Émergence d'une nouvelle espèce bactérienne : Bacillus anthracis, responsable de la maladie du charbon

Mehdi Abdelli

▶ **To cite this version:**

Mehdi Abdelli. Émergence d'une nouvelle espèce bactérienne : Bacillus anthracis, responsable de la maladie du charbon. Biodiversité et Ecologie. Université Paris-Saclay, 2024. Français. NNT : 2024UPASL034 . tel-04699855

HAL Id: tel-04699855
https://theses.hal.science/tel-04699855

Submitted on 17 Sep 2024

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 PARIS-SACLAY

Émergence d'une nouvelle espèce bactérienne : *Bacillus anthracis*, responsable de la maladie du charbon

Emergence of a new bacterial species: Bacillus anthracis, responsible for anthrax

Thèse de doctorat de l'université Paris-Saclay

École doctorale n°577 : Structure et Dynamique des Systèmes Vivants (SDSV)
Spécialité de doctorat : Évolution
Graduate School : Sciences de la vie et santé. Référent : Faculté des sciences d'Orsay

Thèse préparée dans l'unité de recherche **Institute for Integrative Biology of the Cell I2BC, (Université Paris-Saclay, CEA, CNRS)**
sous la direction de **Jacques OBERTO**, directeur de recherche

Thèse soutenue à Paris-Saclay, le 14 juin 2024, par

Mehdi ABDELLI

Composition du Jury

Membres du jury avec voix délibérative

Frédéric CARLIN Directeur de recherche, INRAE, Avignon	Président
Gwennaele FICHANT Professeure des universités, Université Toulouse III Paul Sabatier	Rapporteure & Examinatrice
Guy PERRIÈRE Directeur de recherche, Université Lyon Claude Bernard Lyon 1	Rapporteur & Examinateur
Alice CHÂTEAU Maîtresse de conférences, INRAE, Université d'Avignon	Examinatrice
Didier LERECLUS Directeur de recherche émérite, INRAE, Université Paris-Saclay	Examinateur

THESE DE DOCTORAT

NNT : 2024UPASL034

université PARIS-SACLAY

ÉCOLE DOCTORALE
Structure et dynamique des systèmes vivants (SDSV)

Titre : Émergence d'une nouvelle espèce bactérienne : *Bacillus anthracis*, responsable de la maladie du charbon

Mots clés : Évolution, Métagénomique, *Bacillus anthracis*, Bioinformatique, Plasmide, Espèce

Résumé : *Bacillus anthracis* est la bactérie responsable de la maladie du charbon et demeure une préoccupation majeure de la lutte contre le bioterrorisme. La virulence de cette bactérie hautement pathogène est portée en large part par des plasmides. Son génome résulte en effet d'un assemblage chromosome-plasmides perpétué par l'existence de conditions favorables au maintien de la bactérie pathogène dans un écosystème. Les éléments de connaissance actuels ne permettent pas d'établir le lieu d'origine de cet assemblage. D'où viennent ces plasmides ? Où sont les niches écologiques ? L'existence de plasmides similaires au sein d'espèces appartenant au même groupe *Bacillus cereus* est une indication encourageant la reconstitution de telles chronologies encore très incomplètes. De prime abord, nous avons caractérisé plusieurs lots de souches du groupe *B. cereus* pour déterminer leur pathogénicité, leur potentielle acquisition de plasmides et leur position phylogénétique vis-à-vis de *B. anthracis*. Cela a amené à détecter les souches du groupe *B. cereus* qui sont actuellement les plus proches voisines connues de *B. anthracis*. Cet enrichissement du voisinage a été déterminant dans la suite des travaux, qui ont consisté à s'intéresser à l'émergence de *B. anthracis*. En effet, la meilleure connaissance de son voisinage a permis de situer l'apparition de l'ancêtre de *B. anthracis*. En outre, de nouveaux types de miniréplicons relatifs aux plasmides du groupe *B. cereus* ont été observés dans les proches voisins découverts. Leur analyse a permis d'émettre des hypothèses sur l'acquisition de ces derniers au cours de l'histoire évolutive du groupe *B. cereus*, et *a fortiori* sur l'acquisition de certains facteurs de virulence. Le modèle d'émergence établi pour le pathogène *B. anthracis* s'avère transposable à d'autres espèces clonales, pathogènes pour l'Homme et très proches d'espèces environnementales (par exemple *Yersinia pestis*, *Mycobacterium tuberculosis*, *Brucella* ou *Francisella tularensis*).

L'émergence clonale récente de *B. anthracis* observée dans la première partie des travaux, indiquant une apparition dans une région géographique spécifique (forêt Centre-Africaine) a motivé la recherche de nouvelles souches de *B. anthracis* dans des données métagénomiques. En principe, la démarche est simple, et plusieurs outils bioinformatiques performants permettent d'inventorier une masse de données génétiques. Ces outils utilisent des génomes de références et une taxonomie, pour assigner chacun des tronçons de séquence à un niveau taxonomique. Cependant, les inventaires produits qui, globalement, sont très précis, peuvent être inexacts pour les évènements rares. C'est particulièrement le cas pour la détection de pathogènes, dont *B. anthracis*. Ainsi, certaines études ont conclu à sa présence dans des environnements urbains. Dans ce cas, l'erreur d'interprétation résultait principalement de l'insuffisance des bases de données de référence. Un pipeline de détection dans des échantillons métagénomiques a donc été conçu. Un jeu de données test a également été mis en place, avec l'élaboration de métagénomes environnementaux enrichis par *B. anthracis* et/ou des souches du groupe *B. cereus* et/ou d'autres agents pathogènes, pour tester la sensibilité et la spécificité du pipeline. Cet ensemble de données pourra par ailleurs servir de manière générale à l'évaluation future d'outils de détection de *B. anthracis*. Sur la base de ce nouvel outil, une recherche massive a été opérée dans des échantillons métagénomiques publics. Là encore, la démarche peut être généralisée et ce pipeline de détection est transposable à d'autres pathogènes pour l'Homme.

université PARIS-SACLAY

ÉCOLE DOCTORALE
Structure et dynamique des systèmes vivants (SDSV)

Title : Emergence of a new bacterial species: *Bacillus anthracis*, responsible for anthrax

Keywords : Evolution, Metagenomics, *Bacillus anthracis*, Bioinformatics, Plasmid, Species

Abstract : *Bacillus anthracis* is the bacterium responsible for anthrax, and remains a major concern in the fight against bioterrorism. The virulence of this highly pathogenic bacterium is largely mediated by plasmids. Its genome is the result of a chromosome-plasmid association perpetuated by the existence of conditions favorable to the maintenance of the pathogenic bacterium in an ecosystem. Current knowledge does not allow to determine the origin of this assembly. Where do these plasmids come from? Where are the ecological niches? The existence of similar plasmids within species belonging to the same *Bacillus cereus* group is an indication that our understanding of such chronologies are still very incomplete. As a first step, we characterized several sets of *B. cereus* group strains to determine their pathogenicity, potential plasmid acquisition and phylogenetic position vis-à-vis *B. anthracis*. It led to the detection of the closest known neighbors of *B. anthracis* described so far. This enrichment of the neighborhood was decisive in the subsequent work, which focused on the emergence of *B. anthracis*. The better understanding of the *B. anthracis* neighborhood enabled us to better position the appearance of the *B. anthracis* ancestor. In addition, new types of minireplicons relating to *B. cereus* group plasmids were observed in the close neighbors discovered. Their analysis has led to hypotheses on the acquisition of these plasmids during the evolutionary history of the *B. cereus* group, and on the acquisition of certain virulence factors. The emergence model established for the *B. anthracis* pathogen can be transposed to other clonal species, pathogenic to humans and closely related to environmental species (e.g. *Yersinia pestis*, *Mycobacterium tuberculosis, Brucella* or *Francisella tularensis*).

The recent clonal emergence of *B. anthracis* observed in the first part of the work, indicating an appearance in a specific geographical region (Central African forest), motivated the search for new *B. anthracis* strains in metagenomic data. In principle, the approach is straightforward, and several high-performance bioinformatics tools can be used to inventory a mass of genetic data. These tools use reference genomes and a taxonomy to assign each sequence read to a taxonomic level. However, if the produced inventories are generally highly accurate, they also can be unadapted for rare events. This is particularly the case for the detection of pathogens, such as *B. anthracis*. For example, some studies have concluded that *B. anthracis* is present in urban environments. In this case, the misinterpretation was mainly due to inadequate reference databases. A detection pipeline for metagenomic samples was therefore designed. A test dataset was also set up, with the development of environmental metagenomes spiked by *B. anthracis* and/or *B. cereus* group strains and/or other pathogens, to test the sensibility and the sensitivity of the pipeline. More generally, this dataset could also be used for the evaluation of future similar tools. A massive search was carried out in public metagenomic samples on the basis of this new tool. Again, the approach can be generalized, and this detection pipeline can be transposed to other human pathogens.

Remerciements

Je remercie tout d'abord chaleureusement les membres de mon jury de thèse : les rapporteurs Gwennaele Fichant et Guy Perrière, les examinateurs Frédéric Carlin, Alice Château et Didier Lereclus, d'avoir accepté d'évaluer mon travail et d'en avoir enrichi beaucoup d'aspects de par leurs remarques et idées.

Je remercie également mes supérieurs hiérarchiques, en particulier Bruno Bellier, directeur de DGA Maitrise NRBC, Sandrine Meunier, sous-directrice technique ainsi que Laurent Taysse et Françoise Raynaud, de m'avoir permis de réaliser ce projet de thèse dans les meilleures conditions.

Je remercie grandement Jacques Oberto d'avoir été mon directeur de thèse, dans le format très particulier dans laquelle elle a été menée, pour sa disponibilité, son aide et ces remarques éclairantes pour mes travaux.

Un immense merci à Gilles Vergnaud pour toute l'aide qu'il a pu m'apporter dans le cadre de mes travaux, pour m'avoir aiguillé tout au long de ces derniers et pour sa grande disponibilité et rigueur dans la correction de mon manuscrit. Je lui en suis extrêmement reconnaissant.

Je remercie également Vincent Ramisse pour avoir encadré mes travaux au sein de la DGA, pour ses conseils et sa disponibilité à toute épreuve.

Je remercie chaleureusement Daniel Gautheret pour m'avoir accueilli au sein de son équipe à l'I2BC et pour toute son aide tout au long de cette thèse.

Un grand merci à l'ensemble de la division Biologie de DGA Maitrise NRBC, en particulier à l'équipe bio-informatique (Benjamin, Éloi, Maylis et Xavier) pour tout leur soutien au quotidien.

Un remerciement tout particulier à Timothée et Amélie pour leur soutien permanent et tous les bons moments passés ensemble. Vos noms pourraient être cités dans le paragraphe suivant mais je ne voudrais pas omettre toute l'aide que vous m'avez apportée dans mes travaux. Je vous en suis très reconnaissant.

Je remercie également mes amis, qui m'ont su me ~~déconcentrer~~ soutenir durant ces années de dur labeur. Je citerai pour l'exemple Ali, Arthur, Benjamin, Toufik, Erwan, Vdb, Alice et j'en passe (l'heure tardive à laquelle j'écris ces remerciements m'empêche d'établir une liste exhaustive, mais les absents ne m'en voudront pas car, dans le cas contraire, ils ne seraient pas mes amis, et leur absence serait ainsi justifiée).

Enfin, j'exprime ma profonde reconnaissance à ma famille, et tout particulièrement mes parents, sans qui rien n'aurait été possible dans ma vie. Ce travail vous est dédié.

Table des matières

Table des figures

Liste des tableaux

Abréviations

ADN Acide désoxyribonucléique

ARN Acide ribonucléique

ERIC Consensus Répétitif Intergénique des Entérobactéries

HSP Paire de Segments à Haut Score

IA Intelligence Artificielle

ISS Station Spatiale Internationale

MLST Typage Multilocus de Séquences

MLVA Analyse VNTR sur Plusieurs Loci

MRCA Ancêtre Commun le Plus Récent

NCBI Centre National pour l'Information Biotechnologique

NGS Séquençage de Nouvelle Génération

PCR Réaction en Chaîne par Polymérase

PFGE Électrophorèse en Gel à Champ Pulsé

RAPD Amplification Aléatoire de l'ADN Polymorphe

REP Séquences Palindromiques Extragéniques Répétées

RFLP Polymorphisme de Longueur des Fragments de Restriction

SLST Typage de Séquence à Locus Unique

SNP Polymorphisme de Nucléotide Unique

SNV Variant de Nucléotide Unique

SRA Archive de Lectures de Séquences

VIP Protéine Insecticide Végétative

VNTR Nombre Variable de Répétitions en Tandem

WGS Séquençage du Génome Entier

À mes parents.

Introduction

Bacillus anthracis, l'agent pathogène connu pour causer la maladie du charbon, potentiellement mortelle chez plusieurs espèces, dont l'Homme, a marqué l'Histoire par son rôle dans d'importantes épidémies ainsi que dans l'essor de la recherche en microbiologie. Elle est également notable pour son potentiel d'utilisation comme arme biologique. Cette bactérie possède des attributs remarquables, comme sa capacité à persister dans le sol sous forme de spores. Cette faculté à entrer en état dormant pendant de nombreuses années pose un défi significatif pour la datation précise et le suivi phylogéographique de cette espèce. De plus, la proximité génétique entre *B. anthracis* et d'autres espèces proches au sein du complexe *Bacillus cereus* complique la tâche de distinguer cette bactérie de manière fiable, mettant en lumière les défis inhérents à sa classification, à la définition de son identité propre, et par conséquent, à la compréhension de son évolution. Ces difficultés sont exacerbées dans le contexte de la biodéfense, où une éventuelle incapacité à retracer l'origine précise d'une souche peut encourager l'emploi malveillant de cette bactérie comme arme biologique. En outre, la variabilité de l'efficacité des programmes de vaccination et l'influence des conditions écologiques sur la pathogénicité de la bactérie soulignent l'importance d'une approche holistique pour appréhender les risques associés à *B. anthracis*. À ce titre, le concept « *One Health* » est précisément un exemple de ce type d'approche développé dans l'objectif, entre autres, de lutter contre la menace biologique d'origine anthropique provoquée ou accidentelle. Ainsi, *B. anthracis* continue à jouer un rôle de microorganisme modèle.

Différentes hypothèses ont été formulées quant à l'émergence de *B. anthracis*, chacune s'appuyant sur l'analyse phylogénétique et la caractérisation de nouvelles souches séquencées. Ces hypothèses sont régulièrement confrontées et affinées à la lumière des données génomiques disponibles, soulignant l'importance de poursuivre les efforts dans la détection continue de nouvelles souches. L'analyse comparative de ces souches permet de préciser notre connaissance sur la diversité de *B. anthracis* et aussi de détecter des schémas évolutifs, de tracer les voies de transmission, et d'identifier les potentiels réservoirs environnementaux. Cependant, si la phylogénie de l'espèce *B. anthracis* est bien définie aujourd'hui, il subsiste des interrogations quant à l'apparition de cette espèce, en particulier l'origine de son ancêtre et l'acquisition de son pouvoir de virulence.

À ce propos, la connaissance de son proche voisinage reste parcellaire. Or, celle-ci permettrait de préciser les modèles d'évolution de l'espèce, et, par voie de conséquence, de renforcer la fiabilité des outils de détection actuels.

Pour ce dernier point, des méthodes précises d'identification de la bactérie, en microbiologie, en biologie moléculaire ou encore en immunologie, ont été développées. Cependant, à l'ère du séquençage nouvelle génération, l'exploitation des données métagénomiques dans des environnements et à des dates très variées ouvre des perspectives nouvelles. Les outils actuels de détection métagénomique se heurtent à des défis de fiabilité et de rapidité dans le cadre de la détection de *B. anthracis*, en particulier dans des échantillons environnementaux complexes où la bactérie est présente en faible quantité et est difficilement discernable des espèces voisines.

Cet objectif de meilleure compréhension de l'histoire évolutive de *B. anthracis* a donc guidé mes travaux de thèse à travers la problématique suivante :

PROBLÉMATIQUE GÉNÉRALE

D'où provient l'espèce *B. anthracis* et quels mécanismes ont permis l'acquisition de son pouvoir de virulence ?

La thèse s'organise selon cinq chapitres, les deux premiers étant une étude bibliographique, et les trois derniers présentant des résultats expérimentaux. Leur organisation est la suivante :

— Le chapitre 1 retrace l'histoire de la maladie du charbon et détaille les principales caractéristiques de *B. anthracis*. En particulier, son cycle infectieux et ses facteurs de virulence y sont abordés.

— Le chapitre 2 s'intéresse aux différents éléments soulevant la question de l'histoire évolutive de *B. anthracis*. Y sont détaillées la forte proximité génétique avec les autres espèces du groupe *B. cereus*, auquel est affilié *B. anthracis*, l'existence de souches de ce même groupe causant des affections similaires au charbon ou encore la propagation de *B. anthracis* à travers le monde. Dans un dernier temps, la structure de population de *B. anthracis* est étudiée et un état de l'art quant à l'émergence et la propogation de *B. anthracis* est effectué.

— Le chapitre 3 illustre, à travers deux travaux de caractérisation de souches du groupe *B. cereus*, l'importance d'améliorer notre connaissance actuelle du proche voisinage de *B. anthracis* pour mieux comprendre son histoire. La découverte des souches du groupe *B. cereus* les plus proches voisines de *B. anthracis* connues à ce jour est présentée.

— Le chapitre 4 présente les travaux relatifs à l'émergence de *B. anthracis*, en particulier sur la position de l'ancêtre de l'espèce ainsi que la formation des plasmides de virulence, en tirant profit de la nouvelle connaissance du proche voisinage au sein du groupe *B. cereus*. Entre autres, une datation récente de l'ancêtre est proposée. Cette datation est compatible avec l'hypothèse que la distribution mondiale de *B. anthracis* est le résultat d'évènement d'origine anthropique, et prédit que *B. anthracis* n'existe pas là où les humains ne l'ont pas apporté. Cette prédiction ouvre la voie à l'utilisation de la métagénomique comme nouvelle méthode de détection de *B. anthracis* dans un échantillon environnemental.

— Le chapitre 5 aborde la problématique de la détection de *B. anthracis*. Un état de l'art des différentes méthodes existantes y est réalisé, avec un point d'attention sur les manquements actuels, inhérents à la phylogénie du groupe *B. cereus*. Puis, y est détaillé le développement de l'outil B2FORENSICS, servant à la détection rapide et fiable de traces de *B. anthracis* dans un échantillon métagénomique.

Chapitre 1

Qu'est-ce que *Bacillus anthracis* ?

La maladie du charbon est une zoonose provoquée par la bactérie *B. anthracis*, potentiellement mortelle affectant de nombreuses espèces mammifères, dont l'Homme. Sa forte persistance dans les sols, due à la résistance de ses spores, en fait un danger permanent, tout comme son pouvoir hautement pathogène. Si l'étude de cette bactérie est utile à des fins de suivi épidémiologique ou de recherche médicale, certains faits historiques marquants ont montré qu'il fallait prendre en compte son utilisation dans le risque bioterroriste. Ce premier chapitre introductif vise à dresser une présentation générale de *B. anthracis* et de la maladie du charbon.

OBJECTIFS DU CHAPITRE

1. Présenter les principales caractéristiques de *B. anthracis* et de la maladie causée par cette bactérie, le charbon
2. Définir la composition du génome de *B. anthracis* et en particulier les différents facteurs de virulence
3. Comprendre en quoi cette bactérie, au-delà du contexte sanitaire, peut faire l'objet d'une menace en termes de biodéfense
4. Présenter les différentes méthodes de détection traditionnelles de *B. anthracis*
5. Détailler la diversité génétique au sein de *B. anthracis*

1.1 Historique de la maladie du charbon

L'agent étiologique de la maladie du charbon, une zoonose affectant principalement les herbivores ruminants, est la bactérie à Gram positif *B. anthracis*, qui fait partie du groupe *B. cereus*. La transmission à l'Homme se produit généralement par contact direct avec des animaux infectés par l'intermédiaire d'une plaie, ou par la consommation de viande contaminée. La cinquième plaie d'Egypte (XVème-XIIIème siècle avant notre ère) est souvent considérée comme étant à l'origine de la maladie du charbon (TURNBULL et SHADOMY, 2010). Cependant il faut noter que l'argument principal avancé est que la fièvre charbonneuse est la seule maladie actuellement connue ayant le spectre d'hôtes indiqué dans le passage biblique. De même, la maladie pourrait également être celle évoquée par Homère dans L'Iliade, I. 49-50. L'Iliade aurait été écrite au VIIème siècle avant notre ère, mais la Guerre de Troie qui en est le sujet, aurait eu lieu au XIIIème-XIème siècle avant notre ère (HOMÈRE, s. d.). Les vers qui pourraient évoquer la maladie du charbon sont beaucoup moins riches en information que le texte biblique : « Assis à l'écart, loin des nefs, il [Apollon] lança une flèche, et un bruit terrible sortit de l'arc d'argent. Il frappa les mulets d'abord et

les chiens rapides ; mais, ensuite, il perça les hommes eux-mêmes du trait qui tue. Et sans cesse les bûchers brûlaient, lourds de cadavres ». Parmi les textes anciens identifiés, le plus clairement évocateur est celui des Géorgiques de Virgile (VIRGILE, s. d.). La description faite par Virgile est particulièrement intéressante à divers titres. Un aspect peu souligné est que Virgile fait référence à une « maladie du temps jadis » (*Hic quondam*) survenue dans le Norique (région située dans l'Autriche actuelle, bordant la frontière Nord-Est de l'Italie), ce qui suggère que Virgile lui-même n'aurait pas eu connaissance d'évènements similaires temporellement et géographiquement plus proches, et que cet évènement était donc exceptionnel. La même remarque sur le côté exceptionnel de ces évènements s'applique au récit biblique.

Au cours du Moyen Âge et de la Renaissance, le "poison noir" pourrait avoir été la maladie du charbon et la transmission de la maladie du bétail à l'Homme a été documentée, notamment par Ambroise Paré en 1568, où la cause était souvent attribuée à un "virus" ou à une substance nocive (PARÉ, 1568). Le XVIIIème siècle aurait été marqué par des épidémies qui auraient décimé près de la moitié des moutons en Europe (STERNBACH, 2003). À cette époque, le charbon était considéré avant tout comme une maladie animale, avec de fréquentes épidémies touchant le bétail. En France, la notion de "champs maudits" est apparue, désignant des lieux de pâturage dangereux pour le bétail, comme déjà noté par Virgile. À cette période, l'origine des maladies infectieuses était mal comprise, et on désignait le charbon sous le terme générique de "peste" (BLANCOU, 2000 ; FERRIÈRES, 2002).

La transmission de l'animal à l'Homme a été affirmée par Nicolas Fournier en 1769, reconnaissant ainsi le charbon comme une anthropozoonose[1], en parlant de "pustules malignes" (FOURNIER, 1769). Cette maladie peut affecter l'Homme par le biais de contacts avec des produits animaux contaminés commme la peau ou la laine, ou par ingestion de viande infectée, entraînant des manifestations cutanées et gastro-intestinales (MORENS, 2003).

En 1876, Robert Koch réussit à cultiver *B. anthracis in vitro*, découvrant son cycle de vie (KOCH, 1876), y compris la capacité de la bactérie à former des spores pour survivre dans des conditions défavorables (STERNBACH, 2003). Plus tard, les efforts de Louis Pasteur, qui expérimenta un procédé de vaccination sur des moutons en 1881, les recherches d'Elie Metchnikoff sur la phagocytose, et le développement par Max Sterne d'un vaccin vivant acapsulé dans les années 1930, ont contribué à diminuer l'impact du charbon sur le bétail domestique (SCHWARTZ, 2009).

1.2 Les caractéristiques de la maladie du charbon

Après avoir exploré l'histoire de la maladie du charbon, l'attention se tourne vers les caractéristiques spécifiques de cette maladie. La section suivante décrit les formes qu'elle peut prendre chez l'Homme et les animaux, illustrant la diversité de ses manifestations.

1.2.1 La forme cutanée

L'infection cutanée est le mode de transmission le plus courant chez l'Homme, représentant environ 95% des cas documentés (PURCELL, WORSHAM et FRIEDLANDER, 2006). Elle est également la moins mortelle, avec un taux de survie humain estimé

1. Maladie animale transmissible à l'être humain.

à 80% en l'absence de traitement (SWEENEY et al., 2011). Grâce à des traitements efficaces disponibles, la mortalité peut être réduite à moins de 1%. Les zones les plus couramment touchées sont les mains, les bras, le cou et le visage. Cette infection se produit souvent lorsque les spores s'introduisent à travers la peau à la faveur d'une plaie, en manipulant des produits animaux contaminés, à la suite de piqûres d'insectes ou d'abrasions d'épiderme affectant les follicules pileux (HAHN, SHARMA et SOHNLE, 2005). La période d'incubation est de un à sept jours (rarement jusqu'à 12 jours) (INGLESBY et al., 2002 ; JERNIGAN et al., 2002). Une faible proportion des spores se met à germer localement et provoque une nécrose des tissus mous (escarres) et un œdème. Une autre partie des spores est véhiculée par le système lymphatique jusqu'aux ganglions lymphatiques proximaux, provoquant une lymphadénopathie [2] douloureuse et une lymphangite [3]. Une dissémination dans la circulation sanguine accompagnée de toxémie [4] peut s'ensuivre, mais ce phénomène est rare avec une infection cutanée traitée par antibiothérapie. En général, cette manifestation de la maladie se résorbe en deux à trois semaines, avec la guérison complète de l'escarre, bien que dans certains cas, cela puisse se prolonger jusqu'à six semaines (PURCELL, WORSHAM et FRIEDLANDER, 2006).

FIGURE 1.1 – Exemple de charbon cutané. Ulcération escarrotique de la face postérieure de l'avant-bras. Tiré de WIKIPÉDIA.

1.2.2 La forme digestive

La variante digestive, peu courante chez l'Homme, est marquée par un taux de décès significatif oscillant entre 25% et 60% (ABADIA et al., 2005). Les signes cliniques n'étant pas distinctifs, cela conduit souvent à une sous-évaluation du nombre de cas. La durée d'incubation ressemble à celle de la forme cutanée, c'est-à-dire jusqu'à une semaine. Sans traitement, le taux de survie est inférieur à 50%, mais avec un traitement approprié, il monte à 60% et le rétablissement complet est généralement observé entre 10 et 14 jours (HILMAS et ANDERSON, 2015). Deux formes de charbon digestif sont décrites : la forme gastro-intestinale et la forme oropharyngée. Dans la

2. Atteinte inflammatoire des nœuds lymphatiques (ganglions) qui augmentent de volume et sont douloureux sous l'influence de tous les processus infectieux ou tumoraux.

3. Inflammation, le plus souvent infectieuse, aigüe ou chronique, d'un ou de plusieurs vaisseaux lymphatiques.

4. État pathologique caractérisé par la présence de toxines dans le sang.

forme oropharyngée, la porte d'entrée est orale ou pharyngée (mucus). La maladie commence par un ulcère de la muqueuse, suivi par un œdème localisé et une lymphadénopathie locale. Dans la forme gastro-intestinale, la porte d'entrée est l'ileum terminal[5] ou le cæcum[6] (WORLD HEALTH ORGANIZATION, 2008). Des lésions intestinales se développent et sont suivies par une lymphadénopathie. Un œdème de la paroi intestinale peut se former. Il est à noter que les bacilles peuvent se disséminer dans la circulation sanguine et engendrer une toxémie. La dose infectieuse chez l'Homme est encore méconnue : elle semble pouvoir être très faible, puisque parfois il a été impossible de retrouver l'origine de la contamination (JERNIGAN et al., 2002). L'existence de différences de susceptibilité entre individus est probable. Chez l'animal de laboratoire, cobaye, lapin ou singe rhésus, certaines études n'ont pas permis de déclencher un charbon digestif après ingestion de 10^8 spores administrées par voie orale. Ces animaux sont considérés comme étant moins sensibles que l'Homme à la maladie (BEATTY et al., 2003).

1.2.3 La forme pulmonaire

La variante pulmonaire du charbon est extrêmement mortelle, présentant une période d'incubation d'un à six jours. Une durée d'incubation de 43 jours a été rapportée dans le cas de l'accident de Sverdlovsk ; cette information reste néanmoins sujette à caution compte tenu des circonstances (GAINER, VERGNAUD et HUGH-JONES, 2020). Sans traitement, le taux de mortalité peut grimper jusqu'à 85%, et il reste à 45% même après une intervention médicale intensive (DOGANAY et DEMIRASLAN, 2015 ; CENTERS FOR DISEASE CONTROL AND PREVENTION, 2020). Les spores sont phagocytées puis drainées jusqu'aux ganglions lymphatiques proximaux[7] (WORLD HEALTH ORGANIZATION, 2008). Les spores germent et donnent des formes végétatives, qui quittent les macrophages et se multiplient dans le système lymphatique. Les bactéries envahissent la circulation sanguine et sécrètent des toxines qui provoquent un choc septique avec une méningite hémorragique. Il se produit alors une hémorragie dans les tissus lymphatiques péri bronchiques. Le drainage lymphatique est bloqué et un œdème pulmonaire apparaît. Les effusions pleurales[8] sont courantes et intenses. De vraies pneumonies se développent dans de rares cas ; des lésions pulmonaires nécrosantes sont alors observées. La présence d'effusions pleurales et d'une dilatation du médiastin[9] aide à différencier le charbon d'une simple pneumonie (FRAZIER, FRANKS et GALVIN, 2006). L'insuffisance respiratoire et le choc septique sont les causes les plus fréquentes du décès.

1.2.4 La forme injectionnelle

La forme injectionnelle du charbon est la plus rare. Cependant, elle est associée à une mortalité élevée qui reste au-dessus de 33%, même avec des soins appropriés. Cette forte mortalité est due à une augmentation des complications potentielles,

5. Dernière section de l'iléon, la troisième et dernière partie de l'intestin grêle chez l'homme, située entre le jéjunum et le cæcum.

6. Première partie du gros intestin, située à la jonction entre l'intestin grêle et le gros intestin. Il se présente sous la forme d'une poche aveugle, d'où son nom.

7. Structures nodulaires du système lymphatique situées à proximité immédiate d'une zone du corps subissant une réaction immunitaire, telle qu'une infection ou une inflammation.

8. Accumulation anormale de fluide entre les feuillets de la plèvre, membranes entourant les poumons et tapissant la cavité thoracique.

9. Région anatomique centrale du thorax, située entre les deux poumons, contenant le cœur, l'aorte, l'œsophage, la trachée, ainsi que diverses structures nerveuses et lymphatiques.

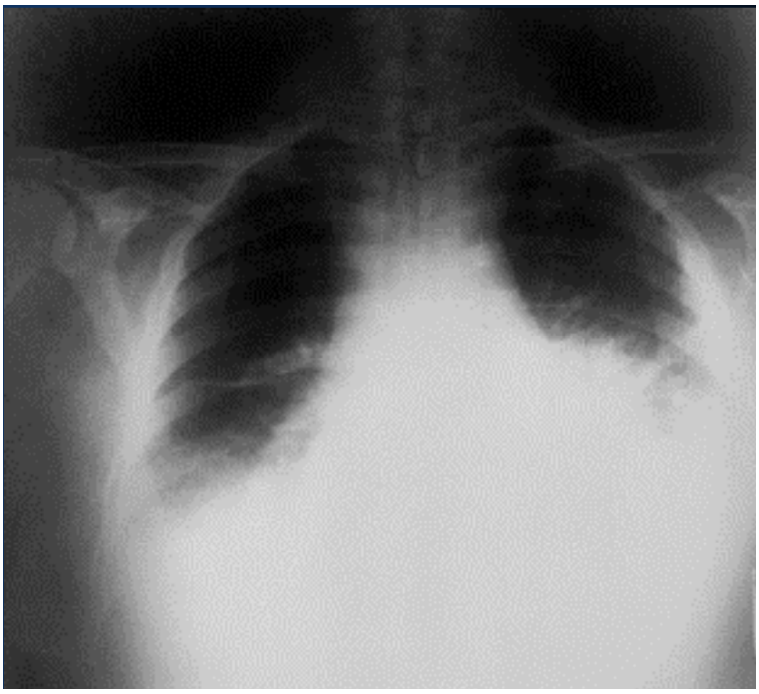

FIGURE 1.2 – Radiographie montrant un épanchement pleural et un élargissement médiastinal caractéristique du charbon pulmonaire. Tiré de CENTERS FOR DISEASE CONTROL AND PREVENTION.

comme les méningites et le choc septique (ZASADA, 2018). Cette forme d'infection résulte de l'injection intraveineuse de lot d'héroïne contaminée. Lorsque l'infection est provoquée par une injection, que ce soit par voie intraveineuse, intramusculaire ou sous-cutanée, elle pénètre profondément dans les tissus. Suite à une période d'incubation pouvant durer jusqu'à dix jours, le rétablissement peut nécessiter une hospitalisation d'une durée pouvant atteindre un mois (ZASADA, 2018). L'infection provoque une sévère atteinte des tissus mous, accompagnée d'un œdème significatif au niveau du site injecté (BERGER, KASSIRER et ARAN, 2014). La dose infectieuse n'a pas pu être déterminée, elle est probablement faible puisque aucune souche n'a pu être cultivée à partir des produits incriminés (BERGER, KASSIRER et ARAN, 2014).

1.2.5 Complications

En complément des manifestations cliniques spécifiées selon la manière dont la bactérie pénètre l'organisme, d'autres symptômes peuvent survenir. Par exemple, le sepsis provoqué par *B. anthracis* et les méningites, qui résultent de la propagation de la bactérie à travers la barrière hémato-méningée, peuvent se développer. Ces complications peuvent survenir indépendamment de la manière dont la bactérie est entrée dans le corps. La méningite est une complication redoutée de la maladie du charbon, bien qu'elle soit peu courante. Cependant, lorsqu'elle survient, son issue est souvent mortelle, avec un taux de décès dépassant 90%. Dans la moitié des situations, elle découle d'une infection pulmonaire initiale. Néanmoins, cette complication peut se manifester à partir de n'importe quelle forme de la maladie du charbon (SEJVAR, TENOVER et STEPHENS, 2005 ; DOGANAY et DEMIRASLAN, 2015). Quant à la septicémie, elle est due à la dispersion de *B. anthracis* via le système lymphatique et sanguin. Cette complication est couramment associée aux formes pulmonaires et

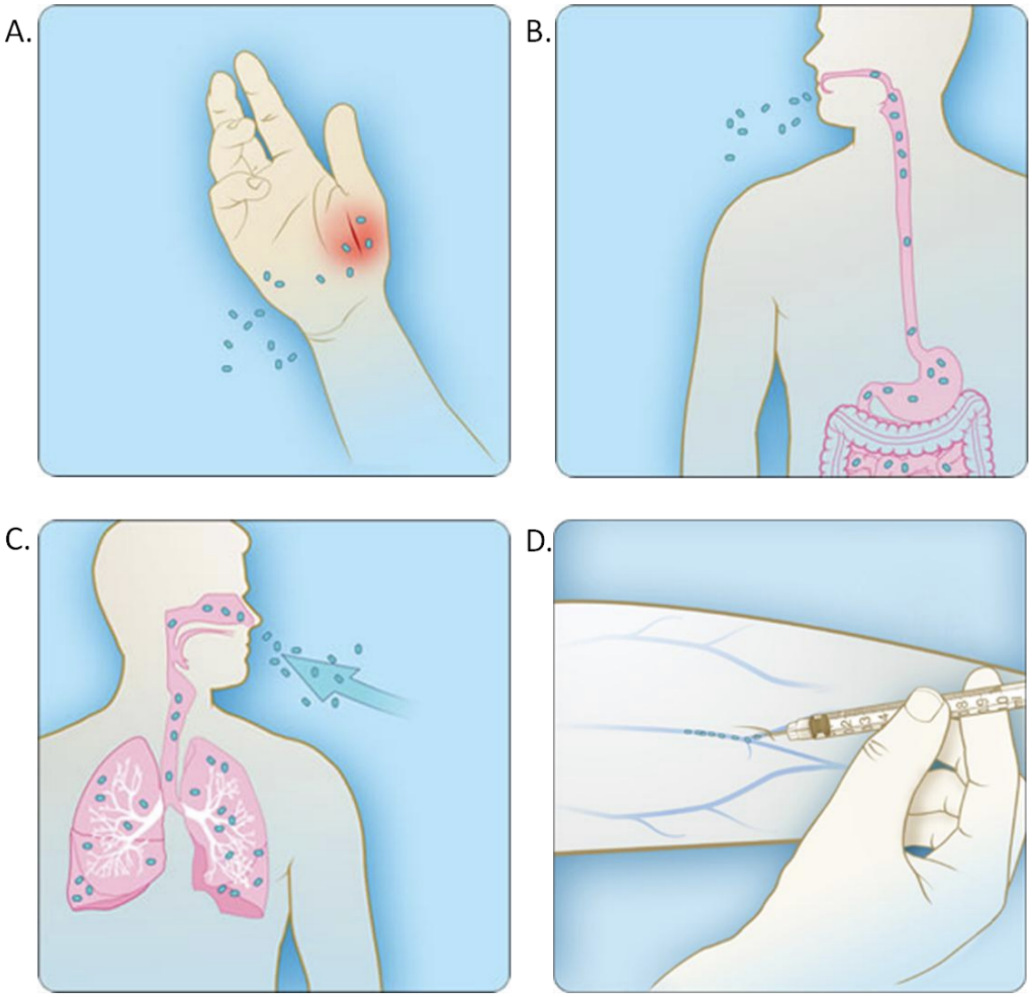

FIGURE 1.3 – Récapitulatif des formes de contamination à *B. anthracis* chez l'Homme. A) Forme cutanée. B) Forme digestive. C) Forme pulmonaire. D) Forme injectionnelle. Tiré de GIRAULT, 2015 (adapté de CENTERS FOR DISEASE CONTROL AND PREVENTION, 2023).

gastro-intestinales de l'infection. Les symptômes majeurs comprennent des difficultés respiratoires, une altération de la conscience et de la fièvre. Étant donné la capacité de la bactérie à produire des toxines et en l'absence de traitements qui ciblent spécifiquement ces toxines, ces complications sont généralement associées à un pronostic défavorable.

1.2.6 Cas de charbon atypique

Le phénomène porte sur des cas humains rapportés entre 1900 et 2004, infectés au charbon, qui n'ont pas eu de points d'entrée cutanés, gastro-intestinaux ou par inhalation connus. Parmi les quarante-deux individus hypothétiquement infectés par inhalation, les symptômes manifestés étaient atypiques (HOLTY, KIM et BRAVATA, 2006). Ces patients présentaient principalement une méningite initiale accompagnée de fièvre et de délire. D'autres symptômes comprenaient un charbon localisé au niveau du larynx, provoquant un œdème du cou et du visage, des saignements nasaux, une obstruction sinusale, et moins de manifestations classiques comme les douleurs

thoraciques et la toux. Après contamination, la maladie s'est localisée dans la partie haute du tractus respiratoire.

1.2.7 Maladie chez les animaux

Le charbon, affectant accidentellement l'Homme, est principalement une maladie animale. Les manifestations de cette maladie diffèrent selon les espèces affectées et reflètent à la fois une toxémie et une septicémie. Les symptômes les plus courants incluent l'apparition d'œdèmes, des troubles dans la coagulation sanguine, une insuffisance respiratoire ainsi que des hémorragies internes qui peuvent affecter la rate, les reins ou la vessie. Chez le mouton par exemple, la forme cutanée de la maladie est foudroyante : le décès survient quelques minutes à quelques heures (cinq minutes à douze heures) après l'apparition des premiers symptômes. Chez les équidés et les bovidés, la maladie évolue plus lentement vers la mort, en 12 à 48 heures après l'apparition des premiers symptômes. Chez le chien et le porc, espèces peu sensibles, la maladie se manifeste par un œdème de la gorge et des troubles gastro-intestinaux (WORLD HEALTH ORGANIZATION, 2008 ; DROMIGNY, 2009).

1.3 Le mode d'action de *Bacillus anthracis*

La section suivante présente la composition génomique de cette bactérie, en soulignant ses facteurs de virulence et son cycle infectieux, pour une compréhension de son mode d'action. L'objectif est de décrire la manière dont *B. anthracis* interagit avec son hôte, éclairant les interactions fondamentales entre l'agent pathogène et son environnement.

1.3.1 Composition du génome de *B. anthracis*

B. anthracis, l'agent pathogène causant la maladie du charbon, possède un génome constitué d'un chromosome circulaire et de deux plasmides. Le chromosome, d'une taille de 5.2 millions de paires de base, renferme la majorité des gènes nécessaires aux fonctions vitales de la bactérie. Le contenu en bases guanine-cytosine (GC) de ce chromosome est d'environ 35%, ce qui est typique des bactéries du genre *Bacillus*. Le nombre de gènes codant des protéines se situe généralement autour de 5,500 à 6,000 incluant des gènes essentiels pour le métabolisme, la réplication de l'ADN, la transcription, la traduction, et d'autres processus cellulaires fondamentaux (READ et al., 2003).

En plus de son chromosome, *B. anthracis* se caractérise par la présence de deux plasmides, pXO1 et pXO2, permettant sa pathogénicité. Le plasmide pXO1, d'une taille de 182 kilobases (kb), contient les gènes codant les composants de la toxine du charbon, qui sont des facteurs de virulence (MIKESELL et al., 1983). Il s'agit des gènes codant le facteur létal (LF, codé par le gène *lef*, ROBERTSON et LEPPLA, 1986), le facteur œdématogène (EF, codé par le gène *cya*, TIPPETTS et ROBERTSON, 1988) et l'antigène protecteur (PA, codé par le gène *pagA*, VODKIN et LEPPLA, 1983). Ces gènes sont contenus dans un îlot de pathogénicité de 44.8 kb comportant en particulier l'opéron *gerX* et les régulateurs *atxA* et *pagR* (ce dernier appartenant à l'opéron *pag*, comportant également le gène *pagA*, HOFFMASTER et KOEHLER, 1999). Le plasmide pXO2, d'une taille de 96 kb, porte des gènes impliqués dans la synthèse de

la capsule, une structure polymérique qui protège la bactérie des défenses immunitaires de l'hôte (UCHIDA et al., 1985). Ce plasmide contient un îlot de pathogénicité d'une taille de 35 kb, incluant l'opéron *capABCDE* et les régulateurs *acpA* et *acpB*.

Ces plasmides sont primordiaux pour la virulence de *B. anthracis*, car ils codent des éléments qui permettent à la bactérie d'infecter son hôte et d'échapper à ses mécanismes de défense. En dehors de ces caractéristiques spécifiques, le génome de *B. anthracis* partage de nombreuses similitudes avec ceux d'autres membres du groupe *B. cereus*, ce qui souligne son appartenance à ce complexe de bactéries étroitement liées.

1.3.2 Les facteurs de virulence associés à *Bacillus anthracis*

Les souches virulentes de *B. anthracis* possèdent les deux plasmides pXO1 et pXO2 qui ne sont pas auto-transmissibles (AUWERA, ANDRUP et MAHILLON, 2005).

Les toxines produites par *B. anthracis* adoptent une structure A-B, où 'A' désigne la composante active et 'B' le domaine de fixation. Cette configuration binaire est couramment observée chez les bactéries pathogènes et toxiques. Pour *B. anthracis*, la partie 'A' se compose soit du facteur d'œdème (EF), soit du facteur létal (LF), tandis que la partie 'B' est représentée par l'antigène protecteur (PA). L'antigène protecteur a la capacité unique de se lier aux deux variantes de la partie 'A', EF ou LF. Chacune de ces protéines, isolément, est inactive tant *in vivo* qu'*in vitro* (THORNE, MOLNAR et STRANGE, 1960 ; BEALL, TAYLOR et THORNE, 1962). Les deux toxines produites par cette configuration sont les suivantes (LEPPLA, 1982 ; LEPPLA, 1988 ; LEPPLA, ARORA et VARUGHESE, 1999) :

— la toxine œdématogène : la combinaison de l'antigène protecteur (PA) et du facteur œdématogène (EF) forme la toxine œdématogène. EF est une adénylate cyclase dont l'activité enzymatique dépend de la présence de la calmoduline, une protéine eucaryote. PA permet l'entrée de EF dans les cellules cibles. Une fois à l'intérieur de la cellule, en présence de la calmoduline, EF catalyse la conversion de l'ATP en AMP cyclique (cAMP), conduisant à divers effets pathologiques, notamment l'œdème (BROSSIER et MOCK, 2001 ; WORLD HEALTH ORGANIZATION, 2008).

— la toxine létale : la toxine létale est formée par l'association de l'antigène protecteur (PA) et du facteur létal (LF). LF est une métalloprotéase à zinc qui cible spécifiquement les MAP kinases kinases (MAPKK1, 2, 3, 4, 6 et 7). Ces enzymes jouent un rôle dans la chaîne d'activation MAPK, permettant le transfert d'informations de l'extérieur de la cellule vers le noyau au niveau transcriptionnel. PA facilite le transport de LF dans le cytosol de la cellule cible, où il peut exercer son activité enzymatique en coupant les MAPKKs, perturbant la signalisation cellulaire et provoquant la mort cellulaire (BROSSIER et MOCK, 2001 ; WORLD HEALTH ORGANIZATION, 2008).

Le mode d'action de ces toxines protéiques, agissant en synergie pour la pathogénèse du charbon, est schématisé sur la figure 1.4.

Le plasmide pXO2 de *B. anthracis* porte les gènes utiles à la synthèse de la capsule, autre facteur de virulence de cette bactérie. Cette capsule est composée d'acide poly-γ-D-glutamique, un polymère qui confère plusieurs propriétés pathogènes à la bactérie (BRUCKNER, KOVÁCS et DÉNES, 1953).

La formation de la capsule est stimulée en présence de dioxyde de carbone (CO2) à une température de 37°C (WORLD HEALTH ORGANIZATION, 2008). La capsule

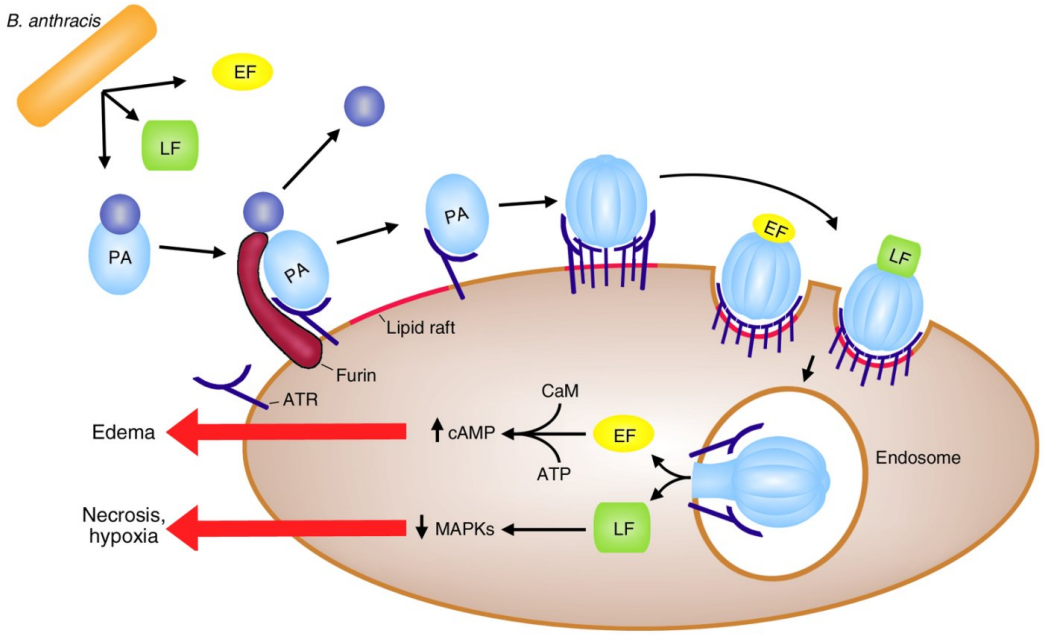

FIGURE 1.4 – Actions des toxines de *B. anthracis*. Tiré de PRINCE et al., 2003.

joue un rôle défensif pour *B. anthracis*, en s'opposant notamment à la phagocytose des cellules végétatives par les cellules immunitaires de l'hôte, permettant par ce biais à la bactérie de se propager et de se multiplier à l'intérieur de l'organisme infecté.

Les gènes *cap*, localisés sur le plasmide pXO2, sont au cœur de la biosynthèse de la capsule. Ces gènes comprennent *capA*, *capB*, *capC*, *capD* et *capE*, chacun étant impliqué différemment dans la formation de la capsule :

— *capA*, *capB*, *capC* : ces gènes codent des enzymes impliquées dans la synthèse de l'acide poly-γ-D-glutamique, formant la structure de base de la capsule (MAKINO et al., 1989).

— *capD* : ce gène code une enzyme nécessaire à l'ancrage de la capsule à la surface de la cellule bactérienne, assurant sa stabilité et sa persistance (UCHIDA et al., 1993 ; MAKINO et al., 2002).

— *capE* : ce gène participe à la synthèse initiale du polymère de la capsule (CANDELA, MOCK et FOUET, 2005).

La structure de la capsule, peu immunogène, permet à *B. anthracis* de se multiplier à l'intérieur de l'hôte sans provoquer de réponse immunitaire significative.

Un schéma récapitulatif des facteurs de virulence associés à *B. anthracis* est disponible à la figure 1.5.

B. anthracis possède également des régulateurs primordiaux dans l'expression de certains facteurs de virulence :

— AtxA, codé par le gène *atxA*, situé sur le plasmide pXO1, permet d'augmenter l'expression de l'antigène protecteur (PA) (KOEHLER, DAI et KAUFMAN-YARBRAY, 1994) et des gènes codant la capsule (GUIGNOT, MOCK et FOUET, 1997).

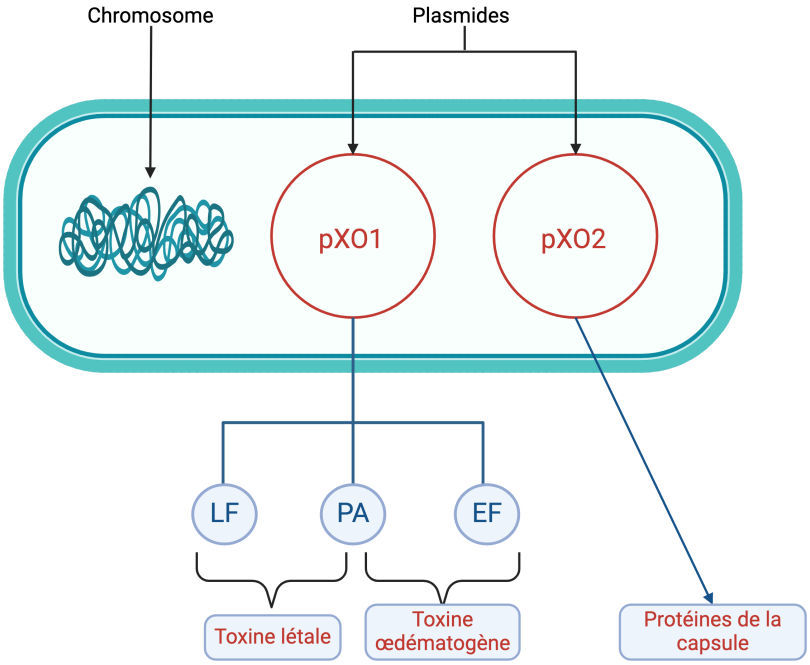

FIGURE 1.5 – Schéma récapitulatif des facteurs de virulence de *B. anthracis*. Créé avec BIORENDER.

— PlcR : chez *B. anthracis*, une mutation affectant l'opéron *plcR* induit une cessation prématurée de sa traduction, entraînant l'inactivité de la protéine résultante (KOLSTØ, TOURASSE et ØKSTAD, 2009). À l'inverse, dans la majorité des souches de *B. cereus* et *Bacillus thuringiensis*, PlcR reste actif et exerce une influence sur l'expression de plusieurs facteurs de virulence chromosomiques. Ces facteurs incluent, entre autres, des phospholipases, des protéases, des hémolysines et des entérotoxines (GOHAR et al., 2002 ; ØKSTAD et KOLSTØ, 2010).

— AcpA et AcpB, situés sur le plasmide pXO2 et codés par les gènes *acpA* et *acpB* respectivement, régulent l'opéron *cap* codant la capsule (DRYSDALE et al., 2004).

— L'opéron GerX, situé sur le plasmide pXO1 et organisé en triple opéron (*gerXB*, *gerXA* et *gerXC*) sert à réguler la germination de *B. anthracis*. Son expression favorise ce processus *in vitro* et *in vivo* (GUIDI-RONTANI et al., 1999b).

— AbrB, situé sur le chromosome, joue un rôle de régulateur de la transcription des gènes spécifiques à la production de toxines lors de leur phase de croissance (KOEHLER, 2002).

Ces processus de régulation sont illustrés sur la figure 1.6.

1.3.3 Cycle infectieux de *Bacillus anthracis*

La propriété de sporulation de *B. anthracis* permet à la bactérie de survivre dans des environnements défavorables, et ainsi de persister pendant des périodes prolongées en créant des réservoirs dans le sol. Par conséquent, dans son cycle de vie naturel, *B. anthracis* alterne entre une forme végétative active et une forme sporulée dormante. Dans la forme végétative, la bactérie se développe et se multiplie à l'intérieur de l'hôte infecté. Les bacilles végétatifs sont relativement grands, avec des

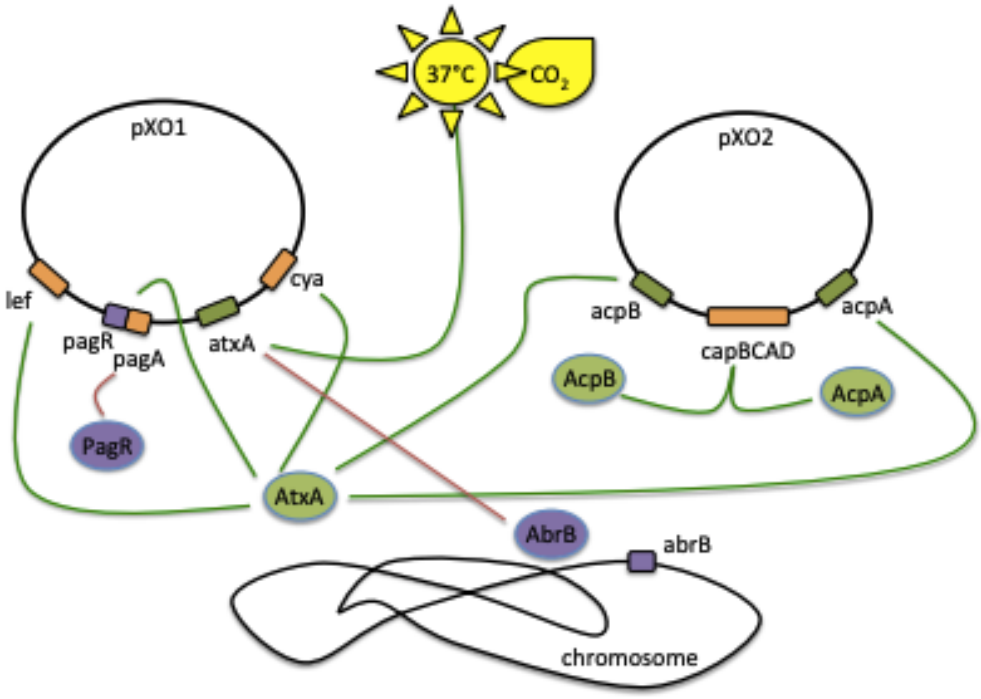

FIGURE 1.6 – Action des régulateurs de *B. anthracis*. En noir : chromosome et plasmides ; en vert : régulateurs positifs ; en violet : régulateurs négatifs, en orange : gènes de virulence. La régulation transcriptionnelle est représentée sous forme de lignes : vertes pour les activations, rouges pour les répressions. Tiré de TESSIER, 2022.

dimensions allant de 1 à 2 µm de diamètre et de 3 à 5 µm de longueur, et ils apparaissent souvent sous forme de chaînes longues et souples ou en tant que cellules individuelles isolées dans les fluides ou les tissus infectés (DROMIGNY, 2009).

Lorsque les conditions environnementales deviennent défavorables, la bactérie se transforme en spores. Les spores de *B. anthracis* sont typiquement oblongues, avec une taille moyenne de 1 µm de diamètre et une longueur de 1 à 2 µm. Cette forme sporulée est responsable de la transmission de la maladie et de sa persistance dans l'environnement en dehors de l'hôte. La spore de *B. anthracis* atteint sa maturité complète à l'intérieur de la cellule bactérienne avant que des enzymes ne lysent cette dernière. La spore ainsi libérée est capable de subsister dans l'environnement sans se reproduire jusqu'à ce qu'elle rencontre des conditions propices à la germination. Les spores de *B. anthracis* sont également enveloppées d'une couche externe nommée exosporium. Principalement protéique, cet exosporium se distingue par sa structure trilaminaire avec un motif hexagonal et est agrémenté de prolongements filamenteux (HACHISUKA, KOZUKA et TSUJIKAWA, 1984). Cet exosporium interagit en premier lieu avec les cellules de l'hôte infecté et joue un rôle de barrière protectrice, empêchant les enzymes extérieures d'endommager les structures internes de la spore, telles que la tunique et le cortex. Il contribue également à l'adhésion des spores aux tissus de l'hôte.

Lorsque ces spores entrent dans un organisme par inhalation, ingestion ou via une lésion cutanée, elles sont rapidement englobées par phagocytose, notamment par les macrophages, comme ceux présents dans les alvéoles pulmonaires (GUIDI-RONTANI et al., 1999a). Au sein de la vésicule de phagocytose et dans le cytoplasme cellulaire, la spore se réveille de sa dormance grâce à la germination : elle augmente

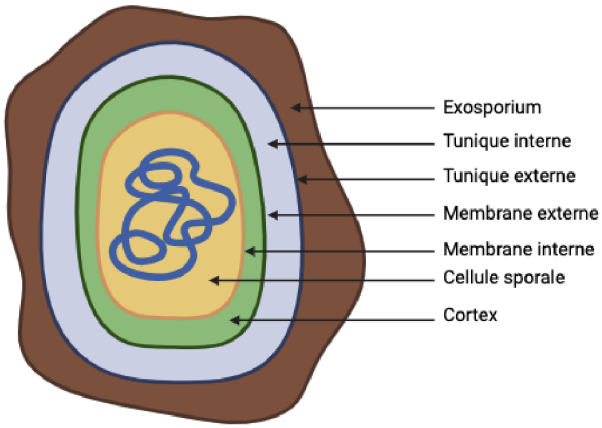

FIGURE 1.7 – Structure schématique d'une spore de *B. anthracis*. Créé avec BIORENDER.

de volume, sa tunique se rompt et son activité métabolique s'initie. Ce processus transforme la spore en une forme végétative, la bactérie étant active. Cette germination, qui prend généralement une trentaine de minutes, coïncide avec la libération des toxines caractéristiques de *B. anthracis* (DIXON et al., 1999). Plus précisément, la germination de la spore se déroule en deux phases séquentielles. Initialement, on observe une diminution de l'indice de réfraction et de la résistance aux contraintes thermiques et autres facteurs environnementaux des spores. Durant cette première phase, il n'y a pas d'accroissement du volume de la spore. La seconde phase est caractérisée par une activité de croissance cellulaire. Elle culmine avec l'émergence de la cellule germinée hors de la tunique de la spore. À ce stade, la cellule exprime les attributs typiques d'une cellule bactérienne active, sans division cellulaire. La germination est optimisée *in vitro* en présence d'ions bicarbonates et à une température de 37 °C, conditions qui imitent l'environnement physiologique de la plupart des mammifères. Ces conditions permettent la transition efficace de la spore au stade bactérien actif, utile à la propagation de l'infection.

B. anthracis, en plus de ses toxines, possède des caractéristiques structurales cellulaires qui contribuent à sa virulence. Ces structures sont des additions aux composants cellulaires standards que l'on trouve chez toutes les bactéries, tels que la membrane cellulaire et la paroi cellulaire. Parmi ces structures, on trouve la capsule, décrite précédemment, et la couche S (ou *S-layer*). Sous-jacente à la capsule, la couche S est une structure cristalline qui offre une protection supplémentaire aux cellules de *B. anthracis* (MESNAGE et al., 1998). Cette couche, relativement rare chez les bactéries capsulées, est composée de deux protéines principales, Sap et EA1. Ces protéines s'assemblent de manière autonome et forment un agencement régulier (ETIENNE-TOUMELIN et al., 1995 ; MESNAGE et al., 1997 ; MESNAGE, HAUSTANT et FOUET, 2001). Bien que la couche S ne modifie pas directement la dose létale de la bactérie, elle est supposée renforcer la résistance contre certaines défenses immunitaires de l'hôte. La synthèse de la protéine Sap se produit initialement pendant la phase de croissance exponentielle de la bactérie. À mesure que la bactérie atteint la phase stationnaire de croissance, la protéine Sap est progressivement remplacée par la protéine EA1, complétant la maturation de la couche S. Ces mécanismes complexes de structure cellulaire et de réponse au cycle de croissance sont primordiaux pour la capacité de *B. anthracis* à infecter et persister dans son hôte (MOCK et FOUET, 2001).

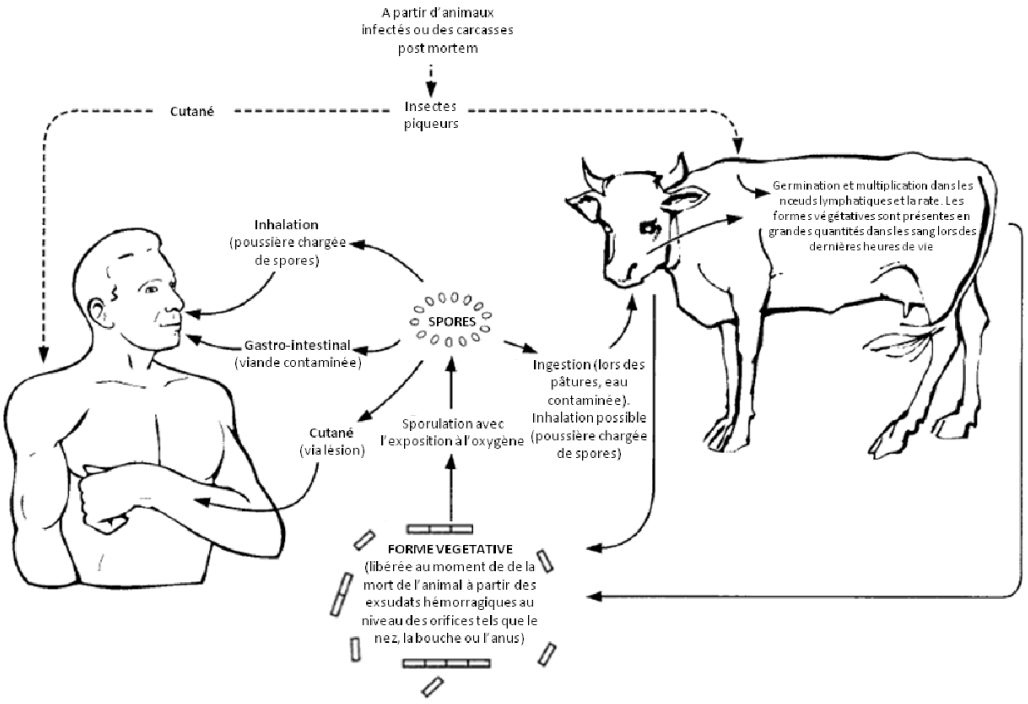

FIGURE 1.8 – Cycle infectieux du charbon. Tiré de GIRAULT, 2015 (adapté de WORLD HEALTH ORGANIZATION, 2008).

1.4 Utilisation de *Bacillus anthracis* à des fins malveillantes

Cette section détaille les principaux usages détournés, avérés ou suspectés, de *B. anthracis* à des fins malveillantes au cours de l'Histoire. Cela met en perspective comment la forte persistance et la simple dissémination de cette bactérie ont pu être exploitées pour un usage bioterroriste.

En termes de menace bioterroriste, les agents pathogènes ont été catégorisés aux États-Unis en trois groupes prioritaires (A, B et C) selon différents critères : leur aptitude à se propager, le taux de mortalité qu'ils engendrent, l'état de préparation des services de santé publique et l'angoisse qu'ils provoqueraient dans la population. *B. anthracis* est classé dans le groupe prioritaire A, en compagnie d'autres pathogènes tels que *Clostridium botulinum* (responsable du botulisme), *Yersinia pestis* (responsable de la peste), *Variola major* (causant la variole), *Francisella tularensis* (causant la tularémie), et les virus tels que Lassa, Machupo, Dengue, Ebola et Marburg (NATIONAL INSTITUTE OF ALLERGY AND INFECTIOUS DISEASES, 2018 ; CENTERS FOR DISEASE CONTROL AND PREVENTION, 2018).

L'usage du charbon comme arme biologique remonterait à des périodes anciennes, bien que les détails précis de ces utilisations soient parfois flous en raison de la limitation des sources historiques. Durant l'Antiquité, aux environs de 1,500 avant notre ère, les Hittites auraient utilisé des animaux infectés par le charbon pour affaiblir et déstabiliser leurs ennemis. Ils auraient envoyé des moutons et autres bêtes malades dans les zones ennemies pour propager la maladie (ROLAND, 2005). Il est à noter que ce type d'événements et l'attribution de l'usage du charbon à des fins de guerre biologique durant l'Antiquité et le Moyen-Âge de manière générale est largement sujette à débat parmi les historiens et les chercheurs en raison de la difficulté

d'identification précise des agents pathogènes avant l'époque moderne.

Durant la Première Guerre Mondiale, l'utilisation d'armes biologiques, y compris le charbon, était encore largement expérimentale, et les informations précises sur leur usage opérationnel sont limitées. Bien que des recherches aient été menées sur le charbon comme potentiel agent de guerre biologique, il n'y a pas de documentation solide sur son utilisation réelle sur le champ de bataille durant ce conflit. Les pays impliqués dans la guerre, tels que l'Allemagne, le Royaume-Uni et les États-Unis, ont exploré le potentiel de divers agents biologiques, mais le consensus historique est que ces armes n'ont pas été déployées à grande échelle. Cela dit, durant la Première Guerre Mondiale, l'Allemagne a été accusée d'avoir utilisé *B. anthracis* et d'autres agents biologiques pour infecter des animaux (chevaux et bétail) expédiés vers des alliés, notamment la Russie, la Roumanie et l'Argentine (dans le cadre d'opérations de type "sabotage"). Si c'est le cas, et au vu de l'absence de preuve, on peut supposer que ces "armes" n'ont pas été très efficaces.

