. (2021). Fusarium Head Blight From a Microbiome Perspective. *Frontiers in Microbiology*, 12. <https://www.frontiersin.org/article/10.3389/fmicb.2021.628373>

- Katoh, K., and Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7 : Improvements in Performance and Usability. *Molecular Biology and Evolution*, 30(4), 772-780. <https://doi.org/10.1093/molbev/mst010>
- Keeling, P. J., and Palmer, J. D. (2008). Horizontal gene transfer in eukaryotic evolution. *Nature Reviews Genetics*, 9(8), 605-618. <https://doi.org/10.1038/nrg2386>
- Kim, Y., Park, Y., Hwang, J., and Kwack, K. (2018). Comparative genomic analysis of the human and nematode *Caenorhabditis elegans* uncovers potential reproductive genes and disease associations in humans. *Physiological Genomics*, 50(11), 1002-1014. <https://doi.org/10.1152/physiol-genomics.00063.2018>
- Kimura, M., Tokai, T., O'Donnell, K., Ward, T. J., Fujimura, M., Hamamoto, H., Shibata, T., and 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 Letters*, 539(1-3), 105-110. [https://doi.org/10.1016/s0014-5793\(03\)00208-4](https://doi.org/10.1016/s0014-5793(03)00208-4)
- King, R., Urban, M., Lauder, R. P., Hawkins, N., Evans, M., Plummer, A., Halsey, K., Lovegrove, A., Hammond-Kosack, K., and Rudd, J. J. (2017a). A conserved fungal glycosyltransferase facilitates pathogenesis of plants by enabling hyphal growth on solid surfaces. *PLoS Pathogens*, 13(10), e1006672. <https://doi.org/10.1371/journal.ppat.1006672>
- King, R., Urban, M., and Hammond-Kosack, K. E. (2017b). Annotation of *Fusarium graminearum* (PH-1) Version 5.0. *Genome Announcements*, 5(2), e01479-16. <https://doi.org/10.1128/genomeA.01479-16>
- King, R., Urban, M., Hammond-Kosack, M. C. U., Hassani-Pak, K., and Hammond-Kosack, K. E. (2015). The completed genome sequence of the pathogenic ascomycete fungus *Fusarium graminearum*. *BMC Genomics*, 16(1), 544. <https://doi.org/10.1186/s12864-015-1756-1>
- Koboldt, D. C., Zhang, Q., Larson, D. E., Shen, D., McLellan, M. D., Lin, L., Miller, C. A., Mardis, E. R., Ding, L., and Wilson, R. K. (2012). VarScan 2 : Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Research*, 22(3), 568-576. <https://doi.org/10.1101/gr.129684.111>

- Krause, K. M., Serio, A. W., Kane, T. R., and Connolly, L. E. (2016). Aminoglycosides : An Overview. *Cold Spring Harbor Perspectives in Medicine*, 6(6), a027029. <https://doi.org/10.1101/cshperspect.a027029>
- Kretschmer, M., Damoo, D., Djamei, A., and Kronstad, J. (2019). Chloroplasts and Plant Immunity : Where Are the Fungal Effectors? *Pathogens (Basel, Switzerland)*, 9(1), E19. <https://doi.org/10.3390/pathogens9010019>
- Krnjaja, V., Stanković, S., Obradović, A., Petrović, T., Mandić, V., Bijelić, Z., and Božić, M. (2018). Trichothecene Genotypes of *Fusarium graminearum* Populations Isolated from Winter Wheat Crops in Serbia. *Toxins*, 10(11), 460. <https://doi.org/10.3390/toxins10110460>
- Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *Journal of Molecular Biology*, 305(3), 567-580. <https://doi.org/10.1006/jmbi.2000.4315>
- Labrador-Herrera, G., Pérez-Pulido, A. J., Álvarez-Marín, R., Casimiro-Soriguer, C. S., Cebrero-Cangueiro, T., Morán-Barrio, J., Pachón, J., Viale, A. M., and Pachón-Ibáñez, M. E. (2020). Virulence role of the outer membrane protein CarO in carbapenem-resistant *Acinetobacter baumannii*. *Virulence*, 11(1), 1727-1737. <https://doi.org/10.1080/21505594.2020.1855912>
- Lapierre, P., and Gogarten, J. P. (2009). Estimating the size of the bacterial pan-genome. *Trends in Genetics: TIG*, 25(3), 107-110. <https://doi.org/10.1016/j.tig.2008.12.004>
- Laslett, D., and Canback, B. (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Research*, 32(1), 11-16. <https://doi.org/10.1093/nar/gkh152>
- Laurent, B., Moinard, M., Spataro, C., Ponts, N., Barreau, C., and Foulongne-Oriol, M. (2017). Landscape of genomic diversity and host adaptation in *Fusarium graminearum*. *BMC Genomics*, 18(1), 203. <https://doi.org/10.1186/s12864-017-3524-x>
- Leal, N. C., Campos, T. L., Rezende, A. M., Docena, C., Mendes-Marques, C. L., de Sá Cavalcanti, F. L., Wallau, G. L., Rocha, I. V., Cavalcanti, C. L. B., Veras, D. L., Alves, L. R., Andrade-Figueiredo, M., de Barros, M. P. S., de Almeida, A. M. P., de Moraes, M. M. C., Leal-Balbino, T. C., Xavier, D. E., and de-Melo-Neto, O. P. (2020). Comparative Genomics of *Acinetobacter baumannii* Clinical Strains From Brazil Reveals Polyclonal Dissemination and Selective Exchange of Mobile Genetic Elements

- Associated With Resistance Genes. *Frontiers in Microbiology*, 11, 1176. <https://doi.org/10.3389/fmicb.2020.01176>
- Lechner, M., Findeiss, S., Steiner, L., Marz, M., Stadler, P. F., and Prohaska, S. J. (2011). Proteinortho : Detection of (co-)orthologs in large-scale analysis. *BMC Bioinformatics*, 12, 124. <https://doi.org/10.1186/1471-2105-12-124>
- Lee, C.-R., Lee, J. H., Park, M., Park, K. S., Bae, I. K., Kim, Y. B., Cha, C.-J., Jeong, B. C., and Lee, S. H. (2017). Biology of *Acinetobacter baumannii* : Pathogenesis, Antibiotic Resistance Mechanisms, and Prospective Treatment Options. *Frontiers in Cellular and Infection Microbiology*, 7. <https://www.frontiersin.org/article/10.3389/fcimb.2017.00055>
- Lee, H.-H., Park, J., Jung, H., and Seo, Y.-S. (2021). Pan-Genome Analysis Reveals Host-Specific Functional Divergences in *Burkholderia gladioli*. *Microorganisms*, 9(6), 1123. <https://doi.org/10.3390/microorganisms9061123>
- Leungtongkam, U., Thummeepak, R., Wongprachan, S., Thongsuk, P., Kittit, T., Ketwong, K., Runchaoren, C., Chantratita, N., and Sitthisak, S. (2018). Dissemination of blaOXA-23, blaOXA-24, blaOXA-58, and blaNDM-1 Genes of *Acinetobacter baumannii* Isolates from Four Tertiary Hospitals in Thailand. *Microbial Drug Resistance*, 24(1), 55-62. <https://doi.org/10.1089/mdr.2016.0248>
- Levasseur, A., Drula, E., Lombard, V., Coutinho, P. M., and Henrissat, B. (2013). Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnology for Biofuels*, 6(1), 41. <https://doi.org/10.1186/1754-6834-6-41>
- Levy-Blitchtein, S., Roca, I., Plasencia-Rebata, S., Vicente-Taboada, W., Velásquez-Pomar, J., Muñoz, L., Moreno-Morales, J., Pons, M. J., Del Valle-Mendoza, J., and Vila, J. (2018). Emergence and spread of carbapenem-resistant *Acinetobacter baumannii* international clones II and III in Lima, Peru. *Emerging Microbes & Infections*, 7(1), 119. <https://doi.org/10.1038/s41426-018-0127-9>
- Li, H. (2018). Minimap2 : Pairwise alignment for nucleotide sequences. *Bioinformatics (Oxford, England)*, 34(18), 3094-3100. <https://doi.org/10.1093/bioinformatics/bty191>
- Li, H., Feng, X., and Chu, C. (2020). The design and construction of reference pangenome graphs with minigraph. *Genome Biology*, 21(1), 265. <https://doi.org/10.1186/s13059-020-02168-z>

- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, 25(16), 2078-2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Li, J., Nation, R. L., Turnidge, J. D., Milne, R. W., Coulthard, K., Rayner, C. R., and Paterson, D. L. (2006). Colistin : The re-emerging antibiotic for multidrug-resistant Gram-negative bacterial infections. *The Lancet. Infectious Diseases*, 6(9), 589-601. [https://doi.org/10.1016/S1473-3099\(06\)70580-1](https://doi.org/10.1016/S1473-3099(06)70580-1)
- Li, R., Li, Y., Zheng, H., Luo, R., Zhu, H., Li, Q., Qian, W., Ren, Y., Tian, G., Li, J., Zhou, G., Zhu, X., Wu, H., Qin, J., Jin, X., Li, D., Cao, H., Hu, X., Blanche, H., ... Wang, J. (2010). Building the sequence map of the human pan-genome. *Nature Biotechnology*, 28(1), 57-63. <https://doi.org/10.1038/nbt.1596>
- Li, Y., Zhou, G., Ma, J., Jiang, W., Jin, L., Zhang, Z., Guo, Y., Zhang, J., Sui, Y., Zheng, L., Zhang, S., Zuo, Q., Shi, X., Li, Y., Zhang, W., Hu, Y., Kong, G., Hong, H., Tan, B., ... Qiu, L. (2014). De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. *Nature Biotechnology*, 32(10), 1045-1052. <https://doi.org/10.1038/nbt.2979>
- Lin, M.-F., and Lan, C.-Y. (2014). Antimicrobial resistance in *Acinetobacter baumannii* : From bench to bedside. *World Journal of Clinical Cases : WJCC*, 2(12), 787-814. <https://doi.org/10.12998/wjcc.v2.i12.787>
- Liu, B., Zheng, D., Jin, Q., Chen, L., and Yang, J. (2019). VFDB 2019 : A comparative pathogenomic platform with an interactive web interface. *Nucleic Acids Research*, 47(Database issue), D687-D692. <https://doi.org/10.1093/nar/gky1080>
- Liu, D., Hunt, M., and Tsai, I. J. (2018). Inferring synteny between genome assemblies : A systematic evaluation. *BMC Bioinformatics*, 19(1), 26. <https://doi.org/10.1186/s12859-018-2026-4>
- Liu, W., Wu, Z., Mao, C., Guo, G., Zeng, Z., Fei, Y., Wan, S., Peng, J., and Wu, J. (2020). Antimicrobial Peptide Cec4 Eradicates the Bacteria of Clinical Carbapenem-Resistant *Acinetobacter baumannii* Biofilm. *Frontiers in Microbiology*, 11, 1532. <https://doi.org/10.3389/fmicb.2020.01532>
- Liu, Y.-Y., Wang, Y., Walsh, T. R., Yi, L.-X., Zhang, R., Spencer, J., Doi, Y., Tian, G., Dong, B., Huang, X., Yu, L.-F., Gu, D., Ren, H., Chen, X., Lv, L., He, D., Zhou, H., Liang, Z., Liu, J.-H., and Shen, J.