L'Unité 731 est créée par l'Empire du Japon en 1937 dans la province occupée du Mandchoukouo, en Chine du Nord-Est. Sous le commandement du Lieutenant-Général Shiro Ishii, cette entité secrète de l'Armée impériale japonaise mène des recherches sur les armes biologiques, dont *B. anthracis*. Dans le cadre de leurs expérimentations, les scientifiques de l'Unité 731 exposent délibérément des prisonniers de guerre au charbon, dans le but d'étudier l'évolution de la maladie et d'évaluer l'efficacité de l'agent en tant qu'arme biologique. Malgré ces tests, les résultats de ces expérimentations ne conduisent pas à l'utilisation opérationnelle de *B. anthracis* sur le champ de bataille pendant la Seconde Guerre mondiale (BARRAS et GREUB, 2014). Toujours dans ce même contexte, le Royaume-Uni explore également l'utilisation potentielle de cette bactérie en tant qu'arme biologique. Les autorités britanniques initient l'Opération Vegetarian, un projet destiné à affaiblir l'économie allemande en décimant son bétail. Pour mettre en œuvre ce plan, des scientifiques britanniques développent des "cakes" de lin contaminés par des spores de *B. anthracis*. Ces tourteaux, conçus pour être largués sur les pâturages allemands et être consommés par le bétail, aurait peut-être pu affecter l'approvisionnement en viande de l'Allemagne. Bien que près de cinq millions de ces "cakes" aient été produits, l'Opération Vegetarian ne sera jamais mise en action, et le stock de cakes sera détruit après la guerre (ALIBEK, LOBANOVA et POPOV, 2009).

L'île de Gruinard, située en Écosse, a été au cœur des expérimentations relatives à la guerre biologique durant la Seconde Guerre mondiale. En 1942, un essai y est conduit par le largage de bombes remplies de spores de *B. anthracis* (MANCHEE et al., 1981). Ce test couvre de spores l'ensemble de l'île, conduisant à la mort rapide des moutons qui y avaient été placés pour l'expérience. Face à la contamination causée par ces essais, l'île est mise en quarantaine afin de prévenir toute dissémination accidentelle sur le territoire écossais. Ce n'est qu'en 1986, plusieurs décennies plus tard, qu'une initiative de décontamination est entreprise. Cette opération inclut l'aspersion de l'île avec une solution de formol, résultant de la dissolution de 280 tonnes de gaz formaldéhyde dans deux millions de litres d'eau de mer (MANCHEE et al., 1983). Afin de confirmer l'efficacité de la décontamination, des moutons sont de nouveau introduits sur l'île et sont observés pendant une période de cinq mois. L'absence de symptômes chez ces animaux atteste de l'efficacité des mesures de décontamination (MANCHEE et al., 1994). 1990 marque la fin officielle de la mise en quarantaine de l'île et la réouverture de l'île au public et à la circulation.

Le 2 avril 1979, dans la ville de Sverdlovsk, située dans l'Union Soviétique (aujourd'hui Ekaterinbourg, en Russie), un événement a mis en évidence les dangers de la production à grande échelle de *B. anthracis*. Cet événement est connu sous le nom

FIGURE 1.9 – Expérimentations au sein d'un laboratoire de l'unité 731. Tiré de WIKIPÉDIA.

FIGURE 1.10 – Expert venu décontaminer l'île de Gruinard (Écosse) en 1986. Tiré de PA MEDIA.

d'incident de Sverdlovsk (MESELSON et al., 1994). Dans une installation de recherche et de production d'armes biologiques de Sverdlovsk, des spores de *B. anthracis* ont été libérées accidentellement dans l'atmosphère. La fuite est attribuée à une erreur humaine à l'occasion d'une opération de maintenance du système de filtration d'air de l'installation (ALIBEK, LOBANOVA et POPOV, 2009). Les spores, libérées dans l'air, ont été transportées par le vent sur une zone d'au moins 50 km^2, exposant des milliers de personnes ainsi que des animaux à l'infection. Au moins 66 personnes sont mortes à la suite de l'exposition aux spores de *B. anthracis*, bien que certains rapports suggèrent que le nombre de décès pourrait être plus élevé. De nombreuses

personnes ont souffert de symptômes graves d'infection au charbon, y compris de fièvre élevée, de difficultés respiratoires et de choc septique. Les victimes étaient principalement des hommes, ce qui serait lié à l'heure à laquelle l'incident est survenu. En plus des décès humains, une grande quantité d'animaux, principalement des bovins, ont également péri à cause de l'exposition. L'Union Soviétique a initialement tenté de dissimuler la véritable nature de l'incident, attribuant les décès à la consommation de viande contaminée. Les autorités ont rapidement procédé à la décontamination de la zone affectée, mais les informations sur l'incident ont été fortement contrôlées. Ce n'est que des années plus tard, après la chute de l'Union Soviétique, que la vérité sur l'incident de Sverdlovsk a été pleinement révélée. Les investigations internationales, menées principalement par des scientifiques et des experts en armes biologiques russes et américains, ont mis en lumière la réalité de l'incident de Sverdlovsk. Les études épidémiologiques, les autopsies et les analyses de sols ont confirmé que la cause des décès était bien une inhalation de spores de *B. anthracis*, et non la consommation de viande contaminée comme l'avaient prétendu les autorités soviétiques (ABRAMOVA et al., 1993).

La même année, en 1979, le Zimbabwe (alors connu sous le nom de Rhodésie), a été le théâtre d'une épidémie de charbon sans précédent. Cette épidémie a coïncidé avec la guerre de libération qui a secoué le pays. L'épidémie de charbon a débuté en 1978 et s'est intensifiée en 1979, touchant principalement les régions rurales et agricoles du pays. *B. anthracis* a infecté à la fois les animaux et les humains, causant la mort de milliers de têtes de bétail et d'environ 200 personnes. 10,738 cas d'infection humaine ont été enregistrés, marquant l'une des plus grandes épidémies de charbon de l'histoire moderne. Certains experts et témoins ont suggéré que l'épidémie pourrait avoir été exacerbée ou délibérément causée dans le cadre d'efforts de guerre biologique par le gouvernement blanc de la Rhodésie, qui luttait contre les forces de libération nationalistes noires. Les soupçons sont alimentés par le fait que l'épidémie a principalement affecté les régions rurales peuplées par des communautés noires, qui étaient largement perçues comme des sympathisants des forces de libération. D'autres explications ont été également proposées, dont l'abandon de la vaccination dans la zone de conflit, et la consommation accrue de viande contaminée en période de pénurie alimentaire (WILSON et al., 2016).

Sous le régime de Saddam Hussein, l'Irak a entamé un programme d'armes biologiques au début des années 1980, notamment pendant la guerre Iran-Irak. La nation aspirait à créer une puissante capacité de dissuasion militaire face à ses adversaires régionaux, avec un accent particulier mis sur *B. anthracis* (ZILINSKAS, 1997). L'Irak a mis en place des infrastructures significatives pour la production de *B. anthracis*. La bactérie a été cultivée en grande quantité dans des fermenteurs industriels avant d'être lyophilisée pour produire des spores sèches. Selon les documents de l'ONU, l'Irak a réussi à produire et à stocker au moins 8,500 litres de spores concentrées de *B. anthracis* en 1991 (« Report of the Secretary-General on the status of the implementation of the Special Commission's plan for the ongoing monitoring and verification of Iraq's compliance with relevant parts of section C of Security Council resolution 687 (1991). » 1994). En vue de leur utilisation militaire, ces spores ont été traitées avec des additifs et des stabilisants pour assurer leur viabilité et leur dissémination efficace. Ces préparations ont été incorporées dans diverses munitions, dont des obus d'artillerie et des bombes aériennes, renforçant l'arsenal militaire irakien avec une capacité de guerre biologique potentielle. Cependant, après la Guerre du Golfe de 1991, l'Irak a été contraint de déclarer et de démanteler son programme d'armes biologiques sous la supervision de l'ONU, en accord avec les résolutions du Conseil de sécurité de l'ONU. Bien que l'Irak ait affirmé avoir détruit tous ses stocks

de *B. anthracis* et d'autres agents biologiques, des doutes ont persisté concernant la véracité de ces affirmations et la possibilité que certaines de ces armes aient échappé à la destruction (SEELOS, 1999).

La secte japonaise Aum Shinrikyo, créée dans les années 1980 par Shoko Asahara, est un exemple marquant de tentative d'utilisation de *B. anthracis* à des fins bioterroristes par une organisation non gouvernementale. La secte de type "apocaliptique", qui était impliquée dans une série d'activités illégales (notamment l'attentat du métro de Tokyo au sarin en 1995), avait une division dédiée à la production d'armes biologiques et avait réalisé plusieurs expérimentations avec divers agents pathogènes, dont *B. anthracis* (OLSON, 1999). En 1993, les membres de la secte ont libéré des spores de *B. anthracis* depuis un immeuble du centre-ville de Tokyo, en utilisant un système de pulvérisation installé sur le toit de l'immeuble. L'attaque n'a pas réussi à provoquer l'infection escomptée, la souche utilisée étant une souche vaccinale (TAKAHASHI, 2004). C'est l'odeur dégagée qui alertera le voisinage.

En 2001, les États-Unis ont été frappés par une attaque bioterroriste impliquant l'utilisation d'enveloppes contenant des spores de la souche virulente *Ames* de *B. anthracis*. L'attaque a commencé quelques jours après les attaques terroristes du 11 septembre, lorsque des enveloppes contenant des spores lyophilisées ont été envoyées à diverses personnalités médiatiques et politiques, déclenchant une crise majeure. La forme utilisée a permis aux particules d'être facilement inhalées, provoquant le charbon pulmonaire, la variante la plus mortelle de la maladie. La qualité de la préparation des spores a indiqué que le ou les responsable(s) de ces attaques possédai(en)t des compétences scientifiques avancées et avai(en)t accès à des laboratoires spécialisés. Les premières enveloppes sont apparues le 18 septembre 2001, adressées à des journalistes de renom de médias nationaux. Dans les jours et les semaines qui ont suivi, d'autres enveloppes ont été envoyées, ciblant également des sénateurs américains. Les spores contenues dans ces enveloppes ont infecté 22 personnes, causant cinq décès (RASKO et al., 2011). La crise a déclenché une réponse massive de la part des autorités sanitaires et de sécurité, en particulier des Centres pour le contrôle et la prévention des maladies (CDC). Initialement, l'hypothèse d'un lien direct entre les attentats par avion et l'attentat bioterroriste était la plus plausible, mais l'équipe de Paul Keim (BHATTACHARJEE, 2009), pionnière dans le domaine, a très rapidement identifié la souche impliquée, grâce à ses travaux menés durant des années sur la problématique de distinction des souches de *B. anthracis* afin de retracer leur origine (par typage MLVA notamment) (BHATTACHARJEE, 2009 ; PBS FRONTLINE, 2011). Les efforts de décontamination des bâtiments affectés ont été intenses, incluant des bureaux de poste, des bureaux de médias et des bâtiments gouvernementaux, s'étalant sur plusieurs mois (RASKO et al., 2011). Le FBI, quant à lui, a mené l'enquête criminelle, initiant l'une des plus grandes et des plus complexes enquêtes de son histoire, intitulée "Amerithrax". L'enquête conduira à conserver en 2008 un unique suspect, le Dr. Bruce Ivins, un scientifique d'un établissement de défense du gouvernement américain. Le suicide du Dr. Ivins mettra fin aux investigations. Cet événement a souligné de manière alarmante la vulnérabilité des systèmes postaux et des infrastructures aux attaques bioterroristes. Alors que l'impact de l'évènement en termes de santé publique aura été modeste, il aura été majeur en termes de désorganisation du fait de la multiplication de l'envoi de canulars à base de poudres diverses dans le monde entier. En France par exemple, où les colis ont été pris au sérieux, environ 3,000 colis suspects ont été traités par les autorités, principalement au sein des établissements de défense (HAGENBOURGER, 2003). En conséquence, il a entraîné une augmentation des investissements dans la biosécurité et la préparation aux urgences de santé publique aux États-Unis principalement et à des degrés

variables dans le reste du monde. Une vision stratégique pour ce type de menaces a par exemple été mise en place dans l'état fédéral américain (THE WHITE HOUSE, 2009). Le développement de la *microbial forensics* (ou "forensique microbienne") a ainsi été encouragé, impliquant l'identification, la caractérisation et la comparaison de preuves microbiologiques dans le but d'établir une source ou une origine de ces agents, de retracer la propagation d'une maladie ou d'un agent biologique, et de contribuer à l'élaboration de stratégies de prévention et de réponse. Plus largement, la stratégie *One Health*, prônant le dépassement des frontières traditionnelles de la science médicale, vétérinaire et agronomique par une collaboration renforcée entre les différents acteurs de la santé publique et animale, a permis la mise en place d'outils aujourd'hui usuels pour le suivi de maladies infectieuses (GIBBS, 2014).

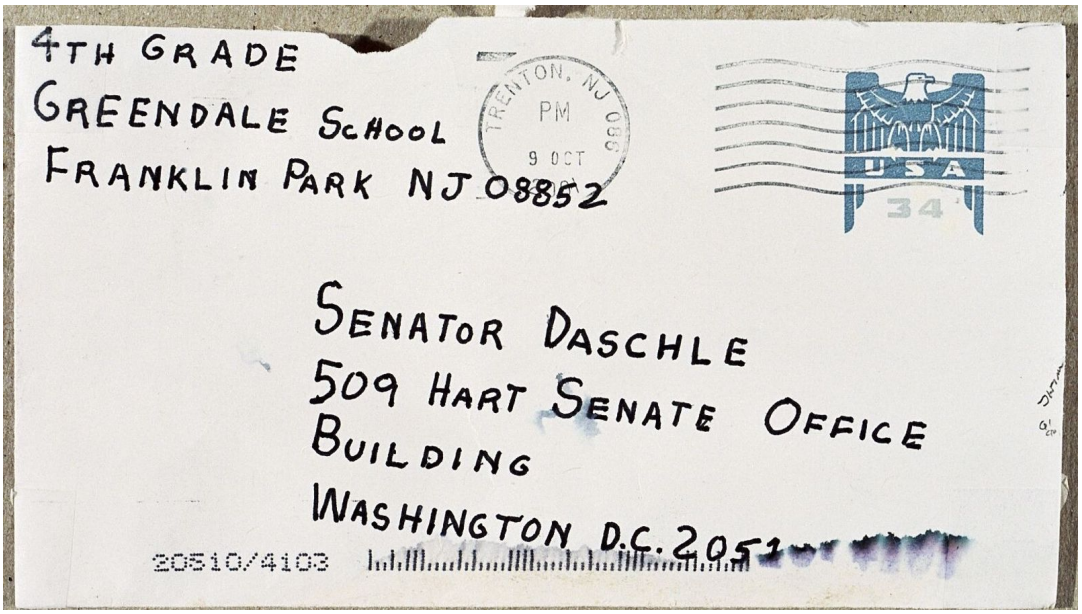

FIGURE 1.11 – Enveloppe contaminée par *B. anthracis* en 2001 aux États-Unis (Source : FBI).

Aujourd'hui, malgré l'adoption de la Convention sur l'interdiction des armes biologiques et à toxines (signée en 1972 et entrée en vigueur en 1975), la menace de l'utilisation de ce type d'armes en particulier à des fins terroristes reste réelle. En effet, l'évolution de la situation internationale ainsi que la multiplication des acteurs (étatiques, non étatiques ou terroristes) pouvant user de cette menace démontre que l'utilisation d'armes biologiques, et en particulier de *B. anthracis*, demeure un risque d'actualité (*Rapport d'information déposé en application de l'article 145 du règlement, par la commission de la défense nationale et des forces armées, en conclusion des travaux d'une mission d'information sur la défense NRBC* 2022). On peut cependant noter que aucun évènement notable n'est survenu depuis 2001, ce qui pourrait laisser penser que les mesures prises ont eu une certaine efficacité et/ou que la mise au point de telles armes n'est pas si simple.

1.5 Méthodes de détection conventionnelles de *Bacillus anthracis*

Pour des raisons de santé publique et de biodéfense, la détection de *B. anthracis* a fait l'objet de l'élaboration de multiples méthodes. Cette section détaille les principales méthodes conventionnelles de détection de *B. anthracis*, en particulier en microbiologie et en biologie moléculaire.

Identification phénotypique Les méthodes microbiologiques traditionnelles, passant généralement par une mise en culture à partir d'un échantillon biomédical, ont fait figure de référence en termes d'identification bactérienne, en particulier pour *B. anthracis*. Ces méthodes impliquent typiquement la culture sur des milieux sélectifs, l'observation de l'absence de mobilité et d'hémolyse, la réalisation de colorations de la capsule, la lyse par le phage γ, la réaction en « collier de perles » (*String of Pearls reaction*), et la vérification de la sensibilité à la pénicilline (ZASADA, 2020).

— L'identification initiale de *B. anthracis* se fait généralement par une coloration de Gram. La détection de bacilles en chaînes, ressemblant à des tiges de bambou, à Gram positif, encapsulés et parfois sporulés, suggère fortement la présence de *B. anthracis*.

— Les observations phénotypiques comme l'absence de mobilité ou d'hémolyse sont également caractéristiques de la présence de la bactérie.

— Les milieux sélectifs utilisés en laboratoire pour la détection de *B. anthracis* contiennent des composés permettant à la bactérie de se développer tout en inhibant les autres microorganismes présents. Le plus utilisé est le milieu PLET agar, composé de polymyxine, lysozyme, EDTA (acide éthylène-diamine-tétraacétique) et d'acétate de thallium (KNISELY, 1966). Ce milieu, bien qu'étant actuellement le plus efficace pour l'isolement de *B. anthracis*, est prohibé dans certains pays car l'acétate de thallium est un composé hautement toxique (TOMASO et al., 2006). Moins sélectif que PLET agar (MARSTON et al., 2008), le milieu ACA (*Anthracis Chromogenic Agar*) (JUERGENSMEYER et al., 2006), composé de 5-bromo-4-chloro-3-indoxyl-myoinositol-1-phosphate et de 5-bromo-4-chloro-3-indoxyl-choline phosphate, est aussi utilisé.

— La coloration de la capsule peut se faire de plusieurs manières : par la réaction de M'Fadyean (coloration au bleu de méthylène), par l'encre de Chine ou par anticorps fluorescent (DROMIGNY, 2009).

— La lysotypie de *B. anthracis* se fait par la lyse de la bactérie par le bactériophage γ (BROWN et CHERRY, 1955), test spécifique à 96% étant donné que certaines souches du groupe *B. cereus* y sont également sensibles (KOLTON et al., 2017).

— *B. anthracis* présente une sensibilité à la pénicilline, et sa culture sur un milieu contenant de faibles niveaux de cet antibiotique conduit à un gonflement des bactéries et à leur organisation en chaîne, évoquant un collier de perles, contrairement à d'autres souches du genre *Bacillus*, qui sont résistantes à la pénicilline (JENSEN, KLEEMEYER et al., 1953).

Cependant, il est à noter que ces caractéristiques phénotypiques ne sont pas exclusives à *B. anthracis* et peuvent se retrouver chez certaines souches du groupe *B. cereus*. Inversement, il existe des souches de *B. anthracis* qui sont hémolytiques, résistantes à la pénicilline et au phage γ (PATRA et al., 1998 ; GIERCZYŃSKI et al., 2004). En

outre ces méthodes nécessitent la mise en culture de la bactérie, dans un laboratoire sécurisé en conséquence.

Détection par PCR ou qPCR L'avènement de l'amplification d'ADN par PCR (*Polymerase Chain Reaction*) a permis de développer d'autres méthodes d'identification de *B. anthracis*. Plusieurs jeux d'amorces permettent cela, ciblant conjointement le chromosome et les plasmides. En effet, certaines souches de *B. anthracis* peuvent perdre un des plasmides et *a contrario*, certaines souches du groupe *B. cereus* ont pu acquérir des plasmides similaires pouvant être ciblés par les amorces PCR. Ainsi, certaines cibles autrefois considérées comme spécifiques se sont avérées présentes dans des souches n'appartenant pas à l'espèce *B. anthracis* : *BA813*, *rpoB*, *gyrA*, *gyrB*, *saspB*, *plcR*, *BA5345* (ANTWERPEN et al., 2008). La qPCR, ou PCR quantitative ou encore PCR en temps réel, est utilisé dans le cadre de la détection de cette bactérie. Des systèmes basés sur cette technique ont été commercialisés. Citons par exemple le FILMARRAY (BioFire, Salt Lake City, UT, USA) qui permet la détection simultanée de 17 agents pathogènes, en utilisant 27 marqueurs génétiques dont trois pour *B. anthracis* (un pour le chromosome, un pour pXO1, un pour pXO2) ou le JBAIS (*Joint Biological Agent Identification and Diagnostic System*).

Immuno-tests Ils s'appuient sur divers marqueurs, selon la forme de la bactérie dans l'échantillon (sous forme végétative ou sous forme sporulée). Des anticorps ciblant plusieurs antigènes sont utilisés : ceux spécifiques à la glycoprotéine BclA, un élément prédominant de l'exosporium des spores, ceux ciblant les motifs oligosaccharidiques associés à BclA, les anticorps ciblant l'antigène extracellulaire EA1, un constituant de la couche S, ceux se fixant à l'antigène protecteur PA, un composant de la toxine du charbon ou encore à la capsule en acide poly-γ-D-glutamique. Les immuno-tests généralement utilisés sont par exemple le test ELISA (KUEHN et al., 2009), la technique FRET (*Frequency Resonance Energy Transfer*) (ZAHAVY et al., 2003) ou la technologie Luminex (TAMBORRINI et al., 2010).

Biocapteurs Un biocapteur se compose de molécules biologiques de reconnaissance immobilisées sur un transducteur de signal, qui convertit le signal en une sortie lisible. Selon le type de transduction de signal, plusieurs modèles de biocapteurs existent : électrochimiques (ampérométriques, potentiométriques et conductométriques), optiques, piézoélectriques et thermiques (GOODING, 2006). Dans le cas de *B. anthracis*, certains biocapteurs ciblent les toxines, les spores ou des marqueurs génétiques (WANG et al., 2021). Les différentes méthodes sont résumées dans la figure 1.12.

MALDI-TOF MS L'augmentation de l'automatisation et des avancées informatiques au cours des années 2000, notamment grâce à l'apparition de vastes bases de données, a conduit à l'adoption de *Matrix-Assisted Laser Desorption/Ionization Time-Of-Flight Mass Spectrometry* (MALDI-TOF MS) comme méthode standard d'identification bactérienne. MALDI-TOF MS est un type de spectrométrie de masse qui repose sur l'utilisation d'une matrice pour aider à la désorption et à l'ionisation des échantillons, suivie d'une mesure du temps de vol des ions pour déterminer leur masse. La détection par cette technologie se fait en plusieurs étapes :

— l'identification commence par l'extraction des protéines de l'échantillon bactérien. Ce processus peut être simple, comme le dépôt direct de colonies bactériennes sur la plaque cible, ou nécessiter une préparation plus élaborée pour

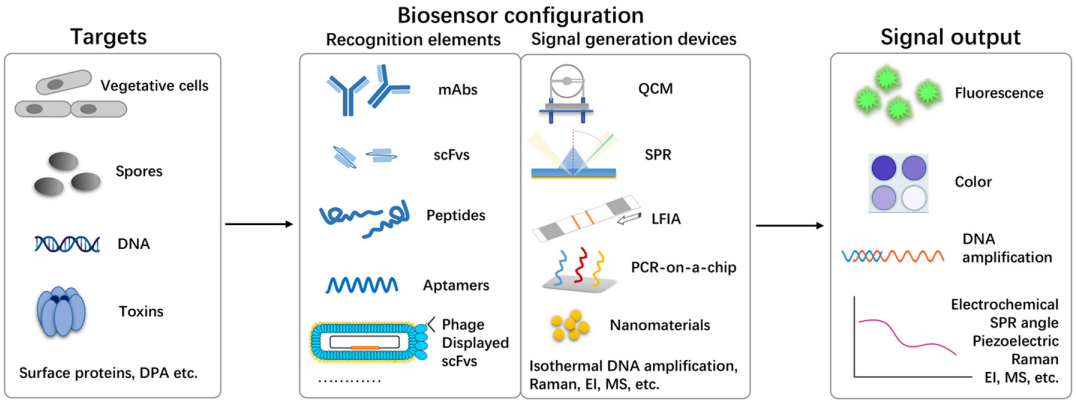

FIGURE 1.12 – Récapitulatif des biocapteurs utilisés pour la détection de *B. anthracis*. Tiré de WANG et al., 2021.

les bactéries ayant des parois cellulaires complexes (par exemple, les mycobactéries).

— une petite quantité de l'extrait protéique est déposée sur une plaque MALDI. Cette plaque est faite d'acier inoxydable et est conçue pour résister aux conditions de haute tension et d'ultravide du spectromètre de masse. La matrice est une substance constituée de petites molécules organiques acides qui facilite l'absorption de l'énergie laser et aide à la désorption et à l'ionisation des protéines. Elle est appliquée sur l'échantillon, généralement en gouttelettes, et permet de cristalliser les protéines. La qualité et la concentration de la matrice peuvent influencer la qualité du spectre. Lors de l'irradiation par le laser, la matrice absorbe l'énergie, facilitant la désorption des protéines qui sont ensuite ionisées.

— les ions protéiques sont accélérés par un champ électrique dans un tube appelé *drift tube*. Leur vitesse dépend de leur masse : les ions légers voyagent plus rapidement que les ions lourds. Un détecteur enregistre le temps de vol de chaque ion, créant le spectre. Le spectre résultant présente des pics distincts correspondant à différentes protéines ou fragments protéiques. Chaque pic représente une molécule d'une masse spécifique.

— une fois le spectre généré, un logiciel analyse ces données, les compare avec une bibliothèque de spectres connus. Le logiciel calcule une valeur numérique de correspondance entre le spectre de l'échantillon et les spectres de la base de données. Une valeur élevée indique une correspondance probable, fournissant une identification de la bactérie en présence.

Le principe de fonctionnement du MALDI-TOF MS est schématisé sur la figure 1.13.

Les différentes méthodes présentées dans cette partie sont comparées dans le tableau 1.1.

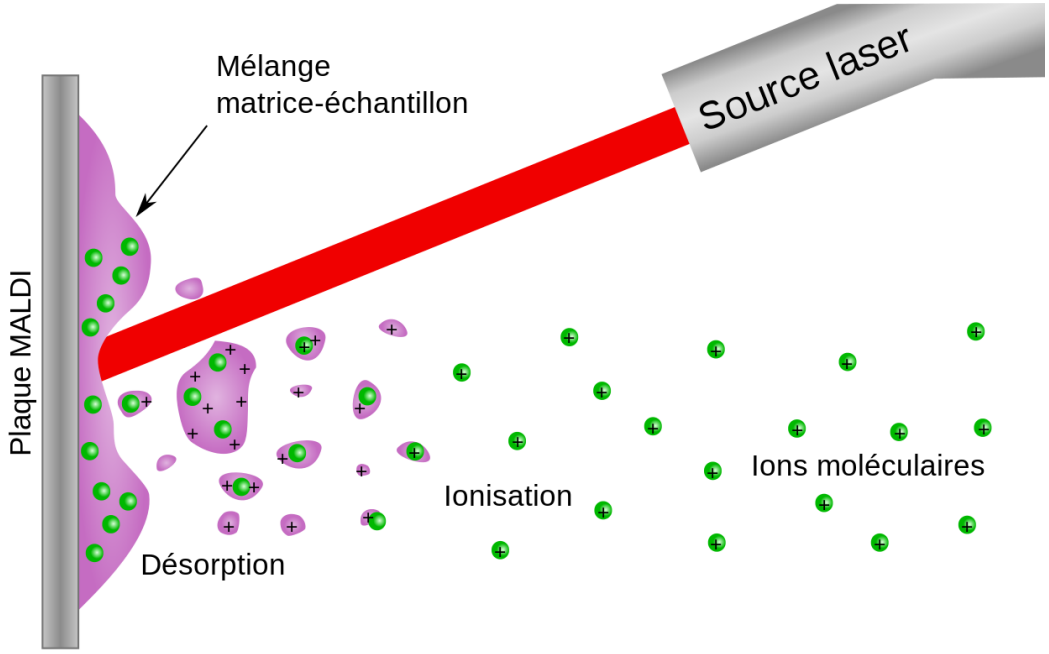

FIGURE 1.13 – Schéma du principe de fonctionnement de la technologie MALDI-TOF. Tiré de WIKIPÉDIA.

Technologie	Sensibilité / Seuil de détection	Avantages	Inconvénients
Culture	Colonie (jusqu'à 10^5 CFU)	Mise en évidence de la viabilité, isolement de souches	Manipulation en laboratoire P3, délai d'incubation
PCR, qPCR	Variable selon les matrices (jusqu'à 10^5 CFU/g)	Mise en évidence de formes végétatives ou sporulées, rapide	Faux positifs, indifférenciation entre forme viable ou non selon la matrice
Immuno-tests ou biocapteurs ciblant les toxines	Jusqu'à 170 pg/mL	Bon marché, utilisable sur le terrain	Nécessité de réactiver les spores risque d'infection
Biocapteurs ciblant les spores	Jusqu'à 10^4 spores/mL	Détection rapide sans réactiver les spores, haute spécificité	Sensibilité moyenne
Biocapteurs ciblant des marqueurs génétiques	Jusqu'à 10^5 CFU/mL ou 1 pg d'ADN selon la méthode	Bon marché, possible détection à l'oeil nu, haute sensibilité	Nécessité de réactiver les spores, risque d'infection
MALDI-TOF MS	Colonie (jusqu'à 10^5 CFU)	Très rapide	Sensibilité faible, méthode invasive

TABLE 1.1 – Tableau comparatif des méthodes conventionnelles de détection de *B. anthracis*. Tiré de *Avis de l'ANSES relatif à la saisine n° 2016-SA-0286* 2016 ; WANG et al., 2021.

1.6 Typage de *Bacillus anthracis*

Les méthodes conventionnelles permettent la détection de *B. anthracis* dans un échantillon mais n'offrent pas la possibilité d'identification de la ou des souches en présence dans un cas positif. Plusieurs méthodes ont successivement été mises en place pour permettre un typage intraspécifique de *B. anthracis*. On peut citer celles principalement utilisées : le typage MLVA, l'analyse SNP et l'approche wgMLST (dont cgMLST).

1.6.1 Typage MLVA

Les génomes bactériens possèdent fréquemment des séquences répétées qui sont de natures différentes : par exemple des répétitions dispersées dans le génome (comme les séquences d'insertion, les séquences REP [10] (STERN, PROSSNITZ et AMES, 1988), les séquences ERIC [11] (HULTON, HIGGINS et SHARP, 1991), les séquences BOX (MARTIN et al., 1992) ou des répétitions en tandem dans un *locus* unique. À la différence des séquences répétitives dispersées, les répétitions en tandem se composent de séquences d'ADN répétées contiguës, homogènes en termes de taille et de séquence. Lorsque le nombre de séquences répétées en tandem à un *locus* donné varie d'une souche bactérienne à une autre, le *locus* est désigné sous le terme de VNTR (*Variable Number of Tandem Repeats*). C'est sur ce principe que se base le typage MLVA, en exploitant une collection de VNTRs pertinents pour différencier les souches au sein d'une espèce.

Dans le cas de *B. anthracis*, la méthode MLVA fut exploitée à partir de 2000 (KEIM et al., 2000). Cela s'avère pertinent dans la mesure où le taux de mutation (10^{-4} à 10^{-5} par génération sur les zones sélectionnées) entraîne une variation du nombre de copies suffisante pour produire des polymorphismes ayant une capacité de discrimination satisfaisante pour typer les souches de *B. anthracis* (KEIM et al., 2000). Des schémas de typage successifs ont été mis en place, permettant un pouvoir de résolution de plus en plus élevé au fur et à mesure que le nombre de VNTRs discriminants augmentaient (KEIM et al., 2000 ; LE FLÈCHE et al., 2001 ; KEIM et al., 2004 ; LISTA et al., 2006 ; VAN ERT et al., 2007 ; BEYER et al., 2012 ; THIERRY et al., 2014). En définitive, 31 *locus* ont été sélectionnés pour le typage MLVA de cette bactérie. La figure 1.2 récapitule ces différents schémas.

1.6.2 Analyse SNP

Parallèlement au MLVA, l'analyse SNP permet également de différencier les souches de l'espèce *B. anthracis*. Un SNP est une substitution de nucléotide à une position donnée dans un génome. Pour *B. anthracis*, ils sont observés à un taux d'environ 10^{-10} pour chaque position, à une génération donnée (VOGLER et al., 2002 ; KEIM et al., 2004). Malgré la rareté d'apparition, les SNPs sont très stables et sont donc utiles à l'établissement de la phylogénie de *B. anthracis*. En effet, leur stabilité suppose une origine unique pour chacun des SNPs et permet donc de retracer l'histoire évolutive de l'espèce. De plus, il s'avère que ces marqueurs génétiques sont fortement corrélés à l'origine géographique des souches, ceci pouvant s'expliquer par le caractère clonal de *B. anthracis* ainsi que par son émergence relativement récente (VAN ERT et al., 2007).

En somme, les SNPs fournissent une quantité limitée d'informations génétiques et leur résolution est relativement faible lorsqu'ils sont utilisés sur un petit nombre de sites. Cependant, l'analyse des SNPs à évolution lente permet de déterminer les principales lignées clonales de *B. anthracis*, tandis que les marqueurs à évolution plus rapide (comme les VNTR utilisés en MLVA) peuvent être utilisés pour discriminer les souches et éventuellement séquencer les plus pertinentes (SHEVTSOV et al., 2021).

Un ensemble de 12 SNPs dits "canoniques", appelés canSNPs, ont été établis pour diviser les souches de *B. anthracis* en trois lignées (A, B et C), elles-mêmes séparées en sept sous-lignées et cinq groupes. Plus tard, un treizième canSNP a été ajouté permettant la division du groupe transeurasien (TEA) A.Br.008/009 en deux groupes

10. *Repetitive Extragenic Palindromic sequences*
11. *Enterobacterial Repetition Intergenic Consensus*

MLVA8	**MLVA15**	**MLVA20**	**MLVA25**	**MLVA31**
pX01	pX01		pX01	pX01
pX02	pX02		pX02	pX02
vrrA	vrrA	vrrA	vrrA	vrrA
vrrB1	vrrB1	vrrB1	vrrB1	vrrB1
vrrB2	vrrB2	vrrB2	vrrB2	vrrB2
vrrC1	vrrC1	vrrC1	vrrC1	vrrC1
vrrC2	vrrC2	vrrC2	vrrC2	vrrC2
CG3	CG3	CG3	CG3	CG3
	bams01	bams01	bams01	bams01
		bams03	bams03	bams03
		bams05	bams05	bams05
		bams07		
		bams13	bams13	bams13
		bams15	bams15	bams15
		bams21	bams21	bams21
		bams22	bams22	bams22
		bams23	bams23	bams23
		bams24	bams24	bams24
		bams25	bams25	bams25
		bams28	bams28	bams28
		bams30	bams30	bams30
		bams31	bams31	bams31
			bams34	bams34
			bams44	bams44
			bams51	bams51
			bams53	bams53
	BaVNTR12			BaVNTR12
	BaVNTR16			BaVNTR16
	BaVNTR17			BaVNTR17
	BaVNTR19			BaVNTR19
	BaVNTR23			BaVNTR23
	BaVNTR35			BaVNTR35

TABLE 1.2 – Tableau récapitulatif des marqueurs génétiques pour les différents profils MLVA. Adapté de GIRAULT, 2015.

distincts : A.Br.008/011 (ou TEA 008/011) et A.Br.011/009 (ou TEA 011) (MARSTON et al., 2011). Au total, il y a donc sept sous-lignées et six groupes :

- Sous-lignées
  - A.Br.Ames
  - A.Br.Aust94
  - A.Br.Vollum
  - A.Br.WNA
  - B.Br.KrugerB
  - B.Br.CNEVA
  - C.Br.A1055

- Groupes
  - A.Br.001/002
  - A.Br.003/004
  - A.Br.005/006
  - A.Br.008/011
  - A.Br.011/009
  - B.Br.001/002

La figure 1.14 illustre la phylogénie de *B. anthracis* à travers la répartition de quelques souches au sein de ces divisions.

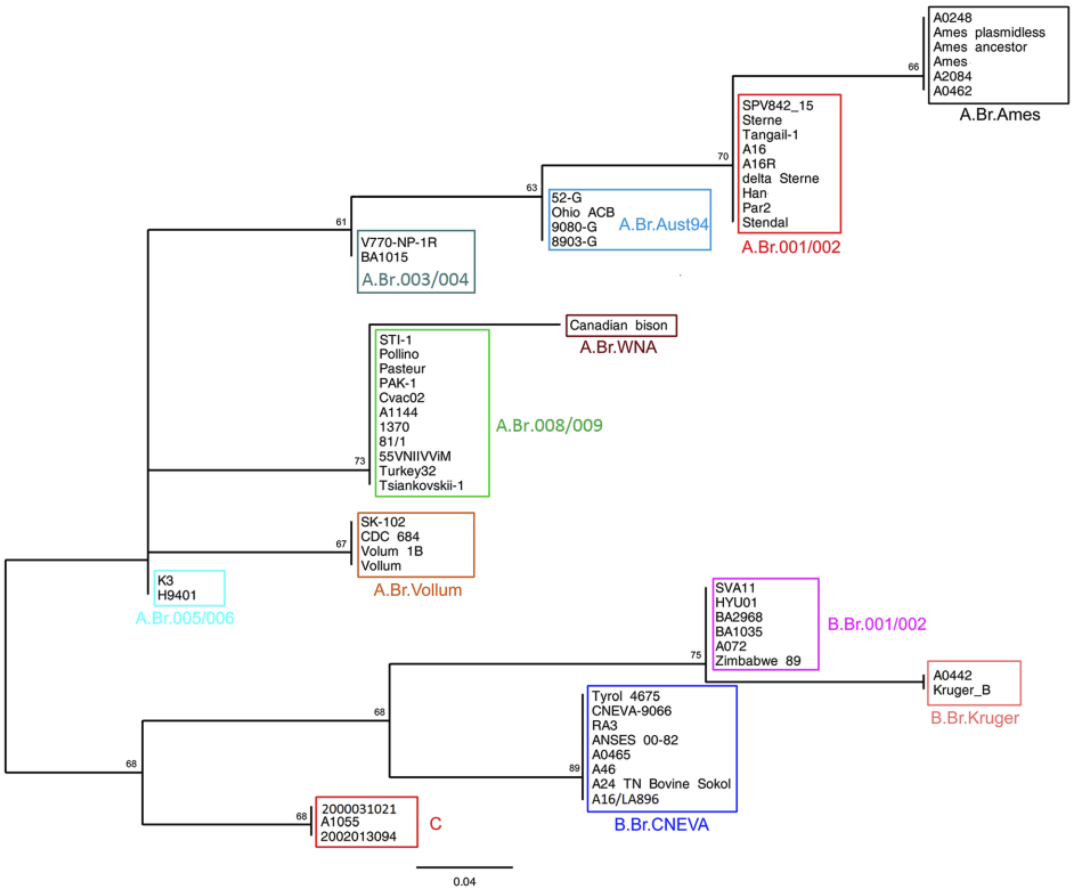

FIGURE 1.14 – Arbre phylogénétique non raciné de 58 souches de *B. anthracis* représentatives des sept sous-lignées et cinq groupes principaux définis par les canSNPs. Tiré de PISARENKO et al., 2019.

La répartition de l'espèce au sein de ces sous-lignées et de ces groupes est très variable :

— La lignée A est la plus répandue à travers le monde, avec une répartition des groupes/sous-lignées dans des zones géographiques assez délimitées. Le groupe A.Br.001/002 se retrouve en Chine (et en Asie) et est structuré sous forme polytomique [12]. La sous-lignée A.Br.Ames a émergé quant à elle aux Etats-Unis, par une importation du groupe A Br.001/002, peut-être de la Mongolie intérieure, région autonome chinoise (SIMONSON et al., 2009). Les souches les plus proches actuellement connues ont été isolées au Kazakhstan (SHEVTSOV

12. Embranchement simultané d'au moins trois branches au sein d'un arbre phylogénétique.

et al., 2021). La sous-lignée A.Br.WNA est présente en Amérique du Nord tandis que le groupe A.Br.003/004 est trouvé en Amérique du Sud. Le groupe A.Br.005/006 est présent principalement sur le continent africain et également en Europe et Australie. La sous-lignée A.Br.Vollum est présente en Afrique du Sud et en Asie du Sud-Ouest (Afghanistan, Pakistan) alors que la sous-lignée A.Br.Aust94 est présente en Australie, en Asie (Inde, Turquie), en Afrique (Namibie), dans le Caucase (Géorgie) (VAN ERT et al., 2007 ; DURMAZ et al., 2012 ; BEYER et al., 2012 ; DERZELLE et THIERRY, 2013 ; KHMALADZE et al., 2014). Enfin, le groupe transeurasien, majoritaire au sein de la lignée A, se retrouve en Europe et en Asie, avec une forte présence en Russie (EREMENKO et al., 2021). Parmi TEA, le groupe A.Br.008/011 est sous forme polytomique, avec sept branches différentes, dont trois font référence à des souches vaccinales : Pasteur, Tsiankovskii et STI (*Sanitary Technical Institute*). L'autre groupe composant TEA, A.Br.011/009, est lui aussi sous forme polytomique, également à sept branches (VERGNAUD et al., 2016 ; VERGNAUD, 2020).

— La lignée B est plus restreinte géographiquement. Le groupe B.Br.001/002 est présent en Afrique du Sud, en Asie du Sud (Corée et Bhutan), dans les pays scandinaves, en Russie (péninsule de Yamal, Sibérie), au Kazakhstan et aux États-Unis (JUNG et al., 2012 ; THAPA et al., 2014 ; LIENEMANN et al., 2018 ; TIMOFEEV et al., 2019 ; PISARENKO et al., 2019 ; SHEVTSOV et al., 2021 ; EREMENKO et al., 2021 ; PISARENKO et al., 2021 ; TIMOFEEV et al., 2023). La sous-lignée B.Br.KrugerB se situe en Afrique du Sud, dans le Parc Kruger plus précisément (VAN ERT et al., 2007). La sous-lignée B.Br.CNEVA est présente en Europe.

— La lignée C est la plus rare et ne comporte que quelques souches connues, isolées en Amérique du Nord exclusivement (PEARSON et al., 2004 ; VAN ERT et al., 2007).

1.6.3 Analyse cgMLST

La méthode de typage moléculaire MLST [13] s'appuie sur l'analyse des séquences d'un nombre restreint de gènes, habituellement sept. Pour chaque variant de séquence d'un gène on attribue un numéro d'allèle spécifique. L'ensemble des numéros d'allèles pour une souche donnée forme un profil génétique distinct, appelé ST [14]. Avant l'avènement des technologies "génome complet", le MLST était reconnu comme la référence en matière de caractérisation moléculaire pour un éventail d'agents pathogènes, bénéficiant d'une nomenclature standardisée et d'une base de données centralisée des allèles accessible via des services web dédiés, tels que PUBMLST (JOLLEY, BRAY et MAIDEN, 2018). Néanmoins, cette technique a ses limites, notamment en termes de résolution, pour *B. anthracis*, qui est caractérisé par un faible polymorphisme génétique. L'évolution des méthodes de séquençage à haut débit a facilité l'acquisition de génomes presque complets, permettant d'élargir le spectre des gènes étudiés, et pour un coût désormais comparable à celui d'un typage MLST traditionnel. Cela a conduit au développement de l'approche "*core genome* [15] MLST" (cgMLST), qui inclut une gamme bien plus large de gènes conservés.

13. *Multilocus Sequence Typing*
14. *Sequence Type*
15. Ensemble de gènes partagés par toutes les souches d'une espèce bactérienne ou d'un groupe de micro-organismes. Il s'oppose au *pangenome*, qui inclut à la fois le *core genome* et les gènes variables présents uniquement dans certaines souches.

Concernant *B. anthracis*, le schéma cgMLST englobe 3,803 gènes conservés, analysés à travers 57 génomes représentatifs de toute la phylogénie de l'espèce (ABDEL-GLIL et al., 2021). Ce schéma permet une identification de *B. anthracis* à la souche près. Il est complémentaire de la méthode de typage basée sur les SNP, facilitant ainsi des comparaisons rapides et standardisées entre laboratoires, une composante clé pour la surveillance globale et l'analyse d'épidémies. À titre d'illustration, la comparaison des phylogénies obtenues selon les deux méthodes pour un panel de souches de *B. anthracis* est disponible à la figure 1.15 : les deux méthodes permettent une même répartition des souches en groupes et sous-lignées, à l'exception de quelques souches du groupe TEA.

Cependant, bien que le cgMLST puisse être efficace pour l'assignation automatisée préliminaire des données de séquences de génomes entiers, une révision complète de la nomenclature actuelle n'offre pas d'amélioration significative en termes de résolution par rapport à une approche dite wgSNP exploitant tous les SNPs disponibles en fonction des souches à analyser (SAHL et al., 2016).

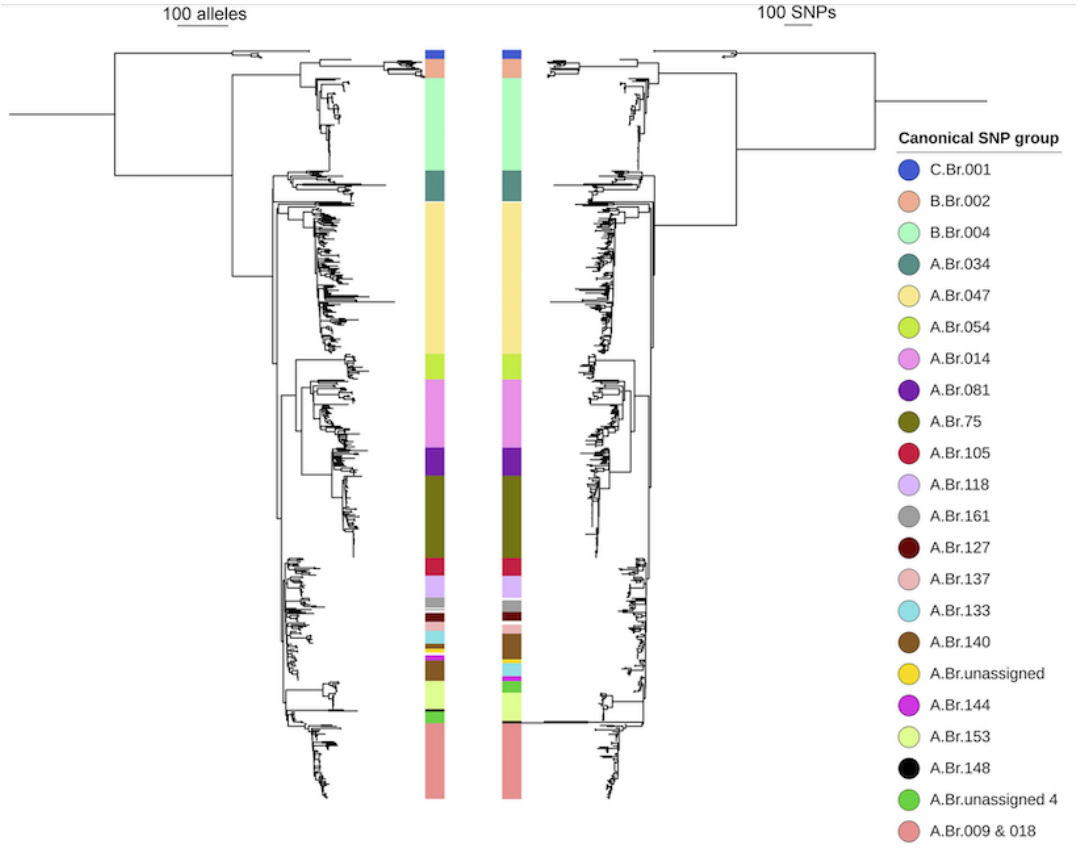

FIGURE 1.15 – Comparaison des phylogénies de *B. anthracis* basées sur l'analyse cgMLST (à gauche) et wgSNP (à droite). Les noms des groupes/sous-lignées sont propres à la figure et ne correspondent pas à ceux détaillés dans la section précédente. Tiré de ABDEL-GLIL et al., 2021.

IDÉES À RETENIR

1. *B. anthracis* est la bactérie responsable de la maladie du charbon, qui est une zoonose affectant principalement les ongulés ; les humains sont également sensibles. Trois formes humaines du charbon existent (cutanée, pulmonaire, digestive), à laquelle s'ajoutent la forme injectionnelle, plus rare, et des cas de charbon atypiques.
2. Les facteurs de virulence de *B. anthracis* sont principalement contenus dans ses deux plasmides pXO1 et pXO2, le premier contenant les gènes nécessaires à la production de toxines (œdématogène et létale), le second contenant ceux pour la synthèse d'une capsule en acide poly-γ-D-glutamique.
3. Sa capacité de persistance couplée à sa propagation facile sous forme de spores a motivé l'utilisation de *B. anthracis* à des fins malveillantes, justifiant entre autres son étude et le développement d'outils de détection.
4. Les méthodes conventionnelles de détection, en microbiologie, en immunologie ou en protéomique, demeurent le "*gold standard*" pour une identification rapide. Leur pouvoir résolutif est cependant limité, particulièrement dans le cas de présence à faible concentration de *B. anthracis* ou pour une identification à la souche près. Elles sont en outre plus particulièrement adaptées à l'analyse d'échantillons biomédicaux, et peuvent requérir une manipulation en laboratoire sécurisé.
5. La diversité génétique de *B. anthracis* est aujourd'hui bien connue et plusieurs outils de typage moléculaire et génomique permettent de caractériser les souches de cette espèce.

Chapitre 2

Comment délimiter l'espèce *Bacillus anthracis* ?

B. anthracis est une espèce très monomorphe [1] au sein d'un complexe plus large, appelé "groupe *Bacillus cereus*". Celui-ci se compose d'autres espèces génétiquement proches de *B. anthracis*, mais la pathogénicité des souches au sein de ce groupe est très variable. Par exemple, certaines sont capables de produire des toxines émétiques ou diarrhétiques, alors que d'autres peuvent sécréter des cristaux insecticides. Certaines souches, dites anthracis-like, sont même capables de provoquer des affections s'apparentant au charbon, en ayant pour autant les attributs d'autres espèces que *B. anthracis*. Enfin, la distribution géographique de *B. anthracis*, largement répandue à la surface du globe, ainsi que les divers facteurs (anthropiques, climatiques par exemple) qui ont permis cette propagation, questionnent sur l'origine de l'espèce.

Ces différents constats amènent à s'interroger sur l'émergence de *B. anthracis* au sein du groupe *B. cereus*. Ce chapitre vise à présenter ce groupe, en mettant l'accent sur les souches anthracis-like, avant de s'intéresser à l'épidémiologie de *B. anthracis* et enfin de détailler les différentes hypothèses à l'étude quant à l'apparition de cette espèce.

OBJECTIFS DU CHAPITRE

1. Présenter le groupe *B. cereus*, auquel appartient *B. anthracis*
2. Détailler les caractéristiques des souches anthracis-like existantes au sein du groupe *B. cereus*
3. Expliquer les différents facteurs de dissémination de cette bactérie et décrire la répartition de celle-ci à travers le monde
4. Faire un état de l'art des hypothèses relatives à l'émergence de *B. anthracis* et définir les zones d'ombre qui persistent à ce sujet

2.1 Le groupe *Bacillus cereus*

Cette section présente le groupe *B. cereus*, incluant *B. anthracis* et d'autres espèces génétiquement proches. Elle se focalise sur la problématique de délimitation de *B. anthracis* au sein de ce complexe, en dépit de leur étroite relation génétique.

1. Dans ce contexte, cela signifie qu'il y a très peu de variabilité génétique au sein des génomes de cette population ; de même pour le phénotype.

Le groupe *B. cereus* appartient au genre *Bacillus*, un ensemble diversifié de bactéries à Gram positif. Ces bactéries, présentes en abondance dans divers environnements terrestres, sont aérobies ou facultativement anaérobies. Une caractéristique majeure des bactéries de ce genre est leur capacité à former des endospores, qui sont des structures dormantes hautement résistantes. Ces endospores permettent aux bactéries de survivre dans des conditions environnementales extrêmes, telles que des températures élevées ou basses, des milieux acides ou basiques, ou en présence de composés toxiques. Ainsi, la formation d'endospores favorise la persistance et la dissémination des bactéries dans des habitats variés.

À l'origine, le groupe *B. cereus* (ou *B. cereus sensu lato*) était constitué de quatre espèces identifiées soit comme agents pathogènes, soit par leurs caractéristiques morphologiques distinctives : *B. anthracis*, *B. cereus* (*sensu stricto*), *Bacillus thuringiensis* et *Bacillus mycoides*. En 1998 furent ajoutées deux autres espèces, grâce à l'utilisation de méthodes moléculaires. Tout d'abord, *Bacillus pseudomycoides*, auparavant considérée comme une variante de *B. mycoides* en raison de leur similitude en termes de morphologie de colonie (rhizoïdale), a été différencié par des techniques d'hybridation ADN-ADN[2], d'analyse des acides gras et de comparaison des séquences d'ADNr 16S. D'autre part, la nouvelle espèce *Bacillus weihenstephanensis* (en référence à la ville allemande où la première souche a été isolée) a été caractérisée grâce à plusieurs marqueurs moléculaires (CARROLL et al., 2022a).

B. anthracis ayant été présentée dans le chapitre 1, seule une description des autres espèces est réalisée ci-dessous, en mettant l'accent sur leurs spécificités en termes de virulence.

2.1.1 Le groupe historique

Bacillus cereus sensu stricto est capable de synthétiser deux types principaux de toxines : émétiques et diarrhéiques (EHLING-SCHULZ, FRICKER et SCHERER, 2004 ; STENFORS ARNESEN, FAGERLUND et GRANUM, 2008). La virulence émétique de *B. cereus sensu stricto* est associée à la production d'une toxine nommée céreulide. Cette toxine se distingue par sa nature robuste ; elle résiste à de nombreuses conditions qui dégradent habituellement d'autres toxines biologiques. Après ingestion d'aliments contaminés par cette toxine, les individus peuvent rapidement présenter des symptômes caractéristiques d'intoxication (nausées, vomissements, douleurs abdominales). Ces manifestations peuvent survenir entre une à cinq heures après la consommation d'aliments contaminés. L'un des défis associés au céreulide est sa résistance. Il est thermostable, ce qui signifie qu'il peut survivre à des températures usuelles de cuisson, rendant difficile son élimination. Sa résistance, son action rapide et ses effets nocifs sur le système gastro-intestinal en font une préoccupation sérieuse en matière de sécurité alimentaire.

Le céreulide est synthétisé par une série d'enzymes de synthèse de peptides non ribosomiques (NRPS[3]), et les gènes responsables de cette biosynthèse sont localisés dans un cluster appelé opéron *ces*. Cet opéron comprend plusieurs gènes, dont *cesA* et *cesB*, qui orchestrent la production du céreulide en assemblant ses précurseurs protéiques (EHLING-SCHULZ et al., 2006). La complexité de ce système enzymatique

2. Technique de biologie moléculaire permettant de mesurer le degré de similitude entre les séquences d'ADN de deux organismes. Cette méthode implique la dénaturation de double-brins d'ADN des organismes à comparer, suivie de leur association pour permettre l'appariement des brins complémentaires. L'hybridation est généralement quantifiée par la température à laquelle les double brins hybrides se désassocient (température de dénaturation), reflétant ainsi la proportion de séquences homologues et fournissant une indication sur la proximité taxonomique des organismes étudiés.

3. *Non-Ribosomal Peptides Synthetase*

NRPS se reflète dans la structure cyclique de la toxine, contribuant à sa stabilité et à sa résistance à la dégradation.

La toxine diarrhéique agit principalement dans l'intestin grêle, perturbant l'équilibre normal des fluides et des électrolytes, ce qui entraîne une sécrétion accrue d'eau dans la lumière intestinale. Cette action provoque des symptômes tels que des diarrhées aqueuses, accompagnées de douleurs abdominales et, dans certains cas, de fièvre. Ils se manifestent plus tardivement que ceux de la toxine émétique, généralement 8 à 16 heures après l'ingestion d'aliments contaminés. Une caractéristique distincte des toxines diarrhéiques est leur sensibilité à la chaleur. Contrairement au céreulide émétique, les toxines responsables de la diarrhée peuvent être inactivées par une cuisson adéquate, ce qui souligne l'importance de la cuisson appropriée des aliments pour prévenir les intoxications (DIETRICH et al., 2021). La virulence diarrhéique de *B. cereus sensu stricto* découle de la production de plusieurs toxines, Hbl, Nhe, et CytK, codées respectivement par les opérons *hbl*, *nhe* (GRANUM, O'SULLIVAN et LUND, 1999 ; MORAVEK et al., 2004) et le gène *cytK* (BRILLARD et LERECLUS, 2004). Les toxines Hbl et Nhe agissent conjointement pour perturber l'équilibre cellulaire de l'intestin, conduisant à une diarrhée aqueuse. La toxine CytK est une hémolysine qui peut endommager les cellules intestinales, augmentant la gravité des symptômes (DIETRICH et al., 2021).

Bacillus thuringiensis est connu pour sa capacité à produire des protéines toxiques pour de nombreux insectes phytophages, en particulier les lépidoptères, mais aussi certains coléoptères et diptères (PALMA et al., 2014). Cette spécificité est due à la production de protéines insecticides, sous forme de cristaux parasporaux, lors de la sporulation. Ceci est le résultat de l'expression de gènes portés majoritairement par des plasmides.

Ces protéines sont communément désignées sous les noms de toxines Cry (cristallines) et Cyt (cytolytiques). Il existe une large diversité des gènes *cry* et *cyt*, qui permet aux souches de *B. thuringiensis* de cibler un éventail varié d'insectes. Par exemple, les protéines Cry1 ont une action insecticide contre les lépidoptères, tandis que les protéines Cry3 sont plus actives contre les coléoptères. Lorsqu'un insecte ingère ces toxines, elles sont activées dans son intestin par des protéases, puis se lient à des récepteurs spécifiques présents sur sa paroi intestinale. Cela provoque la formation de pores dans les cellules intestinales, entraînant une lyse de celles-ci. L'insecte subit alors une paralysie intestinale, cesse de se nourrir et meurt quelques jours plus tard, en raison d'une septicémie (BRETSCHNEIDER, HECKEL et PAUCHET, 2016).

Certaines souches de *B. thuringiensis* produisent d'autres facteurs de virulence, tels que les toxines Vip (*Vegetative insecticidal proteins*) qui sont exprimées durant la phase végétative de la bactérie, et qui élargissent le spectre d'action insecticide de *B. thuringiensis* (CRICKMORE et al., 2020). Cette capacité unique de *B. thuringiensis* à produire ces toxines et leur spécificité pour certains insectes a conduit à son adoption en tant que biopesticide. Les gènes responsables de la production de ces toxines ont été transférés à certaines plantes génétiquement modifiées pour leur conférer une résistance aux ravageurs, réduisant l'usage des pesticides chimiques.

Bacillus weihenstephanensis proche voisin de *B. cereus sensu stricto*, se distingue par sa capacité psychrotrophe, c'est-à-dire sa capacité à croître à des températures basses, typiquement en dessous de 7°C (LECHNER et al., 1998). Cette propriété le

rend particulièrement pertinent pour l'industrie alimentaire, car il peut se développer dans des aliments réfrigérés, provoquant leur altération. D'un point de vue génétique, *B. weihenstephanensis* possède des gènes codant des toxines similaires à ceux de *B. cereus sensu stricto*. Cependant, les manifestations de virulence associées (symptômes gastro-intestinaux chez l'hôte) à cette bactérie sont moins sévères.

Bacillus mycoides* et *Bacillus pseudomycoides peuvent être distingués grâce à leur morphologie lors de la croissance sur les milieux de culture solides. En effet, ces bactéries sont en forme de rhizoïde, formant des filaments entrelacés qui ressemblent à la croissance des mycètes, d'où la dérivation de leurs noms. *B. mycoides* est isolé dans les sols tandis que *B. pseudomycoides* se retrouve dans des sources alimentaires. Malgré la présence de séquences génétiques associées à des toxines dans leur génome, la littérature scientifique n'a pas documenté de cas où *B. mycoides* ou *B. pseudomycoides* ont manifesté une virulence pour l'Homme (GUINEBRETIÈRE et al., 2008). L'identité taxonomique de ces espèces a été confirmée par des analyses moléculaires, notamment le séquençage de l'ARNr 16S (NAKAMURA et JACKSON, 1995; NAKAMURA, 1998).

Plusieurs méthodes ont été élaborées pour distinguer ces différentes espèces, que l'on peut classer en deux catégories : observations phénotypiques et comparaisons génétiques (CARROLL et al., 2022a).

- Observations phénotypiques
  - La β-hémolyse sur gélose au sang de mouton
  - La mobilité
  - La coloration de capsule
  - La lyse par le phage γ
  - La décomposition de la tyrosine
  - La croissance des rhizoïdes
  - L'inclusion parasporale cristalline
  - La composition en acides gras
- Comparaisons génétiques
  - L'hybridation ADN-ADN
  - La comparaison des ADNr 16S et 23S ainsi que des régions intergéniques 16S-23S
  - L'analyse RAPD [4]
  - L'analyse RFLP [5]
  - PFGE [6]
  - La détection de gènes spécifiques par séquençage ciblé
  - Les analyses SLST [7] et MLST [8]

Au-delà de la simple distinction entre ces différentes espèces, des premières tentatives de systèmes de classification phylogénétique des souches au sein du groupe

4. *Random Amplification of Polymorphic DNA*
5. *Restriction Fragment Length Polymorphism*
6. *Pulsed Field Gel Electrophoresis*
7. *Single Locus Sequence Typing*
8. *Multi-Locus Sequence Typing*

Critère	*B. anthracis*	*B. cereus*	*B. mycoides*	*B. pseudomycoides*	*B. thuringiensis*	*B. weihenstephanensis*
Mobilité	-	+/-	-	-	+/-	+/-
Capsule	+	-	-	-	-	-
Hémolyse	-	+	+	?	+	+/-
Lyse par phage γ	+	-	-	?	-	-
Production de cristaux parasporaux	-	-	-	-	+	-
Décomposition de la tyrosine	-	+	+/-	?	+	?

TABLE 2.1 – Tableau comparatif des caractères phénotypiques des six premières espèces du groupe *B. cereus*. + : positif, - : négatif, +/- : certaines souches sont positives, d'autres négatives, ? : critère non connu. Adapté de DROMIGNY, 2009.

B. cereus sont opérées, en particulier par les approches SLST et MLST. Pour la première approche, GUINEBRETIÈRE et al., 2008 distinguent sept groupes phylogénétiques (notés de I à VII) au sein du groupe *B. cereus*, en se basant sur la séquence du *locus panC* (pantoate-β-alanine). Parallèlement, pour la seconde approche, un autre schéma est proposé, se basant sur sept *locus* : *glpF*, *gmk*, *ilvD*, *pta*, *pur*, *pycA* et *tpi* (PRIEST et al., 2004). Il est à noter que ce schéma MLST est congruent avec la classification type *panC*.

En 2013, deux nouvelles espèces furent définies au sein du groupe : il s'agit de *Bacillus toyonensis* et *Bacillus cytotoxicus*. Cela s'est fait grâce à des analyses WGS[9], avec des calculs d'ANI[10] pour différencier ces deux espèces dans le groupe *B. cereus*.

Bacillus toyonensis est une bactérie isolée initialement dans des sols de prairie au Japon. Ce membre du groupe *B. cereus* s'est distingué non pour sa virulence, mais pour son potentiel en tant que probiotique, en particulier dans le domaine de l'alimentation animale (JIMÉNEZ et al., 2013). Il est également suggéré que *B. toyonensis* pourrait jouer un rôle dans le renforcement du système immunitaire, bien que les mécanismes précis restent à déterminer (SANTOS et al., 2018).

Bacillus cytotoxicus a été identifié pour la première fois en 1998 lors d'une étude portant sur des souches toxiques isolées d'échantillons alimentaires. Sa dénomination "*cytotoxicus*" provient de sa capacité à provoquer une cytotoxicité marquée *in vitro*. La particularité de *B. cytotoxicus* réside dans sa production d'une toxine spécifique dénommée toxine cytotoxique (CytK) (GUINEBRETIÈRE et al., 2013). Cette toxine est responsable d'un type de diarrhée nécrosante, sévère et distincte de celle causée par les toxines diarrhéiques habituellement associées à d'autres membres du groupe *B. cereus*. Deux formes de cette toxine, associées à deux gènes distincts, sont connues : CytK-1 et CytK-2. La variante CytK-1 est codée par le gène *cytK-1*, spécifiquement associé à *B. cytotoxicus* (GUINEBRETIÈRE et al., 2013). Cette toxine est responsable des cas sévères de gastro-entérite nécrosante (LUND, DE BUYSER et GRANUM, 2000). Sa puissance cytotoxique est considérablement plus élevée que celle de CytK-2. Le mécanisme d'action de cette toxine est lié à la perturbation de l'intégrité des membranes cellulaires, provoquant la lyse et la nécrose des cellules

9. *Whole Genome Sequencing*
10. *Average Nucleotide Identity*

cibles. Moins virulente que CytK-1, la variante CytK-2 est codée par le gène *cytK-2* et peut être trouvée chez certaines souches du groupe *B. cereus* qui ne sont pas spécifiquement classifiées comme *B. cytotoxicus*.

2.1.2 Le groupe *Bacillus cereus* actuel

Avec l'avènement du séquençage de génomes entiers, la question de définition d'espèces au sens large et *a fortiori* au sein du groupe *B. cereus* a pu être précisée. En effet, l'accès aux génomes permet leur comparaison, par exemple par la méthode ANI. Pour mieux comprendre en quoi le groupe *B. cereus* a suscité (et suscite encore) un débat quant à sa nomenclature et sa taxonomie, il faut revenir plus généralement à la notion d'espèce.

Tout d'abord, seules des considérations phénotypiques ont permis de classifier les bactéries : c'est la classification dite phénétique. La première de ces classifications a été l'association à une maladie, humaine, animale ou végétale. Avec les progrès de la microbiologie et des techniques de mise en culture et d'observation de microorganismes, d'autres critères ont pu être ajoutés. Par exemple, les souches de type bacilles à Gram positif, chimiorganotrophes, aérobies strictes ou aéro-anaérobies facultatives, pouvant sporuler et retrouvées dans les sols ont défini le genre *Bacillus*. Plus précisément encore, l'espèce *B. anthracis* a été définie comme englobant les souches produisant les toxines du charbon tandis que *B. thuringiensis* a été caractérisé par sa capacité à produire des toxines parasporales insecticides. Pour les procaryotes, l'axiome suivant a tout d'abord servi de base à la définition d'espèce : "En bactériologie, une espèce est constituée par sa souche type et par l'ensemble des souches considérées comme suffisamment proches de la souche type pour être incluses au sein de la même espèce" (EUZÉBY, 1997). On peut remarquer que cette définition demeure très ambiguë, le "suffisamment proches" étant sujet à interprétation.

La mise au point des premières techniques de séquençage de l'ADN à la fin des années 70 a naturellement ouvert la voie à de nouvelles approches. Dans un premier temps, les études phylogénétiques bactériennes se sont faites sur la base des séquences d'ARNr 16S, marqueur universel très conservé et de taille compatible avec les technologies disponibles. Cependant, cette approche a commencé à montrer les limites d'une réduction de la notion de proximité à celle de la proximité génétique. En effet, des espèces clairement différentes en termes d'écologie et de phénotype peuvent être très proches en termes de séquence 16S comme c'est le cas pour le groupe *B. cereus*. Puis, le séquençage des génomes complets a permis des comparaisons d'homologies entre des séquences d'ADN : de là provient la notion de *genomospecies*, définissant une espèce sur la base d'une souche de référence et des souches dont le génome dépasse un certain seuil d'homologie avec la souche de référence. Conventionnellement, la méthode d'hybridation ADN-ADN a fixé ce seuil à 70% (en pourcentage d'hybridation). Puis, le séquençage de l'ARNr 16S a ajouté une condition supplémentaire : 97% d'identité avec la séquence d'ARNr 16S de référence. Enfin, le WGS a permis la définition d'un seuil de 95% d'ANI pour définir une espèce (*Systématique en microbiologie. Notion d'espèce. Classification universelle mixte consensuelle et phylogénétique en bactériologie* 2022). De plus, au sein même d'une espèce, divers degrés co-existent, avec des limites plus ou moins définies (figure 2.1). On constate ainsi, et de façon rétrospectivement peu surprenante, que la puissance même de l'approche par séquençage aura conduit à traduire la notion de "suffisamment proches" essentiellement en termes de "proximité génétique".

Dans le cas de souches très proches sur le plan de leur séquence génomique, les seuils fixés pour délimiter les espèces peuvent faire considérablement varier la

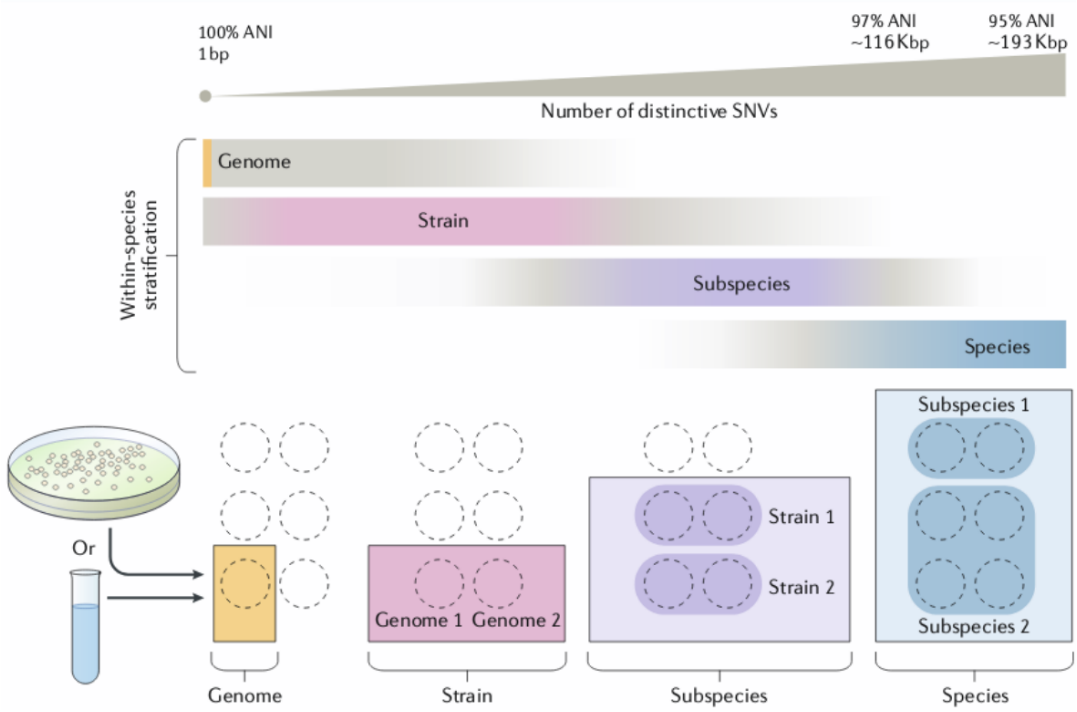

FIGURE 2.1 – Schéma explicatif des différentes appellations au sein d'une espèce bactérienne. Une espèce peut être stratifiée selon le nombre de SNV dans le génome ou les valeurs d'ANI. En couleur : usage recommandé des appellations, en gris : usage courant mais imprécis. Tiré de VAN ROSSUM et al., 2020.

classification (figure 2.2). Par exemple, lorsque neuf nouvelles espèces du groupe *B. cereus* (*Bacillus albus, Bacillus luti, Bacillus mobilis, Bacillus nitratireducens, Bacillus pacificus, Bacillus paramycoides, Bacillus paranthracis, Bacillus proteolyticus* et *Bacillus tropicus*) furent définies en 2017, un seuil de 96% d'ANI a été considéré, ce qui signifie que certaines souches de référence de quelques *genomospecies* partageaient une ANI supérieure à 95% entre elles (LIU et al., 2017). Cela a conduit à une augmentation significative du nombre d'espèces (référencées ou proposées) au sein du groupe *B. cereus*, comme le retrace la frise chronologique disponible sur la figure 2.3.

Si 80% est un seuil d'ANI qui englobe l'ensemble des souches séquencées du groupe *B. cereus*, JAIN et al., 2018 a fixé à 99.9% celui qui serait nécessaire pour définir le *genomospecies* de *B. anthracis*. Ce seuil est cohérent d'un point de vue génétique et écologique mais il ne permet pas de trancher sur le cas des souches du groupe *B. cereus*, n'appartenant pas à ce *genomospecies* mais produisant pourtant les toxines du charbon. Les appellations à leur sujet ont largement varié dans les publications les recensant : initialement "*B. anthracis*" (LEENDERTZ et al., 2004), puis "*B. cereus*" (HOFFMASTER et al., 2004 ; HOFFMASTER et al., 2006 ; MARSTON et al., 2016 ; PENA-GONZALEZ et al., 2017 ; KAMAL et al., 2017 ; SCARFF et al., 2018 ; EHLING-SCHULZ, LERECLUS et KOEHLER, 2019), "*B. cereus* biovar anthracis" (ANTONATION et al., 2016 ; SCARFF et al., 2018 ; EHLING-SCHULZ, LERECLUS et KOEHLER, 2019 ; ROMERO-ALVAREZ et al., 2020) et "*B. cereus* variety anthracis" (KLEE et al., 2010). La section suivante leur est consacrée.

En définitive, il est difficile de différencier *B. anthracis* des populations proches sur la base des séquences de gènes codant des protéines. Cette unique considération de seuils d'ANI ou de caractérisations phénotypiques ne prend pas en compte les

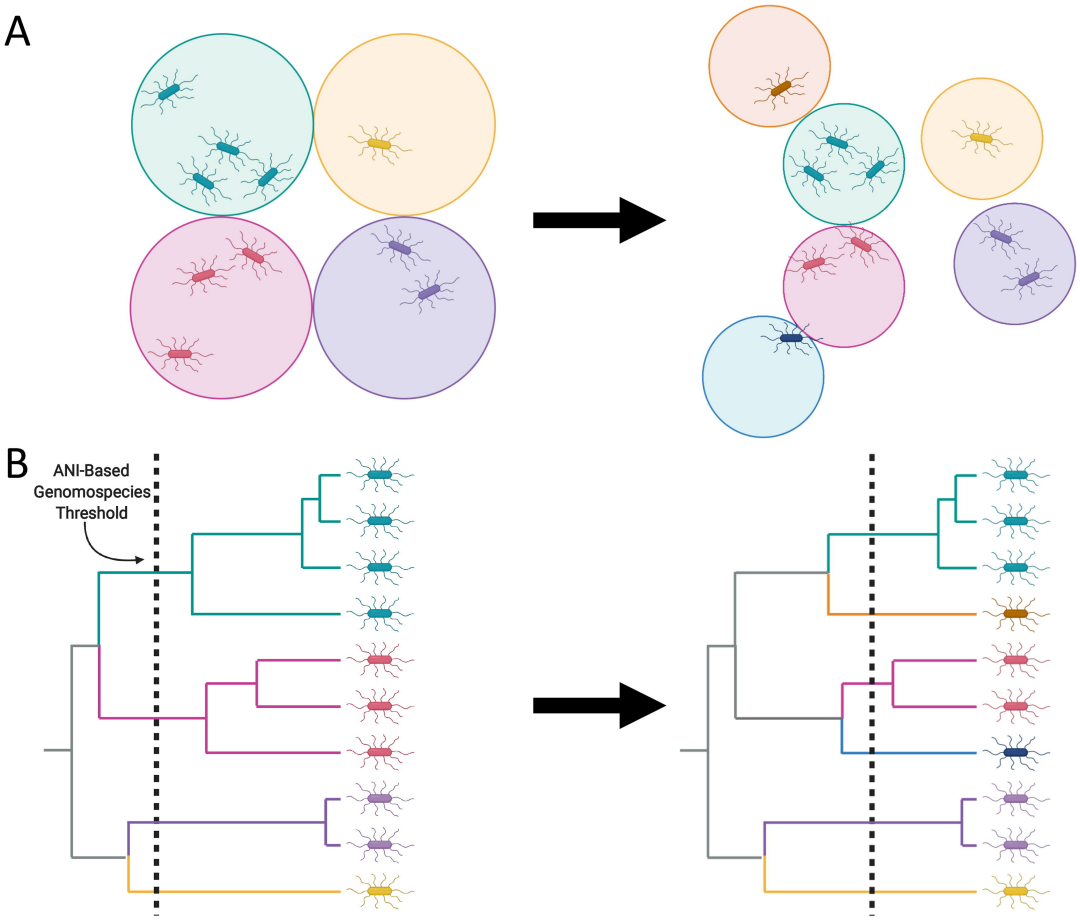

FIGURE 2.2 – La définition de *genomospecies* est dépendante du seuil d'ANI fixé pour les délimiter. A) Représentation schématique d'un changement de seuil d'ANI pour classifier un même ensemble de souches. B) Représentation phylogénétique de la situation. Tiré de CARROLL et al., 2022a.

processus évolutifs inhérents à ces observations. En effet, dans le cas de *B. anthracis*, une nette distinction en termes d'écologie, spécifiquement liée à la virulence, se manifeste par leur mode de vie dans leur environnement naturel. Cette distinction écologique n'est pas facilement réversible, avec le gain ou la perte d'un plasmide par exemple. Par conséquent, il apparaît judicieux de postuler que ces populations continueront à coexister en tant que lignées distinctes (COHAN, 2006). En effet, *B. anthracis* constitue un clone adapté à un nouvel habitat écologique (écosystème), définissant ce que l'on nomme une espèce écologique. Plus précisément, selon le modèle de l'écotype, cette innovation écologique résulte de la sélection naturelle qui a privilégié un mutant adaptatif au sein de l'habitat, surpassant ainsi les autres candidats à l'occupation de cet écosystème. Le mutant adaptatif favorisé est fixé au sein de son écosystème par un processus de balayage sélectif [11], entraînant une réduction locale de la diversité des *locus* au sein de la population de la bactérie considérée. Le mutant adaptatif est ensuite à l'origine d'une diversification du complexe clonal par

11. Processus évolutif par lequel une nouvelle mutation avantageuse se propage rapidement au sein d'une population, entraînant une réduction de la diversité génétique autour du gène muté. Ce phénomène est caractérisé par l'élimination des allèles concurrents dans la région génomique affectée, résultant d'une sélection naturelle forte en faveur de la mutation.

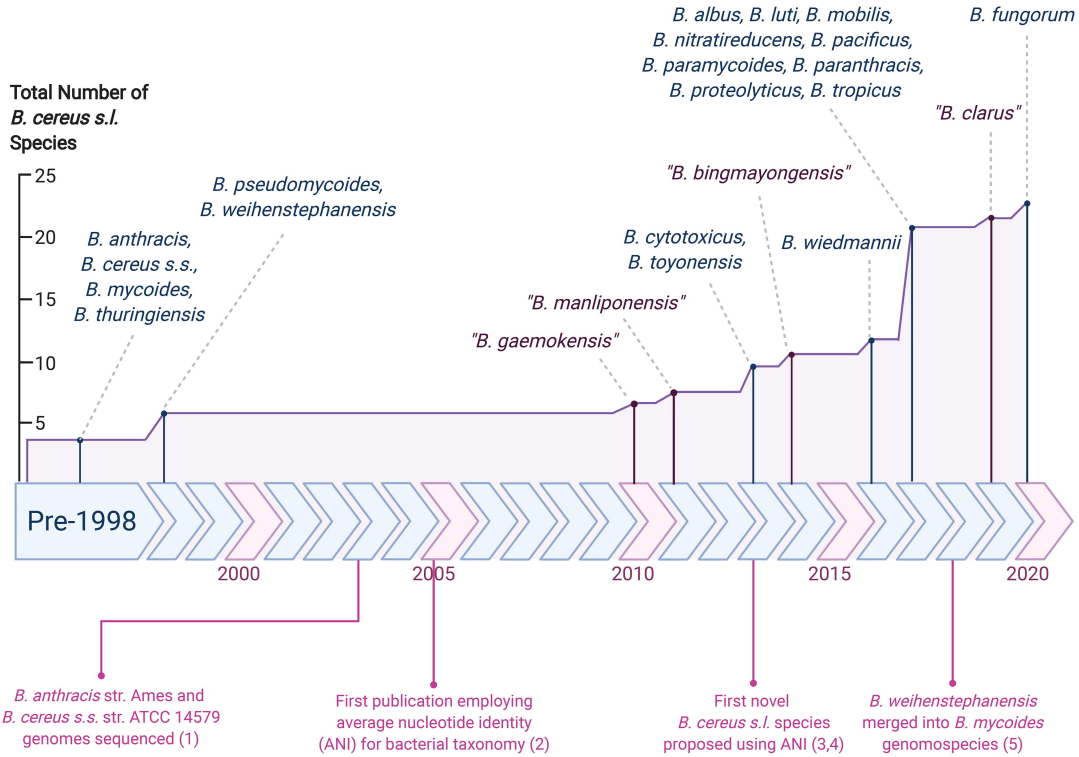

FIGURE 2.3 – Évolution du nombre d'espèces référencées au sein du groupe *B. cereus*. En bleu : les espèces référencées, en violet, les espèces proposées dans la littérature. Tiré de CARROLL et al., 2022a.

dérive génétique [12] principalement, les bactéries clonales se caractérisant par une rareté des phénomènes de recombinaison. Le groupe formé est alors appelé "écotype" (COHAN, 2006 ; SOBRAL, 2012). Le seul critère de "proximité génétique" conduirait à dire que *B. anthracis* ne peut constituer une espèce, parce que ses plus proches voisins montrent une similarité de séquence supérieure à 99%. Ce débat, lié à une forme d'abus de position dominante de l'analyse génétique, n'est pas propre à *B. anthracis*, puisqu'il est illustré par un certain nombre d'autres pathogènes, tels que *Mycobactérium tuberculosis* ou *Yersinia pestis* pour ne citer que les plus connus. Les tenants d'une approche "proximité génétique" argumenteront que le fait que ces bactéries soient pathogènes est justement l'élément historique qui biaise une approche objective. Les tenants d'une approche "phénétique" argumenteront au contraire que c'est précisément parce qu'il s'agit de pathogènes que la compréhension de la notion d'espèce bactérienne a pu être poussée aussi loin et que c'est notre ignorance de l'immense majorité des écosystèmes qui nous empêche de distinguer plus précisément d'autres entités similaires.

12. Terme de biologie évolutive désignant les variations aléatoires de la fréquence des allèles au sein d'une population. Indépendante de la sélection naturelle, la dérive génétique est due au hasard des événements de reproduction et est particulièrement notable dans les petites populations. Elle peut entraîner une modification significative du patrimoine génétique et affecter la diversité biologique en provoquant la fixation ou l'élimination d'allèles, sans considération pour leur valeur sélective.

2.2 Les souches anthracis-like

Cette section explore les souches du groupe *B. cereus* provoquant des affections analogues à la maladie du charbon. Elle détaille comment certaines souches, en dehors de l'espèce *B. anthracis*, manifestent des pathogénicités similaires, remettant en question les frontières traditionnelles de classification basées sur la virulence. L'enjeu est de comprendre la variabilité au sein du groupe *B. cereus* et de discerner les caractéristiques qui contribuent à ces profils pathogènes parallèles.

Des souches atypiques de *B. cereus*, responsables de maladies évoquant le charbon, ont été identifiées chez l'Homme et d'autres mammifères depuis 2004. Ces souches sont caractérisées par la présence d'un ADN chromosomique de type *B. cereus* et d'un ou deux plasmides de virulence très similaires aux plasmides pXO1 et pXO2. Les symptômes et la létalité des maladies provoquées par ces souches récemment découvertes sont comparables à ceux induits par *B. anthracis*. Bien qu'elles puissent engendrer des affections rappelant le charbon, ces souches sont clairement distinctes du lignage *B. anthracis*, de par leurs attributs phénotypiques (ANTONATION et al., 2016).

L'assignation MLST est actuellement la plus commode pour éviter les confusions sur ce sujet qui évolue rapidement. Les souches dites *B. cereus* biovar anthracis (Bcbva) possèdent les deux plasmides, ont été isolées en Afrique, et appartiennent au ST935 au sein du complexe clonal CC365. Les autres souches atypiques appartiennent à d'autres ST, les ST108 et ST11 également au sein du complexe clonal 365, et le ST78 (HOFFMASTER et al., 2006 ; ANTONATION et al., 2016 ; CARROLL et al., 2022b). Par exemple, les souches atypiques BacTX2020a, BacLA2020a et BacLA2020b ont été isolées la même année (2020), chez le même type d'individu (soudeur d'environ une trentaine d'années) avec les mêmes symptômes (pneumonie), dans la même région des Etats-Unis ; pourtant BacTX2020a appartient au ST108 du complexe clonal MLST CC365 alors que les deux souches BacLA2020a et BacLA2020b sont de type ST78 (CARROLL et al., 2022b). On peut noter ici que ces deux souches identiques à un SNP près ont été isolées, l'une chez le patient, et l'autre dans l'environnement (sol du jardin) de ce patient. Cette observation suggère que les souches de ce type pourraient être endémiques au moins en Amérique du Nord, et être capable de causer une infection de type charbon chez des individus sensibilisés par leur exposition professionnelle aux métaux.

Les caractéristiques des souches anthracis-like sont résumées dans le tableau 2.2.

2.2.1 Les souches isolées chez les humains

L'existence de maladies semblables au charbon dues à une infection par *B. cereus* a été mise en lumière lors de l'examen de cas de pneumonies. Les premières descriptions remontent à 1965, concernant des patients en bonne santé (STOPLER, CAMUESCU et VOICULESCU, 1965). Plus tard, des travailleurs de la métallurgie orginaires de Louisiane (en 1994, 1997, 2007, 2020) (MILLER et al., 1997 ; HOFFMASTER et al., 2004 ; PENA-GONZALEZ et al., 2017 ; CARROLL et al., 2022b) et du Texas (2003, 2011, 2020) (AVASHIA et al., 2007 ; WRIGHT et al., 2011 ; CARROLL et al., 2022b) ont présenté des symptômes similaires à ceux du charbon pulmonaire attribués à des souches atypiques de *B. cereus*, notamment la souche *B. cereus* G9241. Sur les sept cas recensés, six se sont avérés mortels. Ce taux de mortalité de 86% est similaire à celui

du charbon pulmonaire provoqué par *B. anthracis*, se situant entre 86 et 89% (KAMAL et al., 2011). Il est possible que d'autres incidents impliquant *B. cereus* aient eu lieu par le passé, mais ils n'ont pas été confirmés (BEKEMEYER et ZIMMERMAN, 1985 ; MILLER et al., 1997 ; HOLTY, KIM et BRAVATA, 2006). Ces métallurgistes, malgré une immunité normale et sans facteurs de risque apparents, étaient plus exposés à une infection par voie respiratoire en raison de leur profession. En effet, leurs conditions de travail, avec une exposition élevée à la poussière métallique, auraient pu renforcer leur prédisposition aux maladies respiratoires de nature infectieuse (ANTONINI, 2003). Par ailleurs, il y a eu deux incidents où *B. cereus* a causé des symptômes évoquant le charbon cutané. L'un s'est produit en Floride chez une personne non exposée particulièrement à des aérosols de poussières métalliques, l'autre était une contamination en laboratoire en Illinois par la souche *B. cereus* G9241.

La souche atypique *B. cereus* G9241 peut présenter une virulence marquée chez certains mammifères. En effet, elle provoque des affections mortelles similaires au charbon chez la souris, immunodéficiente ou immunocompétente, et chez le cobaye. Néanmoins, elle semble inoffensive pour les lapins blancs de Nouvelle-Zélande (WILSON et al., 2011).

2.2.2 Les souches isolées chez les animaux

D'autres mammifères tels que les éléphants, singes, gorilles, chimpanzés, ainsi que du bétail ont été touchés en Afrique de l'Ouest et Centrale par une maladie ressemblant au charbon due à l'infection par *B. cereus* (LEENDERTZ et al., 2004 ; LEENDERTZ et al., 2006 ; KLEE et al., 2006 ; PILO et al., 2011 ; ANTONATION et al., 2016 ; ZIMMERMANN et al., 2017 ; HOFFMANN et al., 2017). Les premiers cas recensés en Côte d'Ivoire ont affecté six chimpanzés en 2001 et 2002 (LEENDERTZ et al., 2004). Par la suite, au Cameroun en 2004 et 2005, trois chimpanzés et un gorille ont été touchés (LEENDERTZ et al., 2006). La bactérie responsable a été initialement classée comme *B. anthracis* par les auteurs (OKINAKA, PEARSON et KEIM, 2006). D'autres cas ont été enregistrés en 2012 et 2013, touchant 32 animaux sauvages en République Centrafricaine et en République Démocratique du Congo (ANTONATION et al., 2016). La majorité des souches de *B. cereus* liées à cette maladie en Afrique sont catégorisées comme Bcbva et sont différentes des souches atypiques observées chez les humains aux États-Unis, tant au niveau phénotypique que génétique, en particulier en termes de plasmides (ANTONATION et al., 2016 ; HOFFMANN et al., 2017). Cependant, une souche singulière, référencée initialement comme *B. anthracis* JF3964 et provenant de bovins au Cameroun, mais dont la séquence génomique n'a pas été rendue publique, se distinguerait des souches Bcbva proches (avec une absence du marqueur chromosomique *Ba813*), même si elle contient les deux plasmides de virulence, pBCXO1 et pBCXO2 (PILO et al., 2011 ; ANTONATION et al., 2016).

Au sein du Parc national de Taï situé en Côte d'Ivoire, les céphalophes [13] sont particulièrement sensibles à la maladie. Les mangoustes et les porcs-épics y sont plus occasionnellement touchés. La transmission se fait principalement par les carcasses d'animaux et les insectes (HOFFMANN et al., 2017). Dans cette même région endémique, une grande proportion (38%) des décès de la faune serait liée à une maladie semblable au charbon provoquée par Bcbva (HOFFMANN et al., 2017). Paradoxalement, seulement 5% de cette faune a montré des signes d'une réponse immunitaire à Bcbva, suggérant que l'infection aurait un taux élevé de mortalité (ZIMMERMANN et al., 2017). Bien qu'aucun cas d'infection humaine par Bcbva n'ait été signalé, des

13. Petite antilope africaine de la famille des bovidés.

anticorps dirigés contre l'antigène spécifique pXO2-60 de Bcbva ont été identifiés chez les populations locales qui seraient donc peu sensibles (DUPKE et al., 2020).

La majorité des souches Bcbva diffèrent des souches anthracis-like isolées chez des humains aux États-Unis par leurs caractéristiques génétiques : elles renferment simultanément des plasmides très similaires à pXO1 et pXO2 (pBCXO1 et pBCXO2). Ces plasmides codent respectivement pour les toxines LF, EF et la capsule d'acide hyaluronique d'une part, ainsi que pour la capsule en acide poly-γ-D-glutamique d'autre part (KLEE et al., 2006 ; KLEE et al., 2010 ; BRÉZILLON et al., 2015). Par ailleurs, des expériences chez la souris et le cobaye ont démontré que deux souches de Bcbva (CI et CA) sont plus virulentes que la souche vaccinale *B. anthracis* Sterne, mais un peu moins que *B. anthracis* 9602 (BRÉZILLON et al., 2015).

Nom	Origine	Hôte	Profil MLST	Disponibilité des données WGS	pXO1, pXO2 ou affiliés	Référence
G9241 FDAARGOS_897	Louisiane (USA)	Humain	ST78	Oui	pBCXO1	HOFFMASTER et al., 2004 SICHTIG et al., 2019
03BB87	Texas (USA)	Humain	ST78	Oui	pBCXO1	HOFFMASTER et al., 2006
03BB102 FDAARGOS_918	Texas (USA)	Humain	ST11	Oui	pBCXO1*, pBCXO2*	HOFFMASTER et al., 2006 SICHTIG et al., 2019
Elc2	Texas (USA)	Humain	ST11	Non	pBCXO1	WRIGHT et al., 2011
FL2013	Floride (USA)	Humain	ST78	Oui	pBCXO1	MARSTON et al., 2016
LA2007	Louisiane (USA)	Humain	ST78	Oui	pBCXO1	PENA-GONZALEZ et al., 2017
LA4726	Louisiane (USA)	Humain	ST78	Non	pBCX01	MARSTON et al., 2016
G9898	Louisiane (USA)	Humain	?	Non	pBCXO1	SUE et al., 2006
BacTX2020a	Texas (USA)	Humain	ST108	Oui	pBCX01	CARROLL et al., 2022b
BacLA2020a	Louisiane (USA)	Humain	ST78	Oui	pBCX01	CARROLL et al., 2022b
BacLA2020b	Louisiane (USA)	Humain	ST78	Oui	pBCX01	CARROLL et al., 2022b
CA	Réserve du Dja (Cameroun)	Chimpanzé, gorille	ST935	Oui	pBCXO1, pBCXO2	KLEE et al., 2006
CI	Parc national de Taï (Côte d'Ivoire)	Chimpanzé	ST935	Oui	pBCXO1, pBCXO2	KLEE et al., 2006
JF3964	Koza (Cameroun)	Bovin	?	Non	pBCXO1, pBCXO2	PILO et al., 2011
BC-AK	Chine	Kangourou	ST78	Oui	pBCXO1, pBCXO2*	DUPKE et al., 2018
UFBc0002	Parc national de Taï (Côte d'Ivoire)	Singe	ST935	Oui	pBCXO1, pBCXO2	NORRIS et al., 2023
UFBc0007	Parc national de Taï (Côte d'Ivoire)	Singe	ST935	Oui	pBCXO1, pBCXO2	NORRIS et al., 2023
UFBc0009	Parc national de Taï (Côte d'Ivoire)	Singe	ST935	Oui	pBCXO1, pBCXO2	NORRIS et al., 2023
UFBc0011	Parc national de Taï (Côte d'Ivoire)	Singe	ST935	Oui	pBCXO1, pBCXO2	NORRIS et al., 2023

TABLE 2.2 – Tableau récapitulatif des caractéristiques des différentes souches anthracis-like. * : les séquences des plasmides pBCXO1 et pBCXO2 ne sont présentes que partiellement. Adapté de SICHTIG et al., 2019 ; BALDWIN, 2020 ; CARROLL et al., 2022b ; NORRIS et al., 2023.

2.3 Épidémiologie de la maladie du charbon

La thématique de la dissémination mondiale est maintenant abordée, par l'analyse des causes de sa propagation et l'identification des réservoirs naturels. Cette section explore comment les interactions environnementales, les pratiques humaines et les caractéristiques biologiques de la bactérie contribuent à l'expansion géographique de la maladie associée. Là encore, ces différents facteurs engagent une réflexion sur l'origine géographique et temporelle de l'espèce.

2.3.1 Réservoirs de *Bacillus anthracis*

Pour *B. anthracis*, l'environnement hydrotellurique représente le principal réservoir permanent. Dans cet habitat, la bactérie est capable de subsister durablement sous forme sporulée. Ces spores, dotées d'une résistance à de nombreux facteurs environnementaux tels que les températures extrêmes, les écarts de pH, la dessiccation, les agents désinfectants et l'irradiation, peuvent demeurer viables dans le sol sur de longues périodes (BRAUN et al., 2022).

Bien que l'eau puisse être contaminée par *B. anthracis*, elle l'est généralement moins que le sol, qui agit comme un filtre naturel. L'eau peut néanmoins jouer un rôle dans la dispersion des spores, en agissant comme vecteur de transport et en concentrant les spores dans certains environnements (DROMIGNY, 2009).

Quant aux animaux, tant domestiques que sauvages, ils ne sont généralement pas considérés comme des réservoirs permanents de cette bactérie. Dans les zones hyperenzootiques du charbon, comme c'est le cas dans certaines régions du Zimbabwe, la faune affectée par la maladie peut augmenter la concentration en spores de *B. anthracis* dans le sol après leur mort, renouvelant ainsi le réservoir environnemental (COETZER, THOMSON et TUSTIN, 1995 ; MUKARATI et al., 2020).

En ce qui concerne les humains, ils ne sont pas reconnus comme un réservoir traditionnel pour le charbon (DROMIGNY, 2009).

2.3.2 Influence du climat

Le charbon est une pathologie dont l'incidence chez les animaux est influencée par les conditions météorologiques. Dans les zones où le charbon est endémique, les flambées de la maladie sont souvent précédées par des périodes de chaleur et de sécheresse, suivant elles-mêmes des épisodes de précipitations importantes ou d'inondations. Bien que les régions chaudes soient plus communément associées à l'enzootie du charbon, des cas ont été documentés dans des climats plus froids, y compris près du cercle arctique, comme chez les rennes dans la péninsule de Yamal (TIMOFEEV et al., 2019).

Les dynamiques climatiques semblent jouer un rôle en modifiant l'exposition des animaux aux spores de *B. anthracis*, surtout pendant les périodes de sécheresse où le fourrage est rare et les animaux peuvent brouter plus près du sol, où les spores sont présentes. Cette observation est pertinente pour les zones arides (TURNBULL et al., 1992) et ne s'applique pas nécessairement aux régions tempérées telles que les États-Unis ou l'Australie, où un important foyer de la maladie en 1997 a coïncidé avec une période d'irrigation intensive (WORLD HEALTH ORGANIZATION, 2008).

L'apparition de la maladie du charbon est également corrélée à l'immunosuppression des animaux due au stress provoqué par divers facteurs environnementaux, en particulier la réduction de l'accès à la nourriture et à l'eau induite par des conditions météorologiques défavorables. Le stress environnemental peut diminuer l'immunité des animaux, réduisant la quantité de spores nécessaire pour provoquer une infection. Par exemple, chez le bison, le regroupement des animaux autour des points d'eau résiduels, ajouté à la chaleur, à une densité accrue d'insectes et à des changements hormonaux liés à la période du rut, peut contribuer à l'augmentation du risque de flambée du charbon (DRAGON et al., 1999 ; DRAGON, RENNIE et ELKIN, 2001 ; DROMIGNY, 2009).

Les eaux de ruissellement pourraient jouer un rôle dans la distribution des spores de *B. anthracis*, en les accumulant dans des zones spécifiques. Lors de périodes de sécheresse, l'évaporation de l'eau pourrait entraîner une concentration des spores,

particulièrement dans les derniers points d'eau affectés par l'évaporation (BAKKEN, 1985). De plus, il est possible que des interactions hydrophobes favorisent l'attachement des spores à la végétation, augmentant la probabilité de contact avec les herbivores vulnérables et, par conséquent, le risque d'infection (DOYLE, NEDJAT-HAIEM et SINGH, 1984).

2.3.3 Influence anthropique

Les activités anthropiques, incluant diverses pratiques industrielles, peuvent contribuer à la propagation de *B. anthracis* dans l'environnement. Des industries telles que les tanneries et les sites d'équarrissage ont été historiquement liées à la dissémination du charbon en raison de la gestion de leurs eaux résiduaires. Des observations ont montré une prévalence du charbon dans les exploitations agricoles situées en aval de telles installations, où les spores de *B. anthracis* peuvent être transportées jusqu'aux champs par les effluents industriels (TURNBULL, 1996). Cette contamination peut survenir même plusieurs années après la fermeture des installations concernées, lors de la remise en culture des terres ou lors de la remobilisation des sédiments fluviaux, comme cela a été noté en Écosse (DROMIGNY, 2009). Le mouvement transrégional du bétail constitue un autre vecteur d'introduction et de dissémination de *B. anthracis*. Par exemple, dans les années 1950, le sud des États-Unis a connu des foyers récurrents du charbon chez le bétail. Il est supposé qu'il a été apporté en Oklahoma depuis les zones côtières affectées de Louisiane et du Texas, à travers le déplacement de troupeaux infectés entre ces États (DROMIGNY, 2009).

De plus, les produits d'origine animale tels que les carcasses, les poils, la laine et les os sont fréquemment transportés sur de grandes distances pour être utilisés dans certains secteurs de l'industrie, ainsi que dans la production alimentaire et l'artisanat. Cette pratique peut entraîner la propagation de *B. anthracis* bien au-delà de sa source initiale d'infection, avec des animaux contractant le charbon à partir de matières premières contaminées importées d'autres continents. De plus, le commerce international des animaux vivants et des produits carnés peut conduire à la dissémination transfrontalière de diverses souches de *B. anthracis*. Par exemple, des études de cas ont révélé des similitudes génétiques entre les souches de *B. anthracis* en France et en Espagne et celles identifiées en Amérique du Nord (VERGNAUD et al., 2016 ; BASSY et al., 2023). En outre, l'expédition de vêtements ou d'étoffes contaminés par des spores de *B. anthracis* a pu être la cause de leur dissémination sur des distances importantes (VERGNAUD, 2020). En Asie centrale, notamment au Kazakhstan, les souches de *B. anthracis* ont été associées historiquement aux routes commerciales telles que la route de la soie, qui facilitaient le transfert de marchandises entre la Chine centrale et la Méditerranée orientale (SHEVTSOV et al., 2021).

En ce qui concerne le charbon humain, bien que la maladie ne soit pas typiquement classée parmi les maladies affectant les voyageurs, des cas d'infections chez les voyageurs ont été documentés dans des zones d'endémie. Par exemple, un cas a été signalé chez un voyageur belge au Botswana, illustrant que les déplacements internationaux peuvent occasionnellement conduire à l'exposition au charbon dans les régions où la maladie est présente (ENDEN, VAN GOMPEL et VAN ESBROECK, 2006).

Des facteurs de risque supplémentaires pourraient influencer la transmission du charbon à l'Homme. Ces facteurs comprennent des pratiques alimentaires à risque, telles que l'abattage de bétail infecté en dehors des contrôles sanitaires réglementaires, ou la consommation de viande issue d'animaux morts de maladie, qui n'ont

pas été soumis à une inspection vétérinaire avant la consommation (WORLD HEALTH ORGANIZATION, 2008 ; DROMIGNY, 2009).

2.3.4 Répartition géographique du charbon

Des études épidémiologiques menées sur une période de 20 ans dans plus de 70 pays ont permis de cartographier la distribution géographique du charbon de 2005 à 2016 (CARLSON et al., 2019). Les régions endémiques comprennent la Chine, le Kazakhstan, l'Amérique du Nord, l'Australie et une large étendue de l'Afrique subsaharienne. Ces zones, souvent en développement, sont confrontées à un manque de ressources pour la vaccination prophylactique du bétail, une stratégie qui pourrait réduire le risque de transmission à l'Homme. En contraste, les données manquent pour évaluer l'incidence de la maladie dans certaines régions, telles que les zones polaires et le Sahara occidental.

En plus des régions susmentionnées, une grande proportion du continent européen, de la péninsule anatolienne et des zones avoisinantes pourraient être propices à la présence de *B. anthracis*. Bien que des pays comme la Turquie et l'Afrique du Sud montrent des taux d'incidence élevés, d'autres comme l'Éthiopie ont présenté une prévalence inférieure aux attentes malgré une charge épidémiologique significative. Cette sous-représentation peut être due à des limites dans les capacités de surveillance, exacerbées par la pauvreté et la prévalence d'autres maladies tropicales et zoonotiques.

En définitive, le charbon affecte une large gamme d'animaux, tant sauvages que domestiques, et pose un risque pour environ 1,83 milliard de personnes vivant dans des régions à risque. Ces régions englobent principalement des communautés rurales comprenant 63,8 millions d'éleveurs et environ 1,1 milliard de têtes de bétail réparties à travers l'Eurasie, l'Afrique et l'Amérique du Nord (CARLSON et al., 2019).

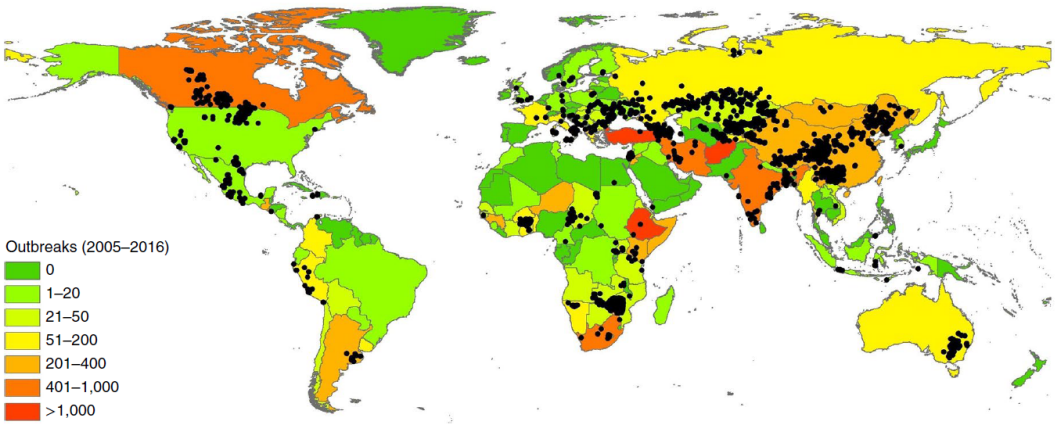

FIGURE 2.4 – Distribution mondiale des épidémies de charbon et localisations géographiques épisodiques (de janvier 2005 à août 2016). Tiré de CARLSON et al., 2019.

2.4 Modèle d'évolution de *Bacillus anthracis*

Les observations exposées dans les trois sections précédentes amènent à se questionner sur l'émergence et l'évolution de *B. anthracis*. Cette section se penche dans un premier temps sur la clonalité de l'espèce *B. anthracis*, avant de s'intéresser aux hypothèses existantes quant à l'émergence et la propagation de *B. anthracis*.

La place particulière qu'occupe *B. anthracis* au sein du groupe *B. cereus* soulève la question de son émergence et de son évolution. En effet, il a été vu *supra* que le groupe *B. cereus* était composé d'espèces très proches génétiquement, avec des caractéristiques phénotypiques diversifiées, provoquant (ou non) des pathologies humaines ou animales de différentes natures et occupant de multiples niches écologiques. Parmi elles, *B. anthracis* occupe une place à part, de par sa capacité à provoquer la maladie du charbon chez ses hôtes (dont les humains) et sa structure clonale qui la différencie du reste du groupe *B. cereus*. Pourtant, il a été observé que des souches s'apparentant génétiquement à d'autres espèces que *B. anthracis* possédant des plasmides de virulence similaires étaient responsables de pathologies s'apparentant au charbon. Ces particularités ont amené à établir des hypothèses quant aux deux problématiques soulevées : l'émergence du caractère clonal de *B. anthracis* (autrement dit, sa différenciation du reste du groupe *B. cereus*) et l'acquisition de son pouvoir de virulence (et en particulier l'origine de ses plasmides pXO1 et pXO2).

2.4.1 Structures de population au sein du groupe *Bacillus cereus*

La figure 2.5 illustre la répartition phylogénétique des souches du groupe *B. cereus*. On y observe une organisation en trois clades distincts. L'ensemble des souches de *B. anthracis* isolées jusqu'à aujourd'hui se situent dans le même clade, le clade 1. Y sont également présentes les souches anthracis-like, même si leur chromosome est plus proche génétiquement de *B. cereus* que de *B. anthracis* au sein de ce clade (ANTONATION et al., 2016), comme illustré sur la figure 2.6. On y retrouve également des souches de *B. cereus* ou *B. thuringiensis*, proche génétiquement de *B. anthracis*, comme *B. cereus* E33L, dite *Zebra killer*, isolée sur une carcasse de zèbre ou *B. thuringiensis* serovar *konkukian* 97-27 isolée sur une plaie nécrotique humaine (HERNANDEZ et al., 1998 ; HAN et al., 2006). Toutes les souches de *B. anthracis* appartiennent au même groupe au sein du groupe *B. cereus*, malgré leur répartition mondiale. Au sein de la lignée *B. anthracis*, aucune autre souche d'une autre espèce n'y est répertoriée. La diversité génétique y est très faible, contrairement au reste du groupe *B. cereus*. Cela est caractéristique d'une structure de population clonale.

Dans certaines communautés bactériennes, on observe une distribution où un petit nombre de génotypes prédomine, accompagnés par une multitude de génotypes plus rares, tous dérivant d'un génotype ancestral commun. Au sein de ces communautés, les fréquences de recombinaison génétique sont généralement basses. On identifie des lignées distinctes constituées de souches fortement apparentées, également connues sous le nom de "clones", qui se diversifient essentiellement par le biais de mutations (FEIL et ENRIGHT, 2004). Ces ensembles de clones forment des complexes clonaux qui peuvent être regroupés en lignées. Ces dernières peuvent à leur tour se regrouper en clades, qui représentent des ensembles très larges de bactéries partageant une origine évolutive commune.

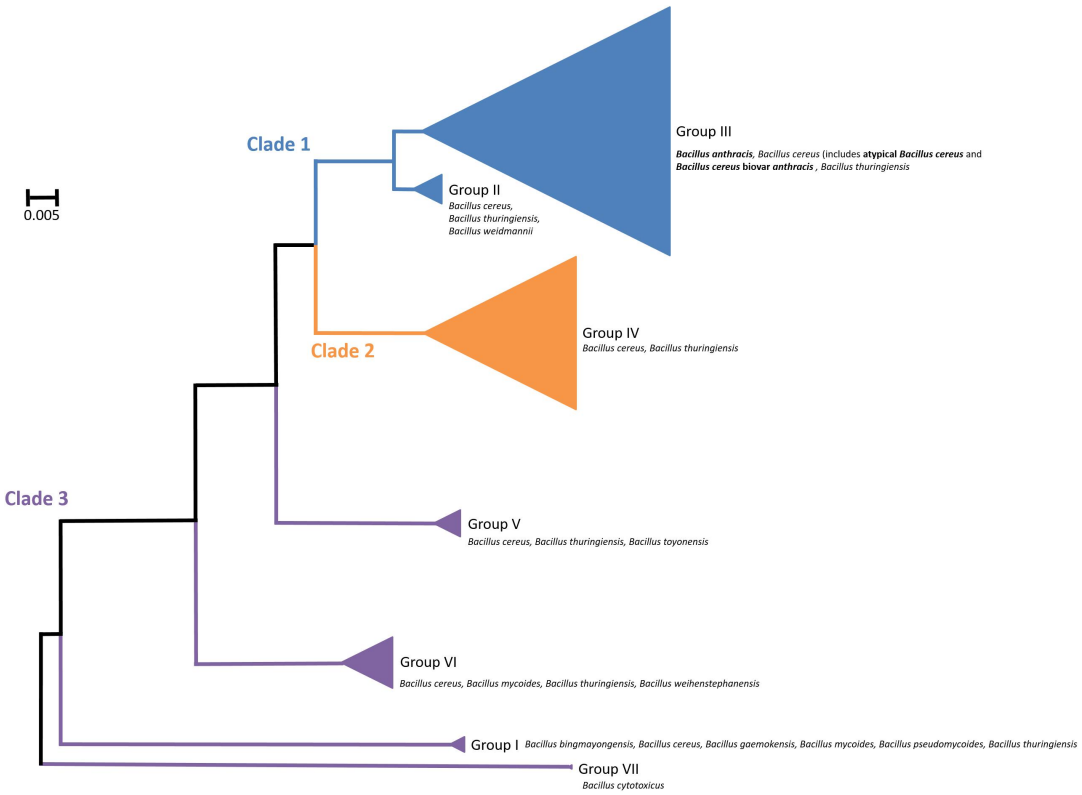

FIGURE 2.5 – Organisation du groupe *B. cereus* en clades. Tiré de BALDWIN, 2020.

Cette situation se reconnaît dans le cas de *B. anthracis*. En outre, d'autres complexes clonaux sont observés au sein des différentes clades du groupe *B. cereus* (PRIEST et al., 2004). Cependant, certaines études ont révélé des phénomènes de recombinaison au sein du groupe (HELGASON et al., 2004).

Un scénario privilégié serait donc une structure de population intermédiaire entre la clonalité et la panmixie [14], dite mixte. Le groupe *B. cereus* aurait évolué de manière clonale à laquelle se seraient mêlés des épisodes de recombinaison. Autrement dit, la diversification des lignées émerge de séparations clonales, lesquelles sont influencées par la sélection naturelle et la dérive génétique. Ces séparations clonales sont la conséquence de l'apparition d'un mutant au sein d'une niche écologique particulière, formant un nouvel écotype [15] (FEIL et ENRIGHT, 2004). Ainsi, les espèces *B. anthracis*, *B. cereus* et *B. thuringiensis* auraient chacune divergé à partir d'un ancêtre commun, chacune exploitant ensuite une niche écologique spécifique (TURNBULL, 1999 ; JENSEN et al., 2003).

L'étude de ce groupe mixte présente donc deux aspects : un accès à de nombreuses souches diversifiées, dû à sa structure panmictique et une faible variation génétique observée au sein d'un complexe clonal donné, due à l'évolution de chacun d'eux. Dans le cas de *B. anthracis*, son émergence est relativement récente, et même si la diversité des souches connues à ce jour est élevée, ces dernières présentent une faible variation génétique (VERGNAUD, 2020). Par conséquent, l'étude de cette population clonale et en particulier la caractérisation de son émergence au sein du groupe

14. Distribution aléatoire d'allèles au sein d'une population, du fait d'échanges génétiques très fréquents.

15. Désigne une population ou un groupe d'organismes qui présente des adaptations génétiques spécifiques à un environnement particulier

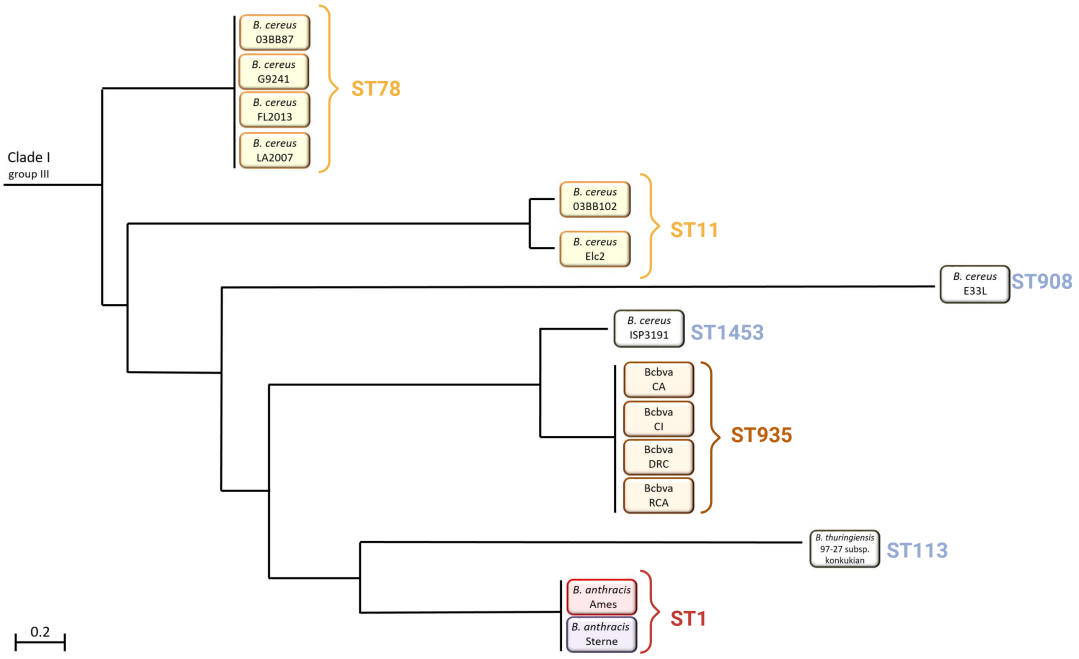

FIGURE 2.6 – Phylogénie simplifiée au sein du clade 1 du groupe *B. cereus*. En jaune : souches atypiques, en orange : souches Bcbva, en rouge/violet : *B. anthracis*. Les profils MLST sont indiqués. Adapté de BALDWIN, 2020.

B. cereus reste un défi. Il est important de garder à l'esprit que l'impact économique et biomédical de la maladie du charbon a conduit à collecter une grande quantité de souches et à étudier la structure de population de *B. anthracis* avec une précision exceptionnelle. L'accroissement des données disponibles concernant les infections alimentaires causées par *B. cereus sensu lato* permettra probablement de faire des comparaisons de plus en plus valides dans un avenir proche (BONIS et al., 2021).

2.4.2 Émergence de l'espèce *Bacillus anthracis*

L'histoire évolutive de *B. anthracis* a fait l'objet de plusieurs constats effectués successivement :

Émergence L'accès progressif au séquençage de génomes complets a permis de confirmer la faible diversité génétique au sein de *B. anthracis* (autrement dit, son caractère monomorphe). Cela va dans le sens d'une émergence récente de l'espèce. Celle-ci proviendrait d'une espèce différente ou d'une population distincte, potentiellement à cause d'une mutation qui a accru sa valeur sélective, menant à une croissance de sa prévalence et de sa distribution géographique (KEIM et WAGNER, 2009). Plus précisément, *B. anthracis* dériverait d'une souche de l'espèce *B. cereus*, ayant subi trois événements évolutifs majeurs : l'acquisition des deux plasmides de virulence pXO1 et pXO2, et une mutation pléiotrope [16] sur l'opéron *plcR* (SLAMTI et al.,

16. Modification génétique qui a la particularité d'affecter simultanément plusieurs traits ou fonctions biologiques différents chez un organisme.

2004 ; EASTERDAY et al., 2005). Ces trois caractéristiques permettraient ainsi à *B. anthracis* de se distinguer de *B. cereus*, en provoquant une maladie animale distincte, le charbon (OKINAKA, PEARSON et KEIM, 2006).

Origine géographique L'Afrique, et plus particulièrement la région subsaharienne, est envisagée comme le berceau de la diversité de la maladie du charbon, ce qui suggère que *B. anthracis* y a probablement émergé. Le continent, caractérisé par une riche population de grands herbivores où la maladie est endémique, révèle une importante diversité génétique de cette bactérie (KEIM et al., 1997 ; SMITH et al., 1999 ; SMITH et al., 2000). Cela signifierait donc selon certains auteurs que le charbon existait naturellement parmi la faune sauvage dans cette région avant l'introduction humaine et celle des animaux domestiques (HUGH-JONES et DE VOS, 2002). Nous reviendrons sur ce point plus loin, pour présenter les interprétations plus actuelles.

Propagation de l'espèce La capacité de dispersion est essentielle pour le succès d'un clone pathogène comme *B. anthracis*, qui bénéficie de ses spores résistantes pour se propager sur de longues distances. Ces adaptations favorisent la colonisation de niches similaires à leur milieu d'origine, souvent non occupées par d'autres agents pathogènes. La variété des hôtes que *B. anthracis* peut infecter facilite également son expansion géographique (KEIM et WAGNER, 2009).

Quant aux causes de la propagation mondiale de *B. anthracis*, l'hypothèse la plus communément admise est que la lignée A, largement majoritaire au sein de l'espèce, s'est répandue grâce aux activités humaines, son émergence coïncidant avec la domestication des animaux (VAN ERT et al., 2007). *A contrario*, les lignées B et C, qui ne représentent qu'une faible part des souches observées à l'échelle mondiale, n'auraient pas autant bénéficié de la dispersion aléatoire influencée par l'Homme (KEIM et WAGNER, 2009). D'autres raisons complémentaires ont été évoquées pour expliquer cette moindre dispersion géographique. En effet, l'analyse des souches de *B. anthracis* dans le parc national Kruger a démontré que la lignée A s'adapte à une plus grande variété d'environnements par rapport à la lignée B, qui est confinée à des niches écologiques limitées (SMITH et al., 2000). Globalement, la moindre abondance des lignées B et C reflète potentiellement les limites de leur spécialisation écologique, due à un coût d'adaptation trop élevé (SMITH et al., 2000 ; KASSEN, LLEWELLYN et RAINEY, 2004). Là encore, des travaux récents permettent d'affiner cette vision.

Datation du MRCA de *B. anthracis* Une première datation des différentes lignées a été opérée par VAN ERT et al., 2007. Selon le modèle établi, la lignée C se serait séparée des lignées A et B entre 12,857 et 25,714 ans avant notre ère. La divergence des lignées A et B serait datée entre 8,746 et 17,493 ans avant notre ère. La radiation initiale de la lignée A de *B. anthracis* aurait eu lieu entre 3,277 et 6,555 ans avant notre ère, pendant l'Holocène Moyen [17].

KENEFIC et al., 2009 apporteront ensuite une correction quant aux estimations annoncées. En effet, la divergence temporelle de la lignée A pourrait être ajustée en tenant compte du taux de génération observé dans le cycle naturel de l'épidémie. Le modèle de VAN ERT et al., 2007 se basait sur une estimation d'une demi à une

17. Période géologique qui s'étend approximativement de 8,200 à 4,200 ans avant notre ère. Elle constitue la phase intermédiaire de l'époque Holocène, marquée par des changements climatiques significatifs, des développements dans les pratiques agricoles humaines et une expansion des sociétés néolithiques.

génération (cycle infectieux) par an, mais des données plus récentes sur des épidémies de charbon au Canada montrèrent un taux inférieur, avec 0.28 génération par an (DRAGON et ELKIN, 2001). Cette fréquence plus faible suggère que les estimations antérieures de divergence, proposées par VAN ERT et al., 2007, étaient probablement sous-estimées. En définitive, ce type d'approche suggérerait que le MRCA [18] de *B. anthracis* soit apparu il y a plusieurs dizaines de milliers d'années. ACHTMAN, 2008 mentionna par exemple une apparition datant de plus de 17,000 ans (en citant les travaux de VAN ERT et al., 2007).

Contributions récentes Les causes de l'apparition de *B. anthracis* dans certaines régions du monde sont sujettes encore à débat avec différents scénarios à l'étude. Par exemple, pour l'émergence de la lignée nord-américaine A.Br.WNA, une première suggestion avait été que *B. anthracis* aurait colonisé cette région à la fin du Pléistocène [19], utilisant pour cela le pont terrestre de Béringie, qui reliait à l'époque l'Alaska à la Sibérie orientale (KENEFIC et al., 2009). Cependant, la découverte d'autres souches a infirmé cette idée, confortant plutôt l'hypothèse d'exportation plus récente, à l'ère de l'émergence du commerce transocéanique (transatlantique et transpacifique), avec une exportation d'une souche de l'Europe de l'Ouest, probablement via la France et des produits d'origine animale contaminés, tels que des vêtements et autres lainages (VERGNAUD et al., 2016 ; VERGNAUD, 2020). Une étude récente de souches de *B. anthracis* issues d'Espagne favoriserait une origine espagnole de cet ancêtre de la lignée A.Br.WNA, d'exportation "post-colombienne" (BASSY et al., 2023).

D'autres travaux concerneront l'émergence du groupe transeurasien A.Br.008/011 (TEA 008/011). Une hypothèse avancée par TIMOFEEV et al., 2019 est que le MRCA de cette polytomie serait lié aux armées mongoles entre le début du XIIIème siècle et le milieu du XVIème siècle. Le XIIIème siècle serait la période la plus probable, avec un transport de *B. anthracis* par les chevaux des armées mongoles durant leurs conquêtes couvrant l'Asie orientale jusqu'à l'Europe de l'Est. Le déclin de l'empire mongol aurait par la suite freiné cette propagation.

L'analyse de souches de *B. anthracis* provenant du Kazakhstan a permis de tester cette hypothèse et d'affiner la datation de l'émergence du groupe TEA (SHEVTSOV et al., 2021). Cette polytomie semble prédominante dans cette région, avec une forte présence des branches "STI", "Heroin" et "Tsiankovskii" (par ordre de prévalence). La branche "STI" pourrait être spécifiquement liée au Kazakhstan, ayant divergé après la chute de l'empire mongol au XIVème siècle et avant l'expansion chinoise au Xin-Jiang au XVIIIème siècle. Quant à la présence de la branche "Tsiankovskii" au Kazakhstan, elle suggérerait une influence russe. Dans cette conception, c'est l'unification économique résultant de la constitution de l'Empire Mongol, et non les armées mongoles elle-mêmes, qui auraient joué un rôle dans la dissémination. Ensuite, la dislocation de l'empire expliquerait la spécificité géographique de certains lignages, tels que le lignage STI.

Enfin, une dernière étude (TIMOFEEV et al., 2023) apporte de nouvelles hypothèses et précisions quant à la propagation de trois groupes de *B. anthracis* : B.Br.001/002, A.Br.001/002 et la branche "Tsiankovskii" (appartenant au groupe A.Br.008/011). De prime abord, le sous-groupe B.Br.001/002 présent dans l'Arctique russe se trouve également dans la steppe du sud de la Sibérie. Pour la polytomie A.Br.001/002, deux

18. *Most Recent Common Ancestor*

19. Période géologique qui a débuté il y a environ 2.58 millions d'années et s'est terminée il y a environ 11,700 ans, précédant l'Holocène. Le Pléistocène fait partie de l'ère Quaternaire et est caractérisé par des cycles de glaciations majeures qui ont affecté les continents du globe, entraînant des modifications profondes des écosystèmes et des habitats.

nouvelles branches identifiées incluent des souches proches en République de Sakha et sur la péninsule de Yamal, distantes de plus de 1,000 km mais dans un même écosystème de toundra, sans obstacles naturels à la migration des rennes, suggérant une possible circulation arctique de ce génotype. Ces souches pourraient aussi provenir d'introductions indépendantes via des activités humaines, comme le commerce de produits animaux contaminés. La proximité génétique entre les souches de diverses régions sibériennes suggère en effet une dissémination d'origine anthropique, avec les fleuves Ob, Yenisei, et Lena servant de corridors pour le déplacement de *B. anthracis* depuis le sud de la Sibérie vers le nord, favorisé par l'usage historique de ces rivières comme voies de transport notamment après la conquête de la Sibérie par la Russie à partir du XVIe siècle. Enfin, la propagation de la branche "Tsiankovskii" serait tracée de la Turquie via la Bulgarie, longeant le nord de la mer Noire et la mer Caspienne, illustrant un autre exemple de propagation géographique influencée par des routes commerciales et mouvements historiques.

En résumé, un scénario global proposé par VERGNAUD, 2020 serait l'apparition du mutant clonal en Afrique centrale. Cette hypothèse découle de la présence des souches anthracis-like dites Bcbva, possédant des plasmides de virulence très similaires à pXO1 et pXO2 (plus de 99.9% d'homologie). Ce mutant clonal émergerait d'une souche du groupe *B. cereus*, possédant des plasmides similaires à pXO1 et pXO2, puis une dissémination de ce mutant en Afrique aurait provoqué des épidémies de charbon, avec l'apparition de nouveaux génotypes propres à chaque niche écologique. La présence des lignées A, B et C ainsi que de souches anthracis-like en Afrique et en Amérique du Nord serait la conséquence d'exportations successives du mutant clonal hors d'Afrique. Dans le scénario proposé par VERGNAUD, 2020, l'évènement déclencheur de la sortie de l'écosystème "forêt d'Afrique Centrale" serait l'arrivée du pastoralisme en Afrique Centrale (PATIN et al., 2017). Toujours selon ce scénario, la lignée actuellement nommée "Ancient A" et trouvée presque exclusivement en Afrique serait le représentant contemporain de l'écotype ancestral.

À travers ces différents exemples, on comprend l'importance de la découverte et de la caractérisation de nouvelles souches pour affiner et confronter les hypothèses proposées.

Acquisition des plasmides pXO1 et pXO2 La phylogénie des plasmides établit des lignées distinctes entre ceux de *B. anthracis* d'une part et ceux des souches Bcbva d'autre part (ANTONATION et al., 2016). Des transferts horizontaux épisodiques sont également mentionnés pour expliquer l'acquisition de ces plasmides de virulence par des souches anthracis-like. Deux arguments appuient cette idée : d'une part, la coévolution observée entre les séquences chromosomiques et plasmidiques de *B. anthracis* (PENA-GONZALEZ et al., 2018 ; BRUCE et al., 2020) ; d'autre part, les possibilités de transfert de pXO1 et pXO2 à *B. anthracis* ou *B. cereus*, qui ne sont pas autotransmissibles, à l'aide d'un plasmide conjugatif, comme pXO12, pXO14 ou pXO16 de *B. thuringiensis* (REDDY, BATTISTI et THORNE, 1987). Plus récemment, la possibilité du transfert de pXO16 vers la souche *B. cereus* CTMA-1571, appartenant au clade de *B. anthracis* (clade 1 selon la figure 2.5), a été démontrée, même s'il s'avère que la transmission de ce plasmide est dépendante de la souche receveuse plutôt que du clade auquel elle appartient (MAKART et al., 2018 ; HINNEKENS et al., 2019 ; HINNEKENS et MAHILLON, 2022). En définitive, aucune certitude n'est établie actuellement quant à l'apparition de pXO1, pXO2 et des plasmides de virulence des souches anthracis-like, et particulièrement leur formation au sein de ces espèces.

IDÉES À RETENIR

1. *B. anthracis* appartient au groupe *B. cereus*, composé de plusieurs espèces génétiquement très proches. Ces espèces ont un pouvoir de virulence varié, dû à la présence (ou non) de toxines spécifiques, pouvant toucher l'Homme comme les animaux.
2. Certaines souches du groupe *B. cereus*, dites anthracis-like, sont capables de provoquer des maladies similaires au charbon. Cela est dû à la présence de plasmides de virulence s'apparentant à pXO1 et/ou pXO2.
3. Plusieurs facteurs (environnementaux et humains) ont amené à la dissémination de *B. anthracis* sur une large partie du globe. Le suivi épidémiologique de cette bactérie peut par ailleurs se trouver compliqué par sa capacité à persister pendant de très longues périodes dans le sol, qui constitue son réservoir principal.
4. Ces différents éléments soulèvent naturellement la question de l'émergence et de l'évolution de cette espèce, notamment l'apparition de l'ancêtre de l'espèce, sa différenciation du reste du groupe *B. cereus*, la datation des différentes lignées ou encore le scénario décrivant l'acquisition de la virulence.
5. Si l'émergence relativement récente de *B. anthracis* en Afrique subsaharienne, la datation approximative du MRCA de cette espèce et la prépondérance des facteurs anthropiques dans la propagation de la bactérie font l'objet d'un consensus aujourd'hui, aucune certitude n'est établie à ce jour quant à l'apparition de l'ancêtre proprement dit ou la formation des plasmides de virulence.

Chapitre 3

Découvrir le proche voisinage de *Bacillus anthracis*

OBJECTIFS DU CHAPITRE

1. Rechercher des souches du groupe *B. cereus*, proches génétiquement de *B. anthracis*
2. Caractériser ces souches, en particulier en déterminant leur séquence complète et leur position phylogénétique vis-à-vis de *B. anthracis*
3. Comprendre en quoi une meilleure connaissance du voisinage de B. anthracis permet de mieux appréhender la problématique de son émergence
4. Illustrer l'importance de la recherche de ce type de souches afin d'alimenter les bases de données de référence servant aux outils de détection actuels de *B. anthracis*

3.1 Contexte de l'étude

Ce premier chapitre expérimental se divise en deux parties, chacune contenant un article décrivant la découverte de nouvelles souches du groupe *B. cereus*, comme étant actuellement les plus proches voisines connues de *B. anthracis*. Dans un premier temps, l'article 1 décrit l'étude d'un panel de 65 souches du groupe *B. cereus* isolées en Slovénie. Il illustre la forte diversité au sein de ce groupe, en particulier par les disparités existantes en termes de virulence. Y est décrite la souche BC38B, s'avérant être la plus proche voisine de *B. anthracis* parmi les souches dont le génome est disponible. Cette souche contient un grand plasmide, dont l'étude est amorcée dans l'article et sera reprise plus en détail dans le chapitre 4. Enfin, les failles de résolution de certains outils utilisés à des fins de détection de *B. anthracis* sont mentionnées, motivant ainsi l'enrichissement continu par de nouvelles souches des bases de données sur lesquelles se réfèrent ces méthodes.

L'article 2 décrit les souches *B. cereus* FFI_BCgr36 et FFI_BCgr46, isolées en Namibie, à proximité d'une carcasse de zèbre. Ces deux souches se trouvent, comme BC38B, dans le proche voisinage de *B. anthracis*. L'une d'elles contient également un grand plasmide, dont l'étude sera détaillée au chapitre suivant dans le cadre de la détermination de la formation des plasmides de virulence de *B. anthracis*.