- (2016). Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China : A microbiological and molecular biological study. *The Lancet. Infectious Diseases*, 16(2), 161-168. [https://doi.org/10.1016/S1473-3099\(15\)00424-7](https://doi.org/10.1016/S1473-3099(15)00424-7)
- Lo Presti, L., Lanver, D., Schweizer, G., Tanaka, S., Liang, L., Tollot, M., Zuccaro, A., Reissmann, S., and Kahmann, R. (2015). Fungal Effectors and Plant Susceptibility. *Annual Review of Plant Biology*, 66(1), 513-545. <https://doi.org/10.1146/annurev-arplant-043014-114623>
- Lonsdale, A., Davis, M. J., Doblin, M. S., and Bacic, A. (2016). Better Than Nothing? Limitations of the Prediction Tool SecretomeP in the Search for Leaderless Secretory Proteins (LSPs) in Plants. *Frontiers in Plant Science*, 7. <https://www.frontiersin.org/article/10.3389/fpls.2016.01451>
- Ma, F., Shen, C., Zheng, X., Liu, Y., Chen, H., Zhong, L., Liang, Y., Liao, K., Xia, Y., Tian, G.-B., and Yang, Y. (2019). Identification of a Novel Plasmid Carrying mcr-4.3 in an *Acinetobacter baumannii* Strain in China. *Antimicrobial Agents and Chemotherapy*, 63(6), e00133-19. <https://doi.org/10.1128/AAC.00133-19>
- Magiorakos, A.-P., Srinivasan, A., Carey, R. B., Carmeli, Y., Falagas, M. E., Giske, C. G., Harbarth, S., Hindler, J. F., Kahlmeter, G., Olsson-Liljequist, B., Paterson, D. L., Rice, L. B., Stelling, J., Struelens, M. J., Vatopoulos, A., Weber, J. T., and Monnet, D. L. (2012). Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria : An international expert proposal for interim standard definitions for acquired resistance. *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases*, 18(3), 268-281. <https://doi.org/10.1111/j.1469-0691.2011.03570.x>
- Mann, R., Rafei, R., Gunawan, C., Harmer, C. J., and Hamidian, M. (2022). Variants of Tn6924, a Novel Tn7 Family Transposon Carrying the blaNDM Metallo- β -Lactamase and 14 Copies of the aphA6 Amikacin Resistance Genes Found in *Acinetobacter baumannii*. *Microbiology Spectrum*, 10(1), e0174521. <https://doi.org/10.1128/spectrum.01745-21>
- Martins, N., Picão, R. C., Cerqueira-Alves, M., Uehara, A., Barbosa, L. C., Riley, L. W., and Moreira, B. M. (2016). A new trilocus sequence-based multiplex-PCR to detect major *Acinetobacter baumannii* clones. *Infection, Genetics and Evolution*, 42, 41-45. <https://doi.org/10.1016/j.meegid.2016.04.031>

- Matsui, M., Suzuki, M., Suzuki, M., Yatsuyanagi, J., Watahiki, M., Hiraki, Y., Kawano, F., Tsutsui, A., Shibayama, K., and Suzuki, S. (2018). Distribution and Molecular Characterization of *Acinetobacter baumannii* International Clone II Lineage in Japan. *Antimicrobial Agents and Chemotherapy*, 62(2), e02190-17. <https://doi.org/10.1128/AAC.02190-17>
- McCarthy, C. G. P., and Fitzpatrick, D. A. (2019). Pan-genome analyses of model fungal species. *Microbial Genomics*, 5(2), e000243. <https://doi.org/10.1099/mgen.0.000243>
- McConnell, M. J., Actis, L., and Pachón, J. (2013). *Acinetobacter baumannii* : Human infections, factors contributing to pathogenesis and animal models. *FEMS Microbiology Reviews*, 37(2), 130-155. <https://doi.org/10.1111/j.1574-6976.2012.00344.x>
- McDowell, J. M. (2013). Genomic and transcriptomic insights into lifestyle transitions of a hemi-biotrophic fungal pathogen. *New Phytologist*, 197(4), 1032-1034. <https://doi.org/10.1111/nph.12141>
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M. A. (2010). The Genome Analysis Toolkit : A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297-1303. <https://doi.org/10.1101/gr.107524.110>
- McMullen, M., Jones, R., and Gallenberg, D. (1997). Scab of Wheat and Barley : A Re-emerging Disease of Devastating Impact. *Plant Disease*, 81(12), 1340-1348. <https://doi.org/10.1094/PDIS.1997.81.12.1340>
- Medini, D., Donati, C., Tettelin, H., Massignani, V., and Rappuoli, R. (2005). The microbial pan-genome. *Current Opinion in Genetics & Development*, 15(6), 589-594. <https://doi.org/10.1016/j.gde.2005.09.006>
- Menke, J., Dong, Y., and Kistler, H. C. (2012). *Fusarium graminearum* Tri12p influences virulence to wheat and trichothecene accumulation. *Molecular Plant-Microbe Interactions: MPMI*, 25(11), 1408-1418. <https://doi.org/10.1094/MPMI-04-12-0081-R>
- Mesterházy, Á. (2002). Role of Deoxynivalenol in Aggressiveness of *Fusarium graminearum* and *F. culmorum* and in Resistance to Fusarium Head Blight. *European Journal of Plant Pathology*, 108(7), 675-684. <https://doi.org/10.1023/A:1020631114063>

- Moffatt, J. H., Harper, M., Harrison, P., Hale, J. D. F., Vinogradov, E., Seemann, T., Henry, R., Crane, B., St Michael, F., Cox, A. D., Adler, B., Nation, R. L., Li, J., and Boyce, J. D. (2010). Colistin resistance in *Acinetobacter baumannii* is mediated by complete loss of lipopolysaccharide production. *Antimicrobial Agents and Chemotherapy*, 54(12), 4971-4977. <https://doi.org/10.1128/AAC.00834-10>
- Montenegro, J. D., Golicz, A. A., Bayer, P. E., Hurgobin, B., Lee, H., Chan, C.-K. K., Visendi, P., Lai, K., Doležel, J., Batley, J., and Edwards, D. (2017). The pangenome of hexaploid bread wheat. *The Plant Journal: For Cell and Molecular Biology*, 90(5), 1007-1013. <https://doi.org/10.1111/tpj.13515>
- Moosavian, M., Ahmadi, K., Shoja, S., Mardaneh, J., Shahi, F., and Afzali, M. (2020). Antimicrobial resistance patterns and their encoding genes among clinical isolates of *Acinetobacter baumannii* in Ahvaz, Southwest Iran. *MethodsX*, 7, 101031. <https://doi.org/10.1016/j.mex.2020.101031>
- Morgante, M., De Paoli, E., and Radovic, S. (2007). Transposable elements and the plant pan-genomes. *Current Opinion in Plant Biology*, 10(2), 149-155. <https://doi.org/10.1016/j.pbi.2007.02.001>
- Murugan, N., Malathi, J., Umashankar, V., and Madhavan, H. N. (2016). Resistome and pathogenomics of multidrug resistant (MDR) *Pseudomonas aeruginosa* VRFPA03, VRFPA05 recovered from alkaline chemical keratitis and post-operative endophthalmitis patient. *Gene*, 578(1), 105-111. <https://doi.org/10.1016/j.gene.2015.12.022>
- Mussi, M. A., Relling, V. M., Limansky, A. S., and Viale, A. M. (2007). CarO, an *Acinetobacter baumannii* outer membrane protein involved in carbapenem resistance, is essential for L-ornithine uptake. *FEBS Letters*, 581(29), 5573-5578. <https://doi.org/10.1016/j.febslet.2007.10.063>
- Naas, T., Oueslati, S., Bonnin, R. A., Dabos, M. L., Zavala, A., Dortet, L., Retailleau, P., and Iorga, B. I. (2017). Beta-lactamase database (BLDB) – structure and function. *Journal of Enzyme Inhibition and Medicinal Chemistry*, 32(1), 917-919. <https://doi.org/10.1080/14756366.2017.1344235>
- Nawrocki, E. P. (2014). Annotating functional RNAs in genomes using Infernal. *Methods in Molecular Biology (Clifton, N.J.)*, 1097, 163-197. https://doi.org/10.1007/978-1-62703-709-9_9
- Nesic, K., Ivanovic, S., and Nesic, V. (2014). Fusarial toxins : Secondary metabolites of *Fusarium* fungi. *Reviews of Environmental Contamination and Toxicology*, 228, 101-120. https://doi.org/10.1007/978-3-319-01619-1_5

- Niehaus, E.-M., Kleigrew, K., Wiemann, P., Studt, L., Sieber, C. M. K., Connolly, L. R., Freitag, M., Güldener, U., Tudzynski, B., and Humpf, H.-U. (2013). Genetic Manipulation of the *Fusarium fujikuroi* Fusarin Gene Cluster Yields Insight into the Complex Regulation and Fusarin Biosynthetic Pathway. *Chemistry & Biology*, 20(8), 1055-1066. <https://doi.org/10.1016/j.chembiol.2013.07.004>
- Nielsen, H. (2017). Predicting Secretory Proteins with SignalP. *Methods in Molecular Biology (Clifton, N.J.)*, 1611, 59-73. https://doi.org/10.1007/978-1-4939-7015-5_6
- Nielsen, H., Petsalaki, E. I., Zhao, L., and Stühler, K. (2019). Predicting eukaryotic protein secretion without signals. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1867(12), 140174. <https://doi.org/10.1016/j.bbapap.2018.11.011>
- Nouri, R., Ahangarzadeh Rezaee, M., Hasani, A., Aghazadeh, M., and Asgharzadeh, M. (2016). The role of gyrA and parC mutations in fluoroquinolones-resistant *Pseudomonas aeruginosa* isolates from Iran. *Brazilian Journal of Microbiology: [Publication of the Brazilian Society for Microbiology]*, 47(4), 925-930. <https://doi.org/10.1016/j.bjm.2016.07.016>
- Obert, C., Sublett, J., Kaushal, D., Hinojosa, E., Barton, T., Tuomanen, E. I., and Orihuela, C. J. (2006). Identification of a Candidate *Streptococcus pneumoniae* core genome and regions of diversity correlated with invasive pneumococcal disease. *Infection and Immunity*, 74(8), 4766-4777. <https://doi.org/10.1128/IAI.00316-06>
- Okonechnikov, K., Conesa, A., and García-Alcalde, F. (2016). Qualimap 2 : Advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics (Oxford, England)*, 32(2), 292-294. <https://doi.org/10.1093/bioinformatics/btv566>
- Orskov, F., and Orskov, I. (1983). From the national institutes of health. Summary of a workshop on the clone concept in the epidemiology, taxonomy, and evolution of the enterobacteriaceae and other bacteria. *The Journal of Infectious Diseases*, 148(2), 346-357. <https://doi.org/10.1093/infdis/148.2.346>
- Pallen, M. J., and Wren, B. W. (2007). Bacterial pathogenomics. *Nature*, 449(7164), 835-842. <https://doi.org/10.1038/nature06248>

- 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 Pathology*, 58(3), 409-424. <https://doi.org/10.1111/j.1365-3059.2009.02039.x>
- Pérez-Pérez, J. M., Esteve-Bruna, D., González-Bayón, R., Kangasjärvi, S., Caldana, C., Hannah, M. A., Willmitzer, L., Ponce, M. R., and Micol, J. L. (2013). Functional Redundancy and Divergence within the Arabidopsis RETICULATA-RELATED Gene Family. *Plant Physiology*, 162(2), 589-603. <https://doi.org/10.1104/pp.113.217323>
- Peter, J., De Chiara, M., Friedrich, A., Yue, J.-X., Pflieger, D., Bergström, A., Sigwalt, A., Barre, B., Freil, K., Llored, A., Cruaud, C., Labadie, K., Aury, J.-M., Istace, B., Lebrigand, K., Barbry, P., Engelen, S., Lemainque, A., Wincker, P., ... Schacherer, J. (2018). Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature*, 556(7701), 339-344. <https://doi.org/10.1038/s41586-018-0030-5>
- Peterson, J. W. (1996). Bacterial Pathogenesis. In S. Baron (Éd.), *Medical Microbiology* (4th éd.). University of Texas Medical Branch at Galveston. <http://www.ncbi.nlm.nih.gov/books/NBK8526/>
- Petre, B., and Kamoun, S. (2014). How Do Filamentous Pathogens Deliver Effector Proteins into Plant Cells? *PLoS Biology*, 12(2), e1001801. <https://doi.org/10.1371/journal.pbio.1001801>
- Petre, B., Lorrain, C., Saunders, D. G. O., Win, J., Sklenar, J., Duplessis, S., and Kamoun, S. (2016). Rust fungal effectors mimic host transit peptides to translocate into chloroplasts. *Cellular Microbiology*, 18(4), 453-465. <https://doi.org/10.1111/cmi.12530>
- Pierleoni, A., Martelli, P. L., and Casadio, R. (2008). PredGPI : A GPI-anchor predictor. *BMC Bioinformatics*, 9, 392. <https://doi.org/10.1186/1471-2105-9-392>
- Pierron, A., Alassane-Kpembi, I., and Oswald, I. P. (2016). Impact of mycotoxin on immune response and consequences for pig health. *Animal Nutrition (Zhongguo Xu Mu Shou Yi Xue Hui)*, 2(2), 63-68. <https://doi.org/10.1016/j.aninu.2016.03.001>
- Plank, J. V. D., and Van Der Plank, J. (1968). *Disease resistance in plants*. Academic Press.