3.2 Article 1

microorganisms

MDPI

Article

Get to Know Your Neighbors: Characterization of Close *Bacillus anthracis* Isolates and Toxin Profile Diversity in the *Bacillus cereus* Group

Mehdi Abdelli [1,2,†], Charlotte Falaise [1,*,†], Valérie Morineaux-Hilaire [1], Amélie Cumont [1], Laurent Taysse [1], Françoise Raynaud [1] and Vincent Ramisse [1,*]

[1] DGA CBRN Defence Center, Biology Division, French Ministry of the Armed Forces, 91710 Vert-le-Petit, France; mehdi.abdelli@intradef.gouv.fr (M.A.); valerie.morineaux-hilaire@intradef.gouv.fr (V.M.-H.); amelie.cumont@intradef.gouv.fr (A.C.); laurent.taysse@intradef.gouv.fr (L.T.); francoise.raynaud@intradef.gouv.fr (F.R.)
[2] Institute for Integrative Biology of the Cell (I2BC), CNRS, Université Paris-Saclay, 91190 Gif-sur-Yvette, France
* Correspondence: charlotte.falaise@intradef.gouv.fr (C.F.); vincent.ramisse@intradef.gouv.fr (V.R.)
† These authors contributed equally to this work.

Abstract: Unexpected atypical isolates of *Bacillus cereus s.l.* occasionally challenge conventional microbiology and even the most advanced techniques for anthrax detection. For anticipating and gaining trust, 65 isolates of *Bacillus cereus s.l.* of diverse origin were sequenced and characterized. The BTyper3 tool was used for assignation to genomospecies *B. mosaicus* (34), *B. cereus s.s* (29) and *B. toyonensis* (2), as well as virulence factors and toxin profiling. None of them carried any capsule or anthrax-toxin genes. All harbored the non-hemolytic toxin *nheABC* and sphygomyelinase *spH* genes, whereas 41 (63%), 30 (46%), 11 (17%) and 6 (9%) isolates harbored *cytK-2*, *hblABCD*, *cesABCD* and at least one insecticidal toxin gene, respectively. Matrix-assisted laser desorption ionization-time of flight mass spectrometry confirmed the production of cereulide (*ces* genes). Phylogeny inferred from single-nucleotide polymorphisms positioned isolates relative to the *B. anthracis* lineage. One isolate (BC38B) was of particular interest as it appeared to be the closest *B. anthracis* neighbor described so far. It harbored a large plasmid similar to other previously described *B. cereus s.l.* megaplasmids and at a lower extent to pXO1. Whereas bacterial collection is enriched, these high-quality public genetic data offer additional knowledge for better risk assessment using future NGS-based technologies of detection.

Keywords: *Bacillus anthracis*; *Bacillus cereus s.l.*; whole genome sequencing (WGS); SNP phylogeny; toxin genes; virulence factors; cereulide; MALDI-TOF MS; biovar Thuringiensis; biovar Emeticus

check for updates

Citation: Abdelli, M.; Falaise, C.; Morineaux-Hilaire, V.; Cumont, A.; Taysse, L.; Raynaud, F.; Ramisse, V. Get to Know Your Neighbors: Characterization of Close *Bacillus anthracis* Isolates and Toxin Profile Diversity in the *Bacillus cereus* Group. *Microorganisms* **2023**, *11*, 2721. https://doi.org/10.3390/microorganisms11112721

Academic Editor: Thomas Proft

Received: 7 October 2023
Revised: 27 October 2023
Accepted: 6 November 2023
Published: 7 November 2023

Copyright: © 2023 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/).

1. Introduction

Anthrax, a bacterial zoonosis transmissible to humans, is feared as a biological weapon or bioterrorism agent. The causative agent, *Bacillus anthracis*, has been listed by the Federal Select Agent Program as having the potential to pose a serious threat to public health [1], thus justifying its detection as confidently as possible even in the most challenging situations, such as complex or degraded samples. As major virulence factors, the tripartite anthrax-toxin genes *cya*, *lef* and *pag*, located on plasmid pXO1 (182 kb), and poly-γ-D glutamic acid capsule genes *capABCDE* on pXO2 (95kb) are common targets for detection [2,3]. With rare exceptions, confirmation of the identity of *B. anthracis* and differentiation from other *Bacillus* species is easy with traditional cultivation techniques for a trained eye [4]. Generic broad-spectrum techniques covering multiple pathogens are increasingly desirable, although significant developments have been accomplished, for example, using mass spectrometry (MS) and nanopore sequencing [5–7]. Thus, it is worthwhile to expand the sampling of strains likely to weaken all these techniques and to assess new strains as opportunities arise.

In addition to its clonal population structure, *B. anthracis* is part of a larger group of species with close phylogeny referred to as the *Bacillus cereus* group or *B. cereus sensu lato* (*B. cereus s.l.*). The description of at least 20 species over the last two decades, accompanied by the expansion of whole genome sequencing (WGS), has led to changes from the traditional "legacy" nomenclature [8,9]. A combined genomospecies–subspecies–biovar nomenclature framework was recently proposed [10]. Note that examples of misinterpretation of *B. cereus* group WGS results call for caution from those who are transitioning to WGS for *B. cereus* group strain characterization [11].

Some *B. cereus s.l.* isolates can produce parasporal crystal proteins (Cry, Cyt) with pesticidal properties used worldwide for pest-control in agriculture. Such strains, historically gathered under the name *B. thuringiensis*, coincide with diverse species of the *B. cereus* group and are now referred to as biovar Thuringiensis with the novel nomenclature framework [10]. Updated nomenclature of crystal protein genes is available [12]. Some other *B. cereus s.l.* isolates are responsible for emetic and/or diarrheal foodborne illness or intoxication [13,14]. Briefly, the heat-stable emetic toxin cereulide encoding gene cluster *cesABCD* is located on a large plasmid [15,16]. The term biovar Emeticus is used in the novel nomenclature framework to describe these cereulide producers [10]. The second syndrome is caused by three thermolabile enterotoxins, hemolysin BL (HBL), non-hemolytic enterotoxin (NHE) and cytotoxin K (CytK), chromosomally encoded by the operons *nheABC* and *hblCDAB* and the genes *cytK-1* or *cytK-2*, respectively. Massive horizontal gene transfer has shaped the evolution of the *B. cereus s.l.* enterotoxin operons *hbl*, *nhe* and *cytK* [17].

B. cereus s.l. lineages have close evolutionary relationships and their possible ecological niches and lifestyles are still under elucidation [18,19]. Naturally occurring isolates belonging to the *B. cereus* group with anthrax-toxin genes are able to cause fatalities even in humans [20] (referred to as biovar Anthracis isolates [10]). Some *B. cereus s.l.* causing anthrax produce an exo-polysaccharide capsule encoded by the plasmidic operon *bpsABCDEFGHX* [20–22] alternatively to the regular poly-γ-D glutamic acid capsule encoded by pXO2-*cap* genes [23,24]. *B. cereus s.l.* G9241 elaborates a hyaluronic acid capsule via plasmidic *hasABC* genes, unlike *B. anthracis* which has a mutation preventing the translation of *hasA* [25]. Recently, a retrospective screening of an anthrax-like disease induced by a strain of *Bacillus tropicus* from Chinese turtles in Taiwan reinforced the idea that the host range and geographic distribution of atypical *B. cereus s.l.* are by far underestimated [26].

Virulence determinants of anthrax and anthrax-like isolates are a tiny part of their genome carried, in the main, by mobile elements. False-positive detection issues are critical in complex environments, particularly because promising technologies are rather generic and increasingly sensitive. For instance, a weak metagenomics detection signal of anthrax in air samples of the New York subway wrongly suggested that the pathogen was present in trace amounts in the normal urban microbiome [27]. The study design did not include sample cultivation, thus hampering any chance to trace back to the probable source of such signal. Similarly, metagenome sequencing of an anthrax-negative soil sample using CRISPR-Cas-based detection showed thousands of reads mapped on strain *B. anthracis* Ames Ancestor. No phylogenetic localization was made, although this would have shed light on the nature of this signal [28].

As a result, confident discriminative detection between harmless or any anthrax-causing isolates is relevant for both public health and biodefence. Trust will depend on our extended sampling efforts within the *B. cereus s.l.* lineages. Even if rare or hypothetically underestimated, such atypical lineages should not be ignored.

Aiming to gain an increased representativeness of *B. cereus* group lineages that may improve our reference collection and subsequent reference databases, we seized the opportunity to characterize a collection of *B. cereus s.l.* isolated from clinical specimens, food including dairy products, and water during the 2007–2015 period in Slovenia. Identification using traditional bacteriology techniques preceded WGS-based species classification and toxin profiling with BTyper3 [29], whereas matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS) provided the production status of cereulide. Phylogeny inferred from whole genome single-nucleotide polymorphism (SNP) analysis placed them in a larger context within the *B. cereus* group in order to highlight those closer to the *B. anthracis* lineage and to improve our exclusivity panel.

2. Materials and Methods

2.1. Strains and Cultivation

Strains listed in Supplementary Table S1 were handled in a biosafety level 2 (BSL-2) laboratory. Sixty-five *Bacillus* isolates were kindly provided by Institute of Microbiology and Immunology, Faculty of Medicine, University of Ljubljana (Ljubljana, Slovenia) and were collected between 2007 and 2015 from multiple isolation sources: thirty isolates from clinical specimens, twenty-one from dairy products, nine from foods and five from water. Some isolates were previously investigated for their antimicrobial susceptibility and their pulsotype diversity [30,31]. Strains were purity-checked upon receipt on tryptic soy agar plates supplemented with 5% sheep blood (TSS-agar; BioMérieux, Marcy l'Etoile, France). The hemolytic activity was assessed on TSS-agar and the phospholipase activity was determined on the selective *B. cereus* group BACARA agar (BioMérieux, Marcy l'Etoile, France). All cultures were incubated for 18 h ± 2 h at 37 °C. Gram staining and microscopic observations were conducted using classical procedures.

2.2. MALDI-TOF Mass Spectrometry

Prior to matrix-assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOF MS) analysis, the isolates were cultivated on TSS agar for 18 ± 2 h at 37 °C. A freshly grown colony sample was picked with a 1 µL sterile loop and a thin film was smeared and left to dry on a 96-polished steel target plate (Bruker Daltonik GmbH, Bremen, Germany). Samples were overlaid with 1 µL of the matrix α-cyano-4-hydroxycinnamic acid (HCCA, Bruker Daltonik GmbH, Bremen, Germany) prepared following the instructions for use and with a final concentration of 10 mg/mL. The matrix was left to crystallize for 10 min at room temperature before mass spectra acquisition of the samples with an MALDI Biotyper® Sirius System (Bruker Daltonik GmbH, Bremen, Germany). The data were processed automatically using MBT Compass 4.1.100 software (Bruker Daltonik GmbH, Bremen, Germany) with default parameters (MBT_AutoX AutoXecute method and MBT_Process processing method). The instrument was calibrated in the range of 3637.8–16,953.3 Da using Bruker Bacterial Test Standard (Bruker Daltonik GmbH, Bremen, Germany). The mass spectra obtained from the isolates were compared with those of known microbial isolates of the commercial libraries provided by Bruker Daltonik, including the MBT Compass Library BDAL (Revision H, 2021) and the MBT Security Related Library 1.0 (SR). For the *B. cereus* group, the BDAL library includes strains of the following species: *B. cereus sensu stricto* (×4), *B. thuringiensis* (×1), *B. mycoides* (×1), *B. pseudomycoides* (×1), *B. weihenstephanensis* (×1) and *B. cytotoxicus* (×4), and the SR library includes *B. anthracis* (×23). The degree of correspondence between the test spectrum and the reference spectra in the database is expressed with log(score) values between 0 and 3.0, with a log(score) ≥ 2.0 indicating that identification could be reliable at the species level of the organism. Each strain was spotted on at least two different spots and the spectrum with the best log(score) was taken into account.

2.3. Genomic DNA Extraction

Bacterial biomass was collected from isolation streaks on a TSA agar plate (BioMérieux SA, Marcy l'Etoile, France) incubated 18 ± 2 h at 37 °C. The biomass was transferred in 200 µL of sterile water and bead-grinded for 45 s at 6000 rpm with a Precellys Evolution homogenizer (Bertin Technologies SAS, Montigny-le-Bretonneux, France). Genomic DNA was extracted with a DNeasy® Blood & Tissue kit (Qiagen, Hilden, Germany) following the manufacturer's recommendations. Eluted DNA solution was sterilized via centrifugation (4 min, 12,000× *g* rpm) on 0.2 µm filter microtubes (Merck KGaA, Darmstadt, Germany). DNA quality and concentration were estimated with a spectrophotometer/fluorometer DS-11 Series (DeNovix, Wilmington, DL, USA) and using a Qubit dsDNA HS Assay kit (Invitrogen, Thermo Fisher Scientific, Waltham, MA, USA).

2.4. Illumina Sequencing, De Novo Assemblies and Ames Ancestor-Reference-Based Assemblies

A paired-end library was constructed using Nextera XT DNA library Prep Kit and sequenced on an Illumina MiSeq platform. Low-quality bases were removed using *Trimmomatic* v0.39 [32] and reads were de novo assembled into contigs using SPAdes v3.13.0 [33] (default parameters). For reference-based assembly, reads were mapped against the chromosome sequence of *B. anthracis* Ames Ancestor A2084 (GCF_000008445.1) using BioNumerics 8.1.1 (BioMérieux, Applied Maths, Sint-Martens-Latem, Belgium) with default parameters and the following options: "perform gapped alignments" (enabled).

2.5. Nanopore Sequencing and De Novo Hybrid Assembly

Genomic DNA of BC38B strain was sequenced for 24 h on a MinION system with a FLO-MIN106 flow cell (R9 version) using a ligation sequencing kit SQK-LSK109 (ONT). Reads were basecalled and demultiplexed using Guppy v6.0.7 (super-accurate mode) [34]. Reads were filtered with a q-score threshold of 10 during guppy basecalling. Adapters were trimmed from the reads using Porechop v0.2.4 [35]. Hybrid (Illumina and MinION reads) de novo assembly was performed using SPAdes v3.13.0. Default parameters were used for these software.

2.6. Nucleotide Sequence Accession Numbers

The assemblies were deposited in DDBJ/ENA/GenBank as BioProjects PRJNA945829 and PRJNA891199.

2.7. Public Genomes

In order to place the strains of this study in a larger phylogenetic context, RefSeq genome assemblies of the *B. cereus* group were downloaded at NCBI (last update: 22 January 2023). Because of different assembly levels (complete, scaffold, contig), data were converted to simulated reads using an in-house script (Python v3.6.2) and then mapped to Ames Ancestor chromosome A2084 (GCF_000008445.1). The resulting alignment spanning set with identical length served for genomic SNP analysis as detailed later in the text. Establishing phylogeny of relatives to anthrax justifies Ames Ancestor as appropriate reference and remains reliable since *B. cereus* group members have close phylogeny. The selection of public genomes included the following: (1) a "*B. cereus* group" panel with 25 public genomes spanning over the genomospecies *B. mosaicus*, *B. cereus s.s.*, *B. luti* and *B. toyonenis* as defined by Carroll et al. [10] (reference genomes of the different species proposed by the NCBI and/or by Carroll et al. [8,10]); (2) a "anthrax toxin gene-harboring genomes" panel with 12 genomes that do not belong to the clonal *B. anthracis* lineage, as defined by Carroll et al. [36]; and (3) a "*B. anthracis*" panel with 24 public genomes representing major lineages and sublineages of *B. anthracis* as described previously [37]. The contents of these panels are summarized in Supplementary Table S2.

In addition, in order to highlight strains from this study that are the closest to *B. anthracis*, all the "*Bacillus cereus* group" assemblies were downloaded from the NCBI database (3950 genomes classified into 24 different species, January 2023). The reconstructed genomes exhibiting at least 70% of nucleotide sequence identity with Ames Ancestor were retained (n = 178) as the overall population neighboring *B. anthracis* (Supplementary Table S3).

2.8. Whole Genome SNPs and Population Clustering Analysis

Whole genome SNP analysis with BioNumerics 8.1.1 used the option "Strict SNP filtering (Closed SNP set)" with default parameters (12 bp inter-SNP), and the detected SNPs were used for population clustering using the "Maximum parsimony tree" (MPT) calculation method with default parameters.

2.9. Genomospecies Assignment and Virulence Factor Detection

An in silico characterization of the draft genomes was performed using BTyper3 tool [29]. It enabled an average-nucleotide-identity (ANI)-based genomospecies assignment, an ANI-based subspecies assignment, an eight-group adjusted *panC* group assignment, the identification of the sequence type (ST) and clonal complex, the screening of the main virulence factors within the *B. cereus* group and the detection of some pesticidal toxins. The BtToxin_Digger tool [38], including all referenced pesticidal toxins on the Bacterial Pesticidal Protein Resource Center database [12], was used to complete Btyper3 screening. Default parameters were used for these software.

In addition, all the draft genomes were uploaded to the Type Strain Genome Server (TYGS) [39–41], in user submission mode, to confirm with digital DNA/DNA hybridization values the closest reference species for each strain obtained by Btyper3 through pairwise genomic analysis.

2.10. Plasmid Analysis

Due to its proximity with the *B. anthracis* cluster, an analysis of the strain BC38B was performed with a particular focus on its plasmid. The NCBI non redundant nucleotide database (last update: 22 April 2023) was used to search similar plasmid sequences in comparison to the BC38B plasmid. The DNA sequences of these closest plasmids were downloaded and compared with BLAST Ring Image Generator (BRIG) [42]. To identify minireplicons contained in the BC38B plasmid, a BLASTP analysis was performed against replication proteins previously described in the megaplasmids of the *B. cereus* group [43]. In order to potentially identify new types of minireplicons, the keywords "replication protein" or "Rep protein" or "Primase" were searched from the annotation file. TubZ protein sequences were also investigated with the same method. In addition, Ori-Finder 2022 (accessed on 23 April 2023) [44] was used to find the predicted origins of replication in the BC38B plasmid.

2.11. MALDI-TOF Cereulide Production Detection

Cereulide toxin peaks were searched at 1175 m/z and 1191 m/z as defined by Ducrest et al. [45]. The production of cereulide was assessed for the 12 strains that clustered close to the emetic *B. cereus* strain AH187. It included 11 strains exhibiting the cereulide synthetase genes *cesABCD* (*cesABCD*+) and one lacking the cereulide gene cluster. The emetic strain AND1407 (*cesABCD*+) was used as a positive control for cereulide production and the *B. cereus s.s.* collection strain ATCC 14579 (*cesABCD*−) as a negative control. Sample preparation of cereulide was performed with the smear method, where a fresh colony sample is directly spotted onto the MALDI target plate and covered by 1 µL of the HCCA matrix (also used for the bacteria identification described above). Each strain was spotted in triplicates. Spectra acquisition was conducted with flexControl Analysis software (Bruker Daltonik GmbH, Bremen, Germany) version 3.4 (Build 207.20). The following parameters were set: random walk shots of partial sample with 100 shots at a raster spot (500 single

spectra accumulation); sample rate and digitizer 0.5 GS/s. The smartbeam laser was set to a linear positive mode in the range of 700–6080 Da with a frequency of 200 Hz. Basic laser settings were high voltage: ion source 1, 10.00 kV; ion source 2, 9.03 kV; Lens, 2.99 kV; pulsed ion extraction set to 130 ns. The external calibration of the instrument was performed using low mass range Peptide Calibration Standard II (Bruker Daltonik GmbH, Bremen, Germany) prepared following the manufacturer's recommendations, covering a mass range of 700–3500 Da. The mass spectra obtained were manually analyzed using flexAnalysis software (Bruker Daltonik GmbH, Bremen, Germany) version 3.4 (Build 79) and each spectrum was subjected to spectral preprocessing procedures: smoothing using the SavitzkyGolay algorithm, baseline subtraction using the TopHat algorithm and peak detection using the centroid algorithm with a signal to noise threshold set at 4.

3. Results

3.1. Strain Isolation and Microbiology

Isolates showed great phenotypical diversity with grayish circular colonies ranging in size from 0.1 cm to 1.3 cm; showing entire or irregular margins; with an elevation either flat, raised or umbonate; and a smooth/glistening or rough surface (Figure S1). Most strains expressed on TSS medium showed strong to weak β-hemolysis (hemolysis clearly extending the colony margin, Figure S1A,B,D, or slight hemolysis below colonies, Figure S1C), but also γ-hemolysis (null hemolysis, Figure S1E) and α-hemolysis (partial greenish hemolysis, Figure S1F). Isolations on TSS also allowed us to separate some strains that were mixed in the original samples (hereinafter referred to as SIBC*XX*A or SIBC*XX*B). It also appears that the subculturing of some strains led to phenotypical variations (Supplementary Table S1 summarizes dimorphic isolates; an example is illustrated in Figure S1E,F). All strains showed typical coral-colored colonies surrounded by an opacification halo on BACARA medium as a result of phospholipase activity. Microscopic observations showed Gram-positive rod-shaped and sporulating cells, isolated or in chains of variable length. None of the strains studied showed microscopic or macroscopic characteristics of *B. anthracis*.

3.2. Identification with MALDI-TOF MS

All isolates were identified using MALDI-TOF MS as belonging to the *B. cereus* group with a log(score) $\geq$ 2.0. As both the BDAL and the SR Bruker libraries were used for the mass spectra comparison, false-positive "*B. anthracis*" identification results were obtained for 53% of the isolates, while the other 47% were identified as "*B. cereus*". Supplementary Table S4 summarizes the identification results and scores obtained for each isolate.

3.3. Phylogenetic Relationships between B. cereus s.l. Isolates Based on wgSNP Analyses

An SNP phylogeny, including all the strains from this study and a selection of public genomes ("*B. cereus* group" panel and "anthrax-toxin gene-harboring genomes" panel), was built and allowed for the taxonomic assignment of the studied isolates (Figure 1). Strains showed great genetic diversity across the three genomospecies *B. toyonensis* (×2), *B. mosaicus* (×34) and *B. cereus s.s* (×29), with isolates close to the reference strains of the species *B. anthracis*, *B. tropicus*, *B. pacificus*, *B. paranthracis*, *B. mobilis*, *B. wiedmannii*, *B. cereus s.s.*, *B. thuringiensis* and *B. toyonensis*.

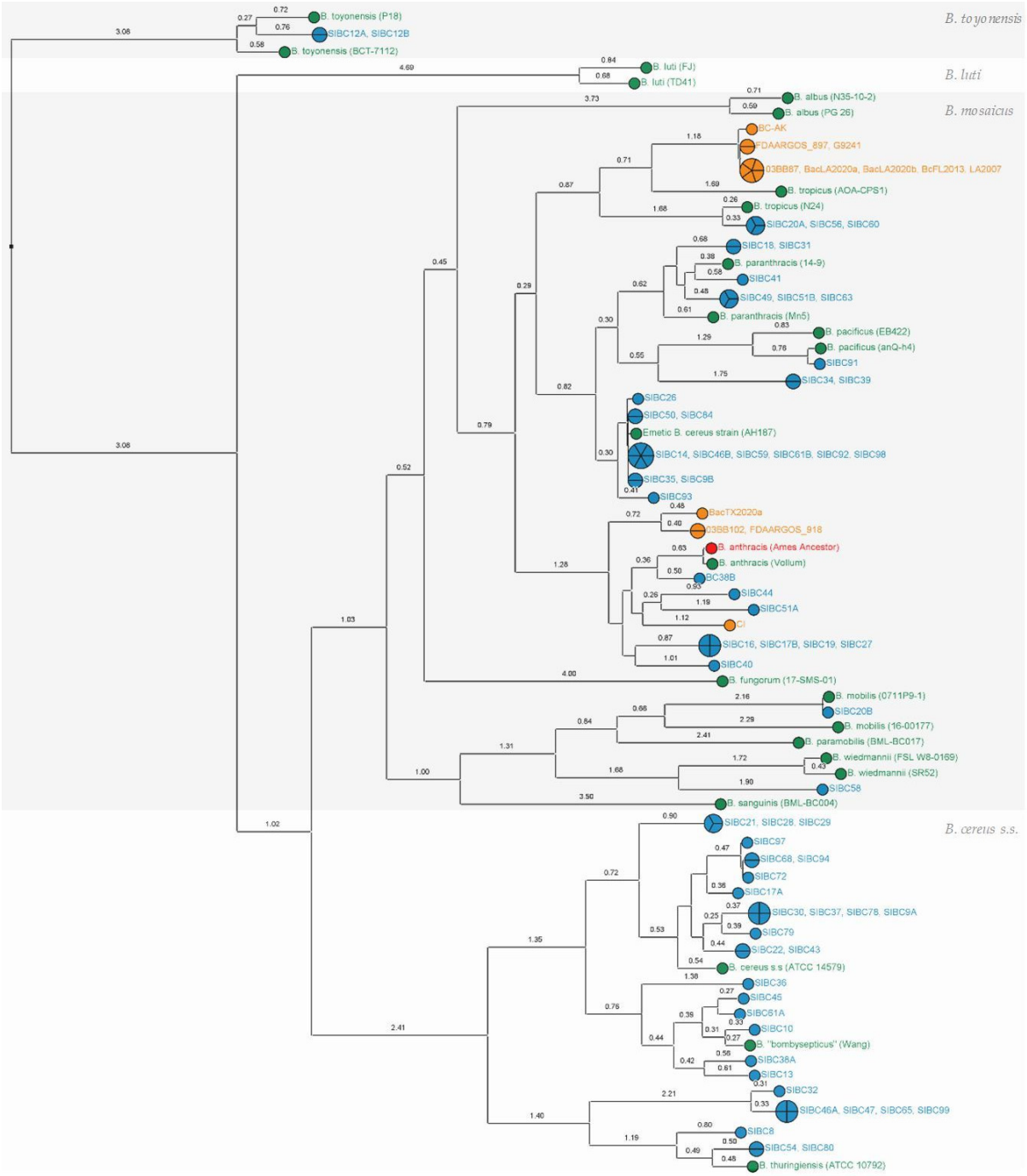

Figure 1. SNP phylogeny including "*B. cereus* group" panel with the public reference genomes of several *B. cereus s.l* species (in green), a panel with public genomes harboring anthrax-like genes (in orange) and strains from this study (in blue). Closely related *B. cereus s.l.* linages can be assigned to different genomospecies (*B. mosaicus*, *B. cereus s.s*, *B. luti* and *B. toyonensis*) as proposed by Carroll et al. [10]. Maximum parsimony tree built from 5929 SNP positions of 102 whole genomes mapped to the chromosomic genome of Ames Ancestor (in red, GCF_000008445.1). Numbers upon branches indicate the number of SNPs ×100. See Supplementary Table S2 for public strain details.

3.4. Close B. anthracis Neighbors

An SNP analysis and MPT clustering were conducted with the public genomes of *B. cereus s.l.* exhibiting ≥ 70% of nucleotide sequence identity with the Ames Ancestor chromosome but that are outside the *B. anthracis* lineage (n = 178). This analysis also included the "*B. anthracis*" panel and the eight isolates of this study close to the *B. anthracis*

lineage (Supplementary Table S3). The closest *B. anthracis* neighbors were selected from the resulting SNP phylogeny and a second SNP analysis was conducted with the 14 closely related public genomes, the isolate BC38B from this study and the "*B. anthracis*" panel (Figure 2). It appeared from this SNP phylogeny that BC38B was the most closely related to the *B. anthracis* lineage in comparison with the public neighbors (4508 SNPs for BC38B vs. 4643 SNPs for the second-closest strain JRS4).

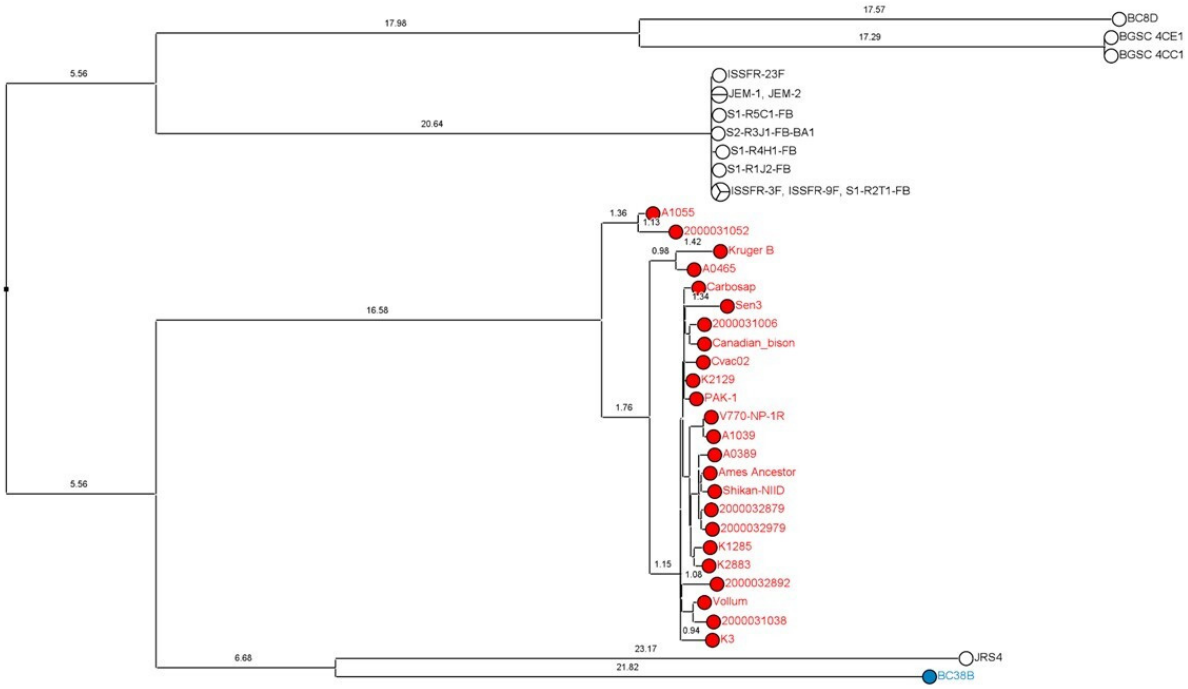

Figure 2. SNP phylogeny including the "*B. anthracis*" panel (in red) and the public genomes (Assembly) most closely related to the *B. anthracis* lineage (in white). The strain BC38B from this study (in blue) shares the fewest SNPs with *B. anthracis* in comparison with the other neighbors and appears to be the closest relative of the species. Maximum parsimony tree built from 13,826 SNP positions of 39 whole genomes mapped to the chromosomic genome of Ames Ancestor (GCF_000008445.1). Numbers upon branches indicate the number of SNPs ×100. See Supplementary Table S3 for strain details.

3.5. Taxonomic Assignment and Virulence Factor Detection

The taxonomic assignment for all strains and the screening for virulence factors are summarized in Table 1 (detailed results obtained with Btyper3, BtToxin_Digger and Type Strain Genome Server are available in Supplementary Table S5).

Strains were assigned as follows: *B. toyonensis* (×2), *B. mosaicus* (×13), *B. mosaicus* biovar Thuringiensis (×6), *B. mosaicus* subsp. *cereus* (×4), *B. mosaicus* subsp. *cereus* biovar Emeticus (×11), *B. cereus s.s.* (×29). Virulence factors involved in the production of diarrheal toxins were found in all isolates. The non-hemolytic enterotoxin genes *nheABC* and the sphingomyelinase *spH* gene were detected in all genomes. The hemolysin BL genes *hblABCD* were harbored by 46% of the isolates and the diarrheal cytotoxin K variant 2 gene *cytk-2* by 63%. The gene coding for the diarrheal cytotoxin variant 1 was not detected in any of the isolates. Anthrax toxin genes *cya*, *lef* and *pagA* were not detected, nor were any of the capsule virulence factors. The cereulide synthetase genes *cesABCD* were harbored by all the isolates close to the reference emetic strain AH187, except for SIBC26. Six strains exhibited pesticidal toxin genes; these are detailed in Table 2.

Table 1. Screening of major toxin-virulence genes among strains of the present study and among the public strains of the "*B. cereus* group" panel (in green) and "anthrax-toxin gene-harboring genomes" panel (in orange) using the BTyper3 tool. "+": presence; "−": absence; "(+)": partially detected (i.e., not all the genes of the cluster were detected). The taxonomic assignment was based on in silico DNA/DNA hybridization and confirmed with Btyper3 tool.

Strain id.	Virulence Factors											Closest Reference Species	Taxonomic Assignment by BTyper3
	nheABC	*hblABCD*	*cesABCD*	*cytK-1*	*cytK-2*	*spH*	*Bt* toxins *	*cya, lef, pagA*	*capABCDE*	*hasABC*	*bpsABCDEFGHX*		
B. toyonensis (BCT-7112), *B. toyonensis* (P18)	+	+	−	−	−	+	−	−	−	−	(+)	*B. toyonensis*	*B. toyonensis*
SIBC12A, SIBC12B	+	+	−	−	−	+	−	−	−	−	(+)	*B. toyonensis*	*B. toyonensis*
B. luti (TD41), *B. luti* (FJ)	+	−	−	−	−	+	−	−	−	−	(+)	*B. luti*	*B. luti*
B. paramobilis (BML-BC017)	+	+	−	−	−	+	−	−	−	−	(+)	*B. paramobilis*	*B. mosaicus*
B. mobilis (16-00177), *B. mobilis* (0711P9-1)	+	−	−	−	−	+	−	−	−	−	(+)	*B. mobilis*	*B. mosaicus*
SIBC20B	+	−	−	−	−	+	−	−	−	−	(+)	*B. mobilis*	*B. mosaicus*
B. wiedmannii (FSL W8-0169)	+	+	−	−	+	+	−	−	−	−	(+)	*B. wiedmannii*	*B. mosaicus*
B. wiedmannii (SR52)	+	+	−	−	−	+	−	−	−	−	(+)	*B. wiedmannii*	*B. mosaicus*
SIBC58	+	+	−	−	−	+	−	−	−	−	(+)	*B. wiedmannii*	*B. mosaicus*
B. sanguinis (BML-BC004)	+	−	−	−	−	+	−	−	−	−	(+)	*B. sanguinis*	*B. mosaicus*
B. tropicus (N24)	+	−	−	−	+	+	−	−	−	−	(+)	*B. tropicus*	*B. mosaicus*
SIBC56, SIBC60, SIBC20A	+	−	−	−	+	+	−	−	−	−	(+)	*B. tropicus*	*B. mosaicus*
G9241, FDAARGOS_897, 03BB87, LA2007, BacLA2020a, BacLA2020b	+	+	−	−	+	+	−	+	−	+	+	*B. tropicus*	*B. mosaicus* biovar Anthracis [(1)]
BcFL2013	+	+	−	−	+	+	−	+	−	+	(+)	*B. tropicus*	*B. mosaicus* biovar Anthracis [(1)]
BC-AK	+	+	−	−	+	+	−	(+)	+	+	(+)	*B. tropicus*	*B. mosaicus* biovar Anthracis [(1)]
B. tropicus (AOA-CPS1)	+	+	−	−	+	+	−	−	−	−	(+)	*B. tropicus*	*B. mosaicus*
B. pacificus (EB422)	+	−	−	−	+	+	−	−	−	−	(+)	*B. pacificus*	*B. mosaicus*

Table 1. *Cont.*

Strain id.	Virulence Factors											Closest Reference Species	Taxonomic Assignment by BTyper3
	nheABC	*hblABCD*	*cesABCD*	*cytK-1*	*cytK-2*	*spH*	*Bt* toxins *	*cya, lef, pagA*	*capABCDE*	*hasABC*	*bpsABCDEFGHX*		
B. pacificus (anQ-h4)	+	–	–	–	–	+	–	–	–	–	(+)	*B. pacificus*	*B. mosaicus*
SIBC91	+	–	–	–	–	+	–	–	–	–	(+)	*B. pacificus*	*B. mosaicus*
SIBC34, SIBC39	+	+	–	–	–	+	+	–	–	–	(+)	*B. paranthracis*	*B. mosaicus* biovar Thuringiensis [3]
B. paranthracis (Mn5)	+	–	–	–	+	+	–	–	–	–	(+)	*B. paranthracis*	*B. mosaicus* subps. *cereus* [4]
SIBC18, SIBC31	+	–	–	–	–	+	–	–	–	–	(+)	*B. paranthracis*	*B. mosaicus* subsp. *cereus* [4]
SIBC49, SIBC63, SIBC51B	+	–	–	–	+	+	+	–	–	–	(+)	*B. paranthracis*	*B. mosaicus* biovar Thuringiensis [3]
B. paranthracis (14-9)	+	–	–	–	–	+	–	–	–	–	(+)	*B. paranthracis*	*B. mosaicus* subps. *cereus* [4]
SIBC41	+	–	–	–	–	+	–	–	–	–	(+)	*B. paranthracis*	*B. mosaicus* subsp. *cereus* [4]
SIBC93	+	–	+	–	–	+	–	–	–	–	(+)	*B. paranthracis* [5]	*B. mosaicus* subsp. *cereus* biovar Emeticus [6]
Emetic *B. cereus* strain (AH187)	+	–	+	–	–	+	–	–	–	–	(+)	*B. paranthracis* [5]	*B. mosaicus* subsp. *cereus* biovar Emeticus [6]
SIBC14, SIBC46B, SIBC59, SIBC61B, SIBC92, SIBC98, SIBC50, SIBC84, SIBC35, SIBC9B	+	–	+	–	–	+	–	–	–	–	(+)	*B. paranthracis* [5]	*B. mosaicus* subsp. *cereus* biovar Emeticus [6]
SIBC26	+	–	–	–	–	+	–	–	–	–	(+)	*B. paranthracis* [5]	*B. mosaicus* subsp. *cereus* [4]
03BB102, FDAARGOS_918, CI	+	–	–	–	–	+	–	+	+	+	(+)	*B. anthracis*	*B. mosaicus* biovar Anthracis [1]
BacTX2020a	+	–	–	–	–	+	–	+	–	+	(+)	*B. anthracis*	*B. mosaicus* biovar Anthracis [1]
B. anthracis (Ames Ancestor), *B. anthracis* (Vollum)	+	–	–	–	–	+	–	+	+	+	(+)	*B. anthracis*	*B. mosaicus* subsp. *anthracis* biovar Anthracis [1,2]

Table 1. *Cont.*

Strain id.	Virulence Factors											Closest Reference Species	Taxonomic Assignment by BTyper3
	nheABC	*hblABCD*	*cesABCD*	*cytK-1*	*cytK-2*	*spH*	*Bt* toxins *	*cya, lef, pagA*	*capABCDE*	*hasABC*	*bpsABCDEFGHX*		
BC38B	+	–	–	–	–	+	–	–	–	–	(+)	*B. anthracis*	*B. mosaicus*
SIBC44, SIBC51A, SIBC16, SIBC19, SIBC27, SIBC17B	+	–	–	–	+	+	–	–	–	–	(+)	*B. anthracis*	*B. mosaicus*
SIBC40	+	–	–	–	–	+	+	–	–	–	(+)	*B. anthracis*	*B. mosaicus* biovar Thuringiensis [3]
B. albus (N35-10-2), *B. albus* (PG 26)	+	+	–	–	–	+	–	–	–	–	(+)	*B. albus*	*B. mosaicus*
B. fungorum (17-SMS-01)	+	(+)	–	–	–	+	–	–	–	–	(+)	*B. fungorum*	*B. mosaicus*
B. cereus s.s. (ATCC 14579)	+	+	–	–	+	+	–	–	–	–	(+)	*B. cereus s.s.*	*B. cereus s.s.*
SIBC21, SIBC28, SIBC29, SIBC22, SIBC43, SIBC30, SIBC37, SIBC78, SIBC9A, SIBC79, SIBC68, SIBC94, SIBC72, SIBC97, SIBC17A	+	+	–	–	+	+	–	–	–	–	(+)	*B. cereus s.s.*	*B. cereus s.s.*
B. "bombysepticus" (Wang)	+	+	–	–	+	+	–	–	–	–	(+)	*B. cereus s.s.*	*B. cereus s.s.*
SIBC13, SIBC38A, SIBC10, SIBC45, SIBC61A, SIBC36	+	+	–	–	+	+	–	–	–	–	(+)	*B. cereus s.s.*	*B. cereus s.s.*
B. thuringiensis (ATCC 10792)	+	+	–	–	+	+	+	–	–	–	(+)	*B. thuringiensis*	*B. cereus s.s.* biovar Thuringiensis [3]
SIBC8, SIBC54, SIBC80, SIBC32	+	+	–	–	+	+	–	–	–	–	(+)	*B. thuringiensis*	*B. cereus s.s.*
SIBC47	+	–	–	–	+	+	–	–	–	–	(+)	*B. thuringiensis*	*B. cereus s.s.*
SIBC65	+	–	–	–	+	+	–	–	–	–	(+)	*B. thuringiensis*	*B. cereus s.s.*
SIBC99	+	–	–	–	+	+	–	–	–	–	(+)	*B. thuringiensis*	*B. cereus s.s.*
SIBC46A	+	–	–	–	+	+	–	–	–	–	(+)	*B. thuringiensis*	*B. cereus s.s.*

* Insecticidal toxins, [1] *B. Anthracis*, [2] *B. anthracis* biovar Anthracis, [3] *B. Thuringiensis*, [4] *B. cereus*, [5] Emetic *B. cereus* reference strain, [6] *B. Emeticus*.

Table 2. List of pesticidal toxins detected among strains of the present study using BtToxin_Digger.

Strain ID	Pesticidal Toxins
SIBC34	Cry28Aa2, Mpp64Aa1, Vpb4Aa1, Spp1Aa1
SIBC39	Cry28Aa2, Mpp64Aa1, Vpb4Aa1, Spp1Aa1
SIBC40	Cry1Ie5, Spp1Aa1
SIBC49	Cry13Aa2, Cry41Ba2, Spp1Aa1
SIBC51B	Cry13Aa2, Cry41Ba2, Spp1Aa1
SIBC63	Cry13Aa2, Cry41Ba2, Spp1Aa1

3.6. BC38B Plasmid Analysis

One megaplasmid of 551,060 kb was successfully reconstructed with hybrid de novo assembly. Minireplicon detection was performed on its sequence. A minireplicon represents the smallest region for plasmid replication and contains the origin of replication and genes encoding replication proteins. Three different origins of replication were determined on the sequence. Additionally, the plasmid exhibited the protein genes *pXO1-14/pXO1-16*-like (DNA-binding protein gene and replication initiator protein gene, respectively), which were first reported to support the replication of *B. anthracis* plasmid pXO1 [46]. A replication-relaxation protein coding gene and a cell division protein FtsZ coding gene were also detected. The first one is essential for plasmid DNA replication, while the second has a main role in cell division and duplication for plasmids [47]. A BLASTP analysis revealed that all these elements were widely distributed in the *B. cereus* group and could be defined as new minireplicons in the large plasmids of the *B. cereus* group. Their locations on the plasmid are shown in Figure 3 and correspond with regions of GC skew sign change, which coincide with the origin or terminus of replication [48].

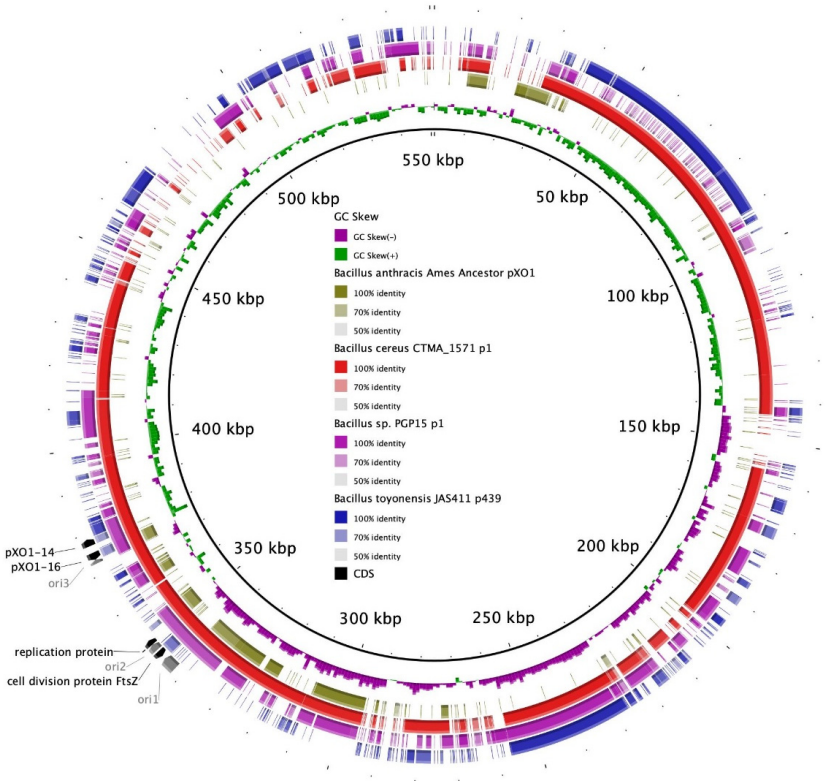

Figure 3. Comparison of BC38B plasmid to homologous plasmids (*B. cereus* strain CTMA_1571 plasmid p1, *B. toyonensis* strain JAS411 plasmid p439, and *Bacillus* sp. PGP15 plasmid p1) and *B. anthracis* Ames

Ancestor pXO1 using Blast Ring Image Generator (BRIG). Rings from the center to the outermost: (1) scale marks; (2) GC skew; (3–6) sequence percentage identity to homologous plasmids; (7) minireplicon-associated protein coding genes (in black), origins of replication (in gray). The locations of shared regions between these plasmids relative to the BC38B plasmid are denoted in color in the figure. The color key corresponding to these designations is provided within the figure itself.

3.7. Detection of Cereulide Production with MALDI-TOF MS

Eleven isolates found positive for *cesABCD* (Table 1) and phylogenetically grouped with the emetic *B. cereus* reference strain AH187 after wgSNP analysis (Figure 2) were verified for cereulide production using MALDI-TOF MS (SIBC14, SIBC46B, SIBC59, SIBC61B, SIBC92, SIBC98, SIBC50, SIBC84, SIBC35, SIBC9B and SIBC93). The isolate SIBC26 *cesABCD* negative (Table 1) within the same phylogroup (Figure 2) was also tested. The two characteristic cereulide peaks (1176 $\pm$ 1 m/z and 1192 $\pm$ 1 m/z) were detected for the 11 isolates that harbored the cereulide synthetase genes only (Figure 4).

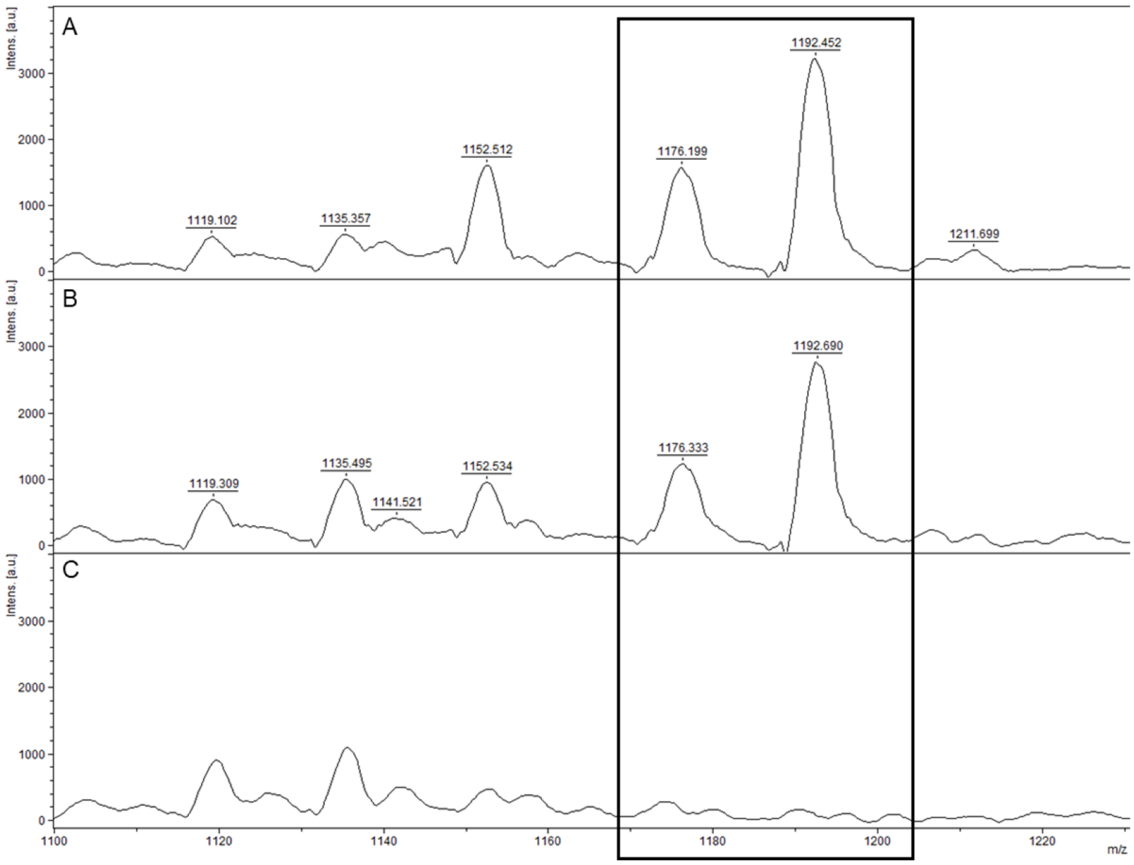

Figure 4. MALDI-TOF mass profiles of three isolates genetically close the emetic-type strain AH187 after wgSNP analysis. Mass spectrum corresponding to the colony smear of (**A**) the emetic collection strain AND1407, exhibiting the cereulide synthetase genes (*cesABCD*+) and showing the two cereulide characteristic peaks at 1176.1 and 1192.4 m/z; (**B**) the strain SIBC50 (*cesABCD*+), also exhibiting cereulide with peaks at 1176.3 and 1192.6 m/z; and (**C**) the strain SIBC26 (*cesABCD*−), lacking cereulide. Results shown are representative spectra of at least three independent experiments.

4. Discussion

4.1. B. cereus s.l. Characterization

Isolates from this study showed notable phenotypic diversity with various colony morphologies among and within species (Table S1, Figure S1). Conventional microbiology techniques such as culturing remain an essential step for bacterial identification and various

selective and/or chromogenic agar media have been developed for the isolation of *B. cereus s.l.* bacteria in complex matrices [49,50]. However, species of the *B. cereus* group cannot be distinguished based on morphological criteria, as characteristics used for taxonomic assignment (e.g., motility, hemolysis) vary within and among species [10,50]. For example, the lack of hemolytic activity on blood agar is a phenotypic characteristic used to discriminate *B. anthracis* from other *B. cereus s.l.*, but such features can vary depending on the *B. anthracis* strain and/or the blood origin [51]. Conversely, some non-*B. anthracis* strains lack hemolytic activity (e.g., isolate SIB79 from this study) and mislead the interpretation. The combination of conventional microbiology methods with advanced technologies such as molecular biology, biosensors or MALDI-TOF MS allows for a reliable identification [3].

Bacterial identification using MALDI-TOF MS is a recommendable first line technique for its speed and ease of use. All strains in this study were identified as members of the *B. cereus* group using MALDI-TOF MS with a high confidence score (log(score) $\geq$ 2.0). Moreover, it successfully identified cereulide-producing isolates. However, more than half were obviously misidentified as *B. anthracis* due to the incompleteness of the commercial library. The identification as either *B. cereus* or *B. anthracis* with MALDI-TOF did not reflect the genetic proximity of the isolate with these two species and it also appeared that different morphotypes of the same isolate could result in different species identifications (see Supplementary Table S3). Discrimination of closely related *B. cereus* group species using MALDI-TOF MS remains challenging, but the detection of species-specific biomarkers [52,53] and the expansion of the current commercial libraries with in-house reference libraries [6,54] could improve the statistical confidence of identification results and would avoid the occurrence of false-positive *B. anthracis* identification.

WGS remains the gold standard for highly precise characterization, and tools such as BTyper3 [29] have been developed to easily assign a taxonomic affiliation and detect virulence factors in *B. cereus* group strains. In the present study, the identification of *B. anthracis* was ruled. Moreover, in silico ANI-based methods and digital DNA/DNA hybridization supported SNP phylogeny to reflect the diversity of the collection studied.

4.2. Unnamed B. anthracis Neighbors

Eight isolates devoid of anthrax-virulence factors according to gene detection were nevertheless grouped proximally to the *B. anthracis* clade (Figure 1). Such genetic proximity is interesting, notably for biodefense, because part of the genome sequence information might constitute a potential source of false-positive signals of anthrax detection with metagenomics. In terms of nomenclature, as discussed by Carroll et al. [10], such isolates should not be referred to as *B. anthracis* to avoid incorrect assumptions of their anthrax-causing capabilities. For instance, these strains can be named after the genomospecies *B. mosaicus* which encompasses *B. anthracis* and other closely related species [8,10], but this taxonomic assignment does not underline the genetic proximity of these isolates with the *B. anthracis* species. It is very likely that the distinction of novel species for *B. anthracis*-close isolates will be described and will provide a better phylogenic characterization. Several strains have already been identified as closely related to *B. anthracis*, including strains that are colonizing the International Space Station [55] or the strain JRS4 isolated in the desert of Saudi Arabia [56,57]. Among the eight closest neighbors from this study, the isolate BC38B was especially interesting as it appeared to be the most closely related to the *B. anthracis* lineage in comparison with the other public isolates described so far (Figure 2).

4.3. Plasmid Analysis

At least six types of minireplicons were discovered in the megaplasmids of the *B. cereus* group [43]. The presence of two or more minireplicons in a *B. cereus* group megaplasmid strongly suggests the integration of several smaller plasmids [43]. This type of integration event might have arisen for the BC38B plasmid. Indeed, each of the closest plasmid neighbors shared different regions with the BC38B plasmid (Figure 3). In particular, the strain CTMA_1571 plasmid p1 (Genbank accession number: CP053657.2) had a high

homology with the BC38B plasmid and possessed the same minireplicon types. Both strains are close to the *B. anthracis* clade [58], although their sequence homology with the pXO1 plasmid is moderate. The BC38B plasmid also had regions of shared homology with plasmid p1 from *Bacillus* sp. PGP15 (isolated in soil from a rhizosphere in China; Genbank accession number: CP095875.1 [59]) and plasmid p439 from the *B. toyonensis* strain JAS411 (isolated in soil from a farmland in Poland; Genbank accession number: CP036114.1). These two strains belong to different clades of the *B. cereus* group, suggesting genetic exchange via horizontal transfer, which is a phenomenon that frequently occurred during the evolutionary history of the *B. cereus* group, even between phylogenetically distant strains [60].

4.4. Virulence Factor Detection and Toxin Profile Diversity

The presence of toxin genes in *B. cereus s.l.* isolates was screened in a multitude of recent studies [8,61–64], and specific tools have been developed to ease their detection [29]. It appeared that the common distribution of virulence factors for the *B. cereus* group strains is the presence of approximately 85–100% *nhe* (*ABC*), 40–70% *hbl* (*CDA*), 40–70% *cytK-2*, very few *ces*+ and typically no *cytK-1*+ [13]. The screening of virulence factors conducted in the present work is totally consistent with this distribution, as *nhe* (*ABC*) were detected in 100% of the isolates, *hbl* (*CDA*) in 46%, *cytK-2* in 63%, the genes *ces* were detected in eleven isolates and none of the isolates were *cytK-1*+. The absence of *cytK-1* detection was expected as this enterotoxin is, for instance, only described in isolates of the species *B. cytotoxicus* [14]. It is also suggested that all *B. cereus* isolates can be categorized into seven different toxin profiles: A (*nhe*+, *hbl*+, *cytK*+), B (*nhe*+, *cytK*+, *ces*+), C (*nhe*+, *hbl*+), D (*nhe*+, *cytK*+), E (*nhe*+, *ces*+), F (*nhe*+) and G (*cytK*+) [65]. The isolates in this study therefore exhibit great toxin profile diversity with the presence of the categories A, C, D, E and F.

The observed toxin profile diversity within the panel of characterized strains highlighted that a nuanced, strain-specific approach to toxin analysis is essential as each isolate can display a unique set of challenges and implications. In a biodefence context, spotting the presence of specific toxins in strains that could be exploited for malevolent purposes is pivotal for the formulation of precise countermeasures and security protocols. For public health, the heterogeneity in toxin profiles directly influences the clinical presentation and outcome of infections. An understanding of the pathogenicity and virulence mechanisms associated with each strain would ensure timely and effective clinical interventions.

4.5. Biovars Anthracis, Emeticus and Thuringiensis

Genomic determinants responsible for some phenotypes are plasmid-mediated (e.g., synthesis of anthrax toxin, bioinsecticide crystal proteins, emetic toxin synthetase proteins) and can be lost or gained within a species. The characterization of biovars in the "genomospecies/subsp./biovar" nomenclature framework proposed by Carroll et al. [8,10] allows one to highlight phenotypes of clinical and/or industrial importance.

The term "biovar Anthracis" is used for isolates that produce anthrax toxin (and/or possess anthrax-toxin encoded genes *cya*, *lef* and *pagA* [8]). *B. anthracis* biovar Anthracis (or *B.* Anthracis) has a long history as a life-threatening infectious agent to humans and animals worldwide [66] and has been extensively studied since the anthrax letter events in 2001 and the subsequent anthrax outbreaks [67,68]. The production of anthrax toxin has long been considered restricted to the *B. anthracis* species, but anthrax-causing strains have been characterized since 2006 outside the *B. anthracis* lineage in humans [8,23], great apes [24], in a kangaroo [8] and, recently, in a soft-shell turtle [26]. These *B. mosaicus* biovar Anthracis strains, previously referred to as "anthrax-*like*" strains, may exhibit different capsular composition [20] and are so far described as close *B. anthracis* neighbors or are related to the *B. tropicus* species [36] (see Figure 1 and Table 1). The isolation of such strains is rare and none were identified in the present study.

Apart from anthrax-causing strains, several other *B. cereus* group isolates are of great concern as they induce food intoxication and toxicoinfection resulting in vomiting, diarrhea

and sometimes death [13,69]. The term "biovar Emeticus" is used for isolates known to produce cereulide (and/or possess the cereulide synthetase plasmid-encoded gene cluster *cesABCD* [8,70]). Emetic *B. cereus* strains (*B.* Emeticus) produce cereulide toxin during growth in food that causes vomiting, a progression referred to as emetic syndrome. Cereulide is a potent toxin, is heat- and acid-stable, and is responsible for an increasing number of foodborne poisonings that have gained public attention in recent years [14]. Cereulide toxin formation was thought to be restricted to a single evolutionary lineage of closely related strains [71], but cereulide-producing isolates have since been characterized across multiple lineages [70] (e.g., *B. mycoides* biovar Emeticus [72,73]). Detection methods for cereulide are improving and/or emerging [74] and the presence of the toxin can be certified using MALDI-TOF MS with rapidity and sensitivity from a colony smear [45,75]. In the present study, the gene cluster *cesABCD* was detected in 11 isolates and the production of cereulide was confirmed for each of them using MALDI-TOF MS (see Table 1 and Figure 4). All these strains were genetically close to the emetic reference strain AH187 (see Figure 1) and no other biovar Emeticus strain was detected in other lineages.

The term "biovar Thuringiensis" is used for isolates that produce one or more *Bt* toxins (and/or possess *Bt* toxin-encoding genes [8]). *Bt* toxins are used as biopesticides in organic agriculture and are considered harmless to humans, although there is a growing concern that residues in food may occasionally cause diarrheal illness [76]. Indeed, diagnostic laboratories generally do not distinguish between *B. cereus s.s.* and *B.* Thuringiensis, potentially leading to misattributed cases of foodborne illnesses. This lack of distinction might result in underestimations of the disease impact associated with *B.* Thuringiensis [16]. The nomenclature of *Bt* toxin is complex and regularly subject to change as new genes involved in *Bt* toxin formation and new proteins with insecticidal properties are frequently discovered [77]. The production of *Bt* toxins is not restricted to isolates genetically close to the *B. thuringiensis* species and insecticidal activities have, for example, been described in strains close to the species *B. toyonensis* (*B. toyonensis* biovar Thuringiensis [78]) and *B. wiedmannii* (*B. mosaicus* biovar Thuringiensis [79,80]). In the present study, *Bt* toxin-encoding genes were detected in six isolates (see Table 1). These *B. mosaicus* biovar Thuringiensis (or *B.* Thuringiensis) strains were genetically close to the *B. paranthracis* or to the *B. anthracis* reference strains. None of the isolates genetically close to the *B. thuringiensis* species reference strain ATCC10792 possessed *Bt* toxin-encoding genes.

5. Conclusions

In this report, we have presented a detailed characterization of 65 strains of *B. cereus s.l.*, completely sequenced and identified as *B. mosaicus*, *B. cereus s.s.* and *B. toyonensis* using bacterial genomic taxonomy. Whatever their origin, clinical, food or water, the identification of *B. anthracis* has been excluded. All strains carried non-hemolytic enterotoxin genes, including those carrying crystal protein genes (biovar Thuringiensis). Such strains, although lacking BL hemolysin, are not completely devoid of diarrheal pathogenic potential, and this raises the issue to what extent biopesticides are free of any risk, especially because additional acquisition via horizontal gene transfer might not be inconceivable. In this way, our work contributed to the surveillance of circulating isolates and any contributions in this direction will be worthwhile. We also described the strain BC38B, carrying a megaplasmid, that was positioned phylogenetically closer to *B. anthracis* than those already described, which constitutes an encouraging additional step towards establishing the age of the *B. anthracis* species or even of the most recent common ancestor of already identified lineages. Overall, this work constitutes a valuable source of information and biological resources intended to better take into account the biological risk, whether in service of public health or biodefense.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/microorganisms11112721/s1, Figure S1: Phenotypic diversity of several *B. cereus s.l* strains on tryptic soy agar + 5% sheep blood after an 18 h incubation at 37 °C. Colony pictures of (A) SIBC91 (*B. pacificus* species); (B) SIBC38A (*B. cereus s.s.*); (C) SIBC17A (*B. cereus s.s.*); (D) SIBC12A (*B. toyonensis*); and (E,F) two morphotypes of SIBC41 (*B. paranthracis*). Scale bars 0.5 cm; Table S1: Summary of the characteristics of the studied strains and their corresponding assemblies; Table S2: Composition of the three public genome datasets used in this study: (1) the "*B. cereus* group" panel with reference/representative strains of several *B. cereus s.l.* species according to the NCBI database and/or according to Carroll et al. [8]; (2) the panel of genomes harboring the anthrax toxin genes as detailed in Carroll et al. [36]; and (3) the "*B. anthracis*" panel with genomes representative of the different clades of the lineage; Table S3: Public genomes out of the *B. anthracis* lineage and genomes from this study exhibiting $\geq$ 70% of nucleotide identity (% NI) with *B. anthracis* strain Ames Ancestor; Table S4: MALDI-TOF identification of the isolates using two BRUKER commercial databases (BDAL and SR) in comparison with the closest reference species determined with SNP analyses and with in silico DNA/DNA hybridization; Table S5: Taxonomic assignation and screening of virulence factors of genomes from this study using BTyper3, Summary of closest reference species of each studied strain with in silico DNA/DNA hybridization using Type Strain Genome Server, Detection of pesticidal toxins for SIBC34, SIBC39, SIBC40, SIBC49, SIBC51B and SIBC63 using BtToxin_Digger.

Author Contributions: Conceptualization, C.F. and M.A.; methodology, M.A., C.F., V.M.-H. and A.C.; software, M.A. and C.F.; validation, C.F. and M.A.; formal analysis, C.F. and M.A.; investigation, M.A. and C.F.; resources, C.F., V.M.-H. and M.A.; data curation, C.F., A.C. and V.R.; writing—original draft preparation, M.A. and C.F.; writing—review and editing, F.R., V.R. and L.T.; visualization, M.A. and C.F.; supervision, F.R., V.R. and L.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The assemblies have been deposited in DDBJ/ENA/GenBank under the accession numbers referenced in Supplementary Table S1.

Acknowledgments: The authors sincerely thank Karmen Godič Torkar from the University of Ljubljana (Slovenia) for the scientific exchanges and for providing to DGA CBRN a great diversity of *B. cereus s.l.* strains that made it possible to carry out the present study.