- Plissonneau, C., Hartmann, F. E., and Croll, D. (2018). Pangenome analyses of the wheat pathogen *Zymoseptoria tritici* reveal the structural basis of a highly plastic eukaryotic genome. *BMC Biology*, 16(1), 5.
Erreur ! Référence de lien hypertexte non valide.
- Pomraning, K.R., Connolly, L.R., Whalen, J.P., Smith, K.M., Freitag, M. (2013). Repeat-Induced Point Mutation, DNA Methylation and Heterochromatin in *Gibberella Zeae* (Anamorph: *Fusarium graminearum*). In *Fusarium Genomics, Molecular and Cellular Biology*; Brown, D., Proctor, R.H., Eds.; Horizon Scientific Press: Norwich, UK, 2013; pp. 93–109.
- Post, V., White, P. A., and Hall, R. M. (2010). Evolution of AbaR-type genomic resistance islands in multiply antibiotic-resistant *Acinetobacter baumannii*. *Journal of Antimicrobial Chemotherapy*, 65(6), 1162-1170. <https://doi.org/10.1093/jac/dkq095>
- Qi, P.-F., Zhang, Y.-Z., Liu, C.-H., Zhu, J., Chen, Q., Guo, Z.-R., Wang, Y., Xu, B.-J., Zheng, T., Jiang, Y.-F., Wang, J.-P., Zhou, C.-Y., Feng, X., Kong, L., Lan, X.-J., Jiang, Q.-T., Wei, Y.-M., and Zheng, Y.-L. (2018). *Fusarium graminearum* ATP-Binding Cassette Transporter Gene FgABCC9 Is Required for Its Transportation of Salicylic Acid, Fungicide Resistance, Mycelial Growth and Pathogenicity towards Wheat. *International Journal of Molecular Sciences*, 19(8), 2351. <https://doi.org/10.3390/ijms19082351>
- Quainoo, S., Coolen, J. P. M., van Hijum, S. A. F. T., Huynen, M. A., Melchers, W. J. G., van Schaik, W., and Wertheim, H. F. L. (2017). Whole-Genome Sequencing of Bacterial Pathogens : The Future of Nosocomial Outbreak Analysis. *Clinical Microbiology Reviews*, 30(4), 1015-1063. <https://doi.org/10.1128/CMR.00016-17>
- Rabouille, C. (2017). Pathways of Unconventional Protein Secretion. *Trends in Cell Biology*, 27(3), 230-240. <https://doi.org/10.1016/j.tcb.2016.11.007>
- Rafei, R., Pailhoriès, H., Hamze, M., Eveillard, M., Mallat, H., Dabboussi, F., Joly-Guillou, M.-L., and Kempf, M. (2015). Molecular epidemiology of *Acinetobacter baumannii* in different hospitals in Tripoli, Lebanon using blaOXA-51-like sequence based typing. *BMC Microbiology*, 15, 103. <https://doi.org/10.1186/s12866-015-0441-5>

- Raffaele, S., and Kamoun, S. (2012). Genome evolution in filamentous plant pathogens : Why bigger can be better. *Nature Reviews. Microbiology*, 10(6), 417-430. <https://doi.org/10.1038/nrmicro2790>
- Rampersad, S. N. (2020). Pathogenomics and Management of *Fusarium* Diseases in Plants. *Pathogens*, 9(5), 340. <https://doi.org/10.3390/pathogens9050340>
- Ramzi, A. B., Me, M. L. C., Ruslan, U. S., Baharum, S. N., and Muhammad, N. A. N. (2019). Insight into plant cell wall degradation and pathogenesis of *Ganoderma boninense* via comparative genome analysis. *PeerJ*, 7, e8065. <https://doi.org/10.7717/peerj.8065>
- Rasko, D. A., Rosovitz, M. J., Myers, G. S. A., Mongodin, E. F., Fricke, W. F., Gajer, P., Crabtree, J., Sebaihia, M., Thomson, N. R., Chaudhuri, R., Henderson, I. R., Sperandio, V., and Ravel, J. (2008). The pangenome structure of *Escherichia coli* : Comparative genomic analysis of *E. coli* commensal and pathogenic isolates. *Journal of Bacteriology*, 190(20), 6881-6893. <https://doi.org/10.1128/JB.00619-08>
- Rauwane, M. E., Ogugua, U. V., Kalu, C. M., Ledwaba, L. K., Woldesemayat, A. A., and Ntushelo, K. (2020). Pathogenicity and Virulence Factors of *Fusarium graminearum* Including Factors Discovered Using Next Generation Sequencing Technologies and Proteomics. *Microorganisms*, 8(2), 305. <https://doi.org/10.3390/microorganisms8020305>
- Rawlings, N. D., Barrett, A. J., and Finn, R. (2016). Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Research*, 44(D1), D343-350. <https://doi.org/10.1093/nar/gkv1118>
- Rawlings, N. D., Barrett, A. J., Thomas, P. D., Huang, X., Bateman, A., and Finn, R. D. (2018). The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Research*, 46(D1), D624-D632. <https://doi.org/10.1093/nar/gkx1134>
- Read, A. F. (1994). The evolution of virulence. *Trends in Microbiology*, 2(3), 73-76. [https://doi.org/10.1016/0966-842x\(94\)90537-1](https://doi.org/10.1016/0966-842x(94)90537-1)
- Read, B. A., Kegel, J., Klute, M. J., Kuo, A., Lefebvre, S. C., Maumus, F., Mayer, C., Miller, J., Monier, A., Salamov, A., Young, J., Aguilar, M., Claverie, J.-M., Frickenhaus, S., Gonzalez, K., Herman, E. K., Lin, Y.-C., Napier, J., Ogata, H., ... Grigoriev, I. V. (2013). Pan genome of the phytoplankton

Emiliania underpins its global distribution. *Nature*, 499(7457), 209-213. <https://doi.org/10.1038/nature12221>

Rech, G. E., Sanz-Martín, J. M., Anisimova, M., Sukno, S. A., and Thon, M. R. (2014). Natural selection on coding and noncoding DNA sequences is associated with virulence genes in a plant pathogenic fungus. *Genome Biology and Evolution*, 6(9), 2368-2379. <https://doi.org/10.1093/gbe/evu192>

Ren, J., Zhang, Y., Wang, Y., Li, C., Bian, Z., Zhang, X., Liu, H., Xu, J.-R., and Jiang, C. (2022). Deletion of all three MAP kinase genes results in severe defects in stress responses and pathogenesis in *Fusarium graminearum*. *Stress Biology*, 2(1), 6. <https://doi.org/10.1007/s44154-021-00025-y>

Rooney, H. C. E., Van't Klooster, J. W., van der Hoorn, R. A. L., Joosten, M. H. A. J., Jones, J. D. G., and de Wit, P. J. G. M. (2005). Cladosporium Avr2 inhibits tomato Rcr3 protease required for Cf-2-dependent disease resistance. *Science (New York, N.Y.)*, 308(5729), 1783-1786. <https://doi.org/10.1126/science.1111404>

Rouli, L., Merhej, V., Fournier, P.-E., and Raoult, D. (2015). The bacterial pangenome as a new tool for analysing pathogenic bacteria. *New Microbes and New Infections*, 7, 72-85. <https://doi.org/10.1016/j.nmni.2015.06.005>

Roy, D., Auger, J.-P., Segura, M., Fittipaldi, N., Takamatsu, D., Okura, M., and Gottschalk, M. (2015). Role of the capsular polysaccharide as a virulence factor for *Streptococcus suis* serotype 14. *Canadian Journal of Veterinary Research*, 79(2), 141-146.

Rumbo, C., Gato, E., López, M., Ruiz de Alegría, C., Fernández-Cuenca, F., Martínez-Martínez, L., Vila, J., Pachón, J., Cisneros, J. M., Rodríguez-Baño, J., Pascual, A., Bou, G., and Tomás, M. (2013). Contribution of Efflux Pumps, Porins, and β -Lactamases to Multidrug Resistance in Clinical Isolates of *Acinetobacter baumannii*. *Antimicrobial Agents and Chemotherapy*, 57(11), 5247-5257. <https://doi.org/10.1128/AAC.00730-13>

Sabotič, J., and Kos, J. (2012). Microbial and fungal protease inhibitors—Current and potential applications. *Applied Microbiology and Biotechnology*, 93(4), 1351-1375. <https://doi.org/10.1007/s00253-011-3834-x>

- Sacristán, S., and García-Arenal, F. (2008). The evolution of virulence and pathogenicity in plant pathogen populations. *Molecular Plant Pathology*, 9(3), 369-384. <https://doi.org/10.1111/j.1364-3703.2007.00460.x>
- Saha, S. B., Uttam, V., and Verma, V. (2015). u-CARE : User-friendly Comprehensive Antibiotic resistance Repository of *Escherichia coli*. *Journal of Clinical Pathology*, 68(8), 648-651. <https://doi.org/10.1136/jclinpath-2015-202927>
- Salloum, T., Tannous, E., Alousi, S., Arabaghian, H., Rafei, R., Hamze, M., and Tokajian, S. (2018). Genomic mapping of ST85 blaNDM-1 and blaOXA-94 producing *Acinetobacter baumannii* isolates from Syrian Civil War Victims. *International Journal of Infectious Diseases: IJID: Official Publication of the International Society for Infectious Diseases*, 74, 100-108. <https://doi.org/10.1016/j.ijid.2018.07.017>
- Santajit, S., and Indrawattana, N. (2016). Mechanisms of Antimicrobial Resistance in ESKAPE Pathogens. *BioMed Research International*, 2016, 2475067. <https://doi.org/10.1155/2016/2475067>
- Sechi, L. A., Karadenizli, A., Deriu, A., Zanetti, S., Kolayli, F., Balikci, E., and Vahaboglu, H. (2004). PER-1 type beta-lactamase production in *Acinetobacter baumannii* is related to cell adhesion. *Medical Science Monitor: International Medical Journal of Experimental and Clinical Research*, 10(6), BR180-184.
- Seemann, T. (2014). Prokka : Rapid prokaryotic genome annotation. *Bioinformatics (Oxford, England)*, 30(14), 2068-2069. <https://doi.org/10.1093/bioinformatics/btu153>
- Selin, C., de Kievit, T. R., Belmonte, M. F., and Fernando, W. G. D. (2016). Elucidating the Role of Effectors in Plant-Fungal Interactions : Progress and Challenges. *Frontiers in Microbiology*, 7. <https://www.frontiersin.org/article/10.3389/fmicb.2016.00600>
- Seong, K.-Y., Pasquali, M., Zhou, X., Song, J., Hilburn, K., McCormick, S., Dong, Y., Xu, J.-R., and Kistler, H. C. (2009). Global gene regulation by *Fusarium* transcription factors Tri6 and Tri10 reveals adaptations for toxin biosynthesis. *Molecular Microbiology*, 72(2), 354-367. <https://doi.org/10.1111/j.1365-2958.2009.06649.x>

- Šeputienė, V., Povilonis, J., and Sužiedėlienė, E. (2012). Novel Variants of AbaR Resistance Islands with a Common Backbone in *Acinetobacter baumannii* Isolates of European Clone II. *Antimicrobial Agents and Chemotherapy*, 56(4), 1969-1973. <https://doi.org/10.1128/AAC.05678-11>
- Sherman, R. M., Forman, J., Antonescu, V., Puiu, D., Daya, M., Rafaels, N., Boorgula, M. P., Chavan, S., Vergara, C., Ortega, V. E., Levin, A. M., Eng, C., Yazdanbakhsh, M., Wilson, J. G., Marrugo, J., Lange, L. A., Williams, L. K., Watson, H., Ware, L. B., ... Salzberg, S. L. (2019). Assembly of a pan-genome from deep sequencing of 910 humans of African descent. *Nature Genetics*, 51(1), 30-35. <https://doi.org/10.1038/s41588-018-0273-y>
- Sherman, R. M., and Salzberg, S. L. (2020). Pan-genomics in the human genome era. *Nature Reviews. Genetics*, 21(4), 243-254. <https://doi.org/10.1038/s41576-020-0210-7>
- Sieber, C. M. K., Lee, W., Wong, P., Münsterkötter, M., Mewes, H.-W., Schmeitzl, C., Varga, E., Berthiller, F., Adam, G., and Güldener, U. (2014). The *Fusarium graminearum* genome reveals more secondary metabolite gene clusters and hints of horizontal gene transfer. *PloS One*, 9(10), e110311. <https://doi.org/10.1371/journal.pone.0110311>
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO : Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics (Oxford, England)*, 31(19), 3210-3212. <https://doi.org/10.1093/bioinformatics/btv351>
- Sinha, D., Gupta, M. K., Patel, H. K., Ranjan, A., and Sonti, R. V. (2013). Cell Wall Degrading Enzyme Induced Rice Innate Immune Responses Are Suppressed by the Type 3 Secretion System Effectors XopN, XopQ, XopX and XopZ of *Xanthomonas oryzae* pv. *Oryzae*. *PLoS ONE*, 8(9), e75867. <https://doi.org/10.1371/journal.pone.0075867>
- Skerniškytė, J., Krasauskas, R., Péchoux, C., Kulakauskas, S., Armalytė, J., and Sužiedėlienė, E. (2019). Surface-Related Features and Virulence Among *Acinetobacter baumannii* Clinical Isolates Belonging to International Clones I and II. *Frontiers in Microbiology*, 9, 3116. <https://doi.org/10.3389/fmicb.2018.03116>
- Snipen, L., Almøy, T., and Ussery, D. W. (2009). Microbial comparative pan-genomics using binomial mixture models. *BMC Genomics*, 10, 385. <https://doi.org/10.1186/1471-2164-10-385>