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

References

1. HHS and USDA Select Agents and Toxins, 7 CFR Part 331, 9 CFR Part 121, and 42 CFR Part 73. Available online: https://www.selectagents.gov/sat/list.htm (accessed on 29 June 2023).
2. Irenge, L.M.; Gala, J.L. Rapid detection methods for *Bacillus anthracis* in environmental samples: A review. *Appl. Microbiol. Biotechnol.* **2012**, *93*, 1411–1422. [CrossRef] [PubMed]
3. Zasada, A.A. Detection and Identification of *Bacillus anthracis*: From Conventional to Molecular Microbiology Methods. *Microorganisms* **2020**, *8*, 125. [CrossRef] [PubMed]
4. World Health Organization. *Anthrax in Humans and Animals*, 4th ed.; WHO Guidelines Approved by the Guidelines Review Committee; WHO: Geneva, Switzerland, 2008.
5. Chen, J.; Xu, F. Application of Nanopore Sequencing in the Diagnosis and Treatment of Pulmonary Infections. *Mol. Diagn. Ther.* **2023**, *27*, 685–701. [CrossRef] [PubMed]
6. Lasch, P.; Stämmler, M.; Schneider, A. Version 4 (20230306) of the MALDI-ToF Mass Spectrometry Database for Identification and Classification of Highly Pathogenic Microorganisms from the Robert Koch-Institute (RKI). 2023. Available online: https://zenodo.org/records/7702375 (accessed on 15 May 2023).
7. Torres-Sangiao, E.; Leal Rodriguez, C.; Garcia-Riestra, C. Application and Perspectives of MALDI-TOF Mass Spectrometry in Clinical Microbiology Laboratories. *Microorganisms* **2021**, *9*, 1539. [CrossRef] [PubMed]
8. Carroll, L.M.; Cheng, R.A.; Wiedmann, M.; Kovac, J. Keeping up with the *Bacillus cereus* group: Taxonomy through the genomics era and beyond. *Crit. Rev. Food Sci. Nutr.* **2022**, *62*, 7677–7702. [CrossRef]
9. Liu, Y.; Lai, Q.; Goker, M.; Meier-Kolthoff, J.P.; Wang, M.; Sun, Y.; Wang, L.; Shao, Z. Genomic insights into the taxonomic status of the *Bacillus cereus* group. *Sci. Rep.* **2015**, *5*, 14082. [CrossRef]
10. Carroll, L.M.; Wiedmann, M.; Kovac, J. Proposal of a Taxonomic Nomenclature for the *Bacillus cereus* Group Which Reconciles Genomic Definitions of Bacterial Species with Clinical and Industrial Phenotypes. *mBio* **2020**, *11*, e00034-20. [CrossRef]

11. Carroll, L.M.; Matle, I.; Kovac, J.; Cheng, R.A.; Wiedmann, M. Laboratory Misidentifications Resulting from Taxonomic Changes to *Bacillus cereus* Group Species, 2018–2022. *Emerg. Infect. Dis.* **2022**, *28*, 1877–1881. [CrossRef]
12. Crickmore, N.; Berry, C.; Panneerselvam, S.; Mishra, R.; Connor, T.R.; Bonning, B.C. Bacterial Pesticidal Protein Resource Center. Available online: https://www.bpprc.org/ (accessed on 10 April 2023).
13. Dietrich, R.; Jessberger, N.; Ehling-Schulz, M.; Martlbauer, E.; Granum, P.E. The Food Poisoning Toxins of *Bacillus cereus*. *Toxins* **2021**, *13*, 98. [CrossRef]
14. Jovanovic, J.; Ornelis, V.F.M.; Madder, A.; Rajkovic, A. *Bacillus cereus* food intoxication and toxicoinfection. *Compr. Rev. Food Sci. Food Saf.* **2021**, *20*, 3719–3761. [CrossRef]
15. Ehling-Schulz, M.; Fricker, M.; Grallert, H.; Rieck, P.; Wagner, M.; Scherer, S. Cereulide synthetase gene cluster from emetic *Bacillus cereus*: Structure and location on a mega virulence plasmid related to *Bacillus anthracis* toxin plasmid pXO1. *BMC Microbiol.* **2006**, *6*, 20. [CrossRef] [PubMed]
16. Stenfors Arnesen, L.P.; Fagerlund, A.; Granum, P.E. From soil to gut: *Bacillus cereus* and its food poisoning toxins. *FEMS Microbiol. Rev.* **2008**, *32*, 579–606. [CrossRef] [PubMed]
17. Bohm, M.E.; Huptas, C.; Krey, V.M.; Scherer, S. Massive horizontal gene transfer, strictly vertical inheritance and ancient duplications differentially shape the evolution of *Bacillus cereus* enterotoxin operons *hbl*, *cytK* and *nhe*. *BMC Evol. Biol.* **2015**, *15*, 246. [CrossRef] [PubMed]
18. Ehling-Schulz, M.; Lereclus, D.; Koehler, T.M. The *Bacillus cereus* Group: *Bacillus* Species with Pathogenic Potential. *Microbiol. Spectr.* **2019**, *7*, 7. [CrossRef] [PubMed]
19. Manktelow, C.J.; White, H.; Crickmore, N.; Raymond, B. Divergence in environmental adaptation between terrestrial clades of the *Bacillus cereus* group. *FEMS Microbiol. Ecol.* **2020**, *97*, fiaa228. [CrossRef]
20. Baldwin, V.M. You Can't *B. cereus*—A Review of *Bacillus cereus* Strains That Cause Anthrax-Like Disease. *Front. Microbiol.* **2020**, *11*, 1731. [CrossRef]
21. Hoffmaster, A.R.; Ravel, J.; Rasko, D.A.; Chapman, G.D.; Chute, M.D.; Marston, C.K.; De, B.K.; Sacchi, C.T.; Fitzgerald, C.; Mayer, L.W.; et al. Identification of anthrax toxin genes in a *Bacillus cereus* associated with an illness resembling inhalation anthrax. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 8449–8454. [CrossRef]
22. Scarff, J.M.; Seldina, Y.I.; Vergis, J.M.; Ventura, C.L.; O'Brien, A.D. Expression and contribution to virulence of each polysaccharide capsule of *Bacillus cereus* strain G9241. *PLoS ONE* **2018**, *13*, e0202701. [CrossRef]
23. Hoffmaster, A.R.; Hill, K.K.; Gee, J.E.; Marston, C.K.; De, B.K.; Popovic, T.; Sue, D.; Wilkins, P.P.; Avashia, S.B.; Drumgoole, R.; et al. Characterization of *Bacillus cereus* isolates associated with fatal pneumonias: Strains are closely related to *Bacillus anthracis* and harbor *B. anthracis* virulence genes. *J. Clin. Microbiol.* **2006**, *44*, 3352–3360. [CrossRef]
24. Klee, S.R.; Ozel, M.; Appel, B.; Boesch, C.; Ellerbrok, H.; Jacob, D.; Holland, G.; Leendertz, F.H.; Pauli, G.; Grunow, R.; et al. Characterization of *Bacillus anthracis*-like bacteria isolated from wild great apes from Cote d'Ivoire and Cameroon. *J. Bacteriol.* **2006**, *188*, 5333–5344. [CrossRef]
25. Oh, S.Y.; Budzik, J.M.; Garufi, G.; Schneewind, O. Two capsular polysaccharides enable *Bacillus cereus* G9241 to cause anthrax-like disease. *Mol. Microbiol.* **2011**, *80*, 455–470. [CrossRef] [PubMed]
26. Tsai, J.-M.; Kuo, H.-W.; Cheng, W. Retrospective Screening of Anthrax-like Disease Induced by *Bacillus tropicus* str. JMT from Chinese Soft-Shell Turtles in Taiwan. *Pathogens* **2023**, *12*, 693. [CrossRef] [PubMed]
27. Afshinnekoo, E.; Meydan, C.; Chowdhury, S.; Jaroudi, D.; Boyer, C.; Bernstein, N.; Maritz, J.M.; Reeves, D.; Gandara, J.; Chhangawala, S.; et al. Geospatial Resolution of Human and Bacterial Diversity with City-Scale Metagenomics. *Cell Syst.* **2015**, *1*, 72–87. [CrossRef] [PubMed]
28. Xu, J.; Bai, X.; Zhang, X.; Yuan, B.; Lin, L.; Guo, Y.; Cui, Y.; Liu, J.; Cui, H.; Ren, X.; et al. Development and application of DETECTR-based rapid detection for pathogenic *Bacillus anthracis*. *Anal. Chim. Acta* **2023**, *1247*, 340891. [CrossRef] [PubMed]
29. Carroll, L.M.; Cheng, R.A.; Kovac, J. No Assembly Required: Using BTyper3 to Assess the Congruency of a Proposed Taxonomic Framework for the *Bacillus cereus* Group with Historical Typing Methods. *Front. Microbiol.* **2020**, *11*, 580691. [CrossRef] [PubMed]
30. Cerar Kišek, T.; Pogačnik, N.; Godič Torkar, K. Genetic diversity and the presence of circular plasmids in *Bacillus cereus* isolates of clinical and environmental origin. *Arch. Microbiol.* **2021**, *203*, 3209–3217. [CrossRef] [PubMed]
31. Torkar, K.G.; Bedenić, B. Antimicrobial susceptibility and characterization of metallo-β-lactamases, extended-spectrum β-lactamases, and carbapenemases of *Bacillus cereus* isolates. *Microb. Pathog.* **2018**, *118*, 140–145. [CrossRef]
32. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef]
33. 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. [CrossRef]
34. Wick, R.R.; Judd, L.M.; Holt, K.E. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol.* **2019**, *20*, 129. [CrossRef]
35. Wick, R.R. Porechop. Available online: https://github.com/rrwick/Porechop (accessed on 18 October 2022).
36. Carroll, L.M.; Marston, C.K.; Kolton, C.B.; Gulvik, C.A.; Gee, J.E.; Weiner, Z.P.; Kovac, J. Strains Associated with Two 2020 Welder Anthrax Cases in the United States Belong to Separate Lineages within *Bacillus cereus sensu lato*. *Pathogens* **2022**, *11*, 856. [CrossRef] [PubMed]

37. Sahl, J.W.; Pearson, T.; Okinaka, R.; Schupp, J.M.; Gillece, J.D.; Heaton, H.; Birdsell, D.; Hepp, C.; Fofanov, V.; Noseda, R.; et al. A *Bacillus anthracis* Genome Sequence from the Sverdlovsk 1979 Autopsy Specimens. *mBio* **2016**, *7*, e01501–e01516. [CrossRef] [PubMed]
38. Liu, H.; Zheng, J.; Bo, D.; Yu, Y.; Ye, W.; Peng, D.; Sun, M. BtToxin_Digger: A comprehensive and high-throughput pipeline for mining toxin protein genes from *Bacillus thuringiensis*. *Bioinformatics* **2021**, *38*, 250–251. [CrossRef] [PubMed]
39. Meier-Kolthoff, J.P.; Auch, A.F.; Klenk, H.P.; Goker, M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinform.* **2013**, *14*, 60. [CrossRef] [PubMed]
40. Meier-Kolthoff, J.P.; Goker, M. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nat. Commun.* **2019**, *10*, 2182. [CrossRef] [PubMed]
41. Meier-Kolthoff, J.P.; Carbasse, J.S.; Peinado-Olarte, R.L.; Goker, M. TYGS and LPSN: A database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. *Nucleic Acids Res.* **2022**, *50*, D801–D807. [CrossRef] [PubMed]
42. Alikhan, N.F.; Petty, N.K.; Ben Zakour, N.L.; Beatson, S.A. BLAST Ring Image Generator (BRIG): Simple prokaryote genome comparisons. *BMC Genom.* **2011**, *12*, 402. [CrossRef] [PubMed]
43. Zheng, J.; Peng, D.; Ruan, L.; Sun, M. Evolution and dynamics of megaplasmids with genome sizes larger than 100 kb in the *Bacillus cereus* group. *BMC Evol. Biol.* **2013**, *13*, 262. [CrossRef]
44. Dong, M.J.; Luo, H.; Gao, F. Ori-Finder 2022: A Comprehensive Web Server for Prediction and Analysis of Bacterial Replication Origins. *Genom. Proteom. Bioinform.* **2022**, *20*, 1207–1213. [CrossRef]
45. Ducrest, P.J.; Pfammatter, S.; Stephan, D.; Vogel, G.; Thibault, P.; Schnyder, B. Rapid detection of *Bacillus* ionophore cereulide in food products. *Sci. Rep.* **2019**, *9*, 5814. [CrossRef]
46. Tinsley, E.; Khan, S.A. A novel FtsZ-like protein is involved in replication of the anthrax toxin-encoding pXO1 plasmid in *Bacillus anthracis*. *J. Bacteriol.* **2006**, *188*, 2829–2835. [CrossRef] [PubMed]
47. Margolin, W. FtsZ and the division of prokaryotic cells and organelles. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 862–871. [CrossRef] [PubMed]
48. Lobry, J.R. Asymmetric substitution patterns in the two DNA strands of bacteria. *Mol. Biol. Evol.* **1996**, *13*, 660–665. [CrossRef] [PubMed]
49. Fuchs, E.; Raab, C.; Brugger, K.; Ehling-Schulz, M.; Wagner, M.; Stessl, B. Performance Testing of *Bacillus cereus* Chromogenic Agar Media for Improved Detection in Milk and Other Food Samples. *Foods* **2022**, *11*, 288. [CrossRef]
50. Tallent, S.M.; Kotewicz, K.M.; Strain, E.A.; Bennett, R.W. Efficient isolation and identification of *Bacillus cereus* group. *J. AOAC Int.* **2012**, *95*, 446–451. [CrossRef] [PubMed]
51. Papaparaskevas, J.; Houhoula, D.P.; Papadimitriou, M.; Saroglou, G.; Legakis, N.J.; Zerva, L. Ruling out *Bacillus anthracis*. *Emerg. Infect. Dis.* **2004**, *10*, 732–735. [CrossRef] [PubMed]
52. Ha, M.; Jo, H.J.; Choi, E.K.; Kim, Y.; Kim, J.; Cho, H.J. Reliable Identification of *Bacillus cereus* Group Species Using Low Mass Biomarkers by MALDI-TOF MS. *J. Microbiol. Biotechnol.* **2019**, *29*, 887–896. [CrossRef] [PubMed]
53. Manzulli, V.; Rondinone, V.; Buchicchio, A.; Serrecchia, L.; Cipolletta, D.; Fasanella, A.; Parisi, A.; Difato, L.; Iatarola, M.; Aceti, A.; et al. Discrimination of *Bacillus cereus* Group Members by MALDI-TOF Mass Spectrometry. *Microorganisms* **2021**, *9*, 1202. [CrossRef]
54. Pauker, V.I.; Thoma, B.R.; Grass, G.; Bleichert, P.; Hanczaruk, M.; Zoller, L.; Zange, S. Improved Discrimination of *Bacillus anthracis* from Closely Related Species in the *Bacillus cereus Sensu Lato* Group Based on Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry. *J. Clin. Microbiol.* **2018**, *56*, e01900–e01917. [CrossRef]
55. Quagliariello, A.; Cirigliano, A.; Rinaldi, T. *Bacilli* in the International Space Station. *Microorganisms* **2022**, *10*, 2309. [CrossRef]
56. Abo-Aba, S.E.; Sabir, J.S.; Baeshen, M.N.; Sabir, M.J.; Mutwakil, M.H.; Baeshen, N.A.; D'Amore, R.; Hall, N. Draft Genome Sequence of *Bacillus* Species from the Rhizosphere of the Desert Plant Rhazya stricta. *Genome Announc.* **2015**, *3*, e00957-15. [CrossRef]
57. Venkateswaran, K.; Singh, N.K.; Checinska Sielaff, A.; Pope, R.K.; Bergman, N.H.; van Tongeren, S.P.; Patel, N.B.; Lawson, P.A.; Satomi, M.; Williamson, C.H.D.; et al. Non-Toxin-Producing *Bacillus cereus* Strains Belonging to the *B. anthracis* Clade Isolated from the International Space Station. *mSystems* **2017**, *2*, e00021-17. [CrossRef]
58. Irenge, L.M.; Bearzatto, B.; Ambroise, J.; Gala, J.L. Complete Genome Sequence of an Environmental *Bacillus cereus* Isolate Belonging to the *Bacillus anthracis* Clade. *Microbiol. Resour. Announc.* **2020**, *9*, e00917-20. [CrossRef] [PubMed]
59. Zhang, Y.; Zhao, S.; Liu, S.; Peng, J.; Zhang, H.; Zhao, Q.; Zheng, L.; Chen, Y.; Shen, Z.; Xu, X.; et al. Enhancing the Phytoremediation of Heavy Metals by Combining Hyperaccumulator and Heavy Metal-Resistant Plant Growth-Promoting Bacteria. *Front. Plant Sci.* **2022**, *13*, 912350. [CrossRef] [PubMed]
60. Zheng, J.; Guan, Z.; Cao, S.; Peng, D.; Ruan, L.; Jiang, D.; Sun, M. Plasmids are vectors for redundant chromosomal genes in the *Bacillus cereus* group. *BMC Genom.* **2015**, *16*, 6. [CrossRef] [PubMed]
61. Bianco, A.; Capozzi, L.; Monno, M.R.; Del Sambro, L.; Manzulli, V.; Pesole, G.; Loconsole, D.; Parisi, A. Characterization of *Bacillus cereus* Group Isolates from Human Bacteremia by Whole-Genome Sequencing. *Front. Microbiol.* **2020**, *11*, 599524. [CrossRef] [PubMed]
62. Fraccalvieri, R.; Bianco, A.; Difato, L.M.; Capozzi, L.; Del Sambro, L.; Simone, D.; Catanzariti, R.; Caruso, M.; Galante, D.; Normanno, G.; et al. Toxigenic Genes, Pathogenic Potential and Antimicrobial Resistance of *Bacillus cereus* Group Isolated from Ice Cream and Characterized by Whole Genome Sequencing. *Foods* **2022**, *11*, 2480. [CrossRef] [PubMed]

63. Frentzel, H.; Kelner-Burgos, Y.; Fischer, J.; Heise, J.; Gohler, A.; Wichmann-Schauer, H. Occurrence of selected bacterial pathogens in insect-based food products and in-depth characterisation of detected *Bacillus cereus* group isolates. *Int. J. Food Microbiol.* **2022**, *379*, 109860. [CrossRef]
64. Huang, Y.; Flint, S.H.; Yu, S.; Ding, Y.; Palmer, J.S. Phenotypic properties and genotyping analysis of *Bacillus cereus* group isolates from dairy and potato products. *LWT* **2021**, *140*, 110853. [CrossRef]
65. Ehling-Schulz, M.; Guinebretiere, M.H.; Monthan, A.; Berge, O.; Fricker, M.; Svensson, B. Toxin gene profiling of enterotoxic and emetic *Bacillus cereus*. *FEMS Microbiol. Lett.* **2006**, *260*, 232–240. [CrossRef]
66. Carlson, C.J.; Kracalik, I.T.; Ross, N.; Alexander, K.A.; Hugh-Jones, M.E.; Fegan, M.; Elkin, B.T.; Epp, T.; Shury, T.K.; Zhang, W.; et al. The global distribution of *Bacillus anthracis* and associated anthrax risk to humans, livestock and wildlife. *Nat. Microbiol.* **2019**, *4*, 1337–1343. [CrossRef] [PubMed]
67. Franz, D.R. Preparedness for an anthrax attack. *Mol. Asp. Med.* **2009**, *30*, 503–510. [CrossRef] [PubMed]
68. Schwartz, M. Dr. Jekyll and Mr. Hyde: A short history of anthrax. *Mol. Asp. Med.* **2009**, *30*, 347–355. [CrossRef] [PubMed]
69. Cui, Y.; Martlbauer, E.; Dietrich, R.; Luo, H.; Ding, S.; Zhu, K. Multifaceted toxin profile, an approach toward a better understanding of probiotic *Bacillus cereus*. *Crit. Rev. Toxicol.* **2019**, *49*, 342–356. [CrossRef] [PubMed]
70. Carroll, L.M.; Wiedmann, M. Cereulide Synthetase Acquisition and Loss Events within the Evolutionary History of Group III *Bacillus cereus Sensu Lato* Facilitate the Transition between Emetic and Diarrheal Foodborne Pathogens. *mBio* **2020**, *11*, 1–16. [CrossRef] [PubMed]
71. Ehling-Schulz, M.; Svensson, B.; Guinebretiere, M.H.; Lindback, T.; Andersson, M.; Schulz, A.; Fricker, M.; Christiansson, A.; Granum, P.E.; Martlbauer, E.; et al. Emetic toxin formation of *Bacillus cereus* is restricted to a single evolutionary lineage of closely related strains. *Microbiology* **2005**, *151*, 183–197. [CrossRef] [PubMed]
72. Hoton, F.M.; Fornelos, N.; N'Guessan, E.; Hu, X.; Swiecicka, I.; Dierick, K.; Jaaskelainen, E.; Salkinoja-Salonen, M.; Mahillon, J. Family portrait of *Bacillus cereus* and *Bacillus weihenstephanensis* cereulide-producing strains. *Environ. Microbiol. Rep.* **2009**, *1*, 177–183. [CrossRef]
73. Thorsen, L.; Hansen, B.M.; Nielsen, K.F.; Hendriksen, N.B.; Phipps, R.K.; Budde, B.B. Characterization of emetic *Bacillus weihenstephanensis*, a new cereulide-producing bacterium. *Appl. Environ. Microbiol.* **2006**, *72*, 5118–5121. [CrossRef]
74. Meng, J.-N.; Liu, Y.-J.; Shen, X.; Wang, J.; Xu, Z.-K.; Ding, Y.; Beier, R.C.; Luo, L.; Lei, H.-T.; Xu, Z.-L. Detection of emetic *Bacillus cereus* and the emetic toxin cereulide in food matrices: Progress and perspectives. *Trends Food Sci. Technol.* **2022**, *123*, 322–333. [CrossRef]
75. Ulrich, S.; Gottschalk, C.; Dietrich, R.; Martlbauer, E.; Gareis, M. Identification of cereulide producing *Bacillus cereus* by MALDI-TOF MS. *Food Microbiol.* **2019**, *82*, 75–81. [CrossRef]
76. Biggel, M.; Jessberger, N.; Kovac, J.; Johler, S. Recent paradigm shifts in the perception of the role of *Bacillus thuringiensis* in foodborne disease. *Food Microbiol.* **2022**, *105*, 104025. [CrossRef]
77. Crickmore, N.; Berry, C.; Panneerselvam, S.; Mishra, R.; Connor, T.R.; Bonning, B.C. A structure-based nomenclature for *Bacillus thuringiensis* and other bacteria-derived pesticidal proteins. *J. Invertebr. Pathol.* **2021**, *186*, 107438. [CrossRef]
78. Sauka, D.H.; Peralta, C.; Pérez, M.P.; Onco, M.I.; Fiodor, A.; Caballero, J.; Caballero, P.; Berry, C.; Del Valle, E.E.; Palma, L. *Bacillus toyonensis* biovar *Thuringiensis*: A novel entomopathogen with insecticidal activity against lepidopteran and coleopteran pests. *Biol. Control* **2022**, *167*, 104838. [CrossRef]
79. Lan, X.; Wang, Q.; Wu, T.; Li, N.; Wang, H.; Zheng, Z. Draft Genome Sequence of *Bacillus wiedmannii* Biovar *thuringiensis* ZZQ-138, Isolated from a Saline Lake. *Microbiol. Resour. Announc.* **2022**, *11*, e0096421. [CrossRef]
80. Lazarte, J.N.; Lopez, R.P.; Ghiringhelli, P.D.; Beron, C.M. *Bacillus wiedmannii* biovar *thuringiensis*: A Specialized Mosquitocidal Pathogen with Plasmids from Diverse Origins. *Genome Biol. Evol.* **2018**, *10*, 2823–2833. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

3.3 Article 2

AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

Check for updates

| Bacteriology | Announcement

Reference genome assembly and annotation of two *Bacillus cereus sensu lato* strains from Etosha National Park, Namibia

Russell J. S. Orr,[1] Ola B. Brynildsrud,[1] Mehdi Abdelli,[2] Vincent Ramisse,[2] Marius Dybwad[1]

AUTHOR AFFILIATIONS See affiliation list on p. 3.

ABSTRACT *Bacillus cereus sensu lato* (*s.l.*) poses health and security issues. Here, we report the reference genome assembly of two *Bacillus cereus s.l.* strains, isolated from Etosha National Park, Namibia (FFI_BCgr36 and FFI_BCgr46). These unique genomes open for better understanding of environmental diversity and improvements in detection of threatening species.

KEYWORDS genome analysis, *Bacillus cereus*, *Bacillus*

Bacillus cereus sensu lato (*s.l.*) includes numerous Gram-positive, spore-forming, and rod-shaped species that are environmentally ubiquitous and pose health and food security issues (1). In particular, *Bacillus anthracis*, the agent of anthrax, the food-borne pathogen *Bacillus cereus sensu stricto,* and the biopesticide *Bacillus thuringiensis* (2).

Two *B. cereus s.l.* strains (FFI_BCgr36 and FFI_BCgr46) were isolated from soil sampled in the vicinity of a *Equus quagga* (plains zebra) carcass (Etosha National Park, Namibia, 2012). Permission was granted from the Ministry of Environment and Tourism of Namibia (permit: 1617/2011). Single-colony isolates were originally cultured from serial dilutions on polymyxin-lysozyme-EDTA-thallous acetate agar plates overnight at 37°C, before storage (3, 4). Isolates were characterized as *B. cereus* using cell and colony morphology, motility, penicillin sensitivity, qPCR, and an MLST scheme (5, 6). Pulsed-field gel electrophoresis characterized plasmid presence and size. Stock was re-cultured overnight at 37°C on tryptic soy agar prior to DNA isolation. DNA was isolated once using Qiagen DNeasy Blood and Tissue kit (Gram-positive protocol). For Illumina, DNA was sheared to 650 bp, and libraries constructed with Nextera XT/DNA prep before MiSeq sequencing. Reads were trimmed using TrimGalore v0.6.10 (7) with –length 80 and –q 30. For Nanopore, libraries were constructed from the same and unsheared genomic DNA with the Rapid Barcoding kit and sequenced on R9.4.1 flow cells, with super-accurate basecalling using Guppy v15.0.0 (8). Reads were trimmed using NanoFilt v2.8.0 (9) with -q 10 -l 5000 and error corrected using FMLRC2 v0.1.7 (10) with cache size 10 and *Kmer* sizes 21, 59, 79, and 127. Hybrid *de novo* assemblies were performed using Unicycler v0.5.0 (11) in "bold" mode allowing circularization of overlapping ends, confirmed with assembly graphs, and rotating assemblies to begin at a consistent starting gene (dnaA). Assemblies were polished with Polypolish v0.5.0 (12) and annotated with PGAP v2022-12-13.build6494 (13). Chromosome and plasmid sequences were compared, at nucleotide level, including closest BLASTn NCBInt hits (January 2023), with nucmer v4.0.0.rc1 (14). Assembly stats were obtained using Bowtie2 v2.5.1 (15) and Minimap2 v2.24-r1122 (16). A chromosome alignment, including closest BLAST hits, was constructed using Parsnp v1.7.4 (17) and phylogenetically inferred with RAxML v8.2.12 (18), employing the GTRCAT model. Topology was the best of 20 heuristic searches and bootstrap from 100 pseudo replicates. The tree was visualized with iTOL (19). Default parameters were used for all software unless otherwise specified.

Editor David Rasko, University of Maryland School of Medicine, Baltimore, Maryland, USA

Address correspondence to Russell J. S. Orr, Russell-John-Scott.Orr@ffi.no.

The authors declare no conflict of interest.

See the funding table on p. 3.

Received 21 June 2023
Accepted 21 September 2023

Published 19 October 2023

Copyright © 2023 Orr et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Downloaded from https://journals.asm.org/journal/mra on 02 December 2023 by 109.12.79.239.

TABLE 1 Assembly statistics, sequence data, and genome annotation[a]

Assembly	FFI_BCgr36 chromosome	FFI_BCgr36 plasmid 1	FFI_BCgr46 chromosome	FFI_BCgr46 plasmid 1	FFI_BCgr46 plasmid 2
Genbank accession	CP125992	CP125993	CP125989	CP125991	CP125990
Size (bp)	5,352,352	12,239	5,206,453	11,735	579,146
GC %	35.3	35.3	35.4	30.1	32.9
Illumina coverage	116×	771×	122×	489×	137×
ONT coverage	74×	2,753×	56×	1,504×	68×
Comparable depth	1×	6.65×	1×	4×	1×
Sequence data	**FFI_BCgr36 genome**		**FFI_BCgr46 genome**		
BioSample accession	SAMN35055520		SAMN35055521		
SRA accession Illumina	SRX20303850		SRX20303852		
Illumina reads (PE)	1,415,409		1,505,971		
SRA accession ONT	SRX20303851		SRX20303853		
ONT reads	322,909		215,326		
ONT N50	12,191		12,311		
Annotation	**FFI_BCgr36 genome**		**FFI_BCgr46 genome**		
Genes (total)	5,558		5,910		
CDSs (total)	5,406		5,757		
Genes (coding)	5,293		5,594		
CDSs (with protein)	5,293		5,594		
Genes (RNA)	152		153		
rRNAs	14, 14, 14 (5S, 16S, 23S)		14, 14, 14 (5S, 16S, 23S)		
Complete rRNAs	14, 14, 14 (5S, 16S, 23S)		14, 14, 14 (5S, 16S, 23S)		
tRNAs	105		106		
ncRNAs	5		5		
Pseudo genes (total)	113		163		
CDSs (without protein)	113		163		
Pseudo genes (ambiguous residues)	0 of 113		0 of 163		
Pseudo genes (frameshifted)	49 of 113		73 of 163		
Pseudo genes (incomplete)	70 of 113		93 of 163		
Pseudo genes (internal stop)	28 of 113		58 of 163		
Pseudo genes (multiple problems)	30 of 113		51 of 163		

[a]Statistics from Bowtie2 and Minimap2 for both the chromosome and plasmid of the two *B. cereus s.l.* strains. Annotation for each complete genome from PGAP.

The two FFI *B. cereus s.l.* strains were assembled to complete circularized chromosome and plasmid sequences (Table 1). FFI_BCgr36 has a 5.35 Mb, and FFI_BCgr46 has a 5.21 Mb chromosome. The annotation (Table 1) and phylogeny (Fig. 1) show strains as highly similar to each other, with the closest public strain being JRS4 *B. cereus* (GCF_001286825.1), isolated from soil, Jeddah (20). *B. cereus* JRS4 has 99.20% BLAST identity, over a 90% query coverage, to both FFI_BCgr36 and FFI_BCgr46 (Fig. 1). For comparison, the FFI strains have a chromosome identity of 99.61% over a 97% query coverage (Fig. 1).

The provided assemblies and annotations permit a better understanding of environmental diversity and improvements in detection of these potentially pathogenic species.

Downloaded from https://journals.asm.org/journal/mra on 02 December 2023 by 109.12.79.239.

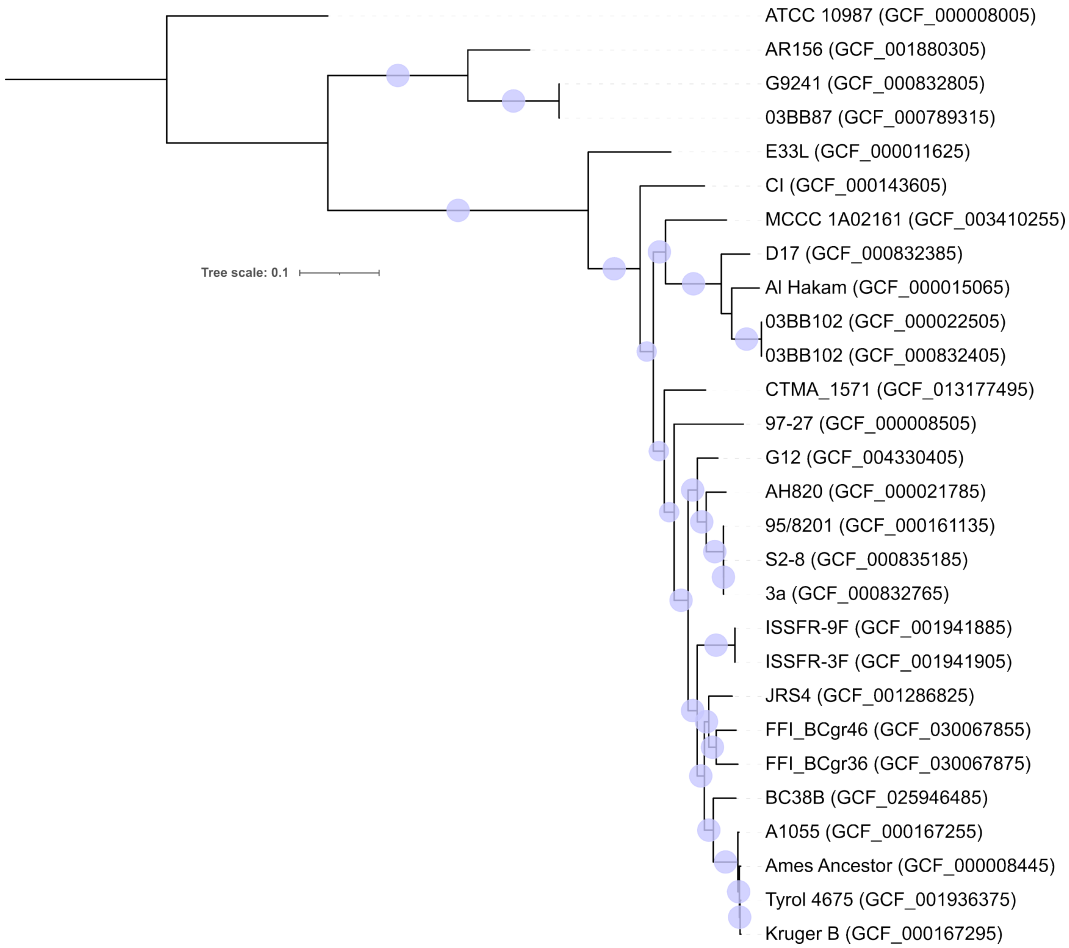

FIG 1 The inferred chromosome phylogeny of multiple *B. cereus s.l*. strains showing the relative position of those presented in this paper. Maximum likelihood phylogeny of 25 *B. cereus s.l*. strains with 190,584 nucleotide characters inferred using RAxML (20 heuristic searches and bootstrap of 100 pseudo replicates). Bootstrap values >90 are shown as blue circles. Assembly accessions for each strain are in brackets.

ACKNOWLEDGMENTS

We thank Tone Aarskaug for culturing and DNA isolation of the strains. We acknowledge Wendy C. Turner and colleagues at the Centre for Ecological and Evolutionary Synthesis (CEES), University of Oslo, Norway, for sampling of the presented strains.

AUTHOR AFFILIATIONS

[1]Total Defence Division, Norwegian Defence Research Establishment FFI, Kjeller, Norway
[2]Division Biologie, DGA Maîtrise NRBC, Vert-le-Petit, France

AUTHOR ORCIDs

Russell J. S. Orr http://orcid.org/0000-0002-1972-1321

FUNDING

Funder	Grant(s)	Author(s)
FFI		Russell J.S. Orr
		Ola B. Brynildsrud
		Marius Dybwad

Downloaded from https://journals.asm.org/journal/mra on 02 December 2023 by 109.12.79.239.

Funder	Grant(s)	Author(s)
DGA		Mehdi Abdelli
		Vincent Ramisse

AUTHOR CONTRIBUTIONS

Russell J. S. Orr, Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Software, Validation, Visualization, Funding acquisition, Writing – review and editing | Ola B. Brynildsrud, Validation, Writing – review and editing | Mehdi Abdelli, Formal analysis, Investigation, Methodology, Visualization, Writing – review and editing | Vincent Ramisse, Conceptualization, Investigation, Writing – review and editing, Funding acquisition | Marius Dybwad, Conceptualization, Writing – review and editing, Funding acquisition

DATA AVAILABILITY

The complete assemblies and annotations of the two *B. cereus s.l.* strains (FFI_BCgr36 and FFI_BCgr46) have been deposited in Genbank under the following accessions: FFI_BCgr36 chromosome: CP125992; FFI_BCgr36 plasmid: CP125993; FFI_BCgr46 chromosome: CP125989; FFI_BCgr46 small plasmid: CP125990; FFI_BCgr46 large plasmid: CP125991. Associated BioSample accessions including SRAs are SAMN35055520 (FFI_BCgr36) and SAMN35055521 (FFI_BCgr46) within BioProject PRJNA971511.

REFERENCES

1. Rouzeau-Szynalski K, Stollewerk K, Messelhäusser U, Ehling-Schulz M. 2020. Why be serious about emetic *Bacillus cereus*: cereulide production and industrial challenges. Food Microbiol 85:103279. https://doi.org/10.1016/j.fm.2019.103279
2. Valseth K, Nesbø CL, Easterday WR, Turner WC, Olsen JS, Stenseth NC, Haverkamp THA. 2017. Temporal dynamics in microbial soil communities at anthrax carcass sites. BMC Microbiol 17:206. https://doi.org/10.1186/s12866-017-1111-6
3. Turner WC, Kausrud KL, Krishnappa YS, Cromsigt JPGM, Ganz HH, Mapaure I, Cloete CC, Havarua Z, Küsters M, Getz WM, Stenseth NC. 2014. Fatal attraction: vegetation responses to nutrient inputs attract herbivores to infectious anthrax carcass sites. Proc Biol Sci 281:20141785. https://doi.org/10.1098/rspb.2014.1785
4. Valseth K, Nesbø CL, Easterday WR, Turner WC, Olsen JS, Stenseth NC, Haverkamp THA. 2016. Draft genome sequences of two *Bacillus anthracis* strains from Etosha National Park, Namibia. Genome Announc 4:e00861-16. https://doi.org/10.1128/genomeA.00861-16
5. Hovland K. 2014. Multilocus sequence typing of close neighbours to *Bacillus anthracis* isolated from soil samples MSc, Norwegian University of Science and Technology, Trondheim, Norway
6. Søndenå M. 2014. Fenotypisk og genetisk kartlegging av *Bacillus cereus*-gruppeisolater fra Etosha National Park, Namibia - isolering og karakterisering av nye stammer med likhet til *Bacillus anthracis* MSc, University of Oslo, Oslo, Norway
7. Krueger F. 2012. Trim galore a wrapper to automate quality and adapter trimming. Available from: https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
8. Wick RR, Judd LM, Holt KE. 2019. Performance of neural network basecalling tools for oxford nanopore sequencing. Genome Biol 20:129. https://doi.org/10.1186/s13059-019-1727-y
9. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. 2018. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics 34:2666–2669. https://doi.org/10.1093/bioinformatics/bty149
10. Mak QXC, Wick RR, Holt JM, Wang JR. 2023. Polishing de novo nanopore assemblies of bacteria and eukaryotes with FMLRC2. Mol Biol Evol 40:msad048. https://doi.org/10.1093/molbev/msad048
11. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13:e1005595. https://doi.org/10.1371/journal.pcbi.1005595
12. Wick RR, Holt KE. 2022. Polypolish: short-read polishing of long-read bacterial genome assemblies. PLoS Comput Biol 18:e1009802. https://doi.org/10.1371/journal.pcbi.1009802
13. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569
14. Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. 2018. MUMmer4: a fast and versatile genome alignment system. PLoS Comput Biol 14:e1005944. https://doi.org/10.1371/journal.pcbi.1005944
15. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923
16. Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094–3100. https://doi.org/10.1093/bioinformatics/bty191
17. Treangen TJ, Ondov BD, Koren S, Phillippy AM. 2014. The harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. Genome Biol 15:524. https://doi.org/10.1186/s13059-014-0524-x
18. Stamatakis A. 2014. RaxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313. https://doi.org/10.1093/bioinformatics/btu033
19. Letunic I, Bork P. 2021. Interactive tree of life (iTOL) V5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res 49:W293–W296. https://doi.org/10.1093/nar/gkab301
20. Abo-Aba SEM, Sabir JSM, Baeshen MN, Sabir MJ, Mutwakil MHZ, Baeshen NA, D'Amore R, Hall N. 2015. Draft genome sequence of *Bacillus* species from the rhizosphere of the desert plant *Rhazya stricta*. Genome Announc 3:e00957-15. https://doi.org/10.1128/genomeA.00957-15

Downloaded from https://journals.asm.org/journal/mra on 02 December 2023 by 109.12.79.239.

IDÉES À RETENIR

1. Les souches BC38B, FFI_BCgr36 et FFI_BCgr46, nouvellement découvertes, constituent le plus proche voisinage de *B. anthracis* actuellement référencé.
2. La caractérisation de ces souches, en particulier de leurs larges plasmides, permettra d'affiner les connaissances actuelles sur l'histoire évolutive de *B. anthracis*.
3. Certains outils de détection (MALDI-TOF par exemple) peuvent s'avérer défaillants dans l'identification d'espèces du groupe *B. cereus*, du fait de bases de données de référence insuffisamment fournies, renforçant ainsi la nécessité de la recherche de nouveaux proches voisins de *B. anthracis*.

Chapitre 4

Un modèle d'émergence de *Bacillus anthracis*

La définition de l'espèce *B. anthracis* est encore l'objet de discussions actuellement, en particulier concernant l'origine et la datation de l'émergence de cette espèce, voire même l'assignation de *B. anthracis* comme espèce. De même, les mécanismes qui ont amené l'espèce à acquérir ses plasmides de virulence demeurent une problématique contemporaine. Ce chapitre vise à tirer profit du proche voisinage de *B. anthracis* nouvellement établi pour aborder ces deux sujets : d'une part en déterminant la position de l'ancêtre de *B. anthracis* le long de la branche liant l'espèce au reste du groupe *B. cereus*, d'autre part en s'intéressant aux réplicons contenus dans les plasmides.

OBJECTIFS DU CHAPITRE

1. Positionner l'ancêtre de *B. anthracis* au sein de la phylogénie de l'espèce, en tirant profit de son proche voisinage et en étudiant les influences des pressions sélectives et de l'homoplasie
2. Comprendre la formation des plasmides de virulence de *B. anthracis* en analysant les séquences plasmidiques proches
3. Affiner les hypothèses existantes quant à l'émergence et l'évolution de *B. anthracis* à la lumière de ces nouveaux éléments

4.1 Contexte de l'étude

B. anthracis, l'agent pathogène responsable de la maladie du charbon, se distingue au sein du groupe *B. cereus*. Ce dernier comprend une variété de souches, allant de celles inoffensives à d'autres pathogènes pour l'Homme, y compris certaines dites anthracis-like, provoquant des maladies animales ou humaines similaires à la maladie du charbon.

L'avènement du séquençage de nouvelle génération a joué un rôle important dans la caractérisation d'un nombre croissant de souches de *B. anthracis*. Cette avancée technologique a permis une identification plus précise des différentes lignées et sous-lignées phylogénétiques de l'espèce, améliorant notre compréhension de la diversité génétique de l'espèce (Van Ert et al., 2007). La clonalité de *B. anthracis* démontrée par ces données implique que cette espèce a émergé d'un événement évolutif unique. L'évolution a ensuite été marquée par une homogénéité génétique à travers le temps.

Cependant, malgré ces progrès significatifs, des questions fondamentales persistent concernant l'émergence de *B. anthracis*. Un point d'interrogation majeur concerne

l'apparition de l'ancêtre des différentes lignées identifiées. Cette incertitude est en partie attribuable à la capacité de *B. anthracis* à former des spores et à subsister dans l'environnement sous une forme dormante (CARLSON et al., 2018). Cela entraîne une difficulté à établir une chronologie de l'émergence et de la dissémination de l'espèce.

En outre, les souches anthracis-like sont réparties dans plusieurs clades du groupe *B. cereus*, présentant de même un défi pour les efforts de classification et de compréhension de leur apparition. L'acquisition de plasmides de virulence par ces souches dites anthracis-like fait l'objet de plusieurs hypothèses : la première étant l'existence d'un ancêtre commun possédant ces plasmides, la seconde étant la divergence des espèces suivie d'un transfert latéral de plasmides entre les souches. Certaines études mettent en lumière l'existence de souches *B. cereus* anthracis-like dont les plasmides de virulence forment un groupe phylogénétique distinct de ceux de *B. anthracis* (ANTONATION et al., 2016). D'autres soulignent la capacité d'échange génétique au sein du groupe *B. cereus*, phénomène bien établi dans divers milieux naturels, en particulier dans le sol de la rhizosphère (SAILE et KOEHLER, 2006). Par exemple, OKINAKA, PEARSON et KEIM, 2006 ont montré que *B. anthracis* est capable d'échanger des séquences génétiques avec *B. cereus*, un processus qui peut se produire pendant son stade végétatif.

Cette étude vise à apporter de nouveaux éléments de compréhension quant aux deux problématiques mentionnées : l'apparition de l'ancêtre de *B. anthracis* et l'acquisition des plasmides de virulence au sein de cette espèce et des souches anthracis-like. La première étape a consisté à élaborer une phylogénie qui inclut *B. anthracis*, ses proches voisins, ainsi que les souches anthracis-like au sein du groupe *B. cereus* et à étudier les pressions de sélection qui s'exercent sur cette phylogénie, tout en examinant les phénomènes d'homoplasie [1]. Cette analyse a permis de placer l'ancêtre de *B. anthracis* sur cette phylogénie. Dans un second temps, les séquences plasmidiques des souches du groupe *B. cereus* proches de *B. anthracis* ont été comparées, mettant en lumière l'existence de plusieurs mêmes types de réplicons en leur sein, dont un commun avec pXO1. Ces observations, intégrées aux diverses hypothèses existantes, ont permis d'établir un modèle global d'émergence de *B. anthracis*.

4.2 Matériel et méthodes

4.2.1 Étude de la phylogénie de *Bacillus anthracis* et des proches voisins

Analyse wgSNP Les génomes pris en compte ont été téléchargés sur NCBI. Ils ont ensuite été convertis en lectures simulées à l'aide d'un script interne (Python v3.6.2), elles-mêmes alignées sur le chromosome de *B. anthracis* Ames Ancestor (GCF_000008445.1), résultant en des alignements de longueur identique. Les SNPs génomiques sont sélectionnés avec BIONUMERICS 8.1.1 (BioMérieux, Applied Maths, Sint-Martens-Latem, Belgique), en utilisant l'option "*Strict SNP filtering (Closed SNP set)*" et les paramètres par défaut (*12 bp inter-SNP*). L'analyse du *clustering* est ensuite réalisée en utilisant la méthode de calcul "*Maximum parsimony tree*" et les paramètres par défaut.

De manière complémentaire, pour confirmer leur bonne cohérence, les arbres phylogénétiques ont également été construits à l'aide de RAXML-NG (v1.2.0) (KOZLOV

1. Désigne la similitude de traits ou de séquences génétiques entre espèces ou lignées qui ne provient pas d'un ancêtre commun mais résulte de l'évolution convergente, de la réversion ou de mutations parallèles.

et al., 2019) avec des alignements multi-FASTA en entrée, générés comme décrit précédemment. Le modèle GAMMAGTR a été utilisé dans chaque cas avec 100 *bootstraps* aléatoires.

Les arbres phylogénétiques créés ont été visualisés avec iTOL v6.8.2 (Letunic et Bork, 2021).

Traitement des recombinaisons La recombinaison doit être exclue de chaque alignement avant la construction d'un arbre phylogénétique, car elle n'est pas liée à l'évolution phylogénétique des souches. Ainsi, les alignements ont été épurés des SNPs agrégés dans une portion de génome de façon statistiquement significative en utilisant le logiciel Gubbins (v3.3.0) (Croucher et al., 2015) avec les paramètres par défaut. Les alignements résultants ont ensuite été utilisés comme entrée pour la construction des arbres phylogénétiques.

Étude de l'homoplasie Pour déterminer et supprimer les SNPs homoplasiques, l'outil SNPPar (v1.2) a été utilisé, avec les paramètres par défaut. L'outil HyPhy a permis ensuite de calculer les ratios dN/dS le long des branches des arbres phylogénétiques, avec la méthode aBSREL (Smith et al., 2015). Seules les séquences codantes ont été conservées en entrée. Les séquences codantes partielles ont été retirées. La visualisation des résultats obtenus s'est faite avec HyPhy-Vision (disponible à `http://vision.hyphy.org/`).

4.2.2 Analyse des séquences plasmidiques

Pour les plasmides étudiés, BLAST (Altschul et al., 1990) (dernière mise à jour : 22 avril 2023) a été utilisée pour rechercher des séquences de plasmides similaires. Les séquences d'ADN des plasmides ont été comparées avec BLAST Ring Image Generator (BRIG) (Alikhan et al., 2011). Pour identifier les minirépliconsⁿ[2] contenus dans les plasmides, une analyse BLASTP a été réalisée contre des protéines de réplication précédemment décrites dans les mégaplasmides[3] du groupe *B. cereus* (Zheng et al., 2013).

Afin d'identifier de nouveaux types de minirépliconsⁿ, les mots-clés "*Replication protein*" ou "*Rep protein*" ou "*Primase*" ont été recherchés dans les annotations. Les séquences protéiques TubZ ont également été recherchées avec la même méthode. Ori-Finder 2022 (consulté le 23 avril 2023) a été utilisé pour trouver les origines de réplication présentes au sein des plasmides (Dong, Luo et Gao, 2022). Pour identifier la ou les origines de réplication actives, les couvertures des lectures de séquençage sur les régions concernées ont été analysées par alignement sur les séquences plasmidiques correspondantes avec Geneious Prime 2022.2.1 (`http://www.geneious.com/`).

Pour détecter d'éventuels systèmes de sécrétion de type IV (T4SS), relatifs au pouvoir conjugatif, une analyse TBLASTN (Altschul et al., 1990) a été faite avec les séquences de VirB4 ayant le statut *reviewed* sur UniProt (`https://www.uniprot.org/`) (accédé en novembre 2023).

2. Plus petite région nécessaire pour la réplication, contenant l'origine de la réplication et des gènes codant les protéines de réplication.

3. Plasmide de grande taille. Dans le cas présent, supérieur à 100,000 paires de bases.

4.3 Résultats

4.3.1 Apparition de l'ancêtre de l'espèce *Bacillus anthracis*

Afin de mieux comprendre l'émergence de l'espèce *B. anthracis* au sein du groupe *B. cereus*, une démarche en plusieurs étapes a été réalisée. Tout d'abord, la connaissance plus fine du voisinage de *B. anthracis* a permis d'établir une phylogénie d'un panel de souches représentatif de la diversité de cette espèce ainsi que des souches du groupe *B. cereus* les plus proches. Sont en particulier présentes dans l'étude la souche BC38B (la plus proche voisine de *B. anthracis* dont la séquence génomique a été publiée à ce jour, ABDELLI et al., 2023), les souches FFI_BCgr36 et FFI_BCgr46 (ORR et al., 2023), la souche JRS4 (ABO-ABA et al., 2015), la souche BC8D, la souche *B. thuringiensis* serovar pulsiensis BGSC_4CC1 (ZWICK et al., 2012), la souche *B. thuringiensis* serovar vazensis BGSC_4CE1 et un panel de souches isolées dans l'*International Space Station* (ISSFR-3F, ISSFR-9F et JEM-2, VENKATESWARAN et al., 2017). A été ajouté un panel de souches anthracis-like, celles étant les plus proches de *B. anthracis* (un minimum de 75% de similtude avec le chromosome de *B. anthracis* Ames ancestor, autrement dit moins de 25% de régions non reconstruites après alignement sur ce chromosome). Il s'agit des souches CI, BacTX2020a, et 03BB102 (alias FDAARGOS_918) (KLEE et al., 2006 ; HOFFMASTER et al., 2006 ; SICHTIG et al., 2019 ; CARROLL et al., 2022b). L'arbre phylogénétique correspondant est représenté à la figure 4.1.

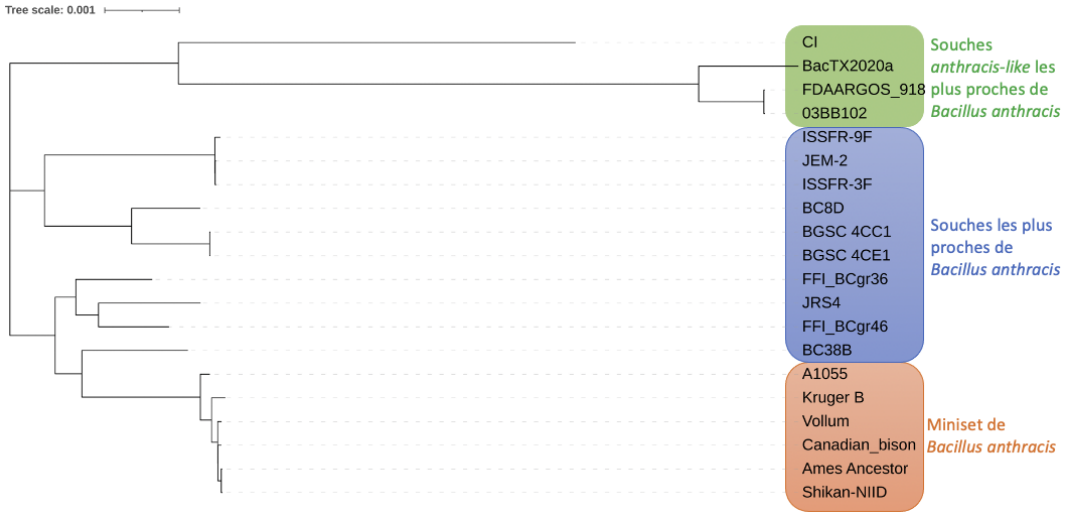

FIGURE 4.1 – Arbre phylogénétique comprenant un miniset de souches de *B. anthracis* et du proche voisinage de l'espèce.

La figure 4.2 montre les ratios dN/dS le long des branches de cet arbre. Trois catégories de valeurs se distinguent :

— Les branches relatives au miniset de souches de *B. anthracis* ont un ratio dN/dS compris entre 0.610 et 0.817.

— Les autres branches de l'arbre montrent un ratio bien inférieur (compris entre 0.0643 et 0.191), y compris celles relatives aux souches anthracis-like.

— à l'exception de celles relatives aux souches de l'*International Space Station* ayant un ratio supérieur à 1 (compris entre 1.06 et 1.96).

Dans la suite de l'étude, nous laisserons de côté les souches relatives à l'ISS, dont la différence en termes de ratios dN/dS est significative d'un isolement dans une

niche écologique bien particulière. Aucun SNP ne distingue les souches ISSFR-3F et JEM-2 et deux SNPs les séparent de la souche ISSFR-9F, suggérant la présence de néomutations [4] survenues dans la station spatiale. Un premier résultat remarquable est que les branches relatives aux souches anthracis-like ont subi des pressions de sélection similaires à celles des souches du groupe *B. cereus*, donc une évolution non clonale *a priori*. C'est le cas en particulier des souches FDAARGOS_918 et 03BB102 du MLST ST11. Un deuxième résultat remarquable est que la branche liant *B. anthracis* à son plus proche voisin (souche BC38B) possède un ratio dN/dS égal à 0.177, soit une valeur caractéristique de la deuxième catégorie.

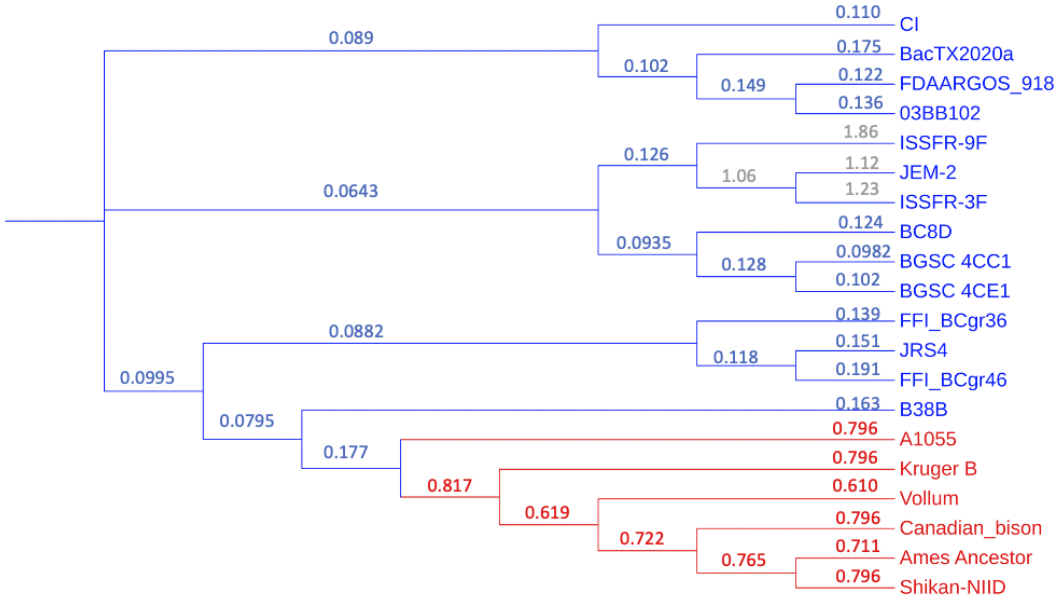

FIGURE 4.2 – Détermination des ratios dN/dS le long des branches de l'arbre phylogénétique de *B. anthracis* et de ses proches voisins. Les longueurs des branches ne sont pas à l'échelle, par souci de lisibilité.

Cette première analyse ne permet pas de déterminer si cette branche a subi des pressions de sélection similaires aux branches relatives aux proches voisins ou si elle a un comportement "hybride" entre *B. anthracis* et ses proches voisins. La figure 4.3 représente l'arbre phylogénétique des souches étudiées, avec la longueur des branches associées. Le taux d'homoplasie correspondant est de 40%. La figure 4.4 représente la même phylogénie, après retrait des SNPs homoplasiques. On remarque une réduction des longueurs des branches relatives à *B. anthracis* d'un facteur moyen égal à 0.92 (réduction de 8%), ce qui est cohérent avec le caractère clonal de cette espèce. En dehors de *B. anthracis*, la réduction des longueurs des branches est d'un facteur 0.36, soit une réduction de 64%, ce qui est cohérent avec la forte homoplasie de l'arbre initial (Figure 4.3). La branche directement en amont de *B. anthracis* a subi une réduction de longueur d'un facteur intermédiaire de 0.75 (25%).

On peut donc émettre l'hypothèse que le caractère clonal de *B. anthracis* a débuté avant la séparation en sous-lignées de l'espèce (position du MRCA). Le ratio permet

4. désigne une mutation qui apparaît pour la première fois dans un individu par modification de l'ADN, et qui n'était présente chez aucun de ses ascendants. Les néomutations peuvent survenir de manière spontanée lors de la réplication de l'ADN ou sous l'influence de facteurs environnementaux mutagènes.

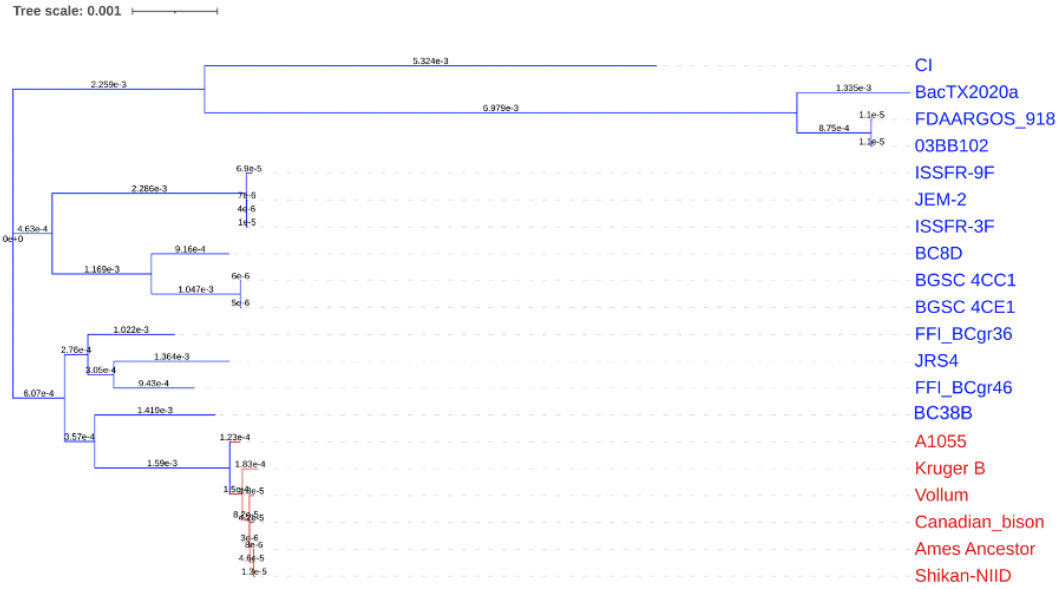

FIGURE 4.3 – Arbre phylogénétique de *B. anthracis* et de ses proches voisins, basé sur 87,369 SNPs. Le taux d'homoplasie associé est de 40%.

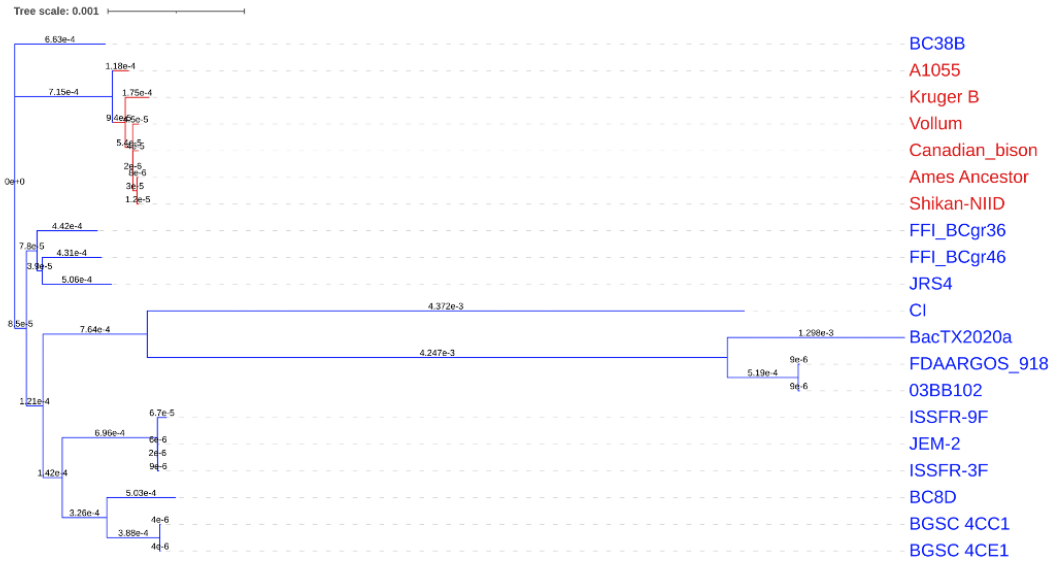

FIGURE 4.4 – Arbre phylogénétique de *B. anthracis* et de ses proches voisins, basé sur 52,472 SNPs, après retrait des SNPs homoplasiques.

d'estimer la proportion de la branche hybride produite sous un régime clonal, autrement dit l'apparition de l'ancêtre de l'espèce, le long de cette branche. Le calcul permettant de l'obtenir est le suivant :

Soit x la proportion de la branche ayant une évolution clonale.

x est solution de l'équation :

$$0,92x + 0,36(1-x) = 0,75$$

soit :

$$x \approx 16\%$$

On en déduit ainsi la position de l'ancêtre de l'espèce *B. anthracis* le long de cette branche, comme illustré sur la figure 4.5. Le niveau dN/dS le long de la branche conduisant à *B. anthracis* (figure 4.2) corrobore cette estimation.

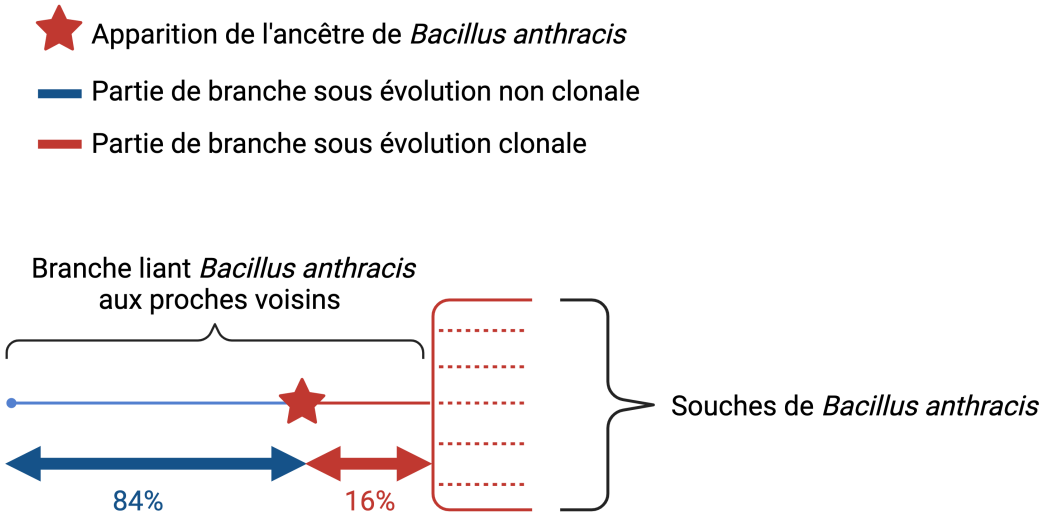

FIGURE 4.5 – Schéma illustratif de l'apparition de l'ancêtre de *B. anthracis* le long de la branche menant aux proches voisins. Créé avec BIORENDER.

Le placement de l'ancêtre de *B. anthracis* le long de la branche liant l'espèce à son plus proche voisin (la souche *B. cereus* BC38B) a été réalisé en usant d'une approche "chromosomique" (l'arbre phylogénétique étudié se basant sur des alignements sur le chromosome de *B. anthracis* Ames ancestor). Sur cette approche "chromosomique", les longueurs de branches figurant sur la figure 4.4 indiquent que la branche au comportement hybride a une longueur de 7,15x10^{-4}. Ainsi la partie de la branche évoluant sous régime clonal aurait une longueur égale à :

$$0,16 * 7,15.10^{-4} = 1,144.10^{-4}$$

Sur ce même arbre, la longueur de la branche relative à la lignée C (représentée par la souche *B. anthracis* A1055) est de 1,18x10^{-4}, valeur du même ordre de grandeur.

On peut donc estimer que l'ancêtre de *B. anthracis* est approximativement deux fois plus" ancien" en termes d'éloignement génétique que le MRCA de l'espèce (figure 4.6).

4.3.2 Formation des plasmides

Trois origines de réplication différentes ont été déterminées sur la séquence du plasmide de la souche BC38B, notées *ori1*, *ori2* et *ori3* sur la figure 4.7. Le *mapping* des lectures de séquençage associées à la souche sur la séquence plasmidique révèle

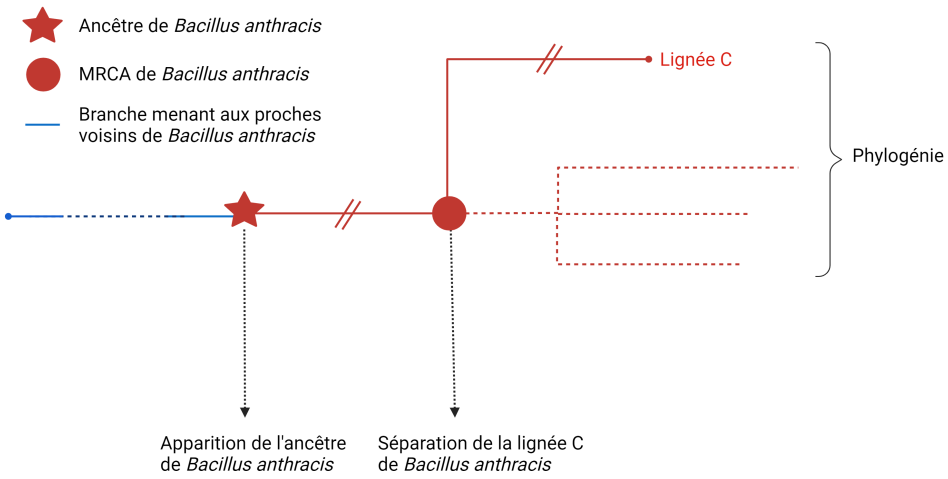

FIGURE 4.6 – Positionnement de l'ancêtre de *B. anthracis*. Il est approximativement deux fois plus ancien que le MRCA de l'espèce. Créé avec BIORENDER.

une couverture maximale de 75x pour *ori3* contre 30x et 33.5x sur *ori1* et *ori2* respectivement. C'est donc l'origine de réplication *ori3* qui est active. Sa position sur le plasmide (figure 4.7) correspond au changement de signe de la déviation GC (*GC skew*), qui coïncident avec l'origine ou le terminus de la réplication. Elle est associée aux gènes codant les protéines pXO1-14/pXO1-16 (protéine de liaison à l'ADN et protéine initiatrice de réplication, respectivement), agissant pour la réplication du plasmide pXO1 de *B. anthracis*.