- Soanes, D. M., Alam, I., Cornell, M., Wong, H. M., Hedeler, C., Paton, N. W., Rattray, M., Hubbard, S. J., Oliver, S. G., and Talbot, N. J. (2008). Comparative Genome Analysis of Filamentous Fungi Reveals Gene Family Expansions Associated with Fungal Pathogenesis. *PLoS ONE*, 3(6), e2300. <https://doi.org/10.1371/journal.pone.0002300>
- Sobrova, P., Adam, V., Vasatkova, A., Beklova, M., Zeman, L., and Kizek, R. (2010). Deoxynivalenol and its toxicity. *Interdisciplinary Toxicology*, 3(3), 94-99. <https://doi.org/10.2478/v10102-010-0019-x>
- Sperschneider, J., Catanzariti, A.-M., DeBoer, K., Petre, B., Gardiner, D. M., Singh, K. B., Dodds, P. N., and Taylor, J. M. (2017). LOCALIZER : Subcellular localization prediction of both plant and effector proteins in the plant cell. *Scientific Reports*, 7, 44598. <https://doi.org/10.1038/srep44598>
- Sperschneider, J., Dodds, P. N., Gardiner, D. M., Singh, K. B., and Taylor, J. M. (2018). Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0. *Molecular Plant Pathology*, 19(9), 2094-2110. <https://doi.org/10.1111/mpp.12682>
- Stamatakis, A. (2014). RAxML version 8 : A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9), 1312-1313. <https://doi.org/10.1093/bioinformatics/btu033>
- Syme, R. A., Tan, K.-C., Rybak, K., Friesen, T. L., McDonald, B. A., Oliver, R. P., and Hane, J. K. (2018). *Pan-Parastagonospora* Comparative Genome Analysis-Effector Prediction and Genome Evolution. *Genome Biology and Evolution*, 10(9), 2443-2457. <https://doi.org/10.1093/gbe/evy192>
- Tahbaz, S. V., Azimi, L., and Lari, A. R. (2019). Characterization of aminoglycoside resistance mechanisms in *Acinetobacter Baumannii* isolates from burn wound colonization. *Annals of Burns and Fire Disasters*, 32(2), 115-121.
- Talavera, G., and Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology*, 56(4), 564-577. <https://doi.org/10.1080/10635150701472164>
- Tan, K.-C., Phan, H. T. T., Rybak, K., John, E., Chooi, Y. H., Solomon, P. S., and Oliver, R. P. (2015). Functional redundancy of necrotrophic effectors – consequences for exploitation for breeding. *Frontiers in Plant Science*, 6, 501. <https://doi.org/10.3389/fpls.2015.00501>

- Tempel, S. (2012). Using and understanding RepeatMasker. *Methods in Molecular Biology (Clifton, N.J.)*, 859, 29-51. https://doi.org/10.1007/978-1-61779-603-6_2
- Tettelin, H., Maignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., Angiuoli, S. V., Crabtree, J., Jones, A. L., Durkin, A. S., Deboy, R. T., Davidsen, T. M., Mora, M., Scarselli, M., Margarit y Ros, I., Peterson, J. D., Hauser, C. R., Sundaram, J. P., Nelson, W. C., ... Fraser, C. M. (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae* : Implications for the microbial « pan-genome ». *Proceedings of the National Academy of Sciences of the United States of America*, 102(39), 13950-13955. <https://doi.org/10.1073/pnas.0506758102>
- Tettelin, H., and Medini, D. (Éds.). (2020). *The Pangenome : Diversity, Dynamics and Evolution of Genomes*. Springer. <http://www.ncbi.nlm.nih.gov/books/NBK558825/>
- Tiku, V., Kofoed, E. M., Yan, D., Kang, J., Xu, M., Reichelt, M., Dikic, I., and Tan, M.-W. (2021). Outer membrane vesicles containing OmpA induce mitochondrial fragmentation to promote pathogenesis of *Acinetobacter baumannii*. *Scientific Reports*, 11, 618. <https://doi.org/10.1038/s41598-020-79966-9>
- Toruño, T. Y., Stergiopoulos, I., and Coaker, G. (2016). Plant-Pathogen Effectors : Cellular Probes Interfering with Plant Defenses in Spatial and Temporal Manners. *Annual review of phytopathology*, 54, 419-441. <https://doi.org/10.1146/annurev-phyto-080615-100204>
- Tralamazza, S. M., Braghini, R., and Corrêa, B. (2016). Trichothecene Genotypes of the *Fusarium graminearum* Species Complex Isolated from Brazilian Wheat Grains by Conventional and Quantitative PCR. *Frontiers in Microbiology*, 7, 246. <https://doi.org/10.3389/fmicb.2016.00246>
- Tranchant-Dubreuil, C., Rouard, M., and Sabot, F. (2019). Plant Pangenome : Impacts on Phenotypes and Evolution. In J. A. Roberts (Éd.), *Annual Plant Reviews online* (1^{re} éd., p. 453-478). Wiley. <https://doi.org/10.1002/9781119312994.apr0664>
- Tretiakova, P., Voegelé, R. T., Soloviev, A., and Link, T. I. (2022). Successful Silencing of the Mycotoxin Synthesis Gene TRI5 in *Fusarium culmorum* and Observation of Reduced Virulence in VIGS and SIGS Experiments. *Genes*, 13(3), 395. <https://doi.org/10.3390/genes13030395>
- Urban, M., Cuzick, A., Seager, J., Wood, V., Rutherford, K., Venkatesh, S. Y., De Silva, N., Martinez, M. C., Pedro, H., Yates, A. D., Hassani-Pak, K., and Hammond-Kosack, K. E. (2020). PHI-base : The

pathogen–host interactions database. *Nucleic Acids Research*, 48(D1), D613-D620.
<https://doi.org/10.1093/nar/gkz904>

Urhan, A., and Abeel, T. 2021. (s. d.). A comparative study of pan-genome methods for microbial organisms : *Acinetobacter baumannii* pan-genome reveals structural variation in antimicrobial resistance-carrying plasmids. *Microbial Genomics*, 7(11), 000690. <https://doi.org/10.1099/mgen.0.000690>

Valiente-Mullor, C., Beamud, B., Ansari, I., Francés-Cuesta, C., García-González, N., Mejía, L., Ruiz-Hueso, P., and González-Candelas, F. (2021). One is not enough : On the effects of reference genome for the mapping and subsequent analyses of short-reads. *PLOS Computational Biology*, 17(1), e1008678. <https://doi.org/10.1371/journal.pcbi.1008678>

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K. V., Altshuler, D., Gabriel, S., and DePristo, M. A. (2013). From FastQ data to high confidence variant calls : The Genome Analysis Toolkit best practices pipeline. *Current Protocols in Bioinformatics*, 43, 11.10.1-11.10.33. <https://doi.org/10.1002/0471250953.bi1110s43>

Van Etten, J., and Bhattacharya, D. (2020). Horizontal Gene Transfer in Eukaryotes : Not if, but How Much? *Trends in Genetics: TIG*, 36(12), 915-925. <https://doi.org/10.1016/j.tig.2020.08.006>

van Esse, H. P., van't Klooster, J. W., Bolton, M. D., Yadeta, K. A., van Baarlen, P., Boeren, S., Vervoort, J., de Wit, P. J. G. M., and Thomma, B. P. H. J. (2008). The *Cladosporium fulvum* Virulence Protein Avr2 Inhibits Host Proteases Required for Basal Defense. *The Plant Cell*, 20(7), 1948-1963. <https://doi.org/10.1105/tpc.108.059394>

Varga, E., Wiesenberger, G., Hametner, C., Ward, T. J., Dong, Y., Schöfbeck, D., McCormick, S., Broz, K., Stückler, R., Schuhmacher, R., Krska, R., Kistler, H. C., Berthiller, F., and Adam, G. (2015). New tricks of an old enemy : Isolates of *Fusarium graminearum* produce a type A trichothecene mycotoxin. *Environmental Microbiology*, 17(8), 2588-2600. <https://doi.org/10.1111/1462-2920.12718>

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J.

- R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., ... Zhu, X. (2001). The sequence of the human genome. *Science (New York, N.Y.)*, 291(5507), 1304-1351. <https://doi.org/10.1126/science.1058040>
- Viehman, J. A., Nguyen, M. H., and Doi, Y. (2014). Treatment Options for Carbapenem-Resistant and Extensively Drug-Resistant *Acinetobacter baumannii* Infections. *Drugs*, 74(12), 1315-1333. <https://doi.org/10.1007/s40265-014-0267-8>
- Vijayakumar, S., Rajenderan, S., Laishram, S., Anandan, S., Balaji, V., and Biswas, I. (2016). Biofilm Formation and Motility Depend on the Nature of the *Acinetobacter baumannii* Clinical Isolates. *Frontiers in Public Health*, 4, 105. <https://doi.org/10.3389/fpubh.2016.00105>
- Villacís, J. E., Bovera, M., Romero-Alvarez, D., Cornejo, F., Albán, V., Trueba, G., Dorn, H. F., and Reyes, J. A. (2019). NDM-1 carbapenemase in *Acinetobacter baumannii* sequence type 32 in Ecuador. *New Microbes and New Infections*, 29, 100526. <https://doi.org/10.1016/j.nmni.2019.100526>
- Voigt, C. A., Schäfer, W., and Salomon, S. (2005). A secreted lipase of *Fusarium graminearum* is a virulence factor required for infection of cereals. *The Plant Journal: For Cell and Molecular Biology*, 42(3), 364-375. <https://doi.org/10.1111/j.1365-313X.2005.02377.x>
- Walkowiak, S., Rowland, O., Rodrigue, N., and Subramaniam, R. (2016). Whole genome sequencing and comparative genomics of closely related Fusarium Head Blight fungi : *Fusarium graminearum*, *F. meridionale* and *F. asiaticum*. *BMC Genomics*, 17(1), 1014. <https://doi.org/10.1186/s12864-016-3371-1>
- Wang, Q., Jiang, C., Wang, C., Chen, C., Xu, J.-R., and Liu, H. (2017). Characterization of the Two-Speed Subgenomes of *Fusarium graminearum* Reveals the Fast-Speed Subgenome Specialized for Adaption and Infection. *Frontiers in Plant Science*, 8, 140. <https://doi.org/10.3389/fpls.2017.00140>
- Wang, Y., Liu, W., Hou, Z., Wang, C., Zhou, X., Jonkers, W., Ding, S., Kistler, H. C., and Xu, J.-R. (2011). A novel transcriptional factor important for pathogenesis and ascosporeogenesis in *Fusarium graminearum*. *Molecular Plant-Microbe Interactions: MPMI*, 24(1), 118-128. <https://doi.org/10.1094/MPMI-06-10-0129>

- Wang-Lin, S. X., Olson, R., Beanan, J. M., MacDonald, U., Balthasar, J. P., and Russo, T. A. (2017). The Capsular Polysaccharide of *Acinetobacter baumannii* Is an Obstacle for Therapeutic Passive Immunization Strategies. *Infection and Immunity*, 85(12), e00591-17. <https://doi.org/10.1128/IAI.00591-17>
- Wawra, S., Belmonte, R., Löbach, L., Saraiva, M., Willems, A., and van West, P. (2012). Secretion, delivery and function of oomycete effector proteins. *Current Opinion in Microbiology*, 15(6), 685-691. <https://doi.org/10.1016/j.mib.2012.10.008>
- Weber, B. S., Harding, C. M., and Feldman, M. F. (2016). Pathogenic *Acinetobacter* : From the Cell Surface to Infinity and Beyond. *Journal of Bacteriology*, 198(6), 880-887. <https://doi.org/10.1128/JB.00906-15>
- 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 Research*, 39(Database issue), D637-D639. <https://doi.org/10.1093/nar/gkq1016>
- Wongsuk, T., Boonsilp, S., Homkaew, A., Thananon, K., and Oonant, W. (2022). Whole genome sequence of pan drug-resistant clinical isolate of *Acinetobacter baumannii* ST1890. *PLOS ONE*, 17(3), e0264374. <https://doi.org/10.1371/journal.pone.0264374>
- World Health Organization. Global Priority List of Antibiotic-Resistant Bacteria to Guide Research, Discovery, and Development of New Antibiotics (2017). Available online: <https://apps.who.int/iris/bitstream/handle/10665/311820/WHO-EMP-IAU-2017.12-eng.pdf> (accessed on 14 November 2021).
- Wu, T. D., Reeder, J., Lawrence, M., Becker, G., and Brauer, M. J. (2016). GMAP and GSNAP for Genomic Sequence Alignment : Enhancements to Speed, Accuracy, and Functionality. *Methods in Molecular Biology (Clifton, N.J.)*, 1418, 283-334. https://doi.org/10.1007/978-1-4939-3578-9_15
- Wyres, K. L., Cahill, S. M., Holt, K. E., Hall, R. M., and Kenyon, J. J. (2020). Identification of *Acinetobacter baumannii* loci for capsular polysaccharide (KL) and lipooligosaccharide outer core (OCL) synthesis in genome assemblies using curated reference databases compatible with Kaptive. *Microbial Genomics*, 6(3). <https://doi.org/10.1099/mgen.0.000339>

- Xu, Q., Tang, C., Wang, X., Sun, S., Zhao, J., Kang, Z., and Wang, X. (2019). An effector protein of the wheat stripe rust fungus targets chloroplasts and suppresses chloroplast function. *Nature Communications*, 10, 5571. <https://doi.org/10.1038/s41467-019-13487-6>
- Xu, X., and Nicholson, P. (2009). Community ecology of fungal pathogens causing wheat head blight. *Annual Review of Phytopathology*, 47, 83-103. <https://doi.org/10.1146/annurev-phyto-080508-081737>
- Yakkala, H., Samantarrai, D., Gribskov, M., and Siddavattam, D. (2019). Comparative genome analysis reveals niche-specific genome expansion in *Acinetobacter baumannii* strains. *PLoS ONE*, 14(6), e0218204. <https://doi.org/10.1371/journal.pone.0218204>
- Yang, F., Xiao, K., Pan, H., and Liu, J. (2021). Chloroplast : The Emerging Battlefield in Plant–Microbe Interactions. *Frontiers in Plant Science*, 12. <https://www.frontiersin.org/article/10.3389/fpls.2021.637853>
- Yang, H., Hu, L., Liu, Y., Ye, Y., and Li, J. (2016). Detection of the plasmid-mediated quinolone resistance determinants in clinical isolates of *Acinetobacter baumannii* in China. *Journal of Chemotherapy (Florence, Italy)*, 28(5), 443-445. <https://doi.org/10.1179/1973947815Y.00000000017>
- Yao, W., Li, G., Zhao, H., Wang, G., Lian, X., and Xie, W. (2015). Exploring the rice dispensable genome using a metagenome-like assembly strategy. *Genome Biology*, 16(1), 187. <https://doi.org/10.1186/s13059-015-0757-3>
- Ye, Y., Ling, N., Gao, J., Zhang, X., Zhang, M., Tong, L., Zeng, H., Zhang, J., and Wu, Q. (2018). Roles of outer membrane protein W (OmpW) on survival, morphology, and biofilm formation under NaCl stresses in *Cronobacter sakazakii*. *Journal of Dairy Science*, 101(5), 3844-3850. <https://doi.org/10.3168/jds.2017-13791>
- Yoon, E.-J., Courvalin, P., and Grillot-Courvalin, C. (2013). RND-type efflux pumps in multidrug-resistant clinical isolates of *Acinetobacter baumannii* : Major role for AdeABC overexpression and AdeRS mutations. *Antimicrobial Agents and Chemotherapy*, 57(7), 2989-2995. <https://doi.org/10.1128/AAC.02556-12>

- Yörük, E., Karlik, E., Gazdagli, A., Kayis, M., Kaya, F., and Albayrak, G. (2015). Expression Analysis of PKS13, FG08079.1 and PKS10 Genes in *Fusarium graminearum* and *Fusarium culmorum*. *Iranian Journal of Biotechnology*, 13(2), 51-55. <https://doi.org/10.15171/ijb.1035>
- Zain, M. E. (2011). Impact of mycotoxins on humans and animals. *Journal of Saudi Chemical Society*, 15(2), 129-144. <https://doi.org/10.1016/j.jscs.2010.06.006>
- Zarrilli, R., Pournaras, S., Giannouli, M., and Tsakris, A. (2013). Global evolution of multidrug-resistant *Acinetobacter baumannii* clonal lineages. *International Journal of Antimicrobial Agents*, 41(1), 11-19. <https://doi.org/10.1016/j.ijantimicag.2012.09.008>
- Zhang, X., Gao, J., Ling, N., Zeng, H., Tong, L., Zhang, M., Zhang, J., Wu, Q., and Ye, Y. (2019). Short communication : Roles of outer membrane protein W on survival, cellular morphology, and biofilm formation of *Cronobacter sakazakii* in response to oxidative stress. *Journal of Dairy Science*, 102(3), 2017-2021. <https://doi.org/10.3168/jds.2018-14643>
- Zhang, X., Nguyen, N., Breen, S., Outram, M. A., Dodds, P. N., Kobe, B., Solomon, P. S., and Williams, S. J. (2016). Production of small cysteine-rich effector proteins in *Escherichia coli* for structural and functional studies. *Molecular Plant Pathology*, 18(1), 141-151. <https://doi.org/10.1111/mpp.12385>
- Zhao, Y., Hu, K., Zhang, J., Guo, Y., Fan, X., Wang, Y., Mensah, S. D., and Zhang, X. (2019). Outbreak of carbapenem-resistant *Acinetobacter baumannii* carrying the carbapenemase OXA-23 in ICU of the eastern Heilongjiang Province, China. *BMC Infectious Diseases*, 19(1), 452. <https://doi.org/10.1186/s12879-019-4073-5>
- Zhao, Z., Liu, H., Wang, C., and Xu, J.-R. (2013). Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. *BMC Genomics*, 14(1), 274. <https://doi.org/10.1186/1471-2164-14-274>
- Zhou, H., Yao, Y., Zhu, B., Ren, D., Yang, Q., Fu, Y., Yu, Y., and Zhou, J. (2019). Risk factors for acquisition and mortality of multidrug-resistant *Acinetobacter baumannii* bacteremia. *Medicine*, 98(13), e14937. <https://doi.org/10.1097/MD.00000000000014937>.

Annexes

Liste des Figures associées aux annexes

Figure A-1 : Développement des technologies de séquençage.....	161
Figure A-2 : Le principe du processus de séquençage Illumina.....	163
Figure A-3 : Workflow de préparation de la librairie d'ADN via Nextera XT.....	166
Figure A-4 : Rapport de la qualité globale obtenue après le séquençage Illumina MiSeq.....	167
Figure A-5 : Format du fichier FastQ	168
Figure A-6 : Les deux approches de l'assemblage <i>de novo</i> du génome	169
Figure A-7 : La mesure de N50 des contigs assemblés	170

Annexe A – Technologies de séquençage actuelles

Dans cette section, nous donnons un aperçu des technologies actuelles de séquençage de l'ADN en mettant l'accent sur les technologies de séquençage utilisées dans les travaux de cette Thèse: la technologie de deuxième génération (plateforme Illumina) et de troisième génération (plateforme PacBio).

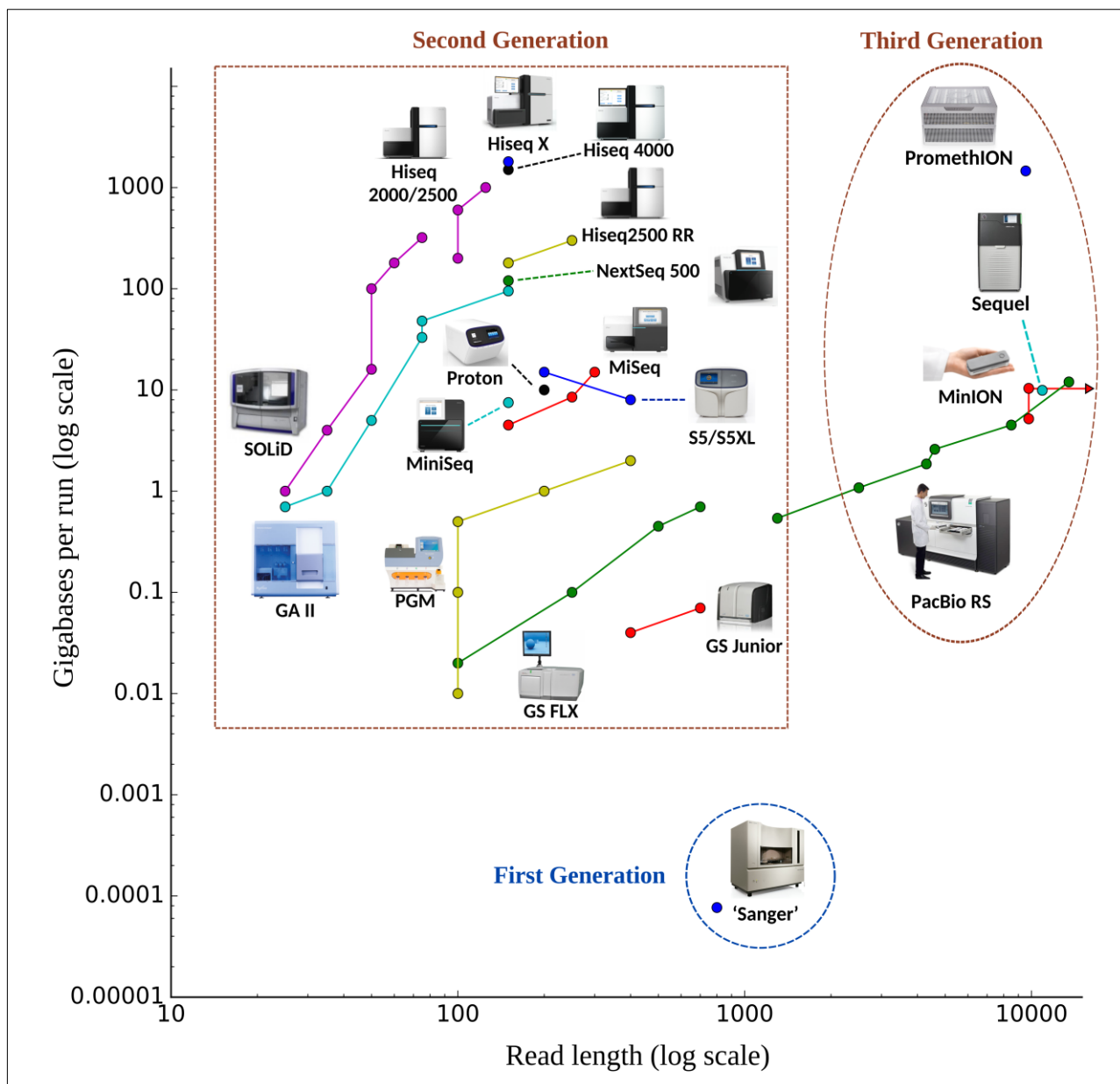


Figure A-1 : Développement des technologies de séquençage. Débit en bases en fonction de la longueur de *read* pour chaque technologie ; de la première génération au séquençage de troisième génération. Chaque point de données connecté représente un développement ultérieur en chimie ou en matériau. Adapté de Nederbragt et al. (2021).

A.1. Séquençage de 1ère génération

Le séquençage de l'ADN a été inventé dans la seconde moitié des années 1970, avec deux méthodes, celle développée par l'équipe de Frederick Sanger en Grande-Bretagne qui est basée sur la synthèse enzymatique (Sanger *et al.*, 1977) et celle développée par l'équipe de Walter Gilbert aux États-Unis, qui est basé sur la dégradation chimique (Maxam *et al.*, 1977). Ces deux méthodes sont considérées comme des séquenceurs de première génération avec un débit plutôt faible (maximum de dizaines à centaines de Kb par exécution). En raison de l'utilisation de produits chimiques toxiques, la méthode Maxam-Gilbert a été abandonnée et le séquençage Sanger est devenu la norme pour le séquençage de l'ADN. En 1977, Sanger a séquencé le premier génome, le bactériophage Φ X 174 avec une longueur de 5375 bases. Ces méthodes et leur histoire ont été discutées en détail précédemment (França *et al.*, 2002). Cette méthode est toujours utilisée et désignée pour être utilisée comme une méthode de référence pour confirmer les résultats des technologies NGS, en particulier pour la validation des variants génétiques en raison de leur grande précision et sensibilité.