Un gène codant une protéine de réplication-relaxation et un gène codant une protéine de division cellulaire FtsZ ont également été détectés. Le premier est associé à la réplication d'ADN plasmidique, tandis que le second a un rôle dans la division cellulaire et la duplication des plasmides. Une analyse BLASTP a révélé que ces éléments étaient largement répartis dans le groupe *B. cereus* et pouvaient être définis comme un nouveau type de miniréplicon dans les mégaplasmides du groupe *B. cereus*. Ces deux gènes se retrouvent par ailleurs dans les séquences plasmidiques des souches *B. cereus* CTMA_1571, FFI_BCgr46 (appartenant au même clade que *B. anthracis*) et des souches *B. toyonensis* JAS411 et *Bacillus sp.* PGP15 (listés dans le tableau 4.1).

La recherche de séquences de VirB4 s'est avérée infructueuse pour l'ensemble des plasmides étudiés.

4.4 Discussion

4.4.1 Apparition de l'ancêtre de *Bacillus anthracis*

On peut comparer les résultats de l'approche "chromosomique" obtenue précédemment à une approche "plasmidique". Cela consiste à établir une phylogénie entre les plasmides pXO1 des différentes lignées de *B. anthracis* et les plasmides similaires

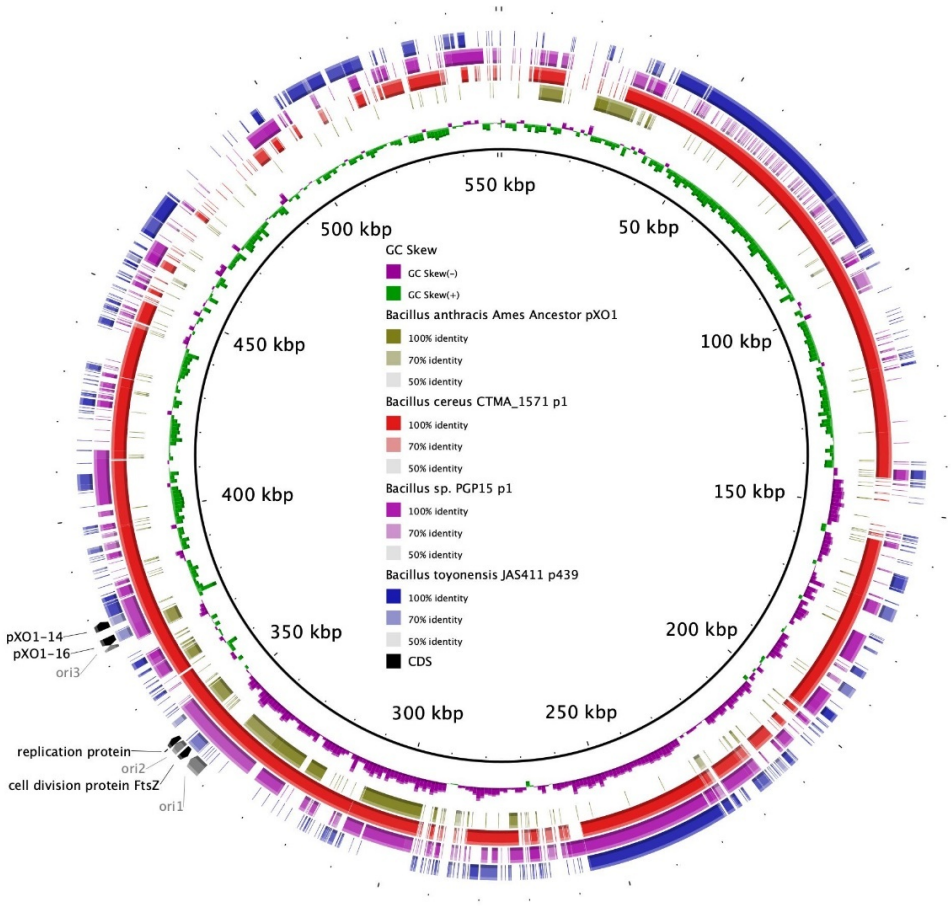

FIGURE 4.7 – Comparaison du plasmide de la souche *B. cereus* BC38B avec les plasmides les plus proches (plasmide p1 de la souche *B. cereus* CTMA_1571, plasmide p439 de la souche *B. toyonensis* JAS411 et plasmide p1 de *Bacillus sp.* PGP15). Anneaux du centre vers l'extérieur : (1) Numérotation des nucléotides ; (2) déviation GC ; (3-6) pourcentage d'identité de séquence avec les autres plasmides selon le code couleur défini sur la figure ; (7) gènes codant des miniréplicons (en noir), origines de réplication (en gris).

Nom de souche	Date de collection	Localisation	Source	*Accession number*	Taille du plasmide
B. anthracis Ames Ancestor	1981	Etats-Unis	Sol	GCA_000008445.1	181,677
B. cereus BC38B	2014	Slovénie	Poivre alimentaire	GCA_025946485.2	551,060
B. cereus FFI_BCgr46	2012	Namibie	Sol	GCA_030067855.1	579,146
B. cereus CTMA_1571	2011	Namibie	Sol	GCA_013177495.2	500,306
B. toyonensis JAS411	2011	Pologne	Sol	GCA_018741625.1	439,318
Bacillus sp. PGP15	2016	Chine	Rhizosphère	GCA_023101325.1	525,898

TABLE 4.1 – Tableau récapitulatif des caractéristiques des plasmides comparés dans cette étude.

des souches anthracis-like, comme illustré sur la figure 4.8 (VERGNAUD, 2020). On y remarque également un positionnement de l'ancêtre de pXO1 à équidistance entre l'émergence de la lignée C et celle des lignées relatives aux souches anthracis-like.

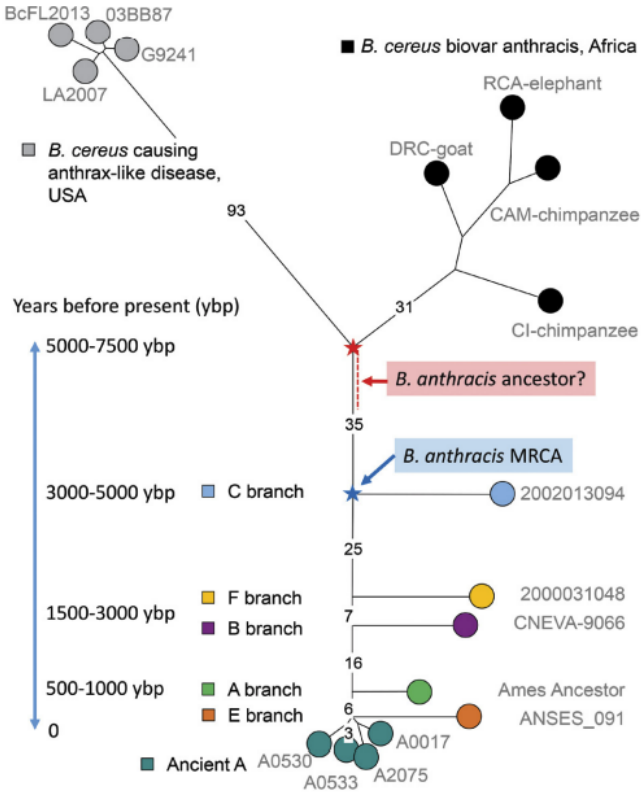

FIGURE 4.8 – Phylogénie des plasmides pX01 de *B. anthracis* et plasmides affiliés des souches anthracis-like. Tiré de VERGNAUD, 2020.

Autrement dit, les deux approches "chromosomique" et "plasmidique" établissent des résultats similaires et complémentaires. Accessoirement, l'étude confirme la co-évolution observée entre le chromosome et le plasmide pXO1 de *B. anthracis* (pas de trace d'échange de plasmides dans la phase "clonale").

Il serait également pertinent d'établir une approche "plasmidique" relative à pXO2 pour observer (ou non) une cohérence des résultats. Peu de séquences de plasmides pBCXO2 (relatives aux souches anthracis-like) sont disponibles à l'heure actuelle, en comparaison aux séquences de plasmides pBCXO1. Nous avons établi néanmoins une phylogénie de quelques séquences de pXO2 de *B. anthracis* (souches Ames ancestor, 2000031039, 2002013094) et de séquences de pBCXO2 de souches anthracis-like (Bcbva CI, CA). La séquence apparentée à pXO2 de la souche *B. cereus* BC-AK (possédant environ 50% de la séquence pXO2) a servi d'*outgroup* pour construire la phylogénie. L'arbre phylogénétique est présenté à la figure 4.9.

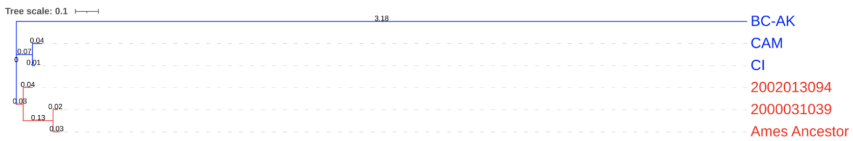

FIGURE 4.9 – Arbre phylogénétique des plasmides pXO2 de *B. anthracis* et plasmides pBCXO2 de souches Bcbva, basé sur 354 SNPs.

À titre de comparaison, on construit la phylogénie en prenant les mêmes souches et leurs séquences de pXO1 et pBCXO1. L'arbre phylogénétique associé est présenté à la figure 4.10.

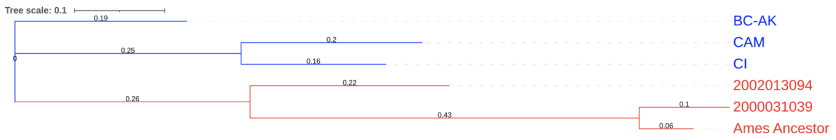

FIGURE 4.10 – Arbre phylogénétique des plasmides pXO1 de *B. anthracis* et plasmides pBCXO1 de souches Bcbva, basé sur 187 SNPs.

On remarque que la branche définie par la souche *B. cereus* BC-AK est beaucoup plus courte pour la phylogénie de pXO1/pBCXO1, relativement aux autres branches de l'arbre (10% du nombre de SNPs total de l'arbre spécifique à pXO1/pBCXO1 contre 89% pour l'arbre spécifique à pXO2/pBCXO2). Cela suggérerait que les plasmides pBCXO1 et pBCXO2 n'ont pas toujours co-évolué.

4.4.2 Formation des plasmides

— Les deux souches CTMA_1571 et FFI_BCgr46 ont été isolées dans le même pays (Namibie), à proximité du berceau hypothétique (forêt d'Afrique centrale) de *B. anthracis* (VERGNAUD, 2020). Quant à la souche BC38B, une recherche bibliographique sur les possibles pays d'exportation de poivre alimentaire n'a pas permis de remonter à son origine.

— Leurs chromosomes sont très proches génétiquement de celui de l'espèce *B. anthracis* (parmi les plus proches connues jusqu'à présent) et leurs plasmides partagent une large région commune entre eux (tout comme le plasmide de la souche BC38B) (ABDELLI et al., 2023 ; ORR et al., 2023).

— Les plasmides des souches CTMA_1571 et FFI_BCgr46 présentent une homologie élevée avec le plasmide de la souche BC38B. De plus, ils possèdent les mêmes types de miniréplicons. Ces trois souches appartiennent au clade I comme *B. anthracis* (Figure 2.5), bien que la similitude de leur plasmide avec le plasmide pXO1 soit modérée (IRENGE et al., 2020).

— Au moins six types de miniréplicons ont été découverts dans les mégaplasmides du groupe *B. cereus*. La présence de deux miniréplicons ou plus dans un mégaplasmide du groupe *B. cereus* suggère fortement l'intégration de plusieurs plasmides plus petits (ZHENG et al., 2013).

— Chacun des plasmides les plus proches de celui de la souche BC38B partage différentes régions avec lui (Figure 4.7). Par exemple, le plasmide p1 de *Bacillus sp.* PGP15 (isolé dans le sol d'une rhizosphère en Chine) ou le plasmide p439 de la souche *B. toyonensis* JAS411 (isolé dans le sol d'une terre agricole en Pologne), dont les deux souches associées appartiennent à des clades différents du groupe *B. cereus*.

Ces observations renforcent l'hypothèse d'une formation du plasmide pXO1 par une fusion spécifique de plusieurs plasmides (ou tronçons de plasmide) de plus petite taille, dont certains se retrouveraient sur d'autres mégaplasmides du groupe *B. cereus*. En particulier, le nouveau type de miniréplicon détecté sur les souches étudiées ici serait caractéristique d'un de ces tronçons ayant mené à la formation de pXO1.

Ce phénomène pourrait se produire aussi pour le plasmide pXO2, qui partage également des régions homologues avec d'autres plasmides du groupe *B. cereus*,

comme par exemple les plasmides pAW63 et pBT9727 des souches *B. thuringiensis subspecies kurstaki* HD73 et *B. thuringiensis* serovar *konkukian* 97-27 respectivement (AUWERA et MAHILLON, 2008). Cette dernière est d'ailleurs contenue dans le clade de *B. anthracis* (Figure 2.6). Ces trois plasmides possèdent une région commune d'environ 40 kb contenant des gènes codant plusieurs composants clés des systèmes de sécrétion conjugatifs de type IV (T4SS) (HU et al., 2009). Cette région de transfert conservée parmi les trois plasmides contient plusieurs gènes de transfert, y compris l'homologue du gène VirD4, mais dans le cas de pXO2, ils sont inactivés par des décalages du cadre de lecture. À l'instar de pXO1, le plasmide pXO2 n'est donc pas auto-transmissible. Cependant, certains des plasmides partageant des régions communes avec pXO2 sont conjugatifs. Cela pourrait expliquer une dissociation dans la formation de ces deux plasmides, et en particulier l'apparition de souches ayant des plasmides partageant des tronçons avec pXO1 mais pas pXO2 ou inversement. Un exemple serait les souches anthracis-like ne possédant que le plasmide pBCXO1 (BALDWIN, 2020).

4.4.3 Cas d'application : Détection de souches pathogènes

La détection du nouveau type de miniréplicon référencé dans les mégaplasmides du groupe *B. cereus* peut s'avérer utile lorsqu'il s'agit de déterminer ou de prévoir l'émergence de souches potentiellement virulentes dans des environnements définis. En effet, si le tronçon plasmidique ancestral contenant ce miniréplicon a permis la formation de pXO1 dans un écotype donné, il n'est pas à exclure que ce phénomène se reproduise dans d'autres niches écologiques. Ici, prenons l'exemple de cinq souches du groupe *B. cereus* isolées en Chine et Taïwan, dont la souche *Bacillus sp.* PGP15 mentionnée précédemment. Leurs caractéristiques sont résumées dans le tableau 4.2.

— La souche *Bacillus sp.* PGP15 a été isolée de la rhizosphère d'un hyperaccumulateur de cadmium [5], *Solanum nigrum*. Elle est étudiée dans le contexte de la phytoremédiation [6] des métaux lourds, en raison de sa capacité à favoriser la croissance des plantes et à résister aux métaux lourds (ZHANG et al., 2022).

— La souche *Bacillus tropicus* JMT105-2, issue des tortues à carapace molle de Chine (*Pelodiscus sinensis*), est associée à une maladie provoquant des symptômes similaires à la maladie du charbon chez ces reptiles. Elle possède un plasmide comportant les gènes *pagA* et *cya* (mais pas *lef*). Les opérons Has et Bps sont également présents dans son génome. Une capsule est produite (TSAI, KUO et CHENG, 2023).

— La souche *Bacillus sp.* SYJ15 est associée à une épidémie touchant des tortues à carapace molle de Chine (*Pelodiscus sinensis*). Elle possède un plasmide comportant les gènes *pagA* et *cya* (mais pas *lef*). L'opéron Bps est aussi présent (YUAN et al., 2020).

— La souche *Bacillus shihchuchen* QF108-045 est associée à une épidémie touchant la même espèce animale. Elle contient un mégaplasmide, très proche de celui de la souche *Bacillus sp.* SYJ15, comportant les gènes *pagA* et *cya* (CHENG et al., 2023).

5. Plante capable d'absorber et d'accumuler dans ses tissus des concentrations de cadmium beaucoup plus élevées que celles généralement trouvées dans la plupart des plantes.

6. Technique de dépollution utilisant des plantes et leurs racines pour absorber, séquestrer ou dégrader des polluants, tels que les métaux lourds, les pesticides ou les composés organiques, présents dans le sol ou l'eau.

— La souche *B. thuringiensis* Bt185, étudiée pour son potentiel en tant qu'agent de biocontrôle d'insectes du sol, possède huit plasmides. L'un d'eux renferme trois gènes *cry8*. Cette caractéristique fait de Bt185 un candidat prometteur pour le contrôle biologique d'insectes nuisibles dans l'agriculture, notamment contre des espèces comme *Holotrichia parallela*, *Holotrichia oblita* et *Anomala corpulenta* (LI et al., 2017).

La tortue à carapace molle chinoise prospère dans les environnements aquatiques. Taïwan, notamment dans les régions de Kaohsiung et Pingtung, est un centre majeur de son élevage en étang. Le climat de cette région permet une longue saison de reproduction, favorisant l'exportation des œufs vers la Chine. Toutefois, des défis tels que la surpopulation des bassins et la qualité médiocre de l'eau ont conduit à une augmentation des épidémies, provoquées particulièrement par des bactéries du groupe *B. cereus*. En 2011, des infections par des souches du groupe *B. cereus* chez les tortues à carapace molle chinoises ont été signalées pour la première fois dans les fermes aquacoles de la province du Guangdong, en Chine (TAN et al., 2011). Par la suite, Taïwan a connu des incidents similaires. En outre, des souches de *B. thuringiensis* ont été identifiés chez ces tortues (CHEN et al., 2014).

Nom de souche	Date de collection	Localisation	Hôte / Source	Taille du plasmide	*Accession number*	Virulence
Bacillus sp. PGP15	2016	Chine	Rhizosphère	GCA_023101325.1	525,898	Inconnue
Bacillus sp. SYJ15	2015	Chine (Zhejiang Province, Huzhou)	*Pelodiscus sinensis*	GCA_004322655.1	218,649	Septicémie
B. shihchuchen QF108-045	2019	Taïwan (Pingtung County)	*Pelodiscus sinensis*	GCA_030272635.1	240,242	Septicémie
B. tropicus JMT105-2	2016	Taïwan	*Tryonyx sinensis*	GCA_027626015.1	229,626	Septicémie
B. thuringiensis Bt185	2002	Chine (Heibei Province)	Sol	GCA_001595725.1	635,508	Insecticide

TABLE 4.2 – Tableau récapitulatif des souches étudiées.

BTYPER3 (CARROLL, CHENG et KOVAC, 2020) a été utilisé pour détecter les gènes de virulence propres au groupe *B. cereus* dans le génome de ces souches. Leur classification taxonomique a été réalisée selon la nomenclature proposée par CARROLL, WIEDMANN et KOVAC, 2020. Les résultats sont résumés dans le tableau 4.3.

Nom de souche	*nheABC*	*hblABCD*	*cesABCD*	*cytK-1*	*cytK-2*	*sph*	Toxines Bt	*cya, lef, pagA*	*capABCDE*	*hasABC*	*bpsABCDEFGHX*	Classification taxonomique
Bacillus sp. PGP15	+	+	-	-	+	+	-	-	-	-	(+)	*B. mosaicus*
Bacillus sp. SYJ15	+	+	-	-	-	+	-	(+)	-	(+)	+	*B. mosaicus* biovar Anthracis
B. shihchuchen QF108-045	+	(+)	-	-	-	+	-	(+)	-	(+)	(+)	*B. mosaicus* biovar Anthracis
B. tropicus JMT105-2	+	+	-	-	-	+	-	(+)	-	(+)	+	*B. mosaicus* biovar Anthracis
B. thuringiensis Bt185	+	+	-	-	+	+	+	-	+	+	(+)	*B. cereus s.s.* biovar Thuringiensis

TABLE 4.3 – Tableau récapitulatif des gènes de virulence des souches étudiées.

Les mégaplasmides de l'ensemble de ces souches ont été comparés à l'aide du logiciel BRIG. Cette comparaison est illustrée sur la figure 4.11. Les plasmides comparés présentent peu de régions communes avec le plasmide p1 de *Bacillus sp.* PGP15. Néanmoins, tous possèdent la région comportant le miniréplicon observé précédemment, y compris la souche *B. thuringiensis* Bt185 collectée bien avant les autres. Ces observations sont cohérentes avec l'hypothèse d'une émergence d'un mutant particulier au sein de cette niche écologique, qui a suivi une évolution similaire à *B. anthracis* dans son écotype, en s'adaptant aux spécificités de son environnement. La détection du miniréplicon nouvellement référencé et, de manière générale, des minirépliconscommuns aux plasmides de *B. anthracis* ou des souches du groupe *B. cereus* étant biovar Anthracis (ou anthracis-like) peut ainsi être un indicateur d'une évolution vers des souches pathogènes dans un environnement donné.

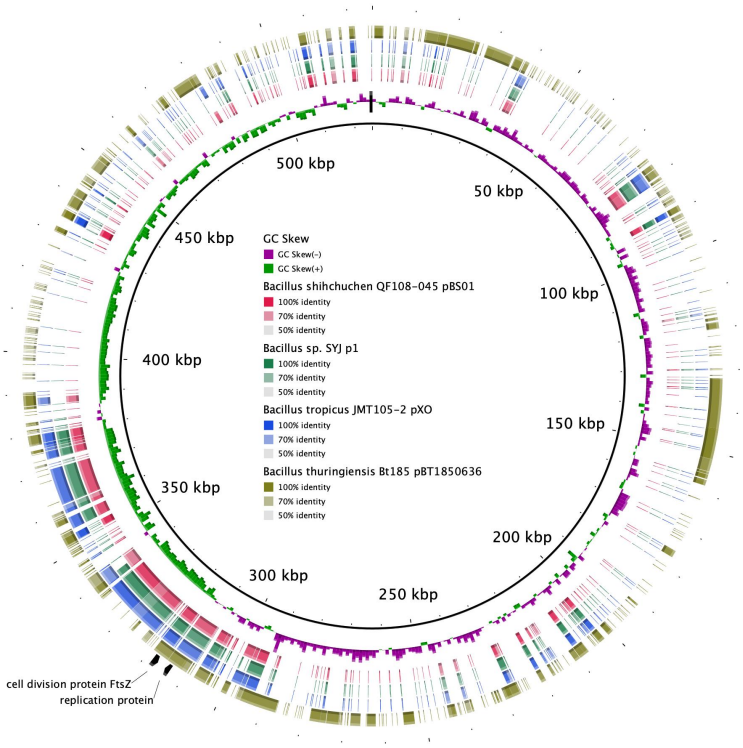

FIGURE 4.11 – Comparaison du plasmide de la souche *Bacillus sp.* PGP15 avec les plasmides des souches *B. shihchuchen* QF108-045, *Bacillus sp.* SYJ15, *B. tropicus* JMT105-2 et *B. thuringiensis* Bt185. Anneaux du centre vers l'extérieur : (1) Numérotation des nucléotides ; (2) déviation GC ; (3-6) pourcentage d'identité de séquence avec les autres plasmides selon le code couleur défini sur la figure ; (7) gènes codant des miniréplicons (en noir).

IDÉES À RETENIR

1. En tirant profit du proche voisinage de *B. anthracis* nouvellement enrichi, et en étudiant les effets des pressions de sélection et d'homoplasie sur la phylogénie de l'espèce, nous avons confirmé l'émergence clonale récente de *B. anthracis* au sein du groupe *B. cereus*.
2. En particulier, nous avons établi la position de l'ancêtre de *B. anthracis* le long de la branche liant l'espèce à ses proches voisins, confirmant et complétant une précédente approche purement plasmidique relative à pXO1.
3. Cela confirme *a minima* la coévolution du chromosome de *B. anthracis* avec pXO1. Un travail similaire, plus qualitatif du fait du faible nombre de séquences disponibles relatives aux souches anthracis-like, suggérerait une absence de co-évolution entre les plasmides pBCXO1 et pBCXO2 du lignage anthracis-like ST78.
4. Un nouveau type de miniréplicon a été détecté au sein des plasmides des proches voisins de *B. anthracis*. La présence simultanée de plusieurs types de miniréplicons au sein de ces plasmides conforte l'hypothèse d'une formation de ces derniers (et de pXO1 et pXO2 *a fortiori*) par l'intégration de plusieurs plasmides (ou tronçons de plasmides) de moindre taille.
5. Si l'absence de région relative à des systèmes de sécrétion conjugatifs de type IV (T4SS) dans les plasmides des proches voisins de *B. anthracis* (eux-mêmes contenant des tronçons communs avec pXO1) a été constatée, le phénomène inverse a déjà été relaté pour pXO2 et des plasmides partageant des régions communes avec ce plasmide. Cela pourrait suggérer des mécanismes différents de formation de pXO1 et pXO2, idée renforcée par l'existence de souches anthracis-like comprenant un plasmide similaire au premier mais pas au second.
6. En somme, un modèle cohérent serait que *B. anthracis* émerge d'un ancêtre commun avirulent du groupe *B. cereus*, ayant acquis les plasmides de virulence. Ces plasmides de virulence se seront préalablement constitués par fusion et/ou transfert horizontal successif de plasmides de plus petite taille dont certains comportant les îlots de pathogénicité qui leurs sont propres. Le même phénomène pourrait expliquer l'émergence des souches anthracis-like, dont la différence principale avec *B. anthracis* serait qu'ils ne se sont pas adaptés à une nouvelle niche écologique en dehors de la forêt Centre-Africaine.

Chapitre 5

À la recherche de *Bacillus anthracis* dans l'environnement

Ce chapitre présente le développement d'un outil de détection métagénomique spécifique à *B. anthracis*, une démarche justifiée par les découvertes antérieures démontrant que l'émergence de cette espèce résulte d'un événement clonal récent. L'analyse indique que *B. anthracis* trouve son origine récente dans une localisation géographique spécifique, où l'ancêtre de l'espèce a été identifié. Cette spécificité géographique et cette origine récente prédisent qu'il n'existe pas de proches précurseurs de *B. anthracis* de façon ubiquitaire sur la planète. Une approche de type "métagénomique" pourrait permettre de tester cette prédiction. En se focalisant sur des zones où l'ancêtre est susceptible d'avoir existé, cet outil ambitionne de contribuer à affiner notre compréhension du modèle d'émergence de *B. anthracis*, en explorant la diversité génétique au sein de son habitat d'origine. Au-delà de cet objectif, un tel outil permettra une détection rapide et fiable de traces de *B. anthracis* dans des échantillons environnementaux, répondant ainsi à un déficit existant.

OBJECTIFS DU CHAPITRE

1. Développer un outil de détection de traces de *B. anthracis* dans un échantillon environnemental, fiable et rapide
2. Développer des données de *benchmarking* permettant d'évaluer les performances de cet outil
3. Utiliser cet outil dans le cadre d'une recherche en routine de *B. anthracis* dans des métagénomes publics
4. Élargir l'utilisation de cet outil pour la recherche d'autres agents pathogènes (*Yersinia pestis*, *Francisella tularensis*, *Brucella*)

5.1 Détection de *Bacillus anthracis* dans un échantillon métagénomique

L'essor du séquençage de nouvelle génération a permis des avancées significatives en métagénomique, améliorant la détection de pathogènes tels que *B. anthracis*. Ce bond en avant, de l'analyse ciblée à l'exploration génomique globale, est mis en lumière dans la section suivante. Elle examine l'évolution de la métagénomique, soulignant l'apport du séquençage massif pour analyser la diversité génétique des échantillons complexes. Sont ensuite décrits les outils d'analyse métagénomique actuels, notamment ceux ciblant *B. anthracis*. Bien que ces outils représentent une avancée, en comparaison avec certaines méthodes conventionnelles décrites précédemment, ils soulèvent également des défis, notamment la distinction difficile entre *B. anthracis* et les espèces génétiquement proches du groupe *B. cereus*, sources de faux positifs.

Avant-propos Le lecteur intéressé par l'historique des différentes méthodes de séquençage nouvelle génération peut se référer à l'annexe A.

5.1.1 Définition générale de la métagénomique

La mise en place du séquençage haut débit et les progrès en capacités de calcul ont permis d'étudier directement la biodiversité présente dans des échantillons environnementaux. Ce concept, appelé analyse métagénomique, consiste à extraire l'ADN d'échantillons biologiques et à en réaliser une analyse pour en déterminer les micro-organismes présents, y compris ceux difficilement cultivables en laboratoire.

Deux types de données métagénomiques peuvent être analysées : ciblée et globale. Dans le premier cas, seuls des segments spécifiques des génomes, appelés "marqueurs phylogénétiques", identifiés comme caractéristiques d'un certain type d'organismes ou associés à un ensemble de fonctions, sont ciblés pour l'amplification puis le séquençage. Ces derniers sont choisis car ils comprennent des zones à la fois hautement conservées et très variables. Les zones bien conservées facilitent la conception d'amorces PCR universelles pour le séquençage du génome, tandis que les zones variables permettent d'identifier distinctement chaque taxon [1]. Parmi les marqueurs existants, l'ARNr 16S fait figure de *gold standard* pour l'identification des bactéries. Après un séquençage ciblé, on obtient des amplicons, dénommés ainsi car résultant d'une amplification. *A contrario* la métagénomique globale ou *shotgun* implique le séquençage de l'intégralité de l'ADN présent dans un échantillon sans ciblage préalable. L'ensemble de l'ADN est fragmenté puis analysé en utilisant des méthodes de séquençage haut-débit conventionnelles. Ceci permet le séquençage complet d'une fraction aléatoire du matériel génétique présent dans l'échantillon.

La métagénomique ciblée vise principalement à évaluer la diversité taxonomique d'un échantillon donné. En séquençant seulement une portion restreinte du génome ou un taxon particulier (ou plusieurs en cas d'utilisation du ciblage par capture avec de multiples sondes, disposées sur une puce par exemple), cette méthode est financièrement plus avantageuse que la métagénomique globale. De plus, elle offre l'avantage d'identifier des organismes moins courants avec le même niveau d'effort

1. Unité de classification utilisée en biologie pour regrouper des êtres vivants partageant des caractéristiques communes.

de séquençage. Souvent, la métagénomique globale capture uniquement une partie de la communauté microbienne existante. À titre d'exemple, pour obtenir un ensemble de données complet de la microfaune d'un gramme de terre, le coût dans l'état actuel des technologies de séquençage pourrait s'élever à plusieurs centaines de millions de dollars (DESAI et al., 2012). Cependant, avec l'évolution constante des techniques de séquençage, cette approche devient de plus en plus accessible. Ceci a donné lieu à la multiplication de vastes initiatives, comme le projet METASOIL (DELMONT et al., 2011) axé sur le microbiome du sol, ou le HUMAN MICROBIOME PROJECT (TURNBAUGH et al., 2007), centré sur l'étude des microbiotes humains.

L'efficacité de l'identification bactérienne dépend étroitement de l'exhaustivité des espèces représentées dans les bases de données de référence existantes. Ces bases de données sont élaborées à partir de l'assemblage de données de séquençage. Pour la méthode ciblée, l'assemblage nécessite peu de lectures de séquençage car seuls les gènes ciblés doivent être assemblés. En revanche, la métagénomique globale nécessite l'assemblage de génomes entiers, ce qui exige une quantité plus importante de séquences et d'analyses informatiques. De ce fait, les bases de données basées sur le séquençage ciblé incluent généralement un spectre plus large d'espèces par rapport aux bases de données de séquençage *shotgun*.

Cependant, il y a une différence de résolution taxonomique entre ces deux méthodes d'analyse. La méthode ciblée, ciblant par exemple un gène spécifique aux bactéries, se restreint à l'identification de celles-ci. À l'inverse, la méthode globale est apte à identifier des organismes appartenant à différents domaines taxonomiques, incluant les bactéries, les eucaryotes et les archées et peut atteindre une résolution au niveau de la souche. La méthode ciblée aboutit généralement à une identification au niveau du genre. Toutefois, l'utilisation conjointe de méthodes de correction d'erreurs, d'amorces spécifiques et d'une base de données de référence optimisée permet parfois une identification jusqu'au niveau de l'espèce.

Enfin, il est à retenir qu'une méthode n'a pas d'avantage absolu sur l'autre et son utilisation dépend surtout des objectifs de l'étude. Par ailleurs, les approches métagénomiques, qu'elles soient globales ou ciblées, ne sont pas mutuellement exclusives et peuvent être utilisées de manière complémentaire.

5.1.2 Premiers développements d'outils de détection métagénomique

Un premier pas vers la métagénomique : BLAST Pour de nombreux échantillons métagénomiques, les espèces, les genres (*genus*) et même certains *phyla* présents dans l'échantillon sont largement inconnus au moment du séquençage, et le but du séquençage est de déterminer cette composition microbienne aussi précisément que possible. Bien sûr, si un organisme est complètement différent de tout ce qui a été vu précédemment, alors sa séquence d'ADN doit être étiquetée comme nouvelle. De nombreuses espèces, cependant, ont une certaine similitude détectable avec une espèce connue, et cette similitude peut être détectée par un algorithme d'alignement sensible. Le plus connu de ces algorithmes est le programme BLAST, ou *Basic Local Alignment Search Tool* (KORF, YANDELL et BEDELL, 2003), qui peut classer une séquence en trouvant le meilleur alignement avec une grande base de données de séquences génomiques.

Même si ce logiciel n'est pas initialement conçu pour des analyses métagénomiques car très lent pour de grandes bases de données, il a été utilisé en première approche pour cela. L'approche suivie par BLAST se décompose en trois étapes

pour trouver des alignements entre de longues séquences données en entrée (typiquement de la taille d'un gène) et des grandes bases de données (typiquement des génomes) :

— les régions à faible complexité sont supprimées de la séquence donnée en entrée.

— l'entrée est découpée en petites sous-séquences d'ADN ou mots de 11 bases par défaut et la base de données du logiciel est interrogée pour trouver les matchs exacts.

— Un alignement de type Smith-Waterman (PEARSON, 1991) est réalisé pour déterminer un *High-scoring segment pair*, ou HSP. Ce dernier consiste en un couple de fragments d'ADN reconnus sur chacune des séquences comparées, de même longueur, et qui possède un score significatif. Un HSP correspond à un segment commun entre deux séquences comparées, le plus long possible, autrement dit une similitude sans insertion-délétion ayant au moins un score supérieur ou égal à un score seuil.

La signification statistique des alignements BLAST est évaluée en utilisant la *E-value*, espérance mathématique.

Premières adaptations de BLAST à la métagénomique Afin d'améliorer l'efficacité de BLAST, qui demeure un "simple" logiciel d'alignement non conçu à l'origine pour les analyses métagénomiques, d'autres méthodes de classification de séquences ont été élaborées, combinant des méthodes d'alignement et de *machine learning*. En premier lieu, avec le programme MEGAN (BAĞCI et al., 2019), une séquence donnée est analysée à travers plusieurs bases de données (sur le principe de BLAST). Suite à cela, le plus petit ancêtre commun (*Lowest Common Ancestor* ou LCA) des meilleurs matchs entre la séquence et chacune des bases de données est assigné à la séquence. Le programme PHYMMBL (BRADY et SALZBERG, 2011) combine les résultats de BLAST et des scores produits par des chaînes de Markov interpolées. Une classification naïve bayésienne [2] est appliquée aux k-mers [3] pour les associer à une région d'un génome. Ces programmes améliorent l'efficacité de BLAST seul mais sont plus lents. Pour pallier ce problème, une autre approche est privilégiée : l'estimation d'abondance.

Une alternative : l'estimation d'abondance Les programmes d'estimation d'abondance consistent tout d'abord à créer un ensemble de bases de données "compressées", plus petit que l'ensemble de tous les génomes utilisés par BLAST. Ces bases de données contiennent des marqueurs génétiques permettant de discriminer certaines espèces ou certains clades. Du fait de la précision moindre de ces bases de données, ce type de programme ne peut classer qu'une faible partie des séquences issues d'un métagénome, à défaut de classer toutes les lectures comme expliqué précédemment. Une estimation d'abondance permet donc de caractériser rapidement de manière globale un métagénome mais n'est pas adaptée dans le cas d'une analyse fine des espèces pouvant le constituer. Par exemple, GRAMMY (XIA et al., 2011), METAPHLAN (SEGATA et al., 2012), GASIC (LINDNER et RENARD, 2013), TAEC (SOHN et al., 2014), CONSTRAINS (LUO et al., 2015) et BRACKEN (LU et al., 2017) se basent sur l'estimation d'abondance.

2. Classification probabiliste se basant sur le théorème de Bayes
3. Ensemble des sous-séquences de longueur k contenues dans la séquence étudiée

5.1.3 Outils de détection métagénomique actuels

Actuellement, plusieurs logiciels existent, chacun usant de méthodes différentes soit pour déterminer la composition globale d'un échantillon, soit pour effectuer une assignation taxonomique de chaque lecture du séquençage étudié. Par exemple, le *mapping* des lectures complètes ou de k-mers, l'utilisation de génomes complets ou seulement de marqueurs génétiques ou encore la traduction de l'ADN en séquences protéiques suivie de leur alignement (BREITWIESER, LU et SALZBERG, 2019). Cette section présente quelques-uns de ces outils, pour illustrer la diversité des méthodes employées.

— KRAKEN2 : logiciel de classification métagénomique rapide et précis qui utilise une bibliothèque de k-mers pour attribuer des séquences à des taxons, en les comparant à une base de données de références génomiques (WOOD, LU et LANGMEAD, 2019).

— METAPHLAN 4 : outil de profilage taxonomique de données métagénomiques qui utilise une base de données étendue, comprenant génomes de référence et assemblages métagénomiques, capable d'identifier des espèces encore non référencées (BLANCO-MÍGUEZ et al., 2023).

— KRAKENUNIQ : classificateur métagénomique, adapté de la première version de KRAKEN (WOOD et SALZBERG, 2014) qui se distingue par son utilisation unique du comptage des k-mers pour des résultats plus spécifiques, c'est-à-dire réduisant le nombre de faux positifs détectés (BREITWIESER, BAKER et SALZBERG, 2018).

— KAIJU : outil destiné à la classification taxonomique rapide et sensible de séquençages métagénomiques ou métatranscriptomiques, en assignant chaque séquence lue à un taxon par comparaison avec une base de données de référence de protéines microbiennes et virales (MENZEL, NG et KROGH, 2016).

— CORE-KAIJU : classificateur métagénomique se basant non sur l'ARNr 16S mais sur la détection de familles de domaines protéiques (essentiellement ribosomiques) pour une estimation fiable du nombre de taxons et de leur abondance dans un échantillon (TOVO et al., 2020).

— CENTRIFUGE : outil permettant de classifier rapidement des séquençages métagénomiques de grande taille en utilisant une indexation compressé basée sur la transformée de Burrows-Wheeler [4] et le FM-index [5] (KIM et al., 2016). Cet outil a été récemment adapté et optimisé pour réduire le nombre de faux positifs (LIU et al., 2023).

Des études ont été réalisées pour comparer les performances de ces différents outils selon des critères de rapidité, de mémoire requise, de sensibilité et de spécificité (MCINTYRE et al., 2017 ; SIMON et al., 2019). Il est à noter que ces comparatifs sont réalisés en testant les outils sur un nombre limité de données simulées et ne sont donc pas des avis absolus. Il s'avère que certains outils sont plus adaptés pour réduire le nombre de faux positifs, d'autres pour une identification plus précise (à la

4. Technique de transformation de texte utilisée principalement dans la compression de données. Elle réarrange les caractères d'une chaîne de texte de manière à regrouper les symboles similaires, facilitant ainsi leur compression ultérieure. Cette méthode est souvent employée dans les algorithmes de compression de fichiers et dans le domaine de la bio-informatique pour l'indexation efficace des séquences d'ADN.

5. Structure de données utilisée pour la recherche de sous-chaînes dans un texte. Basée sur la transformée de Burrows-Wheeler, elle permet une recherche rapide de motifs dans de grands ensembles de données textuelles, comme les séquences génomiques.

souche près plutôt qu'à l'espèce près ou au genre près) ou encore plus efficaces selon le type de séquençage utilisé (*short reads* ou *long reads*). Ces comparaisons doivent être considérées à la lumière de notre objectif de développer un outil capable de détecter avec fiabilité la présence de *B. anthracis* (et de le différencier d'un proche voisin ou d'une souche *anthracis-like* du groupe *B. cereus*). À ce titre, KRAKEN2, du fait de son analyse (relativement rapide) des k-mers et de la possibilité de créer des bases de données de référence à façon, semble un logiciel adéquat sur lequel baser notre étude. La section suivante détaille son principe de fonctionnement.

5.1.4 Présentation du logiciel KRAKEN

Principe de fonctionnement Pour analyser un échantillon métagénomique à l'aide de KRAKEN, l'utilisateur fournit en entrée un ensemble de séquences d'ADN, sous format FASTQ. Pour classer une séquence donnée, KRAKEN produit l'ensemble des k-mers correspondants. Puis, KRAKEN assigne chaque k-mer au plus petit ancêtre commun de tous les génomes contenant exactement le k-mer étudié. L'ensemble des taxa contenant au minimum un k-mer forme ainsi un sous-ensemble de l'arbre phylogénétique du vivant, appelé "arbre de classification". À chacun des noeuds de ce sous-arbre est associé un poids correspondant au nombre de k-mers assignés au taxon en question. Ensuite, à chaque chemin de l'arbre de classification est associé un score égal à la somme des nombres de k-mers associés à tous les taxa composant ce chemin. En définitive, la séquence considérée est assignée à la feuille du chemin obtenant le plus haut score. En cas d'égalité, elle est assignée au plus petit ancêtre commun des feuilles. La séquence est déclarée "*unclassified*" (non-classée) si aucun k-mer n'a pu être associé à un taxon (WOOD et SALZBERG, 2014). Ce processus est schématisé à la figure 5.1.

Création des bases de données de référence Lors de l'assignation taxonomique des k-mers d'une séquence ADN donnée, KRAKEN utilise une base de données interne. Celle-ci a été établie à partir des génomes de référence complets extraits de la librairie *RefSeq* du NCBI [6], incluant bactéries, virus, archées et génome humain. Pour chaque génome de référence, l'ensemble des k-mers le composant (par défaut : $k = 31$) est extrait. Un programme *multithread* de comparaison de k-mers, *Jellyfish* (MARÇAIS et KINGSFORD, 2011), est utilisé pour éliminer les éventuels doublons. Ensuite, la référence taxonomique (*taxonomic ID*) du plus petit ancêtre commun de toutes les espèces est affectée à chaque k-mer. Les références taxonomiques issues de la base de données du NCBI sont utilisées.

Pour optimiser le temps d'assignation taxonomique de chaque k-mer, KRAKEN utilise le concept de minimiseur (ROBERTS et al., 2004). En effet, comparer les nucléotides de deux séquences d'ADN est très consommateur en temps et mémoire. De manière générale, pour résoudre ce problème, on use de la méthode *seed-and-extend* qui consiste à utiliser des sous-séquences courtes appelées *seeds* pour rechercher des correspondances potentiellement plus longues. Dès qu'une correspondance est trouvée entre deux *seeds* de deux séquences comparées, on observe de proche en proche si cette correspondance peut s'étendre à une plus longue séquence. Pour que cette approche soit efficace, il faut donc stocker une grande partie, voire la totalité de l'ensemble des *seeds* possibles des séquences comparées.

Cependant, une aussi grande quantité de données ne peut facilement être stockée dans la mémoire RAM d'un ordinateur. À titre de comparaison, le séquençage du

6. *National Center for Biotechnology Information*

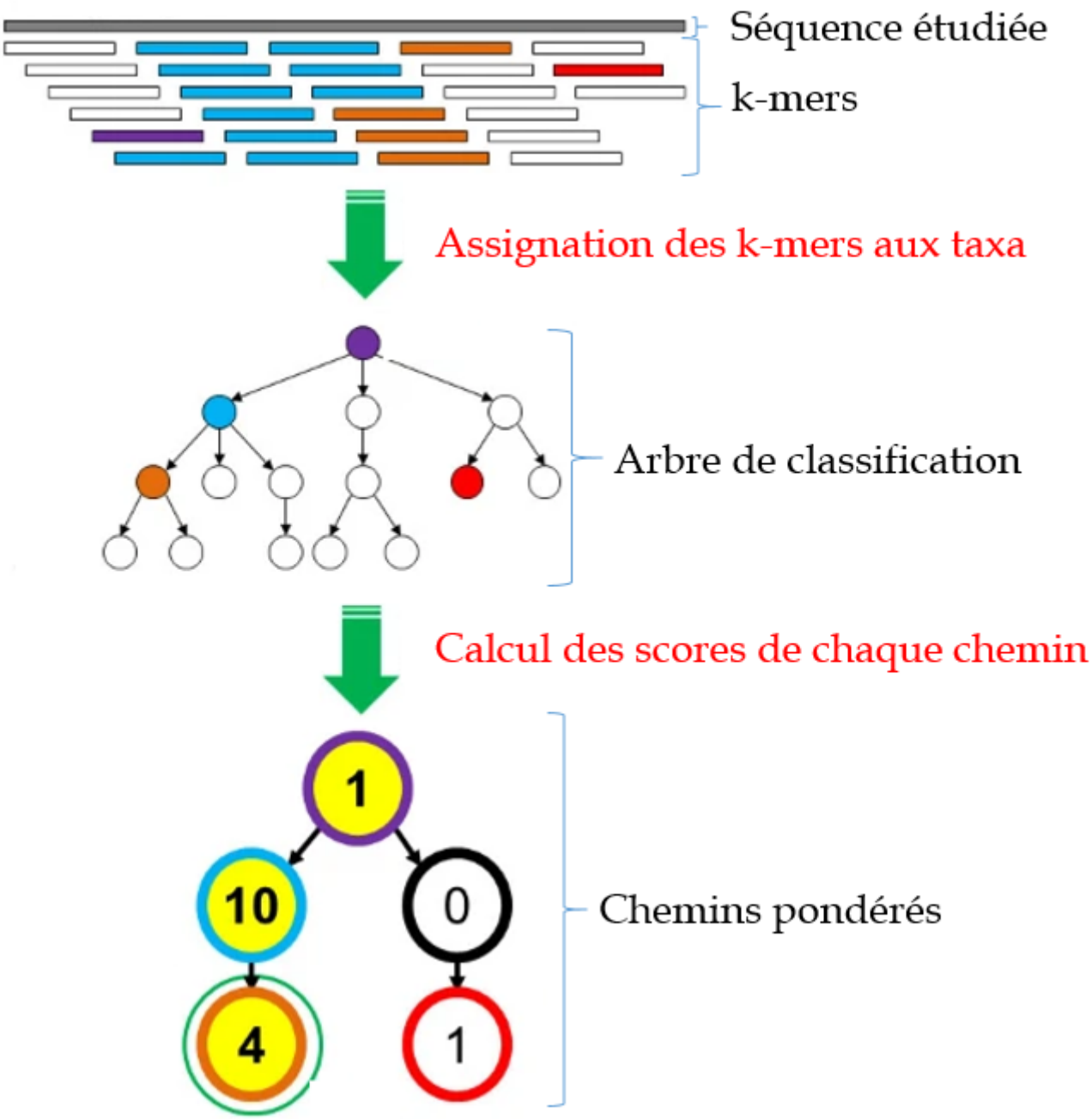

FIGURE 5.1 – Schéma récapitulatif du fonctionnement de KRAKEN. *Dans l'exemple décrit, la séquence étudiée est assignée au taxon orange, entouré en vert.* Adapté de WOOD et SALZBERG, 2014.

génome de référence de *Rattus norvegicus*[7] requiert environ 33×10^6 lectures de 600 nucléotides. Cela donne un total d'environ 2×10^{10} k-mers (avec ici $k = 20$), soit 4×10^{11} lettres. Sachant qu'une lettre est encodée en deux bits, un octet est nécessaire pour stocker quatre lettres. De plus, pour chaque k-mer, il faut stocker la position où il apparaît dans le génome. On arrive donc à un total de 5 octets par k-mer. En définitive, environ 200 GB sont nécessaires pour stocker la base de données de k-mers relative à *Rattus norvegicus*. Il faut donc pouvoir ne stocker qu'une partie des

7. plus communément appelé *rat brun*

seeds, appelés minimiseurs [8], pour mettre en oeuvre l'approche *seed-and-extend* dans le cas de comparaison de grandes séquences.

Pour revenir au cas de KRAKEN, ce ne sont pas de longues séquences qui seront comparées, mais des petits k-mers (avec ici $k = 31$), et l'approche *seed-and-extend* permet d'optimiser l'assignation taxonomique. KRAKEN interroge sa base de données interne pour l'assignation de chaque k-mer et procède de proche en proche. Un k-mer est assigné immédiatement après celui qui lui est adjacent. Pour comprendre l'utilisation des minimiseurs, définissons la représentation canonique d'une séquence d'ADN, disons S, comme la plus petite séquence entre S et le complémentaire de S au sens de l'ordre lexicographique [9]. Déterminer le minimiseur de longueur L d'un k-mer revient à déterminer la plus petite représentation canonique (au sens de l'ordre lexicographique) de tous les L-mers contenus dans le k-mer. Dans la base de données de référence de KRAKEN, tous les k-mers ayant le même minimiseur sont stockés consécutivement. Lors d'une requête dans la base de données de KRAKEN pour un k-mer donné, le logiciel observe les positions des k-mers possédant le même minimiseur que le k-mer donné en entrée puis procède à une recherche dichotomique parmi eux pour repérer le k-mer adéquat. Deux k-mers adjacents ont de manière quasi-systématique le même minimiseur. En enregistrant dans la mémoire cache les données relatives à l'assignation d'un k-mer, on peut procéder à l'assignation du k-mer adjacent de manière beaucoup plus rapide. À titre d'ordre de grandeur, le tableau d'indices des minimiseurs, permettant de répertorier les différents minimiseurs et leurs positions dans la base de données, requiert 8×4^L avec L la longueur des minimiseurs. Par défaut, $L = 15$, ce qui revient à une taille de base de données de 70 GB (contre plus de 600 GB initialement). Pour les utilisateurs ayant moins d'espace disponible, une base "réduite" de moins de 4 GB existe (avec $L = 13$). Ce principe de constitution de bases de données est illustré sur la figure 5.2.

Création de bases de données personnalisées : les *custom databases* Dans la section précédente, nous avons décrit le principe de la base de données de référence de KRAKEN, incluant l'ensemble des génomes de référence du vivant (à quelques exceptions près). Cependant, dans certains cas, seule la détection de certains agents pathogènes d'intérêt est visée, plutôt qu'une analyse complète des espèces contenues dans l'échantillon. Dans ce contexte, une précision élevée dans la reconnaissance de ces agents d'intérêt est requise, ce qui n'est pas le cas en utilisant la base de données par défaut fournie par KRAKEN. En effet, des faux positifs [10] (ou *a contrario* des faux négatifs) peuvent être observés lors de l'analyse de métagénomes.

Concrètement, en sortie d'analyse par KRAKEN, un fichier texte détaillant l'assignation taxonomique de chaque lecture est fourni. Un exemple est donné à la figure 5.3.

Cinq colonnes sont présentes sur ce type de fichier, chaque ligne présentant les résultats d'une lecture.

— La première colonne contient la lettre *C* pour *Classified* ou *U* pour *Unclassified* selon que la lecture en question a été classée ou non par KRAKEN.

8. *minimizers*

9. L'ordre lexicographique est un principe de classement des éléments (mots, nombres, etc.) basé sur l'ordre alphabétique des lettres ou des chiffres qui les composent. Dans cet ordre, les éléments sont comparés caractère par caractère, de gauche à droite, jusqu'à ce qu'une différence soit trouvée. Si un élément est préfixe d'un autre, il est considéré comme inférieur. Ce système est couramment utilisé dans les dictionnaires pour trier les mots, ainsi que dans les systèmes informatiques pour trier des chaînes de caractères ou des séquences numériques.

10. Présence d'une espèce détectée par KRAKEN qui n'est en réalité pas contenue dans l'échantillon

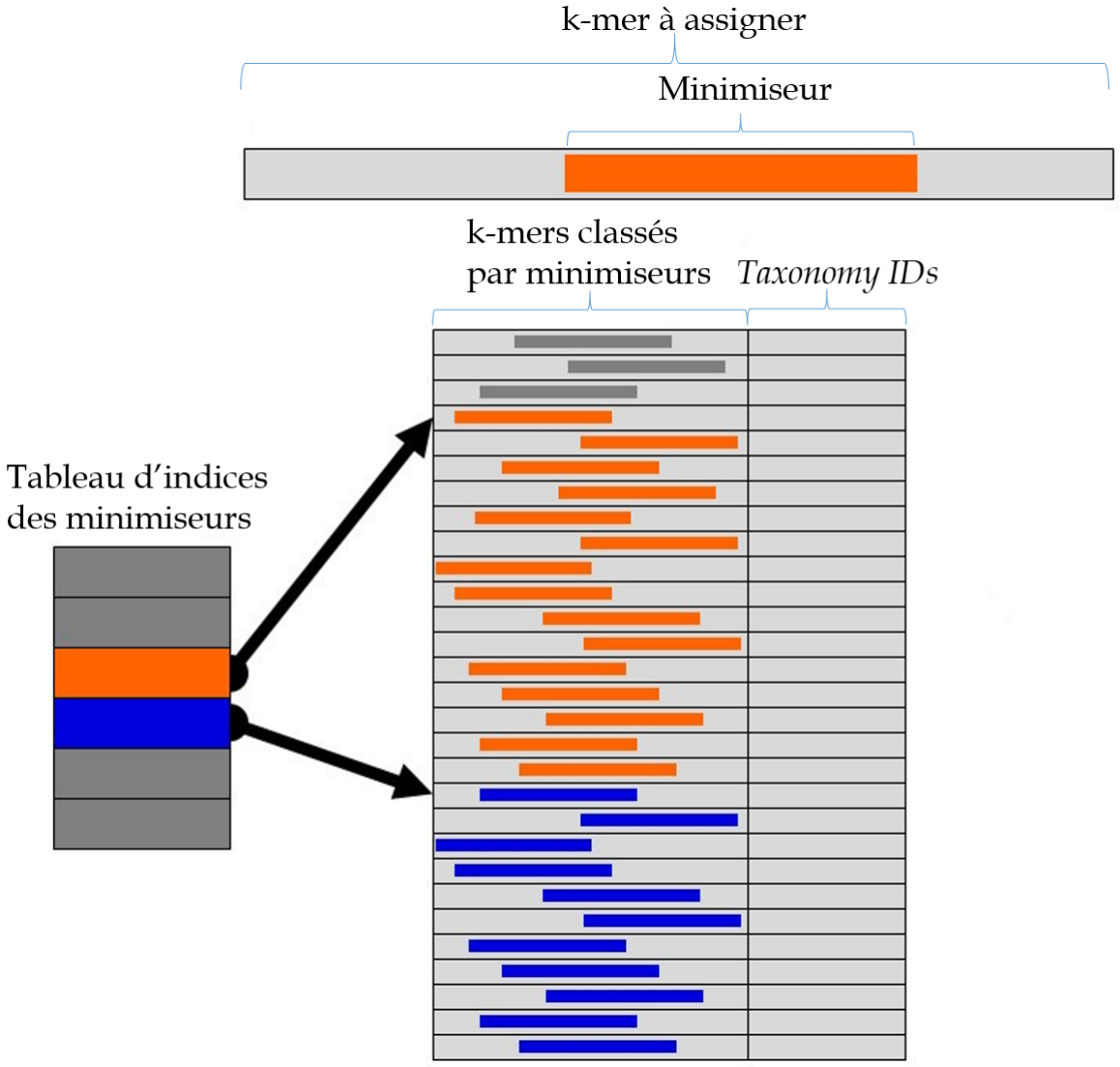

FIGURE 5.2 – Structure de la base de données de référence de KRAKEN. Adapté de WOOD et SALZBERG, 2014.

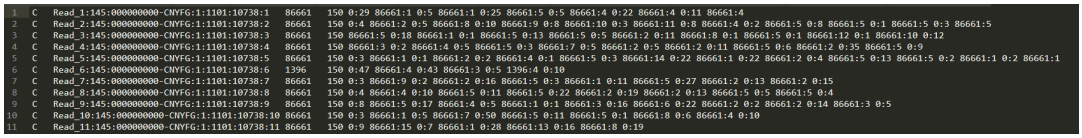

FIGURE 5.3 – Exemple de fichier de sortie de KRAKEN. *Une ligne correspond à une lecture. Première colonne : C pour Classified ou U pour Unclassifiedsd. Deuxième colonne : nom de lecture. Troisième colonne : numéro taxonomique assigné. Quatrième colonne : longueur en paires de bases. Cinquième colonne : nombre de k-mers assignés par taxon.*

— La deuxième colonne correspond au nom (ID) de la lecture étudiée.

— La troisième colonne contient le numéro taxonomique auquel la lecture a été assignée (dans le cas d'une lecture non-classée, KRAKEN assigne par défaut la valeur 0).

— La quatrième colonne correspond à la longueur de la lecture en nombre de paires de bases.

— La cinquième colonne contient une liste détaillant le nombre de k-mers assignés par taxon. Par exemple, la première partie de la liste de la première ligne sur la figure 5.3 *"0 :29 86661 :1 0 :5 86661 :1 0 :25"* signifie que les 29 premiers k-mers ont été non-classés, le suivant a été assigné au taxon 86661, les cinq suivants ont été non-classés et ainsi de suite.

Sur demande de l'utilisateur, KRAKEN fournit un fichier au format *.report* qui est plus explicite. Celui-ci détaille non pas les résultats lecture par lecture, comme le fichier de sortie par défaut, mais taxon par taxon. Un exemple est donné à la figure 5.4.

```
 0.26  2853  2853  U 0 unclassified
99.74  1075197 914 R 1 root
99.65  1074283 3 R1  131567    cellular organisms
99.65  1074277 3333  D 2     Bacteria
99.34  1070914 54  D1  1783272       Terrabacteria group
99.33  1070860 669 P 1239          Firmicutes
99.27  1070191 469 C 91061           Bacilli
99.23  1069722 516 O 1385              Bacillales
99.17  1069148 183 F 186817                Bacillaceae
99.16  1068965 1431  G 1386                  Bacillus
99.02  1067534 913040  G1  86661                   Bacillus cereus group
12.38  133465  132579  S 1392                    Bacillus anthracis
 0.02  231 231 S1  768494                      Bacillus anthracis str. H9401
 0.02  205 205 S1  1452727                     Bacillus anthracis str. Turkey32
 0.01  136 136 S1  260799                      Bacillus anthracis str. Sterne
 0.01  77  77  S1  1412843                     Bacillus anthracis 8903-G
 0.01  60  60  S1  673518                      Bacillus anthracis str. A16R
 0.01  57  57  S1  1412842                     Bacillus anthracis 9080-G
 0.00  53  53  S1  261591                      Bacillus anthracis str. Vollum
 0.00  40  40  S1  1412844                     Bacillus anthracis 52-G
 0.00  21  21  S1  1449979                     Bacillus anthracis str. V770-NP-1R
 0.00  6 6 S1  743835                        Bacillus anthracis str. A16
 1.10  11805 5775  S 1396                    Bacillus cereus
 0.09  960 960 S1  526985                      Bacillus cereus Rock3-42
 0.07  762 762 S1  526992                      Bacillus cereus AH1271
 0.06  673 673 S1  288681                      Bacillus cereus E33L
 0.05  568 0 S1  1179100                     Bacillus cereus biovar anthracis
 0.05  568 568 S2  637380                        Bacillus cereus biovar anthracis str. CI
 0.05  538 538 S1  1003239                     Bacillus cereus C1L
 0.04  409 409 S1  526977                      Bacillus cereus ATCC 4342
 0.03  368 368 S1  526986                      Bacillus cereus Rock3-44
 0.02  260 260 S1  1217984                     Bacillus cereus FRI-35
 0.02  211 211 S1  405535                      Bacillus cereus AH820
 0.02  190 190 S1  572264                      Bacillus cereus 03BB102
 0.01  150 150 S1  526973                      Bacillus cereus m1293
 0.01  101 101 S1  451709                      Bacillus cereus 03BB108
 0.01  94  94  S1  347495                      Bacillus cereus F837/76
 0.01  85  85  S1  222523                      Bacillus cereus ATCC 10987
 0.01  85  85  S1  1454382                     Bacillus cereus D17
 0.01  82  82  S1  526988                      Bacillus cereus Rock4-18
 0.01  75  75  S1  526987                      Bacillus cereus Rock4-2
 0.01  72  72  S1  526970                      Bacillus cereus BGSC 6E1
 0.01  60  60  S1  226900                      Bacillus cereus ATCC 14579
 0.01  60  60  S1  526980                      Bacillus cereus ATCC 10876
 0.00  45  45  S1  405532                      Bacillus cereus B4264
 0.00  32  32  S1  526989                      Bacillus cereus F65185
 0.00  30  30  S1  526984                      Bacillus cereus Rock3-29
 0.00  30  30  S1  526969                      Bacillus cereus m1550
 0.00  30  30  S1  526993                      Bacillus cereus AH1272
 0.00  29  29  S1  526991                      Bacillus cereus AH676
 0.00  19  19  S1  526979                      Bacillus cereus 95/8201
```

FIGURE 5.4 – Exemple de fichier *report*. *Chaque ligne correspond à un taxon. Première colonne : pourcentage de lectures assignées au clade associé au taxon. Deuxième colonne : nombre de lectures assignées au clade associé au taxon. Troisième colonne : nombre de lectures directement associées au taxon. Quatrième colonne : rang. Cinquième colonne : numéro de taxon. Sixième colonne : intitulé du taxon.*

Ici, cinq colonnes sont présentes sur ce type de fichier, chaque ligne correspond à un taxon.

— La première colonne correspond au pourcentage de lectures assignées au clade associé au taxon.

— La deuxième colonne correspond au nombre de lectures assignées au clade associé au taxon.

— La troisième colonne correspond au nombre de lectures directement associées au taxon.

— La quatrième colonne correspond aux rangs [11] : *(U)nclassified*, *(D)omain*, *(K)ingdom*, *(P)hylum*, *(C)lass*, *(O)rder*, *(F)amily*, *(G)enus*, *(S)pecies*.

— La cinquième colonne désigne le numéro du taxon (identique au *taxonomy ID* de NCBI).

— La sixième colonne correspond à l'intitulé scientifique du taxon.

On peut traduire graphiquement ce fichier à l'aide du logiciel KRONA (ONDOV, BERGMAN et PHILLIPPY, 2011) qui permet une visualisation plus simple des résultats. Un exemple est donné à la figure 5.5.

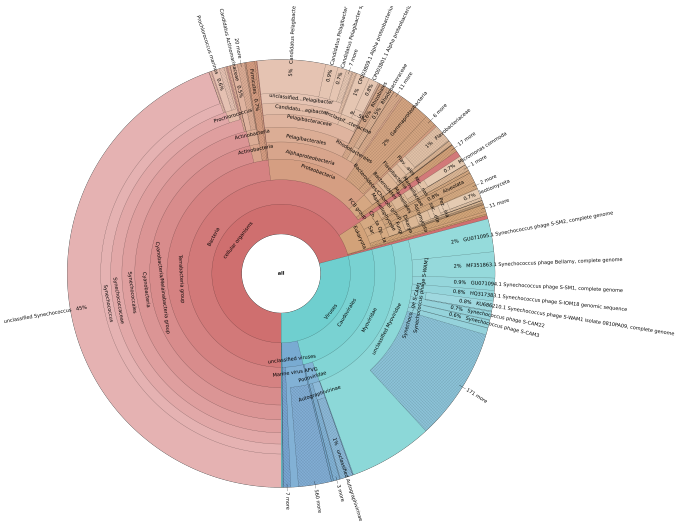

FIGURE 5.5 – Exemple d'assignation taxonomique de lectures sous forme de diagramme circulaire avec le logiciel KRONA.

Une analyse fine d'agents pathogènes d'intérêt est difficilement réalisable à l'aide de la base de données de référence de KRAKEN. Par exemple, sur le cas d'étude présenté précédemment, nous remarquons sur la figure 5.4 la présence de souches du groupe *B. cereus group* avec exactement 99,02% de lectures assignées à ce genre. Cependant, la précision s'arrête grossièrement à ce stade car seulement 12,38% de ces lectures sont assignées à l'espèce *B. anthracis* et moins de 0.1% des lectures sont assignées à des souches de cette espèce.

Pour pallier ce manque de résolution, KRAKEN propose une fonctionnalité de création de *custom databases* (WOOD, LU et LANGMEAD, 2019). Ces dernières sont des bases de données totalement créées par l'utilisateur. Elles sont généralement utilisées comme base de données spécifique à une espèce pour permettre une identification intraspécifique. Pour ce faire, l'utilisateur doit fournir l'arbre phylogénétique

11. Également appelés *rangs taxinomiques*

relatif à l'espèce en question ainsi que les fichiers FASTA [12] de toutes les souches y appartenant.

5.1.5 Outils existants spécifiques à la détection de *Bacillus anthracis*

Si KRAKEN, comme les autres outils présentés précédemment, sont des outils de classification métagénomique globaux, certaines méthodes ont été développées pour le cas spécifique de *B. anthracis*.

Méthode basée sur le *mapping* de lectures Les premières méthodes d'analyse de séquençage pour la détection de *B. anthracis* se sont basées sur le *mapping* des lectures sur des génomes de référence. Par exemple BE et al., 2013 se sont basés sur les génomes de *B. anthracis* Ames ancestor, *B. cereus* biovar CI et *B. thuringiensis* Al Hakam pour aligner les lectures de séquençage et comparer leur nombre respectif. Cette méthode peut permettre dans le cas d'échantillons très concentrés de distinguer *B. anthracis*. Cependant, dans le cas d'échantillons complexes réels, un simple *mapping* ne permet que dans des rares cas de différencier une souche de *B. anthracis* avec une souche du groupe *B. cereus* proche. En effet, un nombre important de lectures alignées sur un génome ne présage pas de la couverture d'alignement. Cet élément est pourtant déterminant, *a fortiori* dans le cas d'espèces génétiquement très proches.

Plus récemment, SAHL et al., 2015 ont développé un pipeline permettant l'assignation phylogénétique de lectures à faible couverture, en ne retenant pour l'analyse que les lectures *mappant* sur des positions spécifiques vis-à-vis d'un génome de référence. Cependant, si cette méthode peut s'avérer adaptée pour des espèces à forte recombinaison, le faible nombre de SNPs discriminants au sein des espèces clonales comme *B. anthracis* limite le pouvoir de résolution de cette méthode.

Méthode avec k-mers À côté des outils de métagénomique globale, d'autres propres à la détection de *B. anthracis* dans des échantillons environnementaux ont été développés. Dans cette section sont présentés quelques-uns d'entre eux, en particulier l'approche "k-mer", ayant la plus grande sensibilité, dont certains résultats expérimentaux seront mis à profit ultérieurement dans nos travaux.

PETIT III et al., 2018 ont développé une méthode de détection se basant sur une approche de type k-mer. Elle repose sur la création de trois sets de k-mers de référence qui sont détectés (ou non) dans l'échantillon métagénomique étudié. Les sets de k-mers sont les suivants :

— *Ba31* : il s'agit de l'ensemble des 31-mers spécifiques au génome de l'espèce *B. anthracis*.

— *BCerG31* : il s'agit de l'ensemble des 31-mers spécifiques au groupe *B. cereus* (en incluant l'espèce *B. anthracis*).

— *lef31* : il s'agit de l'ensemble des 31-mers spécifiques au gène *lef* (porté par le plasmide pXO1 et codant pour le facteur létal).

Le principe de création de ces trois sets de k-mers est détaillé sur la figure 5.6. Pour *Ba31*, plusieurs filtres ont été appliqués. Cela a permis la suppression successive des k-mers communs avec *BCerG31* ou plus largement du genre *Bacillus* puis

12. Également appelé format PEARSON, un fichier FASTA est un fichier texte comprenant des séquences biologiques. À ne pas confondre avec un fichier FASTQ qui stocke des données de séquençage brutes.

communs avec des séquences de ARNr. La dernière étape utilisant BLASTN (avec la base nucléotidique *nt*) a permis d'aligner les k-mers *Ba31* et de supprimer les matchs exacts avec des génomes du groupe *B. cereus* (hors *B. anthracis*). Les mêmes étapes de filtres ont été appliquées au set *BCerG31* pour y supprimer les k-mers communs avec des génomes du genre *Bacillus* (hors groupe *B. cereus*). En définitive, le set *Ba31* est composé de 239,503 31-mers et le set *BCerG31* contient 10,183 31-mers. Le set *lef31* est composé de 2,617 31-mers.

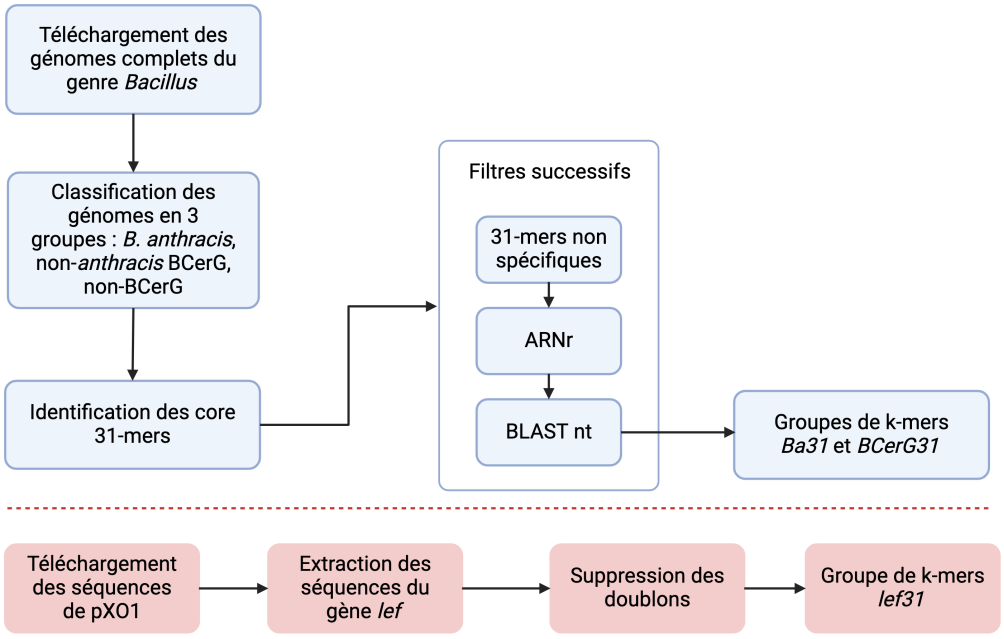

FIGURE 5.6 – Schéma détaillant les étapes successives de la création des trois sets de k-mers de référence. En bleu pour *Ba31* et *BCerG31*, en rouge pour *lef31*. Adapté de PETIT III et al., 2018. Créé avec BIORENDER.

Les sets *Ba31* et *BCerG31* ont été testés sur des séquençages simulés des souches du groupe *B. cereus* (hors *B. anthracis*) les plus proches de *B. anthracis*. Pour cela, des fichiers FASTQ synthétiques à différentes couvertures pour chaque génome étudié ont été créés (en respectant le taux d'erreurs aléatoires du séquençage Illumina). Il s'avère que les nombres de k-mers *Ba31* et *BCerG31* détectés suivent une relation linéaire : pour une unité de couverture [13] de *BCerG31*, il y a en moyenne 172 k-mers *Ba31* détectés. Ce résultat est illustré sur la figure 5.7. Cette détection systématique de faux positifs s'explique par le taux d'erreurs aléatoires de séquençage.

Enfin, le dernier résultat de l'étude basée sur ces sets de k-mers est le suivant : en dessous d'une couverture de génome de *B. anthracis* égale à 0.184x, il est possible de ne pas détecter le gène *lef* dans un séquençage (autrement dit, ne pas détecter un seul 31-mer du set *lef31*).

Tous ces résultats combinés ont permis de définir quatre cas de figure possibles lors de la détection de *B. anthracis* dans un séquençage à l'aide de ces sets de k-mers. Ils sont résumés dans le tableau 5.1.

13. Ici, une unité de couverture d'un set de k-mers est définie comme le rapport entre le nombre de k-mers du set détectés dans le séquençage par le nombre total de k-mers du set.

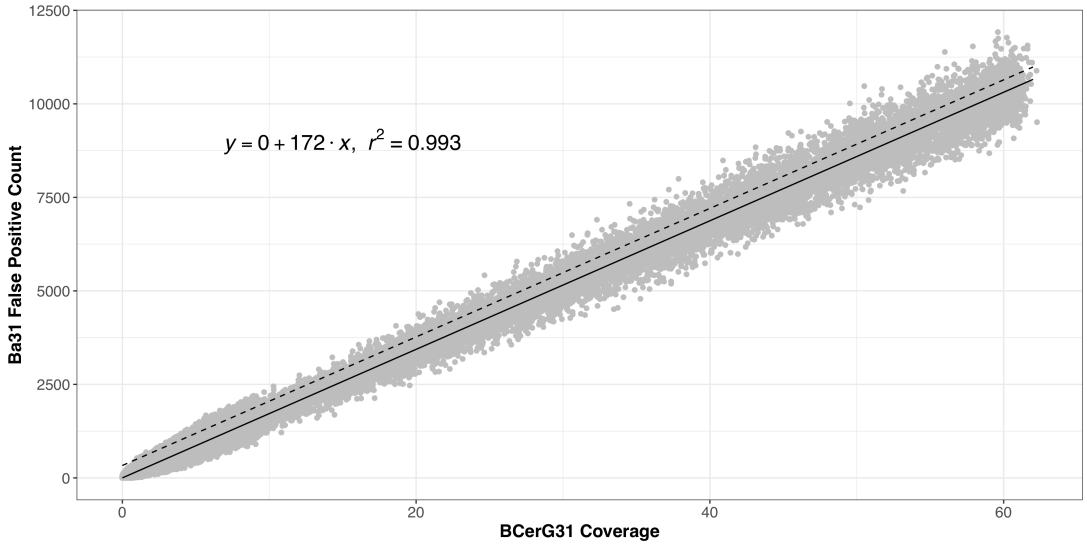

FIGURE 5.7 – Régression linéaire du nombre de faux positifs *Ba31* détectés en fonction de la couverture de *BCerG31*. *Il y a une ordonnée à l'origine de 0, la couverture BCerG31 étant la valeur fixe et le nombre de faux positifs Ba31 étant la variable. La ligne continue montre les valeurs prédites par la régression linéaire et la ligne en pointillé reflète la limite supérieure de l'intervalle de confiance à 99% pour les paramètres*. Tiré de PETIT III et al., 2018.

Cas	*Lef31*	Zone grise	Dépassement de l'intervalle de confiance	Interprétation
1	Oui	Oui ou non	Oui ou non	Détection du facteur létal donc présence d'une souche de *B. anthracis* ou de *B. cereus anthracis-like*.
2	Non	Oui	Oui	Possiblement présence de *B. anthracis* d'après le nombre élevé de k-mers *Ba31* mais la couverture du génome est trop faible pour garantir la présence du gène lef. Nécessite une plus grande couverture et/ou une validation par PCR ou d'autres méthodes.
3	Non	Non	Oui	Le nombre de k-mers *Ba31* dépasse ce qui est attendu par le modèle, mais la couverture du génome est à un niveau où le facteur létal aurait dû être détecté. Une explication probable est qu'il y a présence d'une souche de *B. anthracis* sans pXO1 ou d'une lignée non connue proche de *B. anthracis*.
4	Non	Oui ou non	Non	Le scénario le plus probable est que le bruit de fond du *BCerG* a produit des k-mers *Ba31* par des erreurs aléatoires, mais on ne peut exclure la présence de *B. anthracis* à faible couverture.

TABLE 5.1 – Cas de figure possibles après détection des k-mers *Ba31* dans un échantillon métagénomique. Adapté de PETIT III et al., 2018.

Cet outil spécialisé dans la détection de *B. anthracis*, basé sur les k-mers, revêt plusieurs avantages. Tout d'abord, il permet de résoudre en partie le problème des faux positifs posé par l'approche généraliste des autres outils de détection métagénomique, y compris ceux basés sur les k-mers (comme KRAKEN2 par exemple). En effet, ces derniers sont performants pour déterminer la composition globale d'un échantillon, mais peuvent générer des faux positifs quand il s'agit de détecter des pathogènes ayant une proximité génétique forte avec d'autres espèces (typiquement le cas de *B. anthracis*). De plus, cet outil a permis de quantifier ce phénomène de génération de faux positifs lors de la présence de proches voisins de *B. anthracis*, dans le cas d'un séquençage Illumina.

Cependant, certaines améliorations possibles sont à signaler. Tout d'abord, cet outil ne permet pas de faire la distinction entre la présence d'une souche de *B. anthracis*, d'une souche de *B. cereus* anthracis-like ou d'un mélange de souches. Cela

est principalement dû à la seule détection du gène *lef* qui provoque une perte de résolution conséquente. C'est un outil de détection mais pas de caractérisation. Outre cela, le modèle de génération de faux positifs par rapport au taux de couverture du groupe *B. cereus* est basé sur le taux d'erreur de la technologie de séquençage Illumina. Dans le cadre de l'étude d'un autre type de séquençage (de type *long reads* par exemple), il faudra établir un nouveau modèle. Toutes ces fragilités montrent qu'une confirmation avec des techniques microbiologistes restent nécessaires. De plus, une étude phylogénétique est indispensable pour caractériser la ou les souches en présence le cas échéant, sa puissance de résolution étant plus élevée qu'une détection d'un ou plusieurs gènes spécifiques.

5.2 Présentation de l'étude

Les faiblesses observées en termes de sensibilité ou de spécificité de la part des logiciels de détection métagénomique de *B. anthracis* ont motivé le développement d'un outil adapté. Deux axes principaux d'amélioration ont été recherchés. D'une part, une diminution de l'apparition de faux positifs dans la détection de *B. anthracis*. D'autre part, une augmentation du pouvoir de résolution en proposant une détection intraspécifique (à la ou les souches près), dans le cas où la concentration de *B. anthracis* est suffisamment élevée. Enfin, l'ensemble du processus devra se faire de manière rapide, en comparaison avec les outils bioinformatiques existants. Cette section vise à détailler les différentes étapes qui ont permis le développement d'un tel outil.

Comme décrit dans le chapitre précédent, de nombreux outils bioinformatiques ont été développés pour déterminer la composition globale d'un métagénome, mais aussi pour détecter des agents pathogènes au sein d'un échantillon complexe. Les approches de type k-mer sont particulièrement efficaces par le niveau de résolution qu'elles permettent d'obtenir. Parmi elles, le logiciel KRAKEN2 est régulièrement utilisé pour répondre au premier objectif (composition globale). Cependant, différencier deux espèces ayant une forte homologie (comme *B. anthracis* avec les autres espèces du groupe *B. cereus*) ou détecter de nouvelles espèces et/ou souches peut s'avérer compliqué car ce type d'approche s'appuie sur des bases de données de référence dont dépendent directement le pouvoir résolutif de l'outil. Ainsi, certaines études ont conclu, à tort, à la présence de *B. anthracis* dans des environnements urbains pour toutes ces raisons évoquées. Deux axes d'améliorations en découlent : il faut adapter les outils dans le cadre d'une détection fine de *B. anthracis* et une analyse phylogénétique est indispensable pour conclure à sa présence (ou non) dans un échantillon environnemental.

Pour ce faire, nous avons d'abord développé dans le cadre d'un projet de biodéfense européen un pipeline d'analyse de métagénomes, nommé B2FORENSICS_V1 dédié à la détection de *B. anthracis*. Il s'appuie sur le logiciel KRAKEN2 auquel ont été ajoutées des étapes successives de filtres choisis pour éviter la création de faux positifs. De plus, un ensemble de données tests a été mis en place pour évaluer les

performances de ce pipeline. Ces premiers travaux sont décrits dans l'article suivant [14].

Dans un second temps, une mise à jour du pipeline a été opérée pour gagner en temps d'exécution. En effet, il s'avère que certaines étapes de l'outil requièrent des alignements, qui sont chronophages. Pour répondre à cet objectif de rapidité, tout en conservant la résolution de l'outil, la possibilité qu'offre KRAKEN2 de concevoir des *custom databases* a été exploitée. En outre, une ultime étape d'assignation phylogénétique a été ajoutée.

Avant-propos Les phylogénies mentionnées dans ce chapitre ont toutes été établies suivant la même méthodologie. Les génomes pris en compte ont été téléchargés sur NCBI. Ils ont ensuite été convertis en lectures simulées à l'aide d'un script interne (Python v3.6.2), puis alignés sur le chromosome de *B. anthracis* Ames Ancestor (GCF_000008445.1), résultant ainsi en des alignements de longueur identique. Ce choix se justifie du fait de l'étude focalisée sur l'espèce *B. anthracis* et de la forte proximité génétique existant au sein du groupe *B. cereus*. Les SNP génomiques sont sélectionnés avec BIONUMERICS 8.1.1 (BioMérieux, Applied Maths, Sint-Martens-Latem, Belgique), en utilisant l'option "*Strict SNP filtering (Closed SNP set)*" et les paramètres par défaut (*12 bp inter-SNP*). L'analyse du *clustering* est ensuite réalisée en utilisant la méthode de calcul "*Maximum parsimony tree*" et les paramètres par défaut. Par ailleurs, par souci de simplification, les notions de "alignement des lectures simulées à partir d'un génome sur le chromosome de *B. anthracis* Ames Ancestor" et de "alignement de génome sur le chromosome de *B. anthracis* Ames Ancestor" seront confondues dans ce chapitre. Enfin, les arbres phylogénétiques ont été visualisés à l'aide de l'outil GRAPETREE (ZHOU et al., 2018).

5.3 Observations préliminaires

La stratégie adoptée pour développer le pipeline a été de s'appuyer sur un outil existant, le logiciel KRAKEN2, et d'adapter son utilisation à la détection de *B. anthracis* spécifiquement. Ce choix de KRAKEN2 s'explique pour deux raisons : l'approche k-mer exploitée par KRAKEN2, l'alternative la plus rapide pour traiter d'importants volumes de séquençages (comme c'est le cas pour des métagénomes) et la possibilité qu'offre le logiciel d'utiliser des *custom databases* pour adapter son utilisation.

L'idée première a été de créer une *custom database* spécifique à la bactérie *B. anthracis*, c'est-à-dire appliquer l'assignation taxonomique des lectures de séquençage sur un arbre phylogénétique de cette bactérie. Pour cette étape de validation de concept, un arbre phylogénétique de 181 souches de *B. anthracis*, représentatif de la diversité de l'espèce (lignées A, B et C) a été créé.

Pour tester grossièrement cette démarche, le cas le plus simple à traiter serait le génome complet d'une souche de *B. anthracis*, présente dans l'arbre phylogénétique. Cela a été effectué pour la souche *B. anthracis* ANSES_08-8_20 (numéro d'accession : GCF_000697515.2, assemblage à partir d'un séquençage Illumina). Idéalement, l'assignation taxonomique des lectures synthétiques issues de l'assemblage produirait un chemin entre la racine de l'arbre phylogénétique et la feuille correspondant à la souche en présence, sans aucun artefact, c'est-à-dire sans qu'aucune lecture ne soit

14. Ce travail s'est inscrit dans le cadre du projet B2FORENSICS (pour *Bioforensics for Biodefense*) qui est une collaboration européenne pilotée par l'Agence Européenne de Défense incluant des établissements de défense d'Allemagne, Suède, Italie et France pour détecter la trace de certains agents pathogènes dans des données environnementales.

assignée en dehors de ce chemin principal. Ce cas de figure est illustré sur la figure 5.8. Chaque lecture est assignée par KRAKEN2 soit à une feuille lorsque la séquence est suffisamment spécifique, soit à un noeud interne lorsque l'information contenue ne permet pas de discriminer entre plusieurs souches. On considère qu'une souche réellement présente dans un échantillon séquencé sera révélée par l'assignation des lectures à tous les noeuds reliant cette souche à la racine de l'arbre. L'ensemble du génome devant être recouvert par les lectures, l'échantillon devrait contenir aussi bien des lectures couvrant les SNPs spécifiques à la souche (feuille de l'arbre) que des lectures couvrant une zone commmune de *B. anthracis* (noeud de l'arbre).

Un cas réel (séquençage d'une souche de *B. anthracis*) est présenté sur la figure 5.9. On y observe un chemin certes lisible entre la racine de l'arbre et la feuille correspondant à la souche, mais de nombreux artefacts sont présents en dehors de ce chemin. Les noeuds de l'arbre ayant des lectures qui leur sont assignés sont marqués par un coloris gradué de jaune à rouge, jaune indiquant un nombre faible de lectures et rouge un nombre élevé.

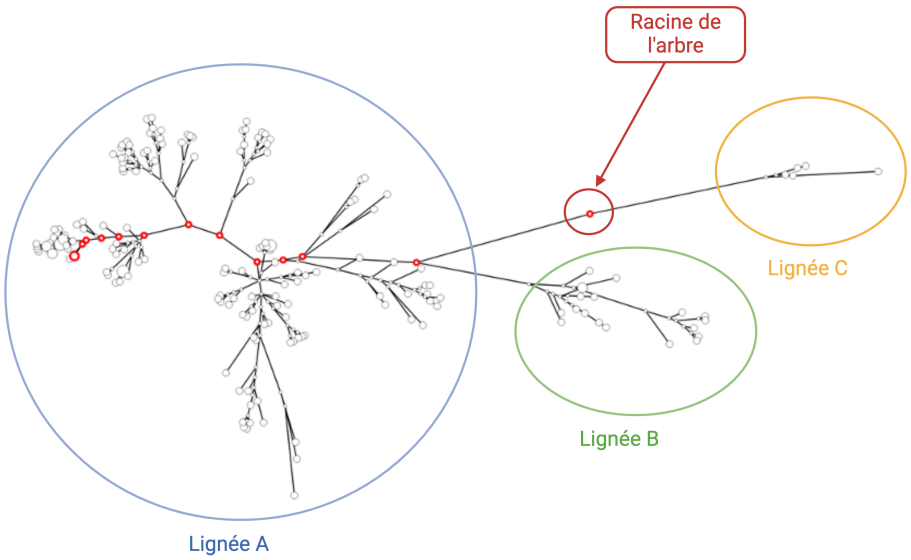

FIGURE 5.8 – Assignation taxonomique idéale des lectures de séquençage d'une souche de *B. anthracis*, c'est-à-dire sans la présence de bruit de fond. Le chemin théorique liant la racine de l'arbre phylogénétique à la souche est surligné en rouge, de droite à gauche.

Un cas plus complexe que le précédent serait cette fois-ci la présence de deux souches dans un échantillon, isolées également. Ce cas a été testé en séquençant avec Illumina MiSeq les souches *B. anthracis* Pasteur et Sterne. L'assignation taxonomique des lectures de séquençage est présentée sur la figure 5.10. Le chemin liant la racine de l'arbre à chacune des deux souches est difficile à discerner par l'utilisateur, du fait de la présence massive d'artefacts.

La présence d'artefacts dans les deux cas de figures présentés peut s'expliquer pour plusieurs raisons, classées par ordre d'importance :

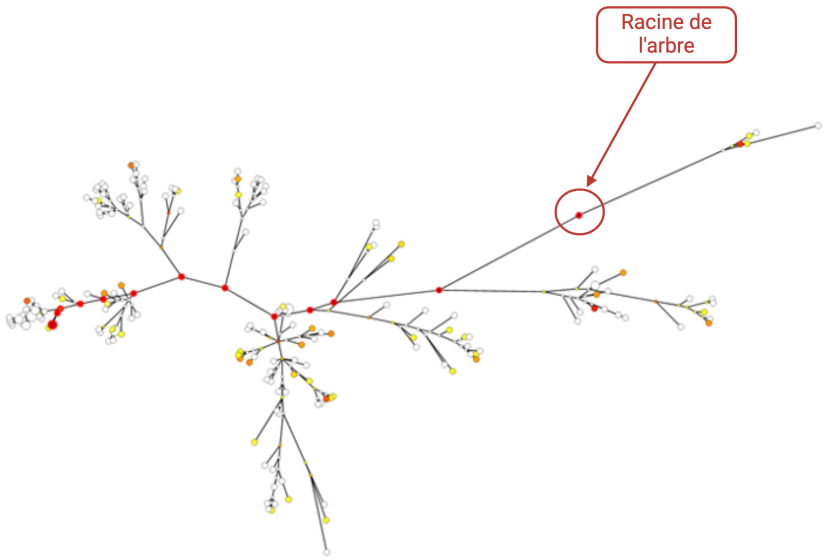

FIGURE 5.9 – Assignation taxonomique réelle des lectures de séquençage d'une souche de *B. anthracis*. Le chemin théorique liant la racine de l'arbre phylogénétique à la souche est surligné en rouge, de droite à gauche.

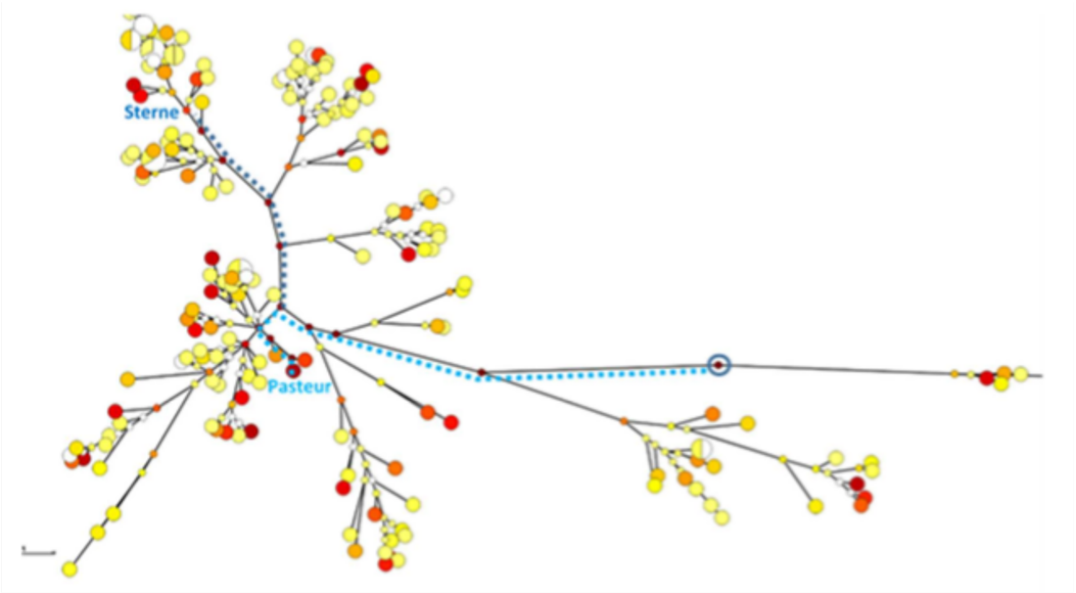

FIGURE 5.10 – Assignation taxonomique des lectures de séquençage d'un mélange de deux souches de *B. anthracis*. En bleu est indiqué le chemin liant la racine de l'arbre phylogénétique aux deux souches en présence, Sterne et Pasteur.

— Utiliser une *custom database* spécifique à *B. anthracis* revient à contraindre l'assignation taxonomique de l'ensemble des lectures à une phylogénie restreinte. En effet, KRAKEN2 utilise par défaut une phylogénie de l'ensemble du vivant

pour attribuer un taxon à chaque lecture. Or, remplacer cette phylogénie du vivant par une phylogénie spécifique à une bactérie revient à assigner toutes les lectures (y compris celles n'étant pas relatives à la ou les souches de *B. anthracis* en présence dans l'échantillon) à la partie de la phylogénie du vivant correspondant à l'espèce *B. anthracis*.

— La fiabilité de l'assignation taxonomique opérée par KRAKEN2 est dépendante de la qualité des assemblages des souches utilisées pour la création de la *custom database*. Autrement dit, si l'assemblage d'une des souches présentes dans l'arbre phylogénétique utilisé est par exemple contaminé, KRAKEN2 associera chaque lecture associée à cette contamination dans un séquençage quelconque à la feuille de la souche concernée.

— Les erreurs de séquençage inhérentes à la technologie utilisée peuvent conduire à une assignation taxonomique faussée de certaines lectures. Cette source d'artefacts est la plus contraignante car indépendante de la conception et de l'utilisation de la *custom database*, contrairement aux deux points précédents.

Ces deux études préliminaires amènent à différentes conclusions pour le développement d'un outil de détection de *B. anthracis* :

— Si l'usage de KRAKEN2 (avec les paramètres par défaut) permet avant tout de déterminer la composition globale d'un métagénome et peut générer de faux positifs pour la détection de *B. anthracis*, la seule utilisation d'une *custom database* spécifique à cette bactérie n'est pas adaptée.

— Pour tirer profit d'une *custom database* avec KRAKEN2, il faut préalablement sélectionner les lectures de séquençage pour lesquelles l'assignation taxonomique sur la phylogénie restreinte serait cohérente. Dans ce cas de figure, la *custom database* spécifique à *B. anthracis* servira à déterminer la ou les souches en présence dans l'échantillon. Le choix des assemblages de génome constituant cette *custom database* est déterminant car le pouvoir de résolution de celle-ci en dépend.

5.4 Développement du pipeline B2FORENSICS_V1 et de données tests : article 3

Cette section présente l'implémentation de B2FORENSICS_V1, un nouvel outil de détection métagénomique conçu spécifiquement pour *B. anthracis*. Face aux défis posés par la détection de cet agent pathogène dans des échantillons environnementaux, B2FORENSICS_V1 se distingue par sa capacité à minimiser significativement les faux positifs, améliorant ainsi la précision des analyses. Pour évaluer l'efficacité de cet outil, une série de tests a été réalisée en utilisant un jeu de données expérimentales, comprenant des échantillons réels enrichis en *B. anthracis*, en espèces proches, et en d'autres agents pathogènes. Cet ensemble de données pourra servir de référence pour la future évaluation et comparaison d'autres outils de détection de *B. anthracis*.

Les fichiers complémentaires (*Supplementary files*) de l'article suivant sont disponibles en Annexe B.

A biodefense-oriented pipeline for analysis of metagenomics data, and associated benchmarking datasets

Abdelli M.[*1,2], Dematheis F.[5], Sundell D.[4], Ramisse V.[1], Christiany D.[2], Vernadet JP.[2], Anselmo A.[3], Faggioni G.[3], Lista F.[3], Walter MC.[5], Sjödin A.[4], Karlsson E.[4], Lista F.[3], Forsman M.[4], Antwerpen MH.[5], Vergnaud G.[2]

[1]DGA CBRN Defence (France)
[2]Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), 91198, Gif-sur-Yvette, France
[3]Army Medical Center (Italy)
[4]Swedish Defence Research Agency (Sweden)
[5]Bundeswehr Institute of Microbiology (Germany)
*corresponding author: mehdi.abdelli@intradef.gouv.fr

Abstract:
Metagenomics refers to the study of total genetic material from a mixed community of organisms and has increasing popularity for deciphering the microbial diversity of environmental samples. Advances in both sample preparation and high throughput DNA sequencing have opened the way for metagenomic detection of pathogens of interest without excluding those unknown to date. It is likely an interesting approach to complement the ultrasensitive PCR, which is the current gold standard for biodetection. PCR suffers some drawbacks, it is poorly multiplexable, limited to targeted pathogens and blind to the presence of unknown ones. Importantly, false positives are difficult to deal with.
The search for pathogens in a biological sample by analysis of global genetic material is simple in principle and several bioinformatic tools allow to inventory a mass of genetic data. Recent tools based on k-mer approaches use reference genomes and a taxonomy to assign each segment of the sequence to a taxonomic level.
However, two major challenges need to be addressed. Firstly, the produced inventories may be inaccurate (specificity). For example, some studies have erroneously concluded that major pathogens such as *Bacillus anthracis* are present in some urban environments. In this case, the misinterpretation was mainly due to the inadequacy of the method used, i.e. the lack of a robust phylogenetic approach allowing the positioning of sequence data at subspecies level. This problem made it difficult to distinguish between *B. anthracis* and its closest relatives. Secondly, the sensitivity of the metagenomics approach might be too low for practical purposes in the presence of a rich biological background.
To address the issue of specificity, we developed pipeline B2forensics_v1, based upon the *Kraken2* software complemented by additional filters which allow to precisely evaluate the phylogenetic position of candidate reads within the *B. anthracis* population. To address the issue of sensitivity, we have developed a realistic benchmark data set by sequencing DNA extracted from wastewater. Prior to DNA extraction and sequencing, half of the samples were spiked with real biological agents or with close neighbours. We then evaluated the performance of B2forensics_v1 using the benchmark dataset. We will present the results we have obtained so far with a focus on *B. anthracis*.

Introduction:

All pathogens contain a specific composition of genetic information. If the genetic content of the environment can be monitored with sufficient speed and efficiency, they can be detected to warn populations in order to avoid any casualties. The common approach for detecting a pathogen is the search for agent-specific targets. Thus, assays have to be developed for each agent of interest independently.
Two approaches are currently in use for direct proof of the presence of microorganisms. The first approach takes advantage of antibodies to recognize the presence of the target. Such antigenic-assays are easy to run, but provide a moderate sensitivity with a significant risk of false positives when testing environmental samples [1]. The second is based on the specific amplification of genetic material by the process called “polymerase chain amplification” (PCR). PCR is more demanding but is very sensitive and specific.

During the past ten years, spectacular progress was achieved in the field of genetic analysis associated with “massively parallel sequencing”. These technologies are revolutionary in terms of volume of data output, speed of data production, cost reduction, and associated software for data analysis. This has allowed the development of large-scale “metagenomics” projects, in which the environment is sampled, genetic material is extracted and then sequenced [2].

In principle, such an approach opens the possibility to search for the presence of any pathogen by analysing the data in a single universal assay (DNA extraction and sequencing) [3]. However, despite the progresses in this field, working with this type of data still represents a big challenge. Indeed, several factors have been identified to affect the outcome of a metagenomic data analyses, namely: sample complexity (species richness and evenness), DNA extraction method, library preparation, sequencing technology (which define the data volume to work with) and software pipeline (which is dependent on computer capacity and software compatibility). Major problems encountered while working with metagenomic data include the ability to distinguish close related microorganisms (e.g. *Bacillus anthracis* and *Bacillus cereus*) from each other (accuracy) and to identify low-abundant members of the environmental specimen (sensitivity) [4].

Essentially two approaches are used to profile the microbial community and/or detect specific members of an environmental sample: assembly-based or non-assembly based. Assembly-based methods are downstream linked to similarity search tool such as BLAST, in order to assign each resulting contig/scaffold to taxa, to which it could be aligned with a statistically-significant similarity score. The major advantage of the method consists in the microbial genome reconstruction which uncovers new insights into the microbial content of antimicrobial resistance (AMR) and virulence acquired genes (VAG), as well as into microbial genetic modifications, resulting from a natural evolution process or from a genetic engineering approach (GMO). However, assembly-based methods are computationally intensive, require long processing time and might produce misassemblies, leading to contig/scaffold misclassification. Furthermore, since different strains or even different related species might share a high degree of genome identity, the use of BLAST for taxonomical identification might cause incorrect read classification as well. Thus, a closer look to the raw BLAST results, to review multiple hits with identical score, is mandatory to validate the final identification.

Differently, free-assembly-based methods like Kraken2 [5], using exact k-mer matches, or sourmash [6] (using MinHashes signatures), have a lower memory requirement, but rely upon a carefully curated reference database (DB). A lack in the DB often results in false positives or false negatives. To avoid this, the DB needs to be customized and kept up to date.

The search for high-risk pathogens in environmental metagenomics data has not yet been extensively investigated, and some early investigations have wrongly reported the presence of *Bacillus anthracis* in environments such as the New York City subway [4], for reasons which have been subsequently explained in detail later [7]. The field appears to be quickly emerging [8].

Building upon these remarkable achievements, we addressed the question of evaluating the specificity and sensitivity of such a free-assembly-based metagenomics approach to detect pathogens. The technological challenge was to be able to distinguish specific genetic signatures from environmental background. In addition, the errors produced by sequencing equipment also needed to be taken into account.

We collected wastewater samples in two locations (Germany and France). Samples were spiked with various controls or real select agents to allow evaluation of specificity and sensitivity in a realistic context. Software previously developed for the fast taxonomic classification of sequence data was combined in a pipeline including additional data filtering to achieve high specificity for biodefense purposes. Using the presented bioinformatics pipeline the amount of only a few hundred bacterial cells in ten millilitres of sewage water can be sufficient to be detectable within about 50 Gb of sequencing data. Whereas bioinformatics pipelines like ours simplify the processing of metagenomics data and identification of pathogens, genetical as well as microbial expertise are still essential for correct phylogenetic classification, characterization of the detected strain and interpretation of the obtained results.

Materials and methods:

Experimental set up

Sewage-in-water from Munich and Paris communal wastewater facility was collected and 10 ml of it were spiked with different amount of four life-threatening bacteria (*Bacillus anthracis, Fransicella tularensis*, *Brucella melitensis* and *Yersinia pestis*) belonging to the European Defence Agency (EDA) strain collection of the European Biodefense Laboratory Network (EBLN), and some close related species.

Two spike-pools were set-up for the experiments: one spike-pool with Select Agents (pool wSA) and one spike-pool with surrogate and Nearest Neighbors (pool sNN). Composition of pool sNN was 10^7 *Francisella hispaniensis* cells, 10^4 *Bacillus cereus* spores and 10^7 *Bacillus thuringiensis subsp. kurstaki* spores, whereas pool wSA harbored 10^3 *Brucella melitensis* 003-00435, 10^3 *Yersinia pestis* strain Dodson, 10^6 *Francisella tularensis* FSC 054, 10^6 *Bacillus anthracis* Ferrara and 10^7 *Bacillus cereus* ATCC 10987. All strains were cultivated under the appropriate biosafety condition (BSL2 or BSL3) and national regulations. All strains of the pool wSA were inactivated using 5 % terralin PAA (Schülke & Mayr, Norderstedt, Germany), whereas spike of pool sNN was not inactivated prior spiking of collected wastewater samples.

Unspiked wastewater samples were prepared as well and used as a control. Pure culture DNA from the strains used as spikes was extracted and sequenced with Illumina MiSeq sequencing technology in order to create spiking reference genomes. Total community (TC) DNA extracted from sewage-in-water was sequenced using NextSeq Illumina technology and obtained data was investigated using the presented bioinformatic pipeline of this study.

Spiking material inoculum and quantification by means of SYBR Green Real-Time PCR

Bacillus cereus ATCC 10987 was cultivated for 48 h on Columbia blood agar at 37°C. Both *Bacillus anthracis* and *Brucella melitensis* strains were cultivated on Columbia blood agar medium at 37°C for 48 h. *Francisella tularensis* strain was grown on Cystine Heart Agar medium at 37°C for 48 h while *Yersinia pestis* strain was grown on Yersinia Selective Agar (CIN) at 28°C for 48 h. Cells and/or spores of each spiking material were washed twice then stored in phosphate-buffered saline (PBS). To determine the concentration (CFU/ml) of the pure bacterial cultures in PBS suspension, 5-fold dilution series of each isolate were prepared and analyzed by qRT-PCR. Universal bacterial primers 27F (5'-AGAGTTTGATCMTGGCTCAG-3') and 340R (5'-CTGCTGCCTCCCGTAGG-3'), targeting the hyper-variable regions of V1 to V3 in the bacterial 16S rRNA gene, were used. PCR mixtures were prepared using 1X SYBR Green I Master Mix (Roche), 1 µM of each primer, 5 µl of cell suspension template and sterile Milli-Q water up to 20 µl PCR reaction volume. PCR amplification was carried out in the LightCycler® 480 instrument (Roche) with the following cycling program: 95°C for 1 min, 45x (95°C for 20 s, 50 °C for 20 s, 72°C for 40 s). The dissociation curve of the amplified product was acquired with the additional melting step: 1x (95°C for 5 s, 40°C for 1 min, 95°C for 0 s). The quantification of the 16S rRNA gene copy number was done using the standard curve method. To assess the concentration of bacterial cells per ml of pure culture, the resulting 16S rRNA gene copy number was normalized by the number of 16S rRNA genes per strain. We divided by three for *Brucella melitensis* 003-00435 and *Francisella tularensis* FSC 054, by six for *Yersinia pestis* strain Dodson, by ten for *Bacillus anthracis* Ferrara and by 14 for *Bacillus cereus* ATCC 10987.

Spiking reference genomes

All five spiking materials from pool wSA used in the current study were sequenced with MiSeq sequencing technology and Nextera DNA library prep kit. The quality of the raw Illumina reads was assessed with FastQC v0.11.9 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and cleaned from adapter contamination, low quality reads and duplicates using Trimmomatic v0.39 [9]. High quality paired-end reads were assembled *de novo* using SPAdes v3.13.0 assembler [10], while

resulting scaffolds were corrected using Pilon v1.24 software [11]. For each sequenced isolate, a high-quality draft genome with an approximately 90-fold depth of coverage per base was generated. Genome comparison between the spiking material and NCBI reference genomes was performed using Mauve v2.4.1 [12]. Only the sequence of *Bacillus anthracis* Ferrara was deposited in the NCBI Sequence Read Archive (SRA) repository (SRR10382389).

Sample Preparation for benchmarking

B2F05

B2F05 corresponds to a wastewater sample collected on May 14th, 2017 (Paris, France). It is total DNA extraction from 10 ml of filtered (glass filter) wastewater sample. 10 ml of wastewater was spiked with pool sNN.

B2F10 and B2F12

B2F10 and B2F12 correspond to wastewater samples collected on May 11th, 2017 (Munich, Germany). DNA was extracted from both samples as explained in the Supplementary file 1.
For B2F10, 10 ml of wastewater was spiked with pool sNN. Sample B2F12 was unspiked as blank control.

B2F13, B2F14, B2F15, B2F16:

Sewage water collected on October 30th, 2017 in Munich was used to produce DNA samples B2F13 to B2F16. B2F13/14 differ from B2F15/16 by the DNA extraction method. B2F13 and B2F15 were unspiked (negative control samples), whereas B2F14 and B2F16 were spiked with pool wSA.

TC-DNA from spiked and unspiked sewage-in water samples was extracted twice using two different DNA extraction methods: the DNeasy Power water kit (Qiagen, Hilden, Germany), used according to the manufacturers´ protocol and a classical phenol-chloroform procedure (Supplementary file 1). Concentration and purity of the TC-DNA were assessed by means of DS-11 FX Spectrophotometers (DeNovix) and Qubit® 3.0 Fluorometer (Thermo Fisher Scientific). The TC-DNA extracted with the DNeasy Power water kit and with the phenol-chloroform protocol will be hereinafter referred to as B2F14, B2F13 and B2F16, B2F15 respectively.

Detection and quantification of the spiking materials in the sewage-in samples by means of TaqMan real-time PCRs

The abundance of spiking materials in sewage-in water samples was determined by means of a qRT-PCR performed either with the Light Cycler 2.0 or the Light Cycler 480 instruments (Roche). *Yersinia pestis* was identified using specific primers for pPCP1 virulence plasmids and targeting the gene *pla* [13,14]. *Brucella melitensis* was detected using primers and TaqMan probe specific for the IS711 element [15]. The chromosomal DNA-marker *dhp61* was used to identify *Bacillus anthracis* [16]. *Francisella tularensis* was detected by means of a qRT-PCR (LightMix®-Kit *Francisella tularensis*) designed by TIB MOLBIOL (Berlin, Germany) and specific for *Francisella tularensis* 16S-rRNA gene.
For the molecular identification of members of the *Bacillus cereus group*, a well-established in-house qRT-PCR, targeting the gene *gyrA* was performed. In particular, the primers *gyrA_f* (5'-AtgTCAgACAATCAASAACAAgC-3'), *gyrA_r* (5'-CgAgAYACgATAACACTCATTgC-3') and the TaqMan probe 5'-6FAM-TATTAgYCATgAAATgCgTACCTC—BBQ were used. PCR reactions were performed using 5 µl of template DNA, 0.5 µM of each primer, 0.25 µM probe, 5 mM MgCl2, 1X LightCycler FastStart DNA Master Hybridization Probes (Roche) and 0.5 U LightCycler Uracil-DNA Glycosylase and nuclease-free water in a final volume of 20 µl. The qRT-PCR was carried out on the Light Cycler 2.0 instrument (Roche) using the following amplification scheme: 40 °C for 10 min, 95 °C for 10 min, 45x (95 °C for 15 s, 62 °C for 20 s, 72 °C for 20 s) and 40 °C for 30 s.

For the specific detection of *Bacillus cereus* ATCC10987 the following primers and probes were designed targeting gene *pdhD*: 5'-TTTAACAACAgAgCTCATAgC-3', 5'-CATATgCTAACTTCggTACAg-3', 5'-6FAM-ATTTCgTCgCCCgCTTCTACTA—BBQ (BcATCC10987-specific) and HEX-gATTTCgTCgCCAgCTTCTACTA—BBQ (other *Bacilli*). PCR reaction mixture was prepared using 5 µl of template DNA, 0.4 µM of each primer, 0.3 µM of each probe and 1X Qiagen Multiplex PCR Master Mix in a 25 µl final volume. The qRT-PCR was carried out on the LightCycler 480 instrument (Roche) with the following thermoprofile: 95 °C for 15 min, 35x (94 °C for 15 s, 60°C for 20 s, 72 °C for 20 s) and 72 °C for 5 min. The specificity of the amplification was validated on several isolates belonging to the *Bacillus cereus* group and including strains of *Bacillus cereus*, *Bacillus anthracis*, *Bacillus thuringiensis* and *Bacillus pumilus* (data not shown).

Sequencing of TC-DNA

Twenty nanograms of each DNA and five PCR cycles were used to produce the Nextera Illumina sequencing library. The average library fragment size was 1100-1400 bp including the 130 bp adapters. Resulting TC-DNA were sequenced using NextSeq 500/550 High Output Kit v2 (150 cycles) on NextSeq Illumina sequencer.

Data availability

The sequences have been deposited in the European Nucleotide Archive (ENA) under the study project PRJEB73343 and are available at https://www.ebi.ac.uk/ena/browser/view/PRJEB73343.

Pipeline B2forensics_v1

The B2forensics_v1 pipeline is written with Snakemake [17] and integrates a collection of Python and shell scripts that are publicly available on GitHub (https://github.com/i2bc/b2forensics). This framework is complemented by several next-generation sequencing (NGS) tools, namely SAMtools [18], BWA [19], Kraken2 [5] and nucleotide BLAST (BLASTN) [20]. An illustrative diagram of the B2forensics_v1 workflow is provided in Figure 1.

Initially, the pipeline processes input data in FASTQ (or zipped FASTQ) format through Kraken2, a system for taxonomic classification of sequencing reads. Kraken2 employs the MiniKraken2 database, which includes a variety of bacterial, viral, archaeal sequences, along with the GRCh38 human genome. Sequences attributed to specific biological agents are isolated for further analysis. The current study focuses on *Bacillus anthracis* due to its significance as a biothreat agent. Then, the selected sequences are mapped against a SILVA database (available at https://www.arb-silva.de/documentation/release-138/). Repetitive sequences and tRNAs are masked and removed from the dataset to eliminate non-specific data. Subsequent to this filtration, the remaining reads are aligned to the genome of *Bacillus anthracis* Ames Ancestor (GCF_000008445.1), selecting for reads that align with no more than one mismatch. Then, the selected sequences are mapped using MegaBLAST [21] (optimized for identification of highly similar sequences and intra-species comparisons) against a local nt database retrieved as of November 26th, 2022. The reads with the highest scores for *Bacillus anthracis* are kept. The final step involves converting these selected reads back into FASTQ format, which then allows detailed phylogenetic analysis and genetic characterization.

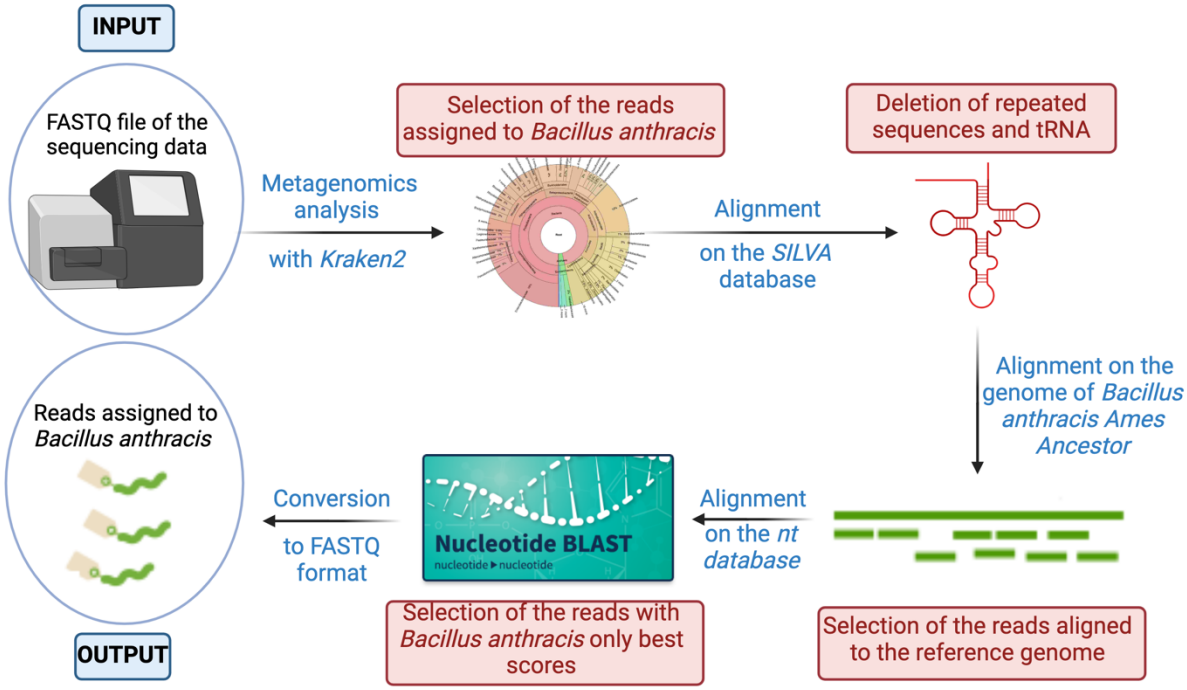

Figure 1: Overview of B2forensics_v1 workflow

Position of the reads at the output of the pipeline with respect to *Bacillus anthracis*

For this study, 181 strains of *Bacillus anthracis* that represent the diversity of the species (with strains from groups A, B, and C) were used to create a phylogenetic tree.

Due to various assembly stages such as complete genomes, scaffolds, and contigs, the data were transformed into simulated reads through a in-house Python script (v3.6.2) and subsequently aligned to the *Bacillus anthracis* Ames Ancestor chromosome (GCF_000008445.1). This alignment, which produced a set of identical length, was utilized for genomic SNP analysis.

Whole genome SNP analysis with BioNumerics 8.1.1 used the option “Strict SNP filtering (Closed SNP set)” with default parameters (12 bp inter-SNP), and the detected SNPs were used for population clustering using the “Maximum parsimony tree” (MPT) calculation method with default parameters. The details of the strains in this database are listed in Supplementary file 2.

To assign the reads to specific parts of the tree, a custom Kraken2 database based on this phylogenetic tree was created.

Results:

Quantification of the pure culture

SYBR Green qRT-PCR targeting the universal bacterial 16S rRNA gene was used to assess the concentration (CFU/ml) of each spiking material cell suspension, separately.

The qRT-PCR revealed a concentration of 1,010 CFU/ml per each strain, except for *Bacillus cereus* ATCC 10987 cell suspension showing about 108 CFU/ml. The melting curve with only one peak around 85 °C for *Francisella tularensis* and 87 °C for all the other strains indicated the specificity of the amplification (data not shown).

Comparison of DNA extraction methods: quality and quantity

Concentrations of TC-DNA extracted with the DNeasy Power water kit and phenol-chloroform methods are reported in Table 1, together with the 260/230 and 260/280 values.

Table 1: Comparison between DNeasy Power water kit and phenol-chlorophorm extraction methods

	DNeasy Power water kit		**Phenol-chloroform**	
	B2F14	***B2F13***	***B2F16***	***B2F15***
DNA concentration	28.6 ng/ul	22.8 ng/ul	14.2 ng/ul	24.0 ng/ul
1.8<260/230<2.2	1.83	1.53	1.37	1.83
1.8<260/280<2	1.93	1.84	1.86	1.93

Sewage-in water deep-sequencing (NextSeq) and computational analyses

TC-DNA from sewage-in water samples was sequenced using NextSeq Illumina technology. A summary of the sequenced data is presented in Table 2. The displayed file size is calculated by adding the sizes of the two associated files R1 and R2.

Table 2: Summary of sequenced data

Sample name	**File size**	**Number of paired reads**
B2F05	43,6 Go	468,968,505
B2F10	48,8 Go	530,117,726
B2F12	21,2 Go	247,293,466
B2F13	48,9 Go	274,828,342
B2F14	52,4 Go	306,218,669
B2F15	59,2 Go	321,871,337
B2F16	62,8 Go	327,550,869

A main approach has been evaluated. It is an assembly-free approach. Individual reads are independently assigned to a taxonomic level.

Assembly-free approach

Specificity

First of all, three samples were analysed with the pipeline B2Forensics_v1: B2F05, B2F10 and B2F12. Table 3 indicates the number of reads for each agent of interest: *Bacillus anthracis, Bacillus thuringiensis, Bacillus cereus, Francisella tularensis, Yersinia pestis, Brucella melitensis* and *Francisella hispaniensis*.

Table 3: Analysis of samples using B2forensics_v1 pipeline. *BT: *Bacillus thuringiensis*, BC: *Bacillus cereus*, FH: *Francisella hispaniensis*, BA: *Bacillus anthracis*, FT: *Francisella tularensis*, BM: *Brucella melitensis*, YP: *Yersinia pestis*

Sample	Spiking*	BT	BC	FH	BA	FT	BM	YP
B2F05	BT-BC-FH	2,030,032	309	13,293	24	0	0	2
B2F10	BT-BC-FH	88,062	176	4,776	1	0	1	0
B2F12	Unspiked	0	0	0	0	0	1	2

A few BA candidate reads are detected in B2F05 and B2F10, in proportion with the BT spiking, reflecting statistically expected sequencing errors among the very high number of BT reads [8]. The *Yersinia pestis* candidates in B2F05 and B2F12 are not specific to *Yersinia pestis*, i.e. they equally match other species in Genbank. The same behaviour is observed for the other unexpected candidates in Table 2.

Sensitivity

The pipeline was applied to the four sequence files spiked or not with real agents (B2F13 to B2F16). No suspicious reads were identified in the unspiked samples, B2F13 and B2F15. Four of the five spikes were detected in B2F14 and B2F16 (Table 3). The relative amount of reads is in agreement with the initial spiking level, with one exception. The *Yersinia pestis* spike is not detected in the sequence data. The qPCR detection confirms that the amount of detectable *Yersinia pestis* DNA is 50 to 100-fold lower than expected (Table 3).

The amount of reads from the spikes is lower in B2F16, by a ratio of two to five. This might be related to the two-fold larger DNA yield in B2F16 as compared to B2F14. If both extraction methods were equivalent in extracting the spikes, method 2 was more efficient in extracting DNA from the biological background. Then the spikes will represent a lower proportion of the sequenced DNA. In the present context, extraction method 1 is providing a higher sensitivity compared to extraction method 2.

Table 3: Detection of the spiking material in sewage water by the pipeline and comparison with qPCR detection

Spike	16S qPCR spike titration on pure culture	Sample	Specific qPCR on extracted DNA from spiked sample	Number of reads detected by B2forensics_v1 pipeline
BC	$1x10^7$	*B2F14*	$1x10^6$	91,663
		B2F16	$2.6x10^6$	48,750
BA	$4.49x10^5$	*B2F14*	$2.72x10^5$	942
		B2F16	$1.25x10^5$	230
FT	$3.05x10^5$	*B2F14*	$9.1x10^4$	541
		B2F16	$2.08x10^5$	104
BM	$1.2x10^3$	*B2F14*	$5.4x10^2$	59
		B2F16	$1.1x10^3$	12
YP	$9.4x10^2$	*B2F14*	13	0
		B2F16	17	0

Position of the reads in B2F14 and B2F05 at the output of the pipeline with respect to the corresponding B-agent

B2Forensics_v1 based upon k-mer filtration and additional alignment-based filtering is not sufficient to solve all false positive issues, and that a final step of phylogenetic assignment of the candidate reads at subspecies level is required. To illustrate this step, we have mapped the reads detected in samples B2F14 to intra-species level, using an adapted Kraken2 custom database. Figure 3 illustrates where the reads fit into the *Bacillus anthracis* tree for B2F14. The assignment is in agreement with the known location of the strain used as spike, *B. anthracis* strain *Ferrara* (run accession SRR10382389). The candidate reads in B2F14 are assigned monophyletically i.e. they are distributed on the same path of the phylogenetic tree. This is consistent with the actual presence of a *Bacillus anthracis* strain in the sample. In contrast, the candidate reads in B2F05 are not monophyletic, as expected for sequencing artefacts, and indicating the absence of *Bacillus anthracis* in the sample (Figure 4).

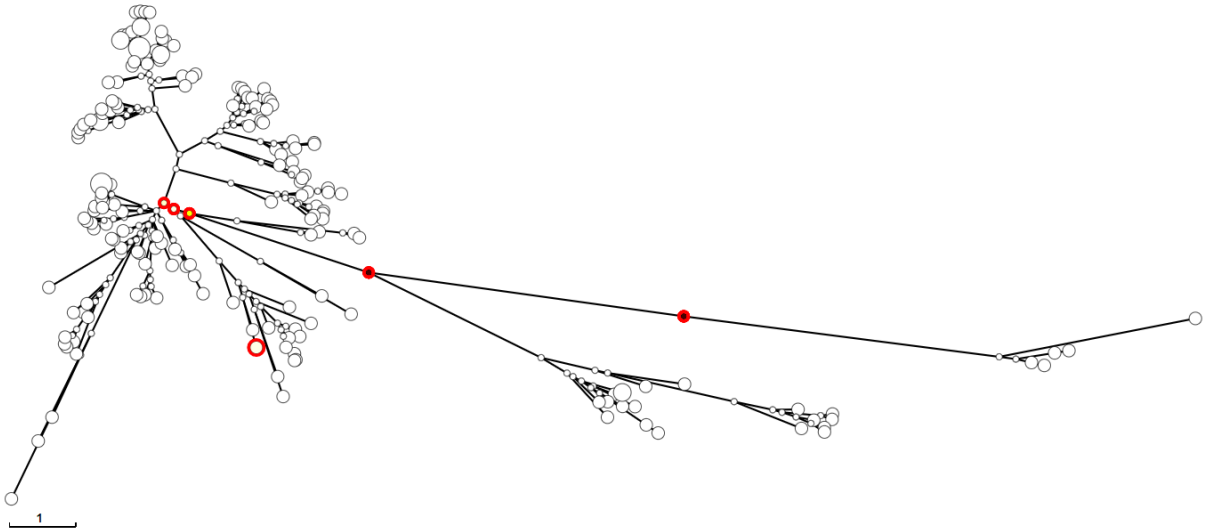

Figure 3: Assignment of the candidate *B. anthracis* reads detected in B2F14 within the *Bacillus anthracis* population (nodes with assigned reads are surrounded in red).

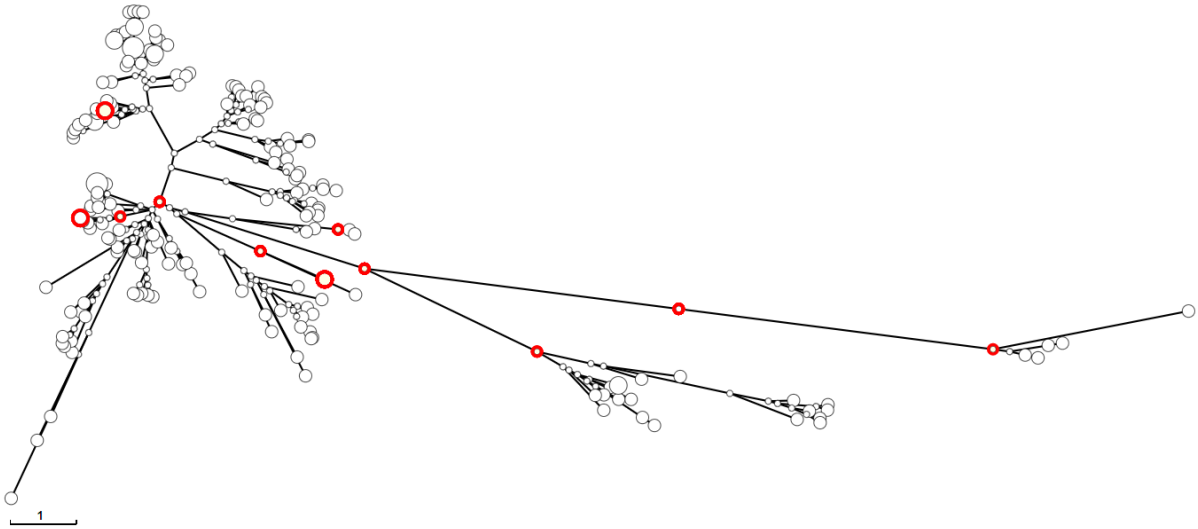

Figure 4: Assignment of the candidate *B. anthracis* reads detected in B2F05 within the B. anthracis population (nodes with assigned reads are surrounded in red).

Discussion and conclusion:

In conclusion, the application of next-generation sequencing (NGS) technologies to the investigation of metagenomic samples represents a significant advancement in the detection and characterization of complex biological agents. This study reaffirms the importance of selecting appropriate DNA extraction methods tailored to the matrix under investigation. Despite the demonstrated efficiency of the two extraction methods employed, which are both suitable for spores, Gram-positive, and Gram-negative organisms, it is crucial to acknowledge the loss of one order of magnitude of the initial material, as evidenced by control qPCR evaluations. To mitigate this reduction in quantity and to obtain a comprehensive genetic profile, it is necessary to generate a higher volume of sequencing data. Such compensation is essential for accurately representing the sample's complexity.

Furthermore, to achieve a detailed examination through sequencing, a substantial output is required, such as that provided by NextSeq or similar platforms. However, this necessitates increased financial investment, which currently limits the feasibility of implementing NGS for routine surveillance on a large scale. Given the cost implications, targeted surveillance of high-risk agents remains a more practical approach under the current technological and economical conditions. This strategy may persist until such time that political decisions or advancements in technology warrant the widespread adoption of NGS in routine surveillance operations.

When investigating NGS, we could show, that using the k-mer approach is an effective way to solve the challenge of detecting traces of an agent of interest in a metagenomic sample. The pipeline *B2Forensics_v1* meets this objective in the specific case of *Bacillus anthracis*. More generally, we can extend the method used in this case to other species and opens up the possibility of fine-scale detection in a metagenome.

This study led to the development of a two-steps approach. The first is the filtering and extraction of candidate reads. The quality of this step is highly dependent upon the knowledge of the nearest neighbours of the pathogenic species of interest. The second is the phylogenetic assignment of the candidates, which is an essential confirmatory test. The quality of this second step is highly dependent upon the quality of the coverage of the pathogen of interest. This last step is crucial to avoid detection of false positives. For instance, Afshinnekoo et al. [4] achieved a massive characterization of metagenome collected in the New York City subway network. The authors produced more than 1,500 surface samples from all the stations, extracted DNA and generated ten billion reads of sequencing data. They suspected the presence of *Bacillus anthracis* and *Yersinia pestis* in two samples. The investigation of the phylogenetic position of these reads with respect to the *Bacillus anthracis* population led to the following result. No reads containing a known intra specific single nucleotide polymorphism (SNP) could be identified, meaning that, in agreement with [8], no read is providing evidence for the presence of a *bona fide B. anthracis* sublineage, and the observation could be indicative of the presence of a close neighbour. Beyond that, even if specific pipelines exist, bioinformatics experts are still needed in combination with agent-specialist for interpretation of data as the results of detection of biothreat agents have wide impact on political decision.

Our study also contributed to the production of a complete benchmark datasets with real agents, which will be useful for benchmarking future developments. It will be very useful to evaluate and compare different analytical tools specific to the detection of a pathogen (*Bacillus anthracis* in our case). In addition, the methodology used to create this dataset can serve as a model for establishing other datasets specific to other pathogens.

References:

1. Kuehn, A.; Kovac, P.; Saksena, R.; Bannert, N.; Klee, S.R.; Ranisch, H.; Grunow, R. Development of antibodies against anthrose tetrasaccharide for specific detection of Bacillus anthracis spores. *Clin Vaccine Immunol* **2009**, *16*, 1728-1737, doi:10.1128/CVI.00235-09.
2. Sunagawa, S.; Acinas, S.G.; Bork, P.; Bowler, C.; Tara Oceans, C.; Eveillard, D.; Gorsky, G.; Guidi, L.; Iudicone, D.; Karsenti, E.; et al. Tara Oceans: towards global ocean ecosystems biology. *Nat Rev Microbiol* **2020**, *18*, 428-445, doi:10.1038/s41579-020-0364-5.
3. Plaire, D.; Puaud, S.; Marsolier-Kergoat, M.C.; Elalouf, J.M. Comparative analysis of the sensitivity of metagenomic sequencing and PCR to detect a biowarfare simulant (Bacillus atrophaeus) in soil samples. *PLoS One* **2017**, *12*, e0177112, doi:10.1371/journal.pone.0177112.
4. Afshinnekoo, E.; Meydan, C.; Chowdhury, S.; Jaroudi, D.; Boyer, C.; Bernstein, N.; Maritz, J.M.; Reeves, D.; Gandara, J.; Chhangawala, S.; et al. Geospatial Resolution of Human and Bacterial Diversity with City-Scale Metagenomics. *Cell Syst* **2015**, *1*, 72-87, doi:10.1016/j.cels.2015.01.001.
5. Wood, D.E.; Lu, J.; Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol* **2019**, *20*, 257, doi:10.1186/s13059-019-1891-0.
6. Brown, C.T.; Irber, L. sourmash: a library for MinHash sketching of DNA. *Journal of open source software* **2016**, *1*, 27.
7. Ackelsberg, J.; Rakeman, J.; Hughes, S.; Petersen, J.; Mead, P.; Schriefer, M.; Kingry, L.; Hoffmaster, A.; Gee, J.E. Lack of Evidence for Plague or Anthrax on the New York City Subway. *Cell Syst* **2015**, *1*, 4-5, doi:10.1016/j.cels.2015.07.008.
8. Petit Iii, R.A.; Hogan, J.M.; Ezewudo, M.N.; Joseph, S.J.; Read, T.D. Fine-scale differentiation between Bacillus anthracis and Bacillus cereus group signatures in metagenome shotgun data. *PeerJ* **2018**, *6*, e5515, doi:10.7717/peerj.5515.
9. 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.
10. 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.
11. Walker, B.J.; Abeel, T.; Shea, T.; Priest, M.; Abouelliel, A.; Sakthikumar, S.; Cuomo, C.A.; Zeng, Q.; Wortman, J.; Young, S.K.; et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **2014**, *9*, e112963, doi:10.1371/journal.pone.0112963.
12. Darling, A.C.; Mau, B.; Blattner, F.R.; Perna, N.T. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* **2004**, *14*, 1394-1403, doi:10.1101/gr.2289704.
13. Riehm, J.M.; Rahalison, L.; Scholz, H.C.; Thoma, B.; Pfeffer, M.; Razanakoto, L.M.; Al Dahouk, S.; Neubauer, H.; Tomaso, H. Detection of Yersinia pestis using real-time PCR in patients with suspected bubonic plague. *Mol Cell Probes* **2011**, *25*, 8-12, doi:10.1016/j.mcp.2010.09.002.
14. Tomaso, H.; Jacob, D.; Eickhoff, M.; Scholz, H.C.; Al Dahouk, S.; Kattar, M.M.; Reischl, U.; Plicka, H.; Olsen, J.S.; Nikkari, S.; et al. Preliminary validation of real-time PCR assays for the identification of Yersinia pestis. *Clin Chem Lab Med* **2008**, *46*, 1239-1244, doi:10.1515/CCLM.2008.251.
15. Hinic, V.; Brodard, I.; Thomann, A.; Cvetnic, Z.; Makaya, P.V.; Frey, J.; Abril, C. Novel identification and differentiation of Brucella melitensis, B. abortus, B. suis, B. ovis, B. canis, and B. neotomae suitable for both conventional and real-time PCR systems. *J Microbiol Methods* **2008**, *75*, 375-378, doi:10.1016/j.mimet.2008.07.002.