A.2. Séquençage de 2ème génération

En 2005, les technologies NGS sont apparues sur le marché et ont été développées pour réduire le temps, le coût du séquençage, et surtout pour augmenter le débit. Ces améliorations ont été rendues possibles autant par les progrès de la biologie moléculaire que par ceux de la bioinformatique. Le premier séquenceur était le 454 Genome Sequencer GS20, qui utilisait le principe de la PCR en émulsion avec pyroséquençage (Heather *et Chain*, 2016). La société 454 a ensuite été reprise par Roche et la machine a été améliorée à plusieurs reprises. Une plate-forme similaire en termes de procédure d'amplification PCR a été développée par Ion Torrent (maintenant ThermoFisher Scientific). Les machines Ion Torrent (par exemple PGM, Proton) agissent comme un pH-mètre en détectant la libération de H⁺ lors de l'incorporation d'une base. Il y avait d'autres techniques de deuxième génération disponibles sur le marché (par exemple SOLiD d'ABI), mais elles ne sont pas aussi populaires par rapport Illumina. Cette technologie d'Illumina est actuellement la plus performante et la plus utilisée (anciennement Solexa) avec sa méthode de séquençage synthétique avec des terminateurs nucléotidiques réversibles marqués (Slatko *et al.*, 2018). Illumina a lancé sa première machine en 2006 et propose la plus large gamme en termes de débit (Figure A-1). Du débit moyen (tel que MiniSeq et MiSeq) au débit actuellement le plus élevé (NovaSeq) générant jusqu'à 6000 Giga base (Gb) par run. La haute précision, le débit et les coûts en constante diminution et leurs diverses applications, font de la plate-forme Illumina la plus appropriée pour cibler de nombreuses questions biologiques.

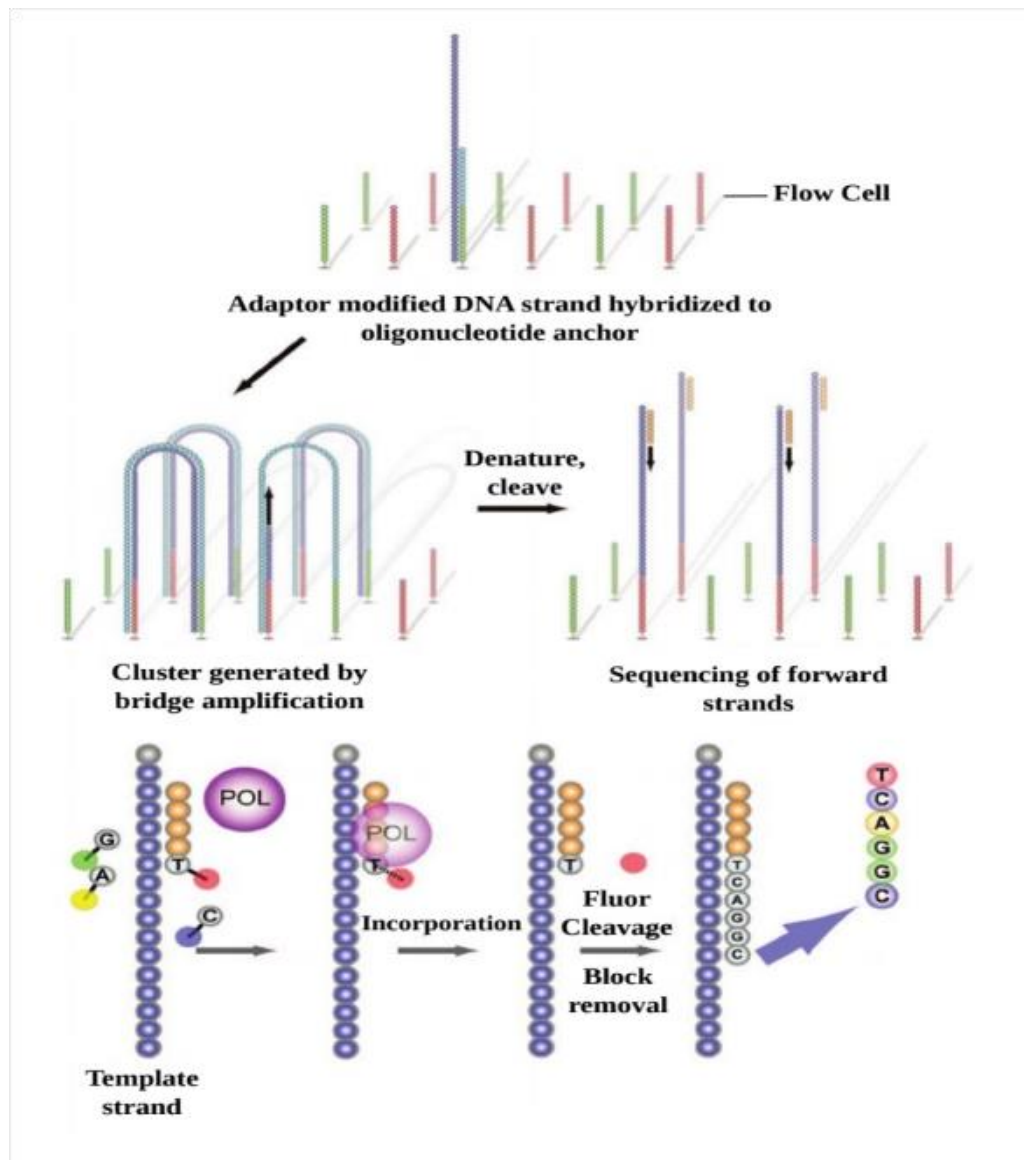


Figure A-2 : Le principe du processus de séquençage Illumina. L'ADN simple brin fragmenté est ligaturé par des adaptateurs. Ces fragments ligaturés sont ajoutés à la Flow Cell et immobilisés par hybridation. L'amplification en pont a généré des amas de séquences identiques amplifiées. Les clusters sont ensuite dénaturés et le séquençage est lancé avec l'ajout d'amorces, de polymérases et de 4 extrémités réversibles marquées aux fluorochromes. La fluorescence post-incorporation de chaque base est enregistrée. Le fluor et le bloc sont ensuite éliminés avant le prochain cycle de synthèse. (Voelkerding *et al.*, 2009)

La première étape du séquençage Illumina est (i) la préparation de librairie. L'échantillon d'ADN est fragmenté et ligaturé avec des adaptateurs aux extrémités 5' et 3'. Une fois l'ADN fragmenté et les adaptateurs ligaturés, les fragments doivent être immobilisés sur une plaque solide (Flow Cell). Cette dernière, contenant préalablement des oligonucléotides complémentaires aux séquences d'adaptateurs permettant la fixation des fragments dans le support (Figure A-2). L'objectif principal de cette première étape (préparation de librairie) est la fixation des adaptateurs dans les extrémités des séquences fragmentées, permettant de jouer ainsi le rôle d'amorces pour la deuxième étape (ii)

« amplification en pont ». Les fragments d'ADN résultants sont donc amplifiés pour générer des clusters de séquences identiques sur la Flow Cell. Ces clusters sont produits pour augmenter la force du signal nécessaire pour (iii) l'étape finale de séquençage par terminateurs réversibles des fragments amplifiés : des nucléotides présentant des caractéristiques spécifiques, vont être incorporés. L'extrémité 3'OH de ces nucléotides est désactivée, et marquée par des fluorochromes spécifiques à chaque nucléotide. Une ADN polymérase va insérer ces nucléotides correspondants aux nucléotides du brin complémentaire des amplicons. Les fluorochromes sont ensuite excités grâce à des lasers permettant l'émission d'un signal lumineux spécifique à chaque nucléotide (Buermans *et al.*, 2014). Des programmes informatiques vont traduire ces signaux en une séquence nucléotidique. Après détection des signaux, le groupement de protection du nucléotide est éliminé par photo-clivage entraînant la restauration du groupement fonctionnel du nucléotide incorporé, ainsi l'ADN polymérase peut incorporer le prochain nucléotide.

Deux types de séquençage existaient pour cette technologie : le séquençage *single-end* (SE, un fragment n'est séquencé qu'à partir de l'extrémité 5' ou 3') adapté aux projets de séquençage tels que les miRNA-Seq (pour *MicroRNA sequencing*), et le séquençage *paired-end* (PE), où les deux extrémités de chaque fragment sont séquencées de telle sorte que deux *reads* (*forward* et *reverse*) sont produits à partir de chaque fragment. Le séquençage PE est adapté aux projets de séquençage du génome entier, RNA-Seq, etc (Head *et al.*, 2014).

A.3. Séquençage de 3ème génération

Séquençage de troisième génération, également appelées techniques de séquençage à molécule unique, sans amplification initiale de l'ADN. Deux fournisseurs sont actuellement en concurrence sur ce marché : Pacific Biosciences (PacBio) avec son séquençage en temps réel d'une seule molécule (SMRT ; première machine introduite en 2011) et Oxford Nanopore, une technologie avec son approche de séquençage des nanopores protéiques (introduite en 2014). Focalisant sur Pacific Biosciences (PacBio RS II et Sequel), cette dernière, utilise une structure composée de cellules SMRT avec des nanostructures appelés Zero-Mode Waveguide (ZMW) et possédant chacune un ADN polymérase immobilisé au fond de puits qui incorpore des dNTP couplés à un fluorophore et dont l'excitation laser entraîne son clivage, permettant une détection en temps réel (Goodwin *et al.*, 2016). En effet, les fragments matrices sont ligaturés à des adaptateurs « en épingle à cheveux » aux deux extrémités, circularisant la molécule. De cette manière, la polymérase traverse un fragment plusieurs fois (chaque séquence répétée est appelée un *subread*). La polymérase incorpore en continu des nucléotides marqués par des fluorophores. Une caméra enregistre la couleur émise et notamment les changements de couleurs traduisant l'incorporation d'une nouvelle base. L'avantage majeur de ces techniques de troisième génération est la longueur de lecture supérieure : PacBio atteint des longueurs

de lecture supérieures à 50 Kb et la plus grande mise en garde, cependant, est un taux d'erreur plus élevé par rapport aux machines de deuxième génération, car des molécules d'ADN uniques sans amplification préalable sont séquencées permettant de couvrir chacune de bases séquencées (Goodwin *et al.*, 2016 ; Slatko *et al.*, 2018).

Annexe B – Préparation de librairie pour le séquençage Illumina MiSeq

Au cours de mes travaux de Thèse, j'ai participé à la préparation de la librairie pour le projet de séquençage des génomes d'*A. baumannii* et d'autres espèces bactériennes (*Mycobacterium tuberculosis*, *Klebsiella pneumonia*, etc.) en utilisant la technologie Illumina MiSeq du laboratoire MedBiotech. Pour *F. graminearum*, les génomes ont été préalablement séquencés par la technologie Illumina HiSeq (six souches) et PacBio Sequel (deux souches) dans le cadre de la collaboration évoqué dans le **chapitre 1**. Dans cette section, nous citant des principales étapes pratiques utilisées pour la préparation de librairie pour le séquençage Illumina. Tout d'abord, le séquençage d'ADN nécessite d'avoir une concentration de 0,2 ng/µl par échantillon selon le kit de préparation de librairies utilisé « *Nextera XT DNA Library Prep V600* ». Il est donc indispensable de doser puis de diluer l'ADNg de toutes les souches, jusqu'à obtenir la concentration requise. La qualité et la quantité d'ADN génomique extrait ont été déterminées respectivement par le spectrophotomètre NanoVue TM Plus et le fluorimètre Quantus TM.

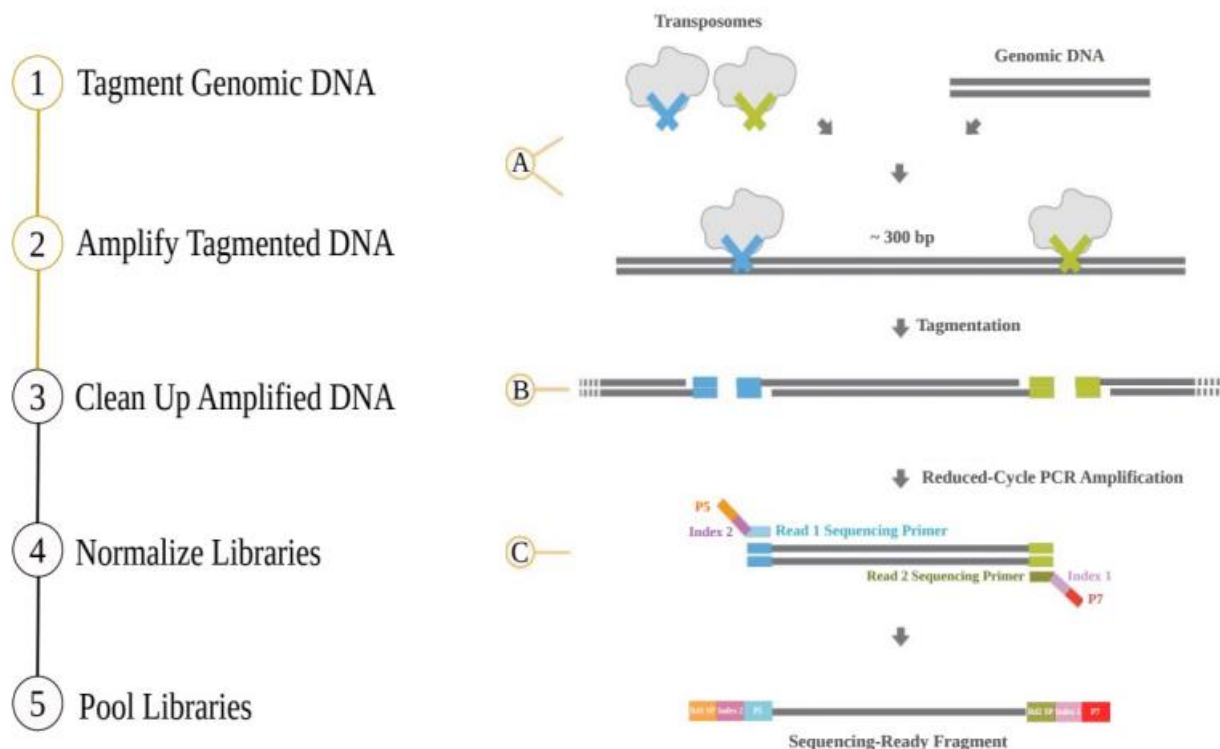


Figure A-3 : Workflow de préparation de la librairie d'ADN via Nextera XT. (A) Transposome Nextera XT avec adaptateurs combinés à l'ADN matrice. (B) Tagmentation pour fragmenter et ajouter des adaptateurs. (C) PCR à cycle limité pour ajouter des séquences d'adaptateurs d'index. https://support.illumina.com/sequencing/sequencing_kits/nextera_xt_dna_kit/documentation.html

La première étape consiste à fragmenter l'ADNg (le NGS ne peut pas fonctionner avec des molécules d'ADN de pleine longueur). Dans cette étape, l'ADN extrait sera fragmenté par la méthode enzymatique en utilisant le complexe transpososome (Figure A-3). Ce transposome comprend une enzyme transposase, fragmente simultanément l'ADN, et un transposase, insère les adaptateurs avec une combinaison de séquences adaptatrices spécifiques pour chaque échantillon (multiplexage). Cette méthode appelé également "*tagmentation Nextera*". L'ADN étiqueté avec les adaptateurs doit être purifié du transposome car peut se lier étroitement aux extrémités de l'ADN et interférer avec les processus en aval s'il n'est pas retiré. (ii) l'ADN étiqueté est amplifié via un programme de PCR à cycle limité. L'étape de PCR ajoute l'index 1 (i7), l'index 2 (i5) dans chaque échantillon, et les séquences nécessaires à l'ancrage dans la Flow Cell (P5 et P7). (iii) Nettoyage de la librairie, cette étape utilise des billes "AMPure XP" pour purifier l'ADN de la librairie et fournit une étape de sélection de la taille en supprimant les fragments très courts. (iv) Normalisation de la librairie, ce processus normalise la quantité la librairie pour garantir une représentation plus égale de chaque échantillon.

Une fois la librairie préparée et placée dans le séquenceur, nous avons créé le *sample sheet*, un fichier (*.csv) contenant les informations nécessaires pour configurer le séquençage, dont la combinaison des indexes utilisée lors de la préparation de librairie afin de séquencer tous les échantillons dans le même run. Au niveau du séquenceur, la librairie finale préparée est chargée dans la Flow Cell. Le séquençage a duré environ deux jours. Nous avons ensuite vérifié la qualité globale de la précision de séquençage des *reads* séquencés via un rapport généré après le séquençage.

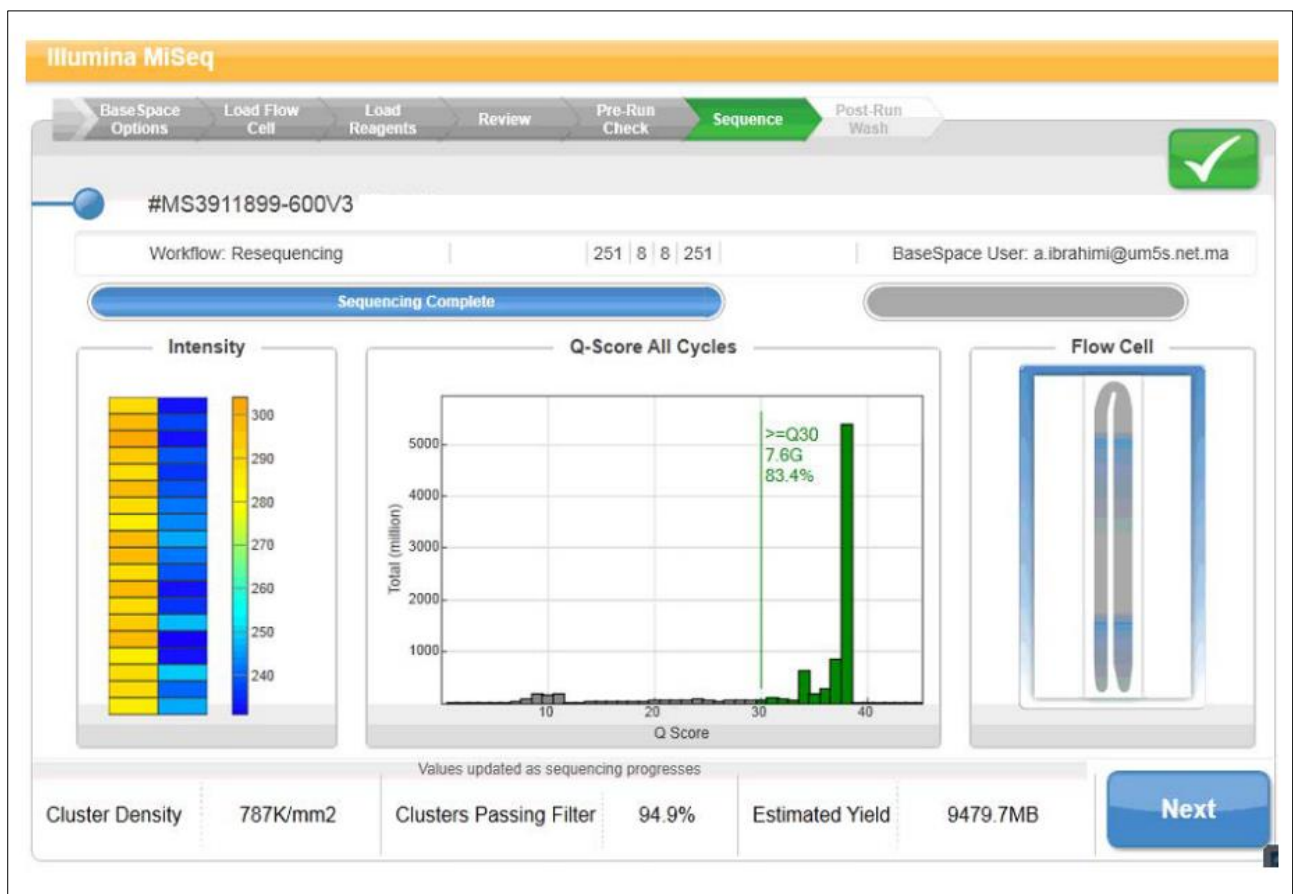


Figure A-4 : Rapport de la qualité globale obtenu après le séquençage Illumina MiSeq

La qualité est mesurée par le score de qualité Phred (Q score) : une mesure utilisée pour évaluer la probabilité d'erreur pour la transition nucléotidique correspondante. Nous avons obtenu 83,4% des données supérieures à Q30 (Figure A-4). Q30 corespond à la probabilité d'avoir une base incorrecte sur 1000 (1/1000).

Les données issues du séquençage :

Une fois le séquençage terminé, les données sont présentées dans un fichier *. FastQ. Ce dernier est un format de fichier texte permettant de stocker des séquences d'acides nucléiques avec leurs scores de qualité. En raison que le séquençage était effectué en *paired-end*, Deux fichiers *.FastQ ont

été obtenus pour chaque échantillon (forward et reverse) comprenant des milliers de *reads* (selon la couverture choisie).

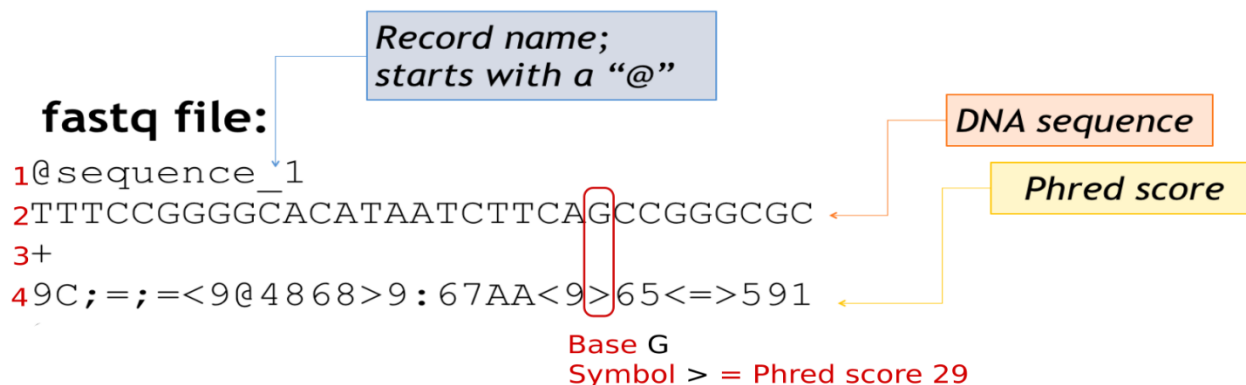


Figure A-5: Format du fichier FastQ. Chaque *read* dans le fichier FastQ est composée de 4 lignes, la première commence par "@" suivi de l'identifiant de la séquence, la 2ème contient les séquences nucléotidiques brutes, la 3ème correspond à l'identifiant de qualité (+), et la 4ème aux scores Phred, indiquant la fiabilité de chaque base, par exemple la base G (encadrée en rouge) avec le symbole « > » correspond au score Phred 29. Ces données ont été analysées à l'aide du workflow d'analyse génomique décrit au **chapitre 1**.

Annexe C – Assemblage et annotation du génome

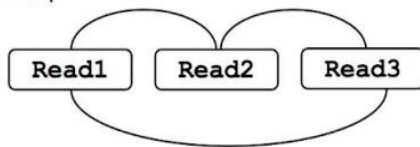
C.1 Assemblage *de novo* de *reads* issus de NGS

L'essence de l'assemblage *de novo* est, de construire de longues séquences appelées contigs/scaffolds à partir d'un grand nombre de *reads* générés par la technologie NGS afin d'obtenir un génome complet ou partiellement complet de l'échantillon séquencé. Il existe deux méthodes principales pour l'assemblage du génome, le graphe de Bruijn (DBJ) et la méthode d'*Overlap-Layout-Consensus* (OLC) (Ayling et al., 2020). En effet, la méthode de DBJ (Figure A-6), consiste dans un premier temps à fragmentées les *reads* en sous séquences de même taille (k) appelées k-mers, puis les graphes sont générés avec des nœuds où chaque nœud représente une séquence k-mer, qui se connectent avec une séquence k-1 en commun. Par exemple, deux k-mers vont pouvoir être assemblés si leur séquence ne diverge que par le premier nucléotide de l'un et le dernier nucléotide de l'autre. De cette façon il est possible d'établir un lien entre chacun des K-mers et extraire les contigs. La

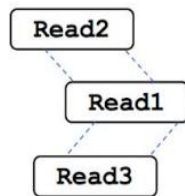
méthode de DBJ est bien adaptée pour l'assemblage de *reads* courts issus principalement par les séquenceurs de 2^{ème} génération comme Illumina. Alors que pour assembler les longs *reads* issues de séquenceurs de 3^{ème} génération, la méthode OLC (Figure A-6) est bien adaptée pour cela et comprend trois étapes complémentaires : la première étape, dite (i) *overlap*, a pour but d'aligner les *reads* deux à deux. A chaque alignement va correspondre un score, qui dépendra du pourcentage d'identité nucléotidique et de la longueur de cet alignement entre deux séquences. Ensuite, l'étape de (ii) *layout* permet de déterminer l'ordre dans lequel les séquences vont être assemblées, les paires de *reads* alignés vont être sélectionnées en fonction de leur score. Pour la dernière étape, (iii) *Consensus*, consiste à établir la séquence consensus des contigs en fusionnant les *reads* alignés. L'étape de chevauchement peut tolérer les mésappariements et les InDels entre les *reads*, et par conséquent fonctionne bien avec des *reads* plus longues même si celles-ci sont sujettes à des erreurs.

(a) Overlap, Layout, Consensus assembly

(i) Find overlaps



(ii) Layout reads



(iii) Build consensus

```

CGATTCTA
  TTCTAAGT
   GATTGTAA
-----
CGATTCTAAGT

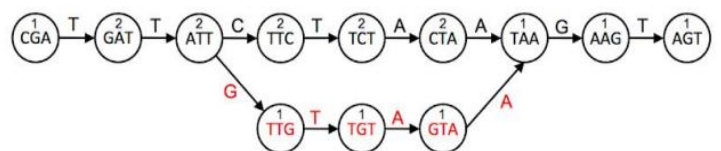
```

(b) De Bruijn graph assembly

(i) Make kmers

Read1: TTCTAAGT	Read2: CGATTCTA	Read3: GATTGTAA
Kmers: TTC	Kmers: CGA	Kmers: GAT
TCT	GAT	ATT
CTA	ATT	TTG
TAA	TTC	TGT
AAG	TCT	GTA
AGT	CTA	TAA

(ii) Build graph



(iii) Walk graph and output contigs

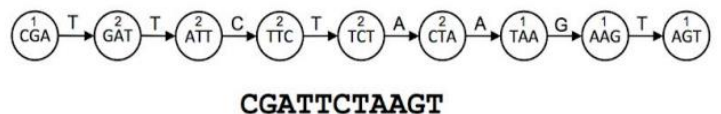


Figure A-6 : Les deux approches de l'assemblage *de novo* du génome : (a) la méthode OLC, (i) des chevauchements sont trouvés entre les *reads* et un graphe de chevauchement est construit. (ii) Les *reads* sont disposés en contigs en fonction des chevauchements. (iii) La séquence la plus probable est choisie pour construire la séquence consensus. (b) La méthode de DBJ, (i) les *reads* sont décomposées en k-mers en faisant glisser une fenêtre de taille k sur les *reads*. (ii) Deux sous-séquences k-mers sont reliées s'il existe une concordance de séquence entre les 3 premiers nucléotides de l'un et les 3 derniers de l'autre. Les polymorphismes (en rouge) forment des branches dans le graphe. Un décompte est tenu du nombre de fois qu'un kmer est présent, indiqué ici sous forme de nombres au-dessus des kmers. (iii) Les contigs sont construits

en parcourant le graphe à partir des *nœuds* de bord, de cette façon il est possible de fournir la séquence consensus (Ayling et al., 2020).

La contiguïté de l'assemblage est souvent mesurée en fonction de la taille et de la précision des contigs (Miller *et al.* 2010), des statistiques simples telles que N50 (Figure A-7) ; est une métrique importante, calcule la longueur du contig qui représente 50% du génome avec les séquences plus longues, et donc 50% de l'assemblage est composé de contigs d'une longueur $\geq$ N50.

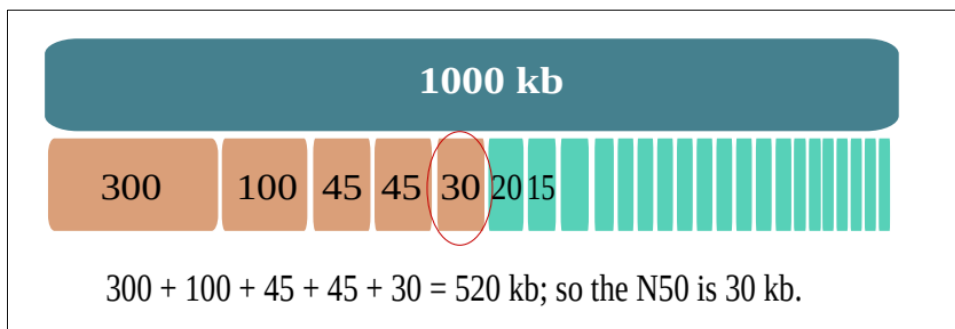


Figure A-7 : La mesure de N50 des contigs assemblés.

Plus la valeur de N50 est élevée, meilleur sera l'assemblage, mais elles ne reflètent pas l'exhaustivité de l'assemblage. Cette dernière, définie par la représentation de l'espace des gènes, est généralement évaluée en étudiant la présence d'un core de gènes orthologues conservé de manière évolutive (détaillé dans le **chapitre 1**).

Une limite fondamentale de l'assemblage des *reads* courts (issus des séquenceurs de 2ème génération) est la faible capacité à résoudre les séquences répétitives naturelles présentes dans le génome, car une répétition proche ou supérieure à la longueur de *read* peut générer des contigs chimériques. Pour cela, le séquençage *paired-end*, permettant d'améliorer considérablement la possibilité d'assembler correctement les *reads* représentant ces régions à plus haute résolution. Alors que pour les génomes connus pour des régions répétitives très longues, l'utilisation de technologies de séquençage de 3ème génération est fortement recommandée en raison de leurs capacités à séquencer les *reads* de grandes tailles.

C.2 Annotation du génome

C.2.1. Annotation du génome eucaryote

Les génomes eucaryotes sont beaucoup plus grands et plus complexes, les stratégies de prédiction et d'annotation diffèrent significativement de celles des procaryotes. Par exemple, dans le génome humain, environ 99 % sont constitués d'ADN codant pour des non-protéines. L'annotation du génome eucaryote commence souvent par l'identification et le masquage de séquences répétitives. En effet, les répétitions sont généralement mal conservées dans le génome et sont donc difficiles à identifier et à masquer. Une détection assez précise du catalogue de répétitions dans le génome

implique dans un premier temps la construction d'une librairie de répétitions spécifique à l'espèce à l'aide d'outils de prédiction *de novo* comme RepeatModeler (Flynn *et al.*., 2020) en combinaison avec des bases de données hébergeant des éléments répétés préalablement identifiés (Bao *et al.*, 2015). La librairie de répétition générée peut ensuite être utilisée comme input pour des outils tels que RepeatMasker (Tempel, 2012) pour masquer les régions de répétition.

Les annotations structurales des gènes se basent sur deux méthodes complémentaires : l'annotation de gènes basée sur les évidences de séquences, appelée extrinsèque et annotation de gènes *ab initio* appelée *de novo* (ou intrinsèque) (Scalzitti *et al.*, 2020). (i) L'annotation des gènes basée sur des évidences repose sur l'alignement précis des séquences transcriptomiques et protéomiques de l'espèce et/ou d'espèces proches sur le génome assemblé. En effet, l'utilisation des données RNA-seq parmi les évidences, a le plus grand potentiel pour améliorer la précision des annotations génétiques en raison de leur capacité à déterminer les exons et les sites d'épissage. De plus, des séquences protéiques provenant d'autres organismes apparentés peuvent ainsi être utilisées dans l'alignement, car elles conservent une similarité de séquence substantielle avec les séquences nucléotidiques. Tandis que (ii) l'annotation de gènes *ab initio* prédit des modèles de gènes basés sur des caractéristiques intrinsèques telles que des approximations sur les limites exon-intron, les codons de départ et d'arrêt, etc. Pour utiliser cette méthode, le génome assemblé cible doit être très étroitement lié au modèle de l'espèce ou de l'espèce proche choisi pour les prédictions génétiques. Etant donné que la longueur des introns et des exons, l'utilisation des codons et la teneur en GC d'organismes étroitement apparentés peuvent différer, par conséquent, pourrait entraîner des inexactitudes importantes des gènes prédits. En revanche, l'utilisation de données RNA-Seq pourrait être utilisée également pour entraîner des outils d'annotation *ab initio* comme SNAP <https://github.com/KorfLab/SNAP>.

Les résultats de ces deux phases d'annotation (intrinsèque et extrinsèque) devraient être intégrés et synthétisés dans un ensemble final d'annotations de gènes.

3.2.2. Annotation du génome procaryote

La prédiction et l'annotation des gènes procaryotes sont nettement plus simples que la prédiction et l'annotation des gènes eucaryotes, compte tenu de la complexité et de la taille du génome eucaryote et de sa constitution génétique. Par exemple, les souches bactériennes n'ont pas d'introns, et les sites d'initiation de la transcription bactérienne et les éléments régulateurs sont typiquement à proximité étroite avec les gènes respectifs. Un pipeline d'annotation génomique typique pour un génome procaryote comprend trois étapes majeures : (i) la prédiction de gène, l'outil de prédiction scanne la séquence du génome assemblé pour les régions qui sont susceptibles de coder des protéines ou de l'ARN fonctionnel. (ii) Les gènes identifiés sont ensuite interrogés dans des bases de données de référence à l'aide d'outils d'alignements (comme BLAST) pour trouver des séquences associées. Si

une séquence associée avec un pourcentage élevé de similitude et d'identité est trouvée, les informations sur leur fonction sont ajoutées à l'annotation. La dernière étape (iii) correspond à une curation manuelle sont utilisées pour valider les résultats de prédiction structurelle et d'annotation fonctionnelles.

Références associées aux annexes

- Ayling, M., Clark, M. D., et Leggett, R. M. (2020). New approaches for metagenome assembly with short reads. *Briefings in Bioinformatics*, 21(2), 584-594. <https://doi.org/10.1093/bib/bbz020>
- Bao, W., Kojima, K. K., et Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA*, 6, 11. <https://doi.org/10.1186/s13100-015-0041-9>
- Buermans, H. P. J., et den Dunnen, J. T. (2014). Next generation sequencing technology : Advances and applications. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1842(10), 1932-1941. <https://doi.org/10.1016/j.bbadis.2014.06.015>
- Flynn, C. M., Broz, K., Jonkers, W., Schmidt-Dannert, C., et Kistler, H. C. (2019). Expression of the *Fusarium graminearum* terpenome and involvement of the endoplasmic reticulum-derived toxosome. *Fungal Genetics and Biology*, 124, 78-87. <https://doi.org/10.1016/j.fgb.2019.01.006>
- França, L. T. C., Carrilho, E., et Kist, T. B. L. (2002). A review of DNA sequencing techniques. *Quarterly Reviews of Biophysics*, 35(2), 169-200. <https://doi.org/10.1017/S0033583502003797>
- Goodwin, S., McPherson, J. D., et McCombie, W. R. (2016). Coming of age : Ten years of next-generation sequencing technologies. *Nature Reviews. Genetics*, 17(6), 333-351. <https://doi.org/10.1038/nrg.2016.49>
- Head, S. R., Komori, H. K., LaMere, S. A., Whisenant, T., Van Nieuwerburgh, F., Salomon, D. R., et Ordoukhanian, P. (2014). Library construction for next-generation sequencing : Overviews and challenges. *BioTechniques*, 56(2), 61-passim. <https://doi.org/10.2144/000114133>
- Heather, J. M., et Chain, B. (2016). The sequence of sequencers : The history of sequencing DNA. *Genomics*, 107(1), 1-8. <https://doi.org/10.1016/j.ygeno.2015.11.003>
- Maxam, A. M., et Gilbert, W. (1977). A new method for sequencing DNA. *Proceedings of the National Academy of Sciences*, 74(2), 560-564. <https://doi.org/10.1073/pnas.74.2.560>

- Modi, A., Vai, S., Caramelli, D., et Lari, M. (2021). The Illumina Sequencing Protocol and the NovaSeq 6000 System. *Methods in Molecular Biology (Clifton, N.J.)*, 2242, 15-42. https://doi.org/10.1007/978-1-0716-1099-2_2
- Nederbragt, L. (2021). *Developments in next generation sequencing* [Jupyter Notebook]. <https://github.com/lexnederbragt/developments-in-next-generation-sequencing> (Original work published 2013)
- Sanger, F., Nicklen, S., et Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 74(12), 5463-5467.
- Scalzitti, N., Jeannin-Girardon, A., Collet, P., Poch, O., et Thompson, J. D. (2020). A benchmark study of ab initio gene prediction methods in diverse eukaryotic organisms. *BMC Genomics*, 21(1). <https://doi.org/10.1186/s12864-020-6707-9>
- Slatko, B. E., Gardner, A. F., et Ausubel, F. M. (2018). Overview of Next-Generation Sequencing Technologies. *Current Protocols in Molecular Biology*, 122(1), e59. <https://doi.org/10.1002/cpmb.59>
- Tempel, S. (2012). Using and understanding RepeatMasker. *Methods in Molecular Biology (Clifton, N.J.)*, 859, 29-51. https://doi.org/10.1007/978-1-61779-603-6_2
- Voelkerding, K. V., Dames, S. A., et Durtschi, J. D. (2009). Next-generation sequencing : From basic research to diagnostics. *Clinical Chemistry*, 55(4), 641-658. <https://doi.org/10.1373/clinchem.2008.112789>