16. Antwerpen, M.H.; Zimmermann, P.; Bewley, K.; Frangoulidis, D.; Meyer, H. Real-time PCR system targeting a chromosomal marker specific for Bacillus anthracis. *Mol Cell Probes* **2008**, *22*, 313-315, doi:10.1016/j.mcp.2008.06.001.
17. Koster, J.; Rahmann, S. Snakemake--a scalable bioinformatics workflow engine. *Bioinformatics* **2012**, *28*, 2520-2522, doi:10.1093/bioinformatics/bts480.
18. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R.; Genome Project Data Processing, S. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078-2079, doi:10.1093/bioinformatics/btp352.
19. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754-1760, doi:10.1093/bioinformatics/btp324.
20. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J Mol Biol* **1990**, *215*, 403-410, doi:10.1016/S0022-2836(05)80360-2.
21. Morgulis, A.; Coulouris, G.; Raytselis, Y.; Madden, T.L.; Agarwala, R.; Schaffer, A.A. Database indexing for production MegaBLAST searches. *Bioinformatics* **2008**, *24*, 1757-1764, doi:10.1093/bioinformatics/btn322.

5.5 Mise en place du pipeline B2FORENSICS_V2

Dans l'étude précédente, la mise en place du pipeline d'analyse B2FORENSICS_V1 a permis de traiter la question de spécificité et de sensibilité à des fins de détection de *B. anthracis*. Dans cette section, la notion de temps d'exécution de ce pipeline va être abordée. En effet, pour permettre la recherche en routine de traces de *B. anthracis* dans un échantillon environnemental, l'outil doit être en mesure de traiter un volume important de données de séquençage en un temps restreint, tout en gardant le même niveau de fiabilité. Cette section vise à présenter les travaux de mise à jour du pipeline B2FORENSICS_V1 à cette fin.

5.5.1 Temps d'exécution sur les données tests

Une comparaison des temps d'analyse pris par B2FORENSICS_V1 et différents outils a été effectuée. Trois échantillons de données tests ont été pris en compte : B2F13, B2F14 et B2F05 dont les caractéristiques sont rappelées dans le tableau 5.2.

Échantillon	Technologie de séquençage	Taille du fichier (*paired-end*)	Taille des lectures	Nombre de lectures	*Spiking* de *B. anthracis* Ferrara	*Spiking* de *B. cereus* ATCC 10987	*Spiking* de *B. thuringiensis var. kurstaki*
B2F13	NGS	48.8 Go	150	274,828,342	0	0	0
B2F14	NGS	52.3 Go	150	306,218,669	10^5	10^7	0
B2F05	NGS	43.6 Go	80	468,968,505	0	10^4	10^7

TABLE 5.2 – Tableau récapitulatif des caractéristiques des échantillons étudiés pour l'évaluation de B2FORENSICS_V2. Les valeurs de *spiking* (ensemencement ou contamination) indiquées correspondent à un nombre de bactéries introduites dans l'échantillon avant l'étape d'extraction d'ADN.

Ces trois échantillons ont été sélectionnés car ils correspondent respectivement à un échantillon témoin (B2F13, absence de *B. anthracis* et d'espèces voisines), un échantillon positif (B2F14, présence de *B. anthracis* et de *B. cereus*) et un échantillon négatif (B2F05, prélèvement indépendant, absence de *B. anthracis* mais présence d'espèces voisines). En définitive, les outils suivants ont été comparés :

— B2FORENSICS_V1

— KRAKENUNIQ (BREITWIESER, BAKER et SALZBERG, 2018), du fait de son adaptation particulière de KRAKEN (WOOD et SALZBERG, 2014) pour réduire le nombre de faux positifs

— le comptage des sets de k-mers *Ba31*, *BCerG31* et *lef31* établi par PETIT III et al., 2018, en utilisant le compteur de k-mers JELLYFISH (MARÇAIS et KINGSFORD, 2011)

Les temps d'exécution sont présentés dans le tableau 5.3.

Plus particulièrement, les temps d'exécution de chacune des étapes du pipeline B2FORENSICS_V1 pour les échantillons B2F13, B2F14 et B2F05 sont détaillés dans le tableau 5.4. Les étapes y sont numérotées de la manière suivante :

— Étape 1 : Assignation taxonomique par KRAKEN2 et sélection des lectures assignées à *B. anthracis*

Échantillon	B2FORENSICS_V1	KRAKENUNIQ	Sets de k-mers (*Ba31*, *BcerG31*, *lef31*)
B2F13	1h 59min	1h 09min	11h 58min
B2F14	3h 38min	1h 12min	11h 36min
B2F05	13h 04min	1h 18min	12h 22min

TABLE 5.3 – Tableau comparatif des temps d'exécution.

— Étape 2 : Alignement sur la base SILVA pour la suppression des séquences répétées, ARNr, ARNt et alignement sur un génome de référence de *B. anthracis* (Ames ancestor)

— Étape 3 : Alignement sur la base *nt* de NCBI à l'aide de MEGABLAST pour la suppression des séquences non relatives à *B. anthracis*.

Échantillon	Etape 1	Etape 2	Etape 3	Total
B2F13	44min	1h 15min	0 min	1h 59min
B2F14	48min	1h 48min	1h 02min	3h 38min
B2F05	55min	5h 22min	6h 47min	13h 04min

TABLE 5.4 – Tableau comparatif des temps d'analyse des échantillons B2F13, B2F14 et B2F05 par le pipeline B2FORENSICS_V1 à chacune des étapes.

Plusieurs observations qualitatives en sont tirées :

— Le comptage des sets de k-mers est chronophage. De plus, elle nécessite d'être réalisée trois fois pour un échantillon donné, pour chacun des sets de k-mers.

— L'étape 1 est la moins longue, du fait de la rapidité de l'assignation taxonomique par l'approche k-mer. En outre, le temps d'analyse par KRAKENUNIQ est plus élevé que celui de KRAKEN2.

— Si le temps d'exécution de B2FORENSICS_V1 est plus élevé pour B2F14 que pour B2F13, cela s'explique par les étapes d'alignements (Etapes 2 et 3) qui doivent être réalisées dans le cas d'un échantillon contenant *B. anthracis*.

— Dans le cas de B2F05, le temps d'exécution est beaucoup plus élevé que B2F13 et B2F14 du fait des étapes d'alignements. Pourtant, cet échantillon ne contient pas *B. anthracis* mais des espèces voisines (*B. cereus* et *B. thuringiensis*). La première étape de sélection après assignation taxonomique par KRAKEN2 semble donc peu sélective.

Un axe d'amélioration majeure est donc de rendre plus sélective la première étape du pipeline afin de réduire le nombre de lectures en sortie qui devront passer les étapes d'alignement. Pour ce faire, nous allons tirer parti de la rapidité d'exécution de KRAKEN2 en ajoutant une étape supplémentaire après l'étape 1, qui sera l'utilisation d'une *custom database*.

5.5.2 Étude 1 : Ajout d'une *custom database* spécifique à *Bacillus anthracis*

L'objectif de cette section est de rendre plus sélective la première étape d'assignation taxonomique du pipeline. Cependant, si le but recherché est un gain significatif de rapidité, il ne faudrait pas qu'il soit atteint au prix d'une perte de sensibilité (c'est-à-dire en supprimant par inadvertance à l'étape 1 des lectures spécifiques à *B. anthracis*).

Pour cela, deux modifications à B2FORENSICS sont opérées :

— L'étape 1 du pipeline est modifiée : les lectures sélectionnées à l'issue de celle-ci ne seront plus celles assignées à *B. anthracis* mais plus généralement au groupe *B. cereus*. La sélection plus fine se fera à l'étape suivante, et conserver cet ensemble plus large de lectures permet de réduire également le risque de faux négatifs.

— Les séquences sélectionnées seront ensuite analysées à l'aide de la *custom database* spécifique à *B. anthracis* (contenant 181 souches). Seules les lectures assignées à un noeud quelconque de l'arbre phylogénétique en aval du MRCA de *B. anthracis* seront conservées pour les étapes suivantes.

Après cette modification, le nombre de lectures à l'issue de chaque étape du pipeline est présenté dans les tableaux 5.5, 5.6 et 5.7 pour les échantillons B2F13, B2F14 et B2F05 respectivement.

Étapes	B2FORENSICS_V1	B2FORENSICS avec *custom database* spécifique à *B. anthracis*
1	1,930	1,930
2	/	519
3	0	0
4	0	0

TABLE 5.5 – Tableau comparatif du nombre de lectures sélectionnées à chaque étape des versions du pipeline B2FORENSICS pour l'échantillon B2F13.

Étapes	B2FORENSICS_V1	B2FORENSICS avec *custom database* spécifique à *B. anthracis*
1	98,190	98,190
2	/	78,021
3	14,810	14,808
4	942	942

TABLE 5.6 – Tableau comparatif du nombre de lectures sélectionnées à chaque étape des versions du pipeline B2FORENSICS pour l'échantillon B2F14.

Les résultats sont les suivants :

Étapes	B2FORENSICS_V1	B2FORENSICS avec *custom database* spécifique à *B. anthracis*
1	8,743,815	8,743,815
2	/	5,793,129
3	409,576	409,570
4	24	24

TABLE 5.7 – Tableau comparatif du nombre de lectures sélectionnées à chaque étape des versions du pipeline B2FORENSICS pour l'échantillon B2F05.

— En sortie des deux versions du pipeline, le nombre de lectures est identique pour chacun des échantillons. De plus, après vérification, les lectures sélectionnées sont bien identiques. Il n'y a donc pas perte de sensibilité.

— L'étape supplémentaire ajoutée (étape 2 dans les tableaux ci-dessus) réduit le nombre de lectures sélectionnées avant les étapes d'alignement. Quantitativement, cela correspond à une réduction de 73%, 20% et 34% pour les échantillons B2F13, B2F14 et B2F05 respectivement, en comparaison avec B2FORENSICS_V1.

— Il reste encore un nombre élevé de lectures à traiter après l'utilisation de la *custom database* pour l'échantillon B2F05, ne comportant pourtant pas de *B. anthracis* (mais des espèces voisines). De même pour l'échantillon B2F14, pour lequel la réduction des lectures sélectionnées à l'issue de l'étape 2 est modérée.

L'enjeu est donc maintenant de modifier la *custom database* utilisée afin qu'elle permette une élimination plus efficace des lectures spécifiques aux espèces voisines de *B. anthracis*.

5.5.3 Étude 2 : Optimisation avec une *custom database* élargie

Pour l'échantillon B2F14, on observe sur la figure 5.11 l'assignation taxonomique des lectures sélectionnées lors de l'utilisation de la *custom database* spécifique à *B. anthracis*. Il s'avère que sur les 78,021 lectures retenues à l'issue de cette étape, 72,078 d'entre elles sont assignées à la racine (MRCA) de l'arbre phylogénétique de *B. anthracis*, soit un total de 92%.

Pour vérifier si ces lectures sont spécifiques à *B. anthracis* ou sont également communes à d'autres espèces voisines, il faut intégrer des souches du groupe *B. cereus* proches de *B. anthracis* dans l'arbre phylogénétique utilisé pour la *custom database*.

Une nouvelle *custom database* est donc créée : nous la nommerons "*custom database* élargie 1". Elle contient :

— un miniset de souches de *B. anthracis*, composé de 22 souches représentatives de la diversité de l'espèce (groupes A, B et C en proportions de leur abondance)

— les souches du groupe *B. cereus*, dont les génomes ont été téléchargés sur NCBI (en janvier 2023) et dont l'alignement sur le chromosome de *B. anthracis* Ames Ancestor (GCF_000008445.1) présente au moins 70% de similitude (autrement dit, moins de 30% de régions non reconstruites)

Les résultats obtenus pour les échantillons étudiés sont présentés dans les tableaux 5.8, 5.9 et 5.10.

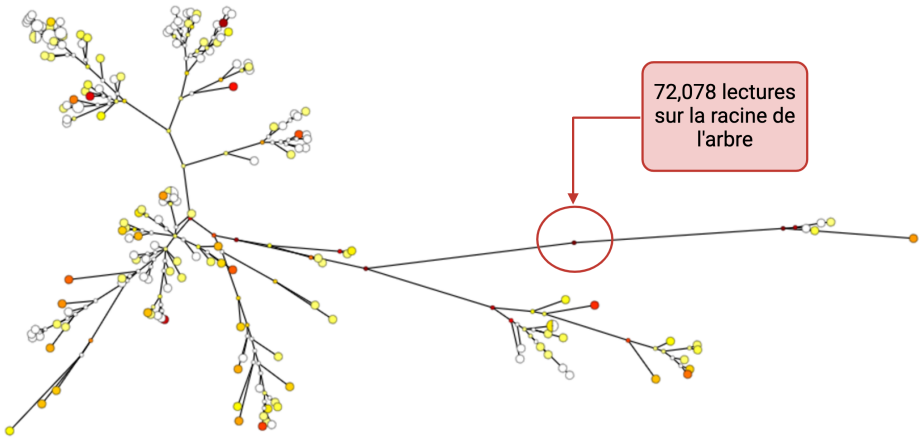

FIGURE 5.11 – Assignation taxonomique des lectures de l'échantillon B2F14 lors de l'étape 2 du pipeline.

Étapes	B2FORENSICS_V1	B2FORENSICS avec *custom database* spécifique à *B. anthracis*	B2FORENSICS avec *custom database* élargie 1
1	1,930	1,930	1,930
2	/	519	9
3	0	0	0
4	0	0	0

TABLE 5.8 – Tableau comparatif du nombre de lectures sélectionnées à chaque étape des versions du pipeline B2FORENSICS pour l'échantillon B2F13.

Étapes	B2FORENSICS_V1	B2FORENSICS avec *custom database* spécifique à *B. anthracis*	B2FORENSICS avec *custom database* élargie 1
1	98,190	98,190	98,190
2	/	78,021	2,112
3	14,810	14,808	978
4	942	942	942

TABLE 5.9 – Tableau comparatif du nombre de lectures sélectionnées à chaque étape des versions du pipeline B2FORENSICS pour l'échantillon B2F14.

Étapes	B2FORENSICS_V1	B2FORENSICS avec *custom database* spécifique à *B. anthracis*	B2FORENSICS avec *custom database* élargie 1
1	8,743,815	8,743,815	8,743,815
2	/	5,793,129	200,011
3	409,576	409,570	17,306
4	24	24	19

TABLE 5.10 – Tableau comparatif du nombre de lectures sélectionnées à chaque étape des versions du pipeline B2FORENSICS pour l'échantillon B2F05.

Plusieurs remarques sont à faire :

— En termes de résultats en sortie de pipeline : pour l'échantillon B2F14 positif à *B. anthracis*, le nombre de lectures en sortie du pipeline ne varie pas et les lectures sont identiques à celles obtenues en sortie précédemment. Il n'y a donc pas de perte en sensibilité. Pour l'échantillon négatif B2F13, il n'y a pas de faux positif en sortie. Pour l'échantillon négatif B2F05 (contenant des espèces voisines), le nombre de lectures en sortie reste sensiblement constant (une baisse de 24 à 19 lectures), toujours en cohérence avec le modèle de génération de faux positifs de *B. anthracis* en fonction de la couverture du groupe *B. cereus* dans l'échantillon (PETIT III et al., 2018).

— En termes de réduction de lectures candidates sélectionnées à l'issue de l'étape 2 : il y a une réduction de 99.5%, 97.8% et 97.7% entre le nombre de lectures retenues à l'issue de l'étape 1 et de l'étape 2 pour les échantillons B2F13, B2F14 et B2F05 respectivement. Pour rappel, ces taux étaient de 73%, 20% et 34% avec la *custom database* contenant seulement des souches de *B. anthracis*.

5.5.4 Étude 3 : Comparaison de *custom databases* élargies

Une dernière adaptation est évaluée. Ici, la *custom database* élargie 1 contient les souches du groupe *B. cereus* présentant au moins 70% de similitude avec le chromosome de *B. anthracis* Ames ancestor. Testons à présent le pipeline d'analyse avec deux autres *custom databases* élargies à deux degrés plus élevés, en incluant les souches du groupe *B. cereus* ayant au moins 50% et 30% de similitude respectivement avec le même génome de référence de *B. anthracis*. Nous nommerons ces *custom databases* de la manière suivante : *custom database* élargie 2 (pour le seuil de 50%) et *custom database* élargie 3 (pour le seuil de 30%).

L'objectif de cette évaluation est de comprendre jusqu'à quel seuil nous pouvons augmenter le nombre de souches du groupe *B. cereus* au sein de la *custom database* élargie, pour réduire le nombre de lectures sélectionnées à l'issue de l'étape 2 du pipeline (et ainsi gagner en temps d'exécution) tout en ne perdant pas en performance (en termes de sensibilité et spécificité).

Les résultats pour les deux échantillons B2F14 et B2F05 sont présentés dans les tableaux 5.11 et 5.12.

On en déduit donc :

— Pour la *custom database* élargie 2 : après analyse de l'échantillon B2F14 contenant *B. anthracis*, il y a une perte de 43% de lectures spécifiques à *B. anthracis*

Étapes	B2FORENSICS avec *custom database* élargie 1	B2FORENSICS avec *custom database* élargie 2	B2FORENSICS avec *custom database* élargie 3
1	98,190	98,190	98,190
2	2,112	817	1,505
3	978	544	1,422
4	942	536	298

TABLE 5.11 – Tableau comparatif du nombre de lectures sélectionnées à chaque étape des versions du pipeline B2FORENSICS pour l'échantillon B2F14.

Étapes	B2FORENSICS avec *custom database* élargie 1	B2FORENSICS avec *custom database* élargie 2	B2FORENSICS avec *custom database* élargie 3
1	8,743,815	8,743,815	8,743,815
2	200,011	36,225	33,002
3	17,306	13,405	4,101
4	19	17	7

TABLE 5.12 – Tableau comparatif du nombre de lectures sélectionnées à chaque étape des versions du pipeline B2FORENSICS pour l'échantillon B2F05.

en sortie de pipeline, ces lectures étant incluses dans l'ensemble des 942 lectures précédemment obtenues en sortie. Il y a donc perte de sensibilité. Pour confirmer ce résultat et infirmer l'hypothèse que ce serait l'inverse (c'est-à-dire que sur les 942 lectures de séquençage, certaines ne sont pas spécifiques à *B. anthracis*), une vérification qualitative a été effectuée en enrichissant *in silico* l'échantillon négatif B2F13. Pour ce faire, des lectures de séquençages ont été simulées avec l'outil ART (HUANG et al., 2012), avec les paramètres par défaut, et le taux d'erreur de la technologie Illumina. Un *mapping* du set de k-mers *Ba31* a été effectué sur les lectures simulées et une sélection aléatoire de 1,000 lectures parmi celles dont un nombre non nul de k-mers s'y alignaient a été ajouté dans le séquençage de B2F13. Après analyse par les trois versions du pipeline, la *custom database* élargie 1 permet la détection de l'intégralité des lectures, contrairement aux deux autres cas.

— Pour la *custom database* élargie 3 : il y a cette fois-ci une réduction de 68% de lectures en sortie de pipeline pour l'échantillon B2F14. Cette fois-ci, ces lectures ne sont pas toutes incluses dans l'ensemble des 942 lectures obtenues. Pour l'échantillon B2F05, ce phénomène se reproduit : les sept lectures obtenues en sortie de pipeline ne sont pas incluses dans les 19 obtenues précédemment. Cela s'explique par le fait que l'arbre phylogénétique associé à la *custom database* n'est plus représentatif de la véritable phylogénie de *B. anthracis* et de son proche voisinage. En effet, le nombre de souches de *B. anthracis* (22) est

trop faible en comparaison du nombre de souches voisines incluses. Ce phénomène est observable sur la figure 5.12 sur laquelle on peut lire qualitativement l'accroissement important des souches à considérer selon les seuils de 30%, 50% et 70%. De plus, les séquences ayant servies à établir l'arbre sont basées sur un alignement avec la référence *B. anthracis* Ames ancestor. Dans ce cas, les pourcentages de nucléotides indéterminés dans les alignements des souches prises en compte (directement corrélés au seuil de similitude fixé trop largement) peuvent atteindre 70% de la taille du génome. Ainsi, l'assignation taxonomique des lectures de séquençage effectué par KRAKEN2 s'en retrouve faussée.

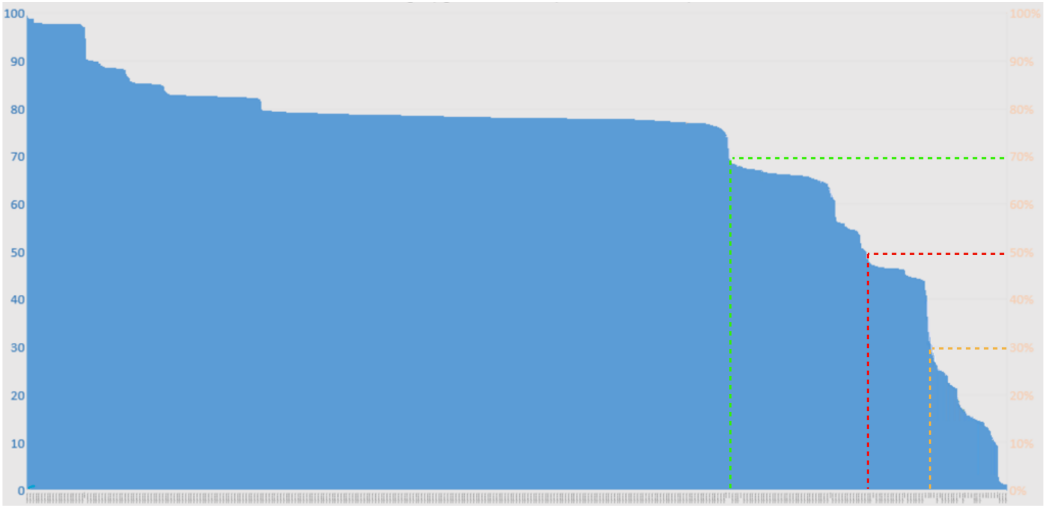

FIGURE 5.12 – Histogramme des pourcentages de N (nucléotides indéterminées) des génomes du groupe *B. cereus* après alignement sur le chromosome de *B. anthracis* Ames ancestor. En orange : limite de 30% de N ; en rouge : limite de 50% de N ; en vert : limite de 70% de N. Les souches à plus de 99% de N ne sont pas représentées par souci de lisibilité.

5.5.5 Bilan : le pipeline B2FORENSICS_V2

En définitive, la *custom database* 1 est la mieux adaptée pour gagner en temps d'exécution lors de l'analyse de séquençage par le pipeline B2FORENSICS, sans perte de performance en termes de spécificité ou sensibilité. Ce pipeline mis à jour est nommé B2FORENSICS_V2.

La comparaison définitive de temps d'exécution pour les trois échantillons B2F13, B2F14 et B2F05 est présenté dans le tableau 5.13.

Échantillon	B2FORENSICS_V1	B2FORENSICS_V2
B2F13	1h 59min	49min
B2F14	3h 38min	57min
B2F05	13h 04min	2h 06min

TABLE 5.13 – Tableau comparatif des temps d'exécution des deux versions du pipeline B2FORENSICS.

Le pipeline B2FORENSICS_V2 est un outil de détection de *B. anthracis* dans un métagénome, rapide en comparaison des logiciels de détection existants. En outre, il permet une réduction du risque de faux positifs et sa sensibilité est du même ordre de grandeur qu'une PCR dans l'échantillon biologique considéré. Cependant, selon la concentration d'une espèce voisine de *B. anthracis* dans un échantillon, les erreurs de séquençage peuvent entraîner la détection de certaines lectures comme étant à tort *B. anthracis*. De plus, lorsque la présence de *B. anthracis* est établie, le pipeline ne permet pas de donner l'identité de la ou des souches en présence. Pour ces deux raisons, une ultime étape d'étude phylogénétique des lectures en sortie du pipeline B2FORENSICS_V2 doit être réalisée.

5.5.6 Analyse phylogénétique complémentaire

Une nouvelle *custom database* est créée. Elle est composée de l'ensemble des assemblages (au nombre de 291) et SRA [15] (au nombre de 1,028) de *B. anthracis*, téléchargés sur *RefSeq* (février 2022). Une phylogénie est établie, et les doublons au sein de l'arbre résultant sont supprimés, amenant à un total de 556 souches. En cas d'un nombre non nul de lectures en sortie du pipeline B2FORENSICS_V2, une assignation taxonomique des lectures sélectionnées après l'étape 2 (sélection après analyse par la *custom database* élargie) est opérée. Le choix de sélectionner ces lectures plutôt que seulement celles en sortie de pipeline s'explique par le souhait de ne pas perdre en sensibilité en supprimant des lectures du *core genome* de *B. anthracis* par exemple.

Cette démarche a été testée pour l'échantillon B2F14, contenant la souche *B. anthracis* Ferrara. Le résultat de l'assignation taxonomique est présenté sur la figure 5.13.

On visualise un chemin clair, sans artefact, de la racine de l'arbre vers la polytomie transeurasienne (TEA). Une précision supérieure ne peut être obtenue, du fait de la concentration de *B. anthracis* dans l'échantillon (2,112 lectures, représentant une couverture de 0.008% du génome total).

15. *Sequence Read Archive*

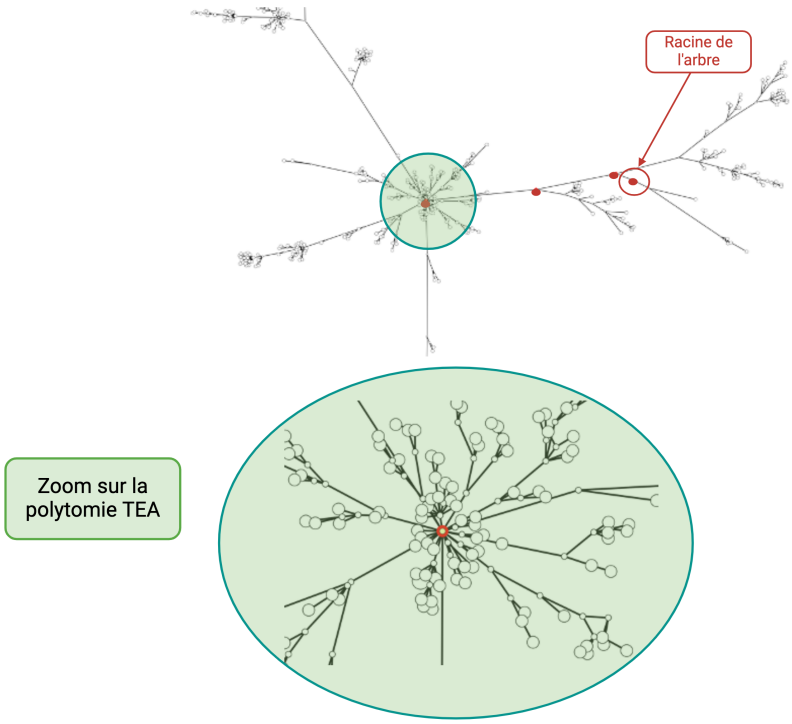

FIGURE 5.13 – Assignation taxonomique des 2,112 lectures assignées à *B. anthracis* de l'échantillon B2F14 pour détermination de la souche en présence. Les noeuds ayant des lectures qui leur sont assignées sont surlignées en rouge.

5.6 Élargissement du pipeline à d'autres agents de la menace et recherche de nouvelles souches

Dans cette section, nous aborderons l'intégration en routine du pipeline B2FORENSICS_V2 pour l'analyse à grande échelle de données métagénomiques, ciblant spécifiquement la détection de *B. anthracis*. Avant son déploiement, le pipeline a été adapté pour inclure la détection d'autres agents pathogènes partageant des traits évolutifs et des enjeux de biodéfense similaires à *B. anthracis*. Cela offre une solution robuste pour la surveillance d'agents pathogènes d'intérêt comparable.

La méthode employée pour la détection de *B. anthracis* est transposable à d'autres espèces posant des problématiques similaires en termes d'identification. *Yersinia pestis* est la bactérie à Gram négatif responsable de l'anthropozoonose appelée peste. Elle fait partie du genre *Yersinia pseudotuberculosis*, composée d'autres espèces pathogènes pour l'Homme (*Yersinia pseudotuberculosis*, *Yersinia enterocolitica*) et d'espèces environnementales, potentiellement pathogènes. De façon comparable à *B. anthracis* et *B. cereus*, *Yersinia pestis* et *Yersinia pseudotuberculosis* sont très proches génétiquement. Un pipeline d'analyse spécifique à la détection de *Y. pestis*, sur le même modèle que B2FORENSICS peut ainsi être envisagé.

Ce pipeline a été réalisé en adaptant B2FORENSICS_V2 au niveau de deux étapes :

— Lors de l'étape 3, l'alignement sur le génome de référence se fait non pas sur *B. anthracis* Ames ancestor mais sur *Y. pestis* CO92 (GCF_000009065.1).

— Lors de l'étape 2, la *custom database* utilisée se base sur un arbre phylogénétique composé de l'ensemble des génomes de *Y. pestis* disponibles sur NCBI (mars 2021), après suppression des doublons. Elle se compose de 370 souches.

Un autre exemple est celui de la bactérie *Francisella tularensis*, responsable de la tularémie, zoonose pouvant se transmettre à l'Homme et également pathogène de catégorie A en termes de menace biologique. Elle se compose en trois sous-espèces : *tularensis* et *holartica* d'une part (pathogènes pour l'Homme) et *mediasiatica* (non pathogène pour l'Homme). *Francisella tularensis* a également émergé récemment d'une espèce environnementale voisine, *Francisella novicida*, avant de suivre une évolution clonale. Le pipeline de détection a été adapté, avec le génome de *Francisella tularensis subsp. tularensis* SCHU S4 (GCF_000008985.1) comme référence et une *custom database* composée de 829 souches de *Francisella tularensis* (génomes extraits de NCBI en mars 2021).

Enfin, une dernière espèce, suscitant un intérêt dans un contexte de biodéfense entre autres, est jointe à cet élargissement. Il s'agit de *Brucella*, la bactérie responsable de la brucellose, zoonose transmissible à l'Homme comme dans les cas mentionnés précédemment (pathogène classé en catégorie B). Ici, pour l'adaptation au pipeline d'analyse, le génome pris pour référence est *Brucella melitensis* bv. 1 str. 16M (GCF_000007125.1). La *custom database* associée est composée de 700 souches (génomes extraits de NCBI en mars 2021).

Pour les trois espèces ajoutées, ce sont donc seulement des souches appartenant à l'espèce en question qui composent la *custom database*, et non pas un miniset de souches auquel est ajouté son proche voisinage. Cela ferait l'objet d'une amélioration possible à des fins de gain en temps d'exécution comme démontré précédemment.

Ces pipelines nouvellement constitués ont été utilisés pour détecter d'éventuelles traces de ces agents pathogènes dans des échantillons métagénomiques. Nous avons récupéré des données de séquençage publiques à l'aide de MGNIFY (RICHARDSON et al., 2023). Les métagénomes environnementaux, séquencés par la technologie Illumina (accès en novembre 2022) ont été récupérés par téléchargement par vagues successives et analysés. À ce jour, 2,331 jeux de données ont été traités, représentant un volume de données d'environ 20 To. Aucun signal positif n'a été détecté en sortie de pipeline pour *B. anthracis*, tout comme les autres espèces étudiées. Sur 2,292 données (soit 98% des données étudiées), aucune lecture de séquençage ne figurait en sortie dès la deuxième étape du pipeline (analyse par la *custom database*). Pour les données restantes, au plus douze lectures de séquençage étaient sélectionnées en sortie de la deuxième étape (dont l'ensemble sur la racine de l'arbre dans le cas des espèces autres que *B. anthracis*), et aucune en sortie de la troisième étape. Ces résultats négatifs sont en adéquation avec les prévisions attendues, c'est-à-dire la distribution limitée de *B. anthracis* dans l'environnement. Cependant, un faible nombre de données métagénomiques provenant d'Afrique subsaharienne (de sol subsaharien *a fortiori*) a pu être analysé (11 au total), justifiant ainsi un futur effort à ce sujet. Par ailleurs, ce travail de détection doit être poursuivi en continu sur les autres données de séquençage disponibles sur la plateforme.

IDÉES À RETENIR

1. L'émergence récente de *B. anthracis*, étudiée dans le chapitre précédent, motive la recherche de nouvelles souches de l'espèce dans une niche écologique restreinte, préférentiellement en Afrique équatoriale, lieu supposé de son apparition. De ce fait, une approche métagénomique pour cette recherche, à des fins de meilleure compréhension de l'histoire évolutive de l'espèce, s'avère cohérente.
2. Afin de compenser le manque existant dans la détection de traces de *B. anthracis* dans des échantillons métagénomiques, nous avons développé une stratégie d'analyse se basant sur deux étapes.
3. Nous avons mis en place un pipeline d'analyse B2FORENSICS_V2, permettant de traiter rapidement de gros volumes de données (de l'ordre de plusieurs dizaines de Go). De plus, le risque de faux positifs est fortement atténué et la sensibilité est de l'ordre de grandeur d'une technique PCR dans le type d'échantillon considéré, un milieu biologiquement riche (eaux usées). Ce pipeline s'appuie sur une base de données de référence qui inclut non seulement des souches de *B. anthracis* mais également son proche voisinage (les souches du groupe *B. cereus* ayant au moins 70% de similitude avec un génome de référence de cette espèce). Il est ainsi important d'enrichir cette base de données en y ajoutant de nouvelles souches voisines de *B. anthracis*, au fur et à mesure qu'elles seront découvertes.
4. Pour évaluer les performances de ce pipeline, un ensemble de données tests a été mis en place. Il s'agit de prélèvements réels d'eaux usées enrichies ou non avec des agents d'intérêt (*B. anthracis* et espèces voisines, entre autres).
5. Une précision intraspécifique a pu être obtenue en étudiant l'assignation taxonomique des lectures considérées comme *B. anthracis* en sortie de pipeline. Cette étape permet également d'éluder d'éventuels signaux persistants de faux positifs, en cas de présence non négligeable d'une espèce voisine dans l'échantillon. Cette identification intraspécifique est dépendante de la couverture du génome d'intérêt au sein du séquençage de l'échantillon. Là encore, une connaissance accrue de la phylogénie de l'espèce étudiée (ici, *B. anthracis*) ainsi que de son proche voisinage est primordiale.
6. Ce pipeline développé peut être mis en analyse de routine sur les bases de données publiques de métagénomes pour la détection de nouvelles souches de *B. anthracis*. Il a également été adapté pour la détection d'agents pathogènes d'intérêt en biodéfense.
7. À ce jour, aucune trace de *B. anthracis* n'a été détectée dans les données publiques analysées, en cohérence avec la localisation restreinte de l'espèce dans des écotypes particuliers. On pourrait anticiper des résultats positifs lorsqu'un nombre plus conséquent de données métagénomiques publiques liés à l'environnement africain équatorial sera disponible.

Conclusion et Perspectives

Conclusion générale Ces travaux de thèse se sont portés sur l'étude de *B. anthracis*, responsable de la maladie du charbon. Cette bactérie, appartenant au groupe *B. cereus* constitué d'espèces très proches génétiquement, soulève de nombreuses questions quant à son émergence, son évolution et sa propagation. Ces problématiques touchent des domaines qui partagent une forte zone d'adhérence : santé publique, phylogénomique ou encore biodéfense. En effet, définir précisément l'espèce *B. anthracis*, notamment en termes de localisation et de datation de son émergence récente, de l'acquisition de ses mécanismes de virulence, et de son évolution au sein de sa niche écologique, permet ensuite de répondre aux enjeux inhérents à ces différents domaines. Une compréhension approfondie de ces aspects permettrait d'améliorer les stratégies de prévision épidémiologique, en anticipant les conditions favorables aux résurgences de la maladie du charbon et en adaptant les mesures de contrôle et de prévention. De même, cette connaissance renforcée faciliterait l'adaptation des outils de détection de *B. anthracis*, rendant les systèmes de surveillance plus réactifs et précis face à des menaces biologiques potentielles.

Ainsi, mes travaux ont été menés afin de répondre à la question centrale suivante :

PROBLÉMATIQUE GÉNÉRALE

D'où provient l'espèce *B. anthracis* et quels mécanismes ont permis l'acquisition de son pouvoir de virulence ?

La première partie de nos travaux a permis de dater l'apparition de l'ancêtre de *B. anthracis*, à une période deux fois plus ancienne que celle de l'apparition du MRCA de l'espèce. Ce positionnement de l'ancêtre par rapport au MRCA renforce l'hypothèse proposée récemment (VERGNAUD, 2020) indiquant que la spéciation de *B. anthracis* soit le résultat de l'arrivée du pastoralisme et de l'agriculture en Afrique équatoriale, peut-être lors des migrations Bantoues à travers la forêt équatoriale il y a 4,000 à 5,000 ans (PATIN et al., 2017). En outre, de nouvelles hypothèses quant à la formation des plasmides pXO1 et pXO2 ont été apportées.

Pour arriver à ces résultats, nous nous sommes penchés sur la caractérisation des souches du groupe *B. cereus* les plus proches de *B. anthracis*, génétiquement parlant. Les trois souches nouvellement découvertes, *B. cereus* BC38B, FFI_BCgr36 et FFI_BCgr46, ont été utilisées pour affiner la phylogénie du proche voisinage de *B. anthracis*. En particulier, l'étude des différences de pression de sélection et de phénomènes homoplasiques entre *B. anthracis* et son voisinage a permis d'établir le positionnement de l'ancêtre de l'espèce. Ce positionnement, congruent avec une approche strictement plasmidique, renforce l'hypothèse d'une émergence récente d'un ancêtre possédant les plasmides de virulence. Cet ancêtre se serait établi de manière clonale dans une niche écologique située très vraisemblablement en Afrique centrale, pour former un écotype particulier, se différenciant ainsi du reste du groupe *B. cereus*. La reconstruction complète des séquences plasmidiques des proches voisins

a révélé des tronçons communs avec pXO1 d'une part, et à d'autres plasmides du groupe *B. cereus* d'autre part. Un nouveau type de miniréplicon est présent dans ces plasmides des proches voisins. Ce phénomène avait été décrit auparavant pour le plasmide pXO2 et certains plasmides conjugatifs. Ces observations renforcent l'idée d'une formation des plasmides de virulence pXO1 et pXO2 par intégration successive de plus petits plasmides, dont certains comportaient les îlots de pathogénicité qui leur sont propres. Une coévolution entre le chromosome de *B. anthracis* et ces plasmides se serait ensuite opérée dans sa niche écologique.

À ce titre, les observations effectuées dans le cadre de nos travaux signalent la particularité de *B. anthracis* vis-à-vis de ses proches voisins. *B. anthracis* constitue une espèce bactérienne à part dans le groupe *B. cereus*, ayant suivi une évolution singulière au sein de ce groupe. Ainsi, la proposition de CARROLL, WIEDMANN et KOVAC, 2020 de considérer *B. anthracis* comme une sous-espèce de *B. mosaicus* (un complexe plus large incluant *B. anthracis* et ses proches voisins), en le nommant *B. mosaicus* supsc. anthracis, n'est pas cohérente avec cette conception d'espèce bactérienne que constitue *B. anthracis*.

L'émergence récente du mutant clonal ancestral à l'espèce *B. anthracis* indique que cette dernière est localisée dans des niches écologiques restreintes. Cette distribution naturelle limitée motive ainsi une approche métagénomique afin de découvrir de nouvelles souches qui affineront le modèle d'émergence de l'espèce.

Pour cette partie, nous avons développé un outil de détection spécifique à la détection de *B. anthracis* dans des échantillons métagénomiques. Cet outil répond à un manquement actuel de l'existant, qui permet certes l'identification de *B. anthracis* lorsqu'il est isolé d'une part ou la composition globale d'un métagénomique d'autre part, mais ne rend pas possible une détection rapide et fiable de *B. anthracis* en faible quantité dans un échantillon environnemental. Cet outil permet de répondre à cet objectif, tout en apportant une précision intraspécifique en cas de signal positif. Pour concevoir ce pipeline d'analyse, nous avons tiré profit de la différenciation entre *B. anthracis* et son proche voisinage, critère déterminant pour réduire le risque de faux positifs et gagner en temps d'exécution. Cela souligne l'importance de caractériser en routine de nouvelles souches du groupe *B. cereus* pour y détecter des voisines de l'espèce étudiée. Enfin, ce pipeline a été adapté à d'autres agents pathogènes, dont la structure de population et la proximité avec des espèces environnementales font écho à la situation de *B. anthracis*. L'outil conçu B2FORENSICS_V2 permet de rechercher de traces de ces bactéries dans les données de séquençage publiques (en particulier dans l'ADN ancien), rendant possible leur exploitation à cette fin de compréhension de l'évolution des espèces.

Perspectives d'étude Ces travaux de thèse ont ouvert la voie à d'autres projets en cours de réalisation ou à mener dont voici les enjeux :

— La situation de *B. anthracis* au sein d'un groupe génétiquement très homogène n'est pas exclusive. *Yersinia pestis*, *Francisella tularensis*, *Mycobacterium tuberculosis* et *Brucella* en sont des exemples connus. Effectuer la même approche serait judicieux tant en termes de compréhension de leur émergence qu'à des fins de détection. À ce sujet, VERGNAUD ET AL. (2024) (non publié) ont ainsi placé la position de l'ancêtre de l'espèce *Francisella tularensis* en tirant profit du proche voisinage de l'espèce environnementale *Francisella novicida*.

— Cette démarche employée n'est possible qu'en élargissant les efforts non seulement sur la détermination de nouvelles souches des espèces en question mais

aussi de leurs proches voisinages. À ce titre, la caractérisation en routine de souches environnementales est indispensable et cet effort doit être poursuivi tant pour *B. anthracis* que pour les espèces susmentionnées par exemple. L'outil B2FORENSICS_V2 peut être adapté à cette utilisation en modifiant certaines étapes de sélection de lectures de séquençage, afin de garder non pas uniquement celles relatives à *B. anthracis*, mais celle du voisinage.

— Nous avons vu au cours des travaux présentés que l'approche type k-mer permet d'obtenir une résolution élevée dans le cadre d'une analyse d'échantillon complexe. Certains outils existants présentés usent de cette approche pour différencier *B. anthracis* et le reste du groupe *B. cereus*, s'appuyant sur leurs différences génétiques fondamentales (en particulier les k-mers chromosomiques signatures de *B. anthracis* et ceux spécifiques au gène *lef*). Nous avons essayé de reproduire cette stratégie pour distinguer les souches de *B. anthracis* et les souches anthracis-like du groupe *B. cereus*. Cependant, cet essai s'est révélé infructueux du fait du nombre insuffisant de séquences plasmidiques relatifs au souches anthracis-like disponibles et aux trop faible différences génétiques existants avec celles disponibles avec les plasmides pXO1 et pXO2.

— Dans le cadre d'une utilisation en routine de l'outil B2FORENSICS, un travail similaire à celui effectué sur *B. anthracis* devra être opéré en incluant les souches voisines des espèces ajoutées (*Y. pestis*, *F. tularensis*, *Brucella*) dans les *custom databases* correspondantes.

— Une optimisation avec des techniques d'intelligence artificielle (IA) permettra de faciliter la lecture des chemins obtenus après analyse des données de séquençage par les *custom databases* de KRAKEN2. En effet, comme décrit *supra*, la présence de nombreux artefacts peut gêner la détermination d'un chemin clair vers une ou des souches en cas de présence de *B. anthracis*. Deux méthodes entraînées par apprentissage profond ont commencé à être évaluées en parallèle avec ce projet de thèse, la première utilisant une représentation séquentielle des données, la seconde utilisant une représentation des données sous forme de graphe. Ces méthodes ont été entraînées et évaluées sous différents contextes simples (présence ou absence de *B. anthracis*, présence ou absence d'espèces voisines). Les résultats préliminaires sont prometteurs. Afin d'améliorer les performances globales des modèles, il faudra mener des études sur l'augmentation des données d'entraînement afin de contrebalancer le déséquilibre des classes et sur la transformation des données afin d'anticiper des cas de mélanges de souches non vus dans le jeu de données d'entraînement.

Annexe A

Historique du séquençage

A.1 Principe du séquençage

A.1.1 Genèse du séquençage

Le procédé de séquençage introduit par Fred Sanger en 1977 a ouvert la voie à la compréhension de la composition et de la fonction des génomes bactériens, en révélant leurs séquences sous forme de fragments de quelques centaines de bases, connus sous le nom de lectures (figure A.1). Parallèlement, en cette même année, l'ARN ribosomique a été identifié comme un indicateur pour la classification taxonomique des organismes. Citons par exemple la découverte du troisième domaine du vivant (les Archées) par Carl Woese en 1977 (WOESE et FOX, 1977). En s'appuyant sur ces découvertes, Pace a suggéré en 1985 d'analyser l'ARN ribosomique directement à partir de l'environnement, évitant ainsi la nécessité de cultiver les bactéries. Grâce à cette innovation, il a été possible de contourner les limitations liées à la capacité de culture et de dévoiler une portion significative de la diversité microbienne auparavant méconnue (RAPPÉ et GIOVANNONI, 2003). HANDELSMAN et al., 1998 introduit en 1998 le concept de métagénomique, faisant référence à l'analyse directe de l'ADN au sein d'un environnement, permettant d'accéder potentiellement à l'ensemble des génomes d'une communauté.

A.1.2 Le séquençage de deuxième génération

Bien que la méthode de Sanger ait révolutionné le séquençage, elle présente des contraintes, notamment en raison du nombre restreint de lectures qu'elle produit et du coût élevé associé aux grands projets de séquençage. Le séquençage de deuxième génération (également appelé "séquençage haut débit" ou "séquençage NGS" pour *Next Generation Sequencing*) désigne les technologies permettant de séquencer un grand nombre de fragments d'ADN. Ces nouvelles techniques, telles que le séquençage Illumina, Roche 454, ion Torrent (Proton / PGM) ou encore le séquençage SOLiD, apportent des améliorations conséquentes en termes de coût et de temps de réalisation par rapport au processus initial de séquençage Sanger. A titre d'illustration, le projet *Human Genome*, destiné à séquencer entièrement le génome humain, a coûté 3,5 milliards de dollars en 13 ans (entre 1990 et 2003) avec des séquenceurs de type Sanger (VENTER et al., 2001). De nos jours, cette opération prend environ une journée et coûte à peu près 1,000 dollars (GOLDFEDER et al., 2017).

Ces techniques permettent l'analyse simultanée de nombreuses molécules, mais ne fournissent pas directement la séquence intégrale d'un génome. L'idée principale sur laquelle repose cette technologie est la parallélisation du séquençage. En effet, séquencer un fragment d'ADN revient à déterminer la suite de nucléotides constituant celui-ci. La technique Sanger permet de lire environ 800 paires de bases en une

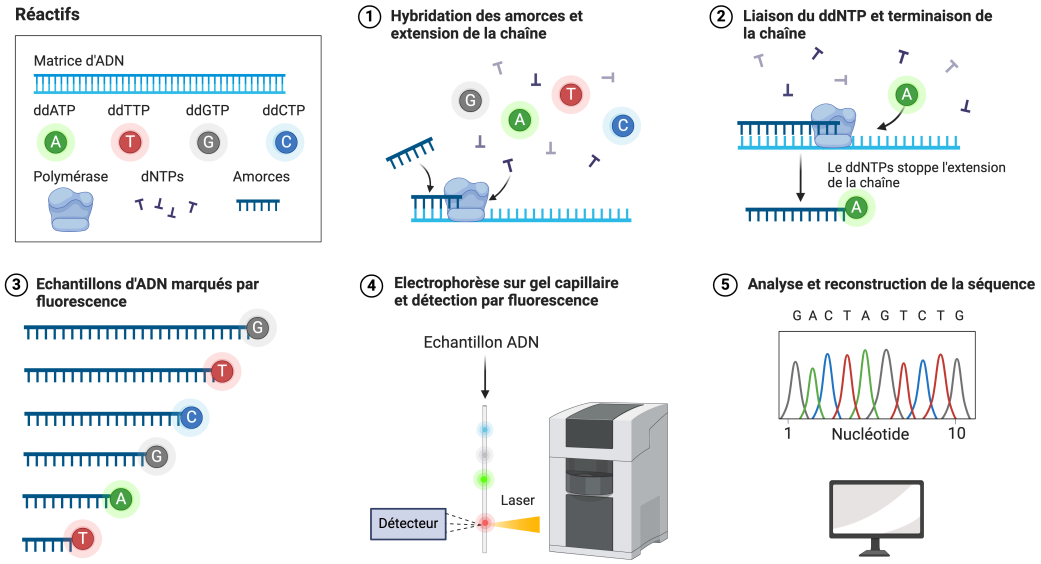

FIGURE A.1 – Schéma du principe du séquençage Sanger. Adapté de ONA, 2020b.

heure, ce qui nécessiterait plus de 400 ans pour séquencer le génome humain avec une seule machine et une seule piste de lecture. L'idée du séquençage NGS est de séquencer des petits fragments d'ADN en même temps. Les séquenceurs haut débit de deuxième génération peuvent par exemple lire jusqu'à 20 milliards de fragments de 150 à 300 paires de bases, ce qui réduit considérablement le temps d'exécution. Après le séquençage, on obtient des fragments de séquences ne dépassant généralement pas quelques centaines de bases, provenant de diverses zones du génome. L'origine d'une lecture dans le génome est aléatoire.

Il est essentiel de générer de nombreuses lectures pour réaliser un séquençage complet du génome. Cela garantit que le génome est couvert à une certaine profondeur. Par exemple, une profondeur de couverture de 20x signifie que chaque portion du génome est représentée en moyenne vingt fois. Ces séquences peuvent contenir des erreurs, telles que des substitutions de nucléotides ou des erreurs d'omission et d'ajout de nucléotides (délétions et insertions).

A.1.3 Les étapes du séquençage

Préparation de la librairie de séquençage Le principe général du séquençage NGS repose sur plusieurs étapes. La première est l'élaboration d'une librairie constituée des fragments d'ADN que l'on devra ensuite séquencer. Pour ce faire, il est possible de découper aléatoirement la molécule d'ADN initiale (le génome) en plusieurs fragments : cela s'appelle la méthode "*shotgun*", utilisée lorsque l'objectif est de séquencer un génome entier (ou l'ensemble de l'ADN présent dans un échantillon). Afin de découper le fragment d'ADN, on peut utiliser des enzymes de restriction ou encore envoyer des ultrasons à une certaine fréquence. Si au contraire l'objectif n'est pas d'obtenir la séquence complète du fragment étudié mais seulement une partie, on utilise la méthode dite ciblée qui permet de sélectionner uniquement les fragments d'ADN souhaités. Pour cela, deux techniques équivalentes sont mises en

œuvre : l'enrichissement par capture qui consiste à filtrer les fragments d'ADN souhaités qui s'hybrideront à des brins complémentaires disposées sur une plaque et l'enrichissement par PCR en amplifiant les régions souhaitées par PCR.

Le séquençage La librairie de séquençage effectuée, il reste à déterminer les séquences nucléotidiques des fragments. Plusieurs méthodes existent ; seule la technique de séquençage par synthèse (Illumina Solexa), qui est la plus largement répandue, est présentée.

— Tout d'abord, une banque d'ADN double-brin est générée à partir de l'échantillon à analyser en fractionnant aléatoirement en morceaux de 200 paires de bases.

— des adaptateurs spécifiques sont ajoutés aux extrémités des fragments.

— l'ADN est dénaturé en simple-brin.

— l'extrémité des ADN simples-brins se fixe aléatoirement sur la surface d'une *flow cell* sur laquelle les différentes étapes seront réalisées.

— une amplification en pont en phase solide est effectuée : de l'ADN double-brin est donc créé. S'ensuit une dénaturation de cet ADN et une formation de clusters denses de fragments amplifiés.

— le séquençage à proprement parler a lieu. Les quatre terminateurs (les nucléotides A, G, T, C), les amorces et l'ADN polymérase sont ajoutés pour procéder au premier cycle de séquençage.

— pour lire la première base des séquences des fragments, on utilise un laser pour exciter les clusters et on analyse la fluorescence émise par ces derniers.

— on procède au deuxième cycle de séquençage en ajoutant une nouvelle fois les quatre terminateurs.

— on lit ainsi à chaque cycle une nouvelle base de la séquence de chaque fragment. Par ailleurs, la lecture des bases de l'ensemble des clusters est faite en parallèle.

Ce procédé est illustré à la figure A.2.

A.1.4 Le séquençage troisième génération

La technologie Illumina, bien que performante, présente une limitation liée à la longueur de ses lectures, ce qui peut être un obstacle pour la caractérisation complète de certaines zones du génome humain. Afin de surmonter ce défi, diverses alternatives ont vu le jour. Parmi celles-ci, on peut mentionner les lectures liées (*linked-reads*) proposées par 10X Genomics ou Hi-C, ainsi que les lectures longues (*long reads*) développées par des entreprises comme Oxford Nanopore Technologies ou Pacific Biosciences.

La technologie *linked-reads* permet de relier des lectures provenant de la même région génomique. Les techniques de séquençage de longues lectures se distinguent quant à elles principalement par deux caractéristiques majeures. Premièrement, elles se basent sur le séquençage d'une unique molécule d'ADN, appelé *Single Molecule Sequencing*. Deuxièmement, comme le suggère le terme "*long-reads*", ces technologies génèrent des lectures de longueur nettement supérieure à celles produites par les techniques *short-reads*. En moyenne, ces lectures dépassent 1 kb, et en la pratique, la longueur moyenne des lectures produites par certaines de ces technologies gravite autour de 5 kb ou plus. Un autre avantage des technologies *long reads* est leur

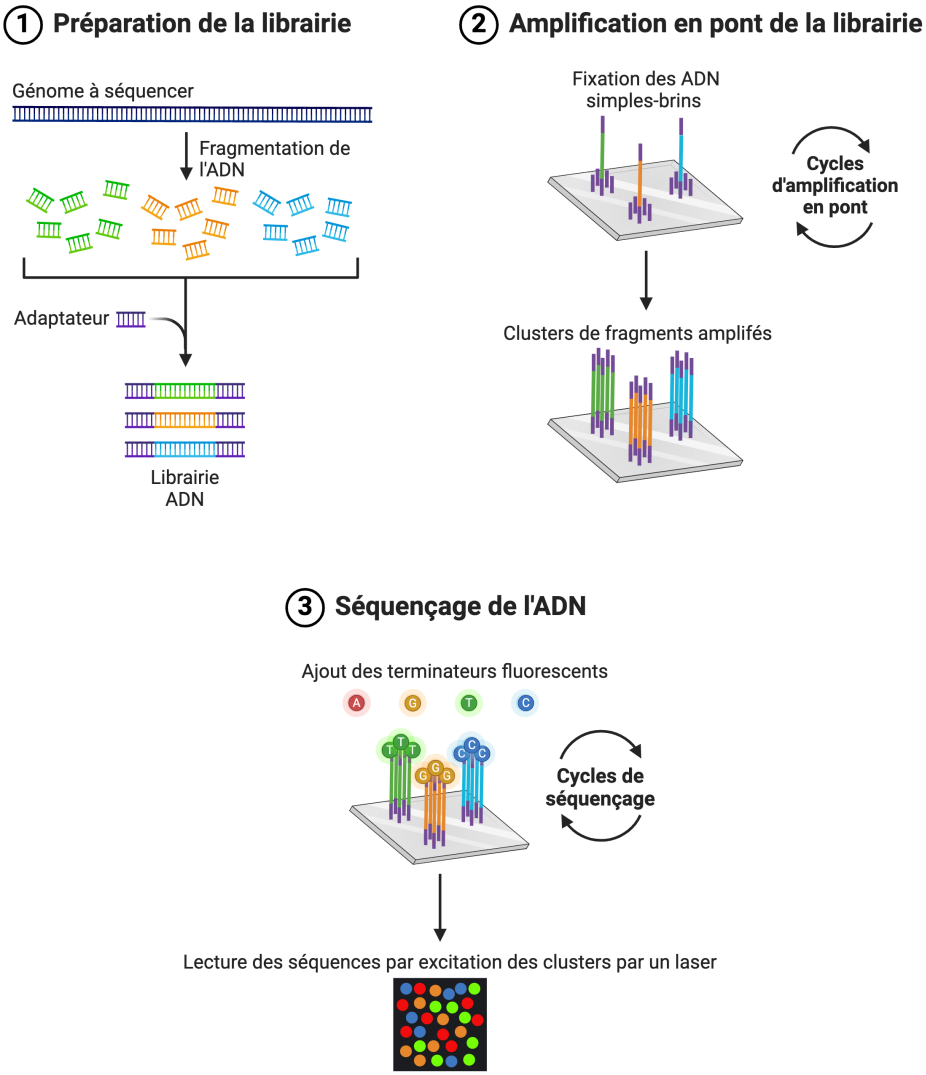

FIGURE A.2 – Schéma du principe du séquençage deuxième génération (Illumina). Adapté de ONA, 2020a.

rapidité. Le temps nécessaire pour séquencer peut s'étaler de quelques minutes à quelques jours, en fonction de la technologie employée et de l'échantillon étudié. Un avantage de ces méthodes est qu'elles éliminent le besoin d'amplification de l'ADN. L'étape d'amplification, utilisée dans d'autres techniques, peut entraîner des erreurs de réplication. En évitant cette étape, les techniques de séquençage *long reads* réduisent le risque d'introduire des biais dans les données de séquençage. Les techniques de séquençage *long reads* offrent des solutions prometteuses aux défis associés à la métagénomique. Grâce à leur capacité à produire des lectures plus longues, elles améliorent significativement l'assemblage métagénomique. Des méthodes innovantes, comme le Hi-C, peuvent être utilisées pour identifier et différencier les lectures issues de différents organismes (BURTON et al., 2014). Toutefois, en termes de débit, ces techniques *long reads* n'égalent pas encore les capacités du séquençage NGS, et elles requièrent naturellement une excellente qualité d'ADN. Étant donné

que les études métagénomiques demandent une quantité significative de séquençage pour capter la complexité des communautés microbiennes, l'adoption généralisée de ces nouvelles techniques dépendra en grande partie de leur évolution et de leurs améliorations futures.

Une des méthodes de séquençage *long-reads* est celle développée par Oxford Nanopore Technologies (ONT). Elle utilise la détection de variations de courant lors du passage d'une molécule d'ADN à travers un nanopore biologique intégré dans une membrane synthétique. Ce nanopore, agissant comme un capteur d'ions, détecte les changements causés par différentes structures moléculaires de l'ADN qui le traversent, permettant ainsi de séquencer de longs fragments d'ADN, y compris ceux avec des régions répétitives ou riches en GC (LOGSDON, VOLLGER et EICHLER, 2020). Le principe est décrit dans la figure A.3.

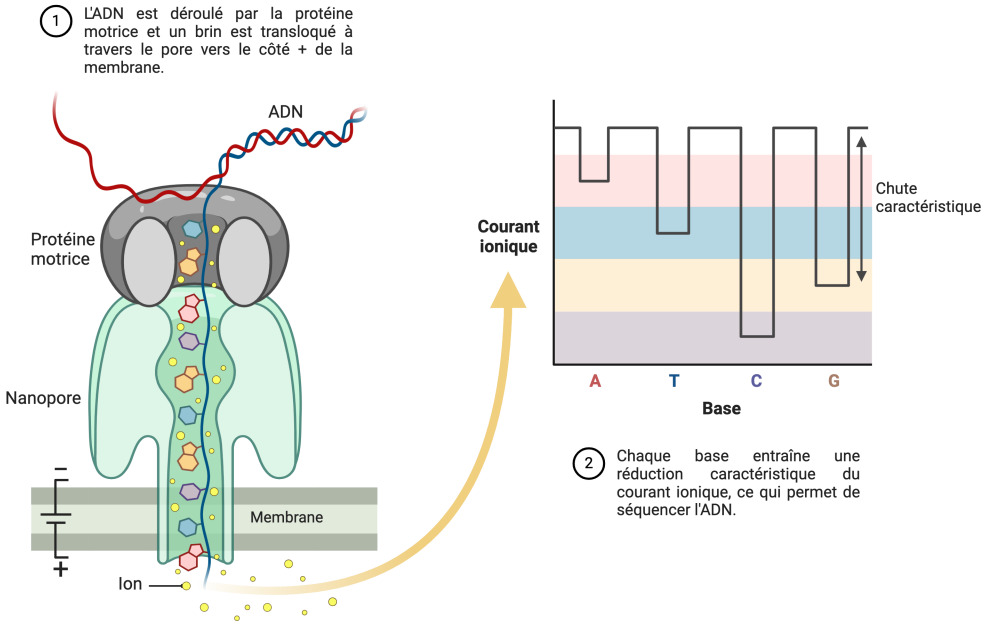

FIGURE A.3 – Schéma du principe du séquençage par nanopores. Adapté de BIORENDER, 2020.

Le tableau A.1 récapitule les principales caractéristiques de quelques technologies de séquençage nouvelle génération. On remarque l'apport du séquençage troisième génération en termes de longueur de lecture et de temps d'exécution et l'avantage du séquençage de deuxième génération quant à la précision.

En somme, la figure A.4 retrace les différentes étapes du séquençage de la technique Sanger à nos jours.

A.1.5 Techniques d'assemblage

Avec la progression des technologies de séquençage de nouvelle génération, des méthodes algorithmiques pour agencer les millions de lectures produites ont dû être

Technologie	Longueur maximale des lectures (en nucléotides)	Durée de séquençage	Précision
Illumina	300	Entre 4h et 56h	99.9%
PacBio	300,000	Jusqu'à 30h	87-92%
Oxford Nanopore	$4x10^6$	De quelques minutes à 72h	87-98%

TABLE A.1 – Tableau comparatif des techniques de séquençage nouvelle génération. Adapté de HU et al., 2021.

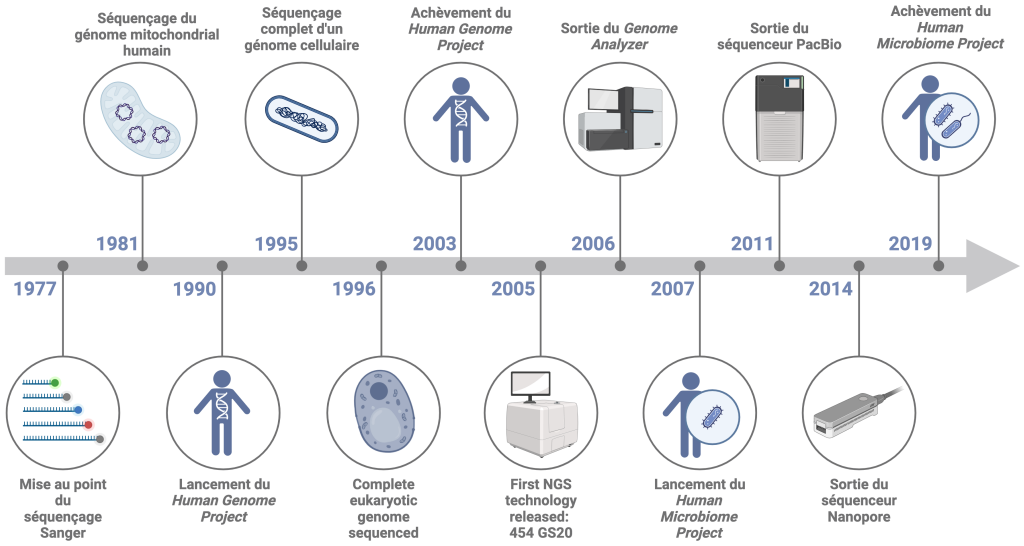

FIGURE A.4 – Historique du séquençage. Adapté de HUANG, 2023.

conçues. L'objectif est de déterminer l'ordonnancement des nucléotides dans la séquence d'ADN du génome étudié à partir de la séquence de tronçons aléatoires. Avant de procéder à l'assemblage des lectures obtenues lors du séquençage, il est nécessaire d'éliminer les lectures de faible qualité ou celles qui ne sont pas pertinentes pour l'espèce dont on veut déduire la séquence d'ADN. Ce processus est appelé *trimming* (élagage) des lectures.

L'assemblage peut s'effectuer selon deux méthodes distinctes : la première est l'approche par *mapping*, où les lectures sont alignées en utilisant une séquence de référence préexistante ; la seconde est l'approche *de novo*, qui construit l'assemblage exclusivement à partir des lectures, sans se référer à une séquence préalablement connue.

La stratégie d'assemblage *de novo* s'articule généralement autour de plusieurs phases. On commence par rectifier certaines erreurs présentes dans les lectures, puis on rassemble les lectures qui se chevauchent pour former des séquences étendues dénommées *contigs*. Enfin, on élabore des *scaffolds* en disposant les *contigs* dans un ordre précis, séparés par des espaces qui symbolisent les zones que l'assemblage n'a pas réussi à identifier. La gestion des séquences répétitives, bien qu'elle ne soit pas toujours considérée comme une étape à part entière, reste néanmoins importante. Ces répétitions sont des séquences dupliquées de manière identique ou quasi-identique dans la section d'ADN cible. Dans chacune des lectures, on ne conserve pas l'information de l'emplacement exact de chaque duplication sur la séquence d'ADN d'origine. De plus, une même duplication peut être présente dans diverses lectures, en totalité ou en fragments. La tâche est donc de distinguer les lectures

contenant une copie spécifique de la répétition de celles portant d'autres copies, qu'elles soient similaires ou exactes, positionnées distinctement dans la séquence d'ADN cible. Cette démarche est particulièrement compliquée si le génome renferme des structures répétitives complexes, comme une répétition nichée à l'intérieur d'une autre ou dépassant la longueur d'une lecture. C'est ici que le séquençage à longues lectures présente un avantage majeur, en permettant de reconstituer des génomes complets.

Les deux méthodes d'assemblage sont illustrées dans la figure A.5.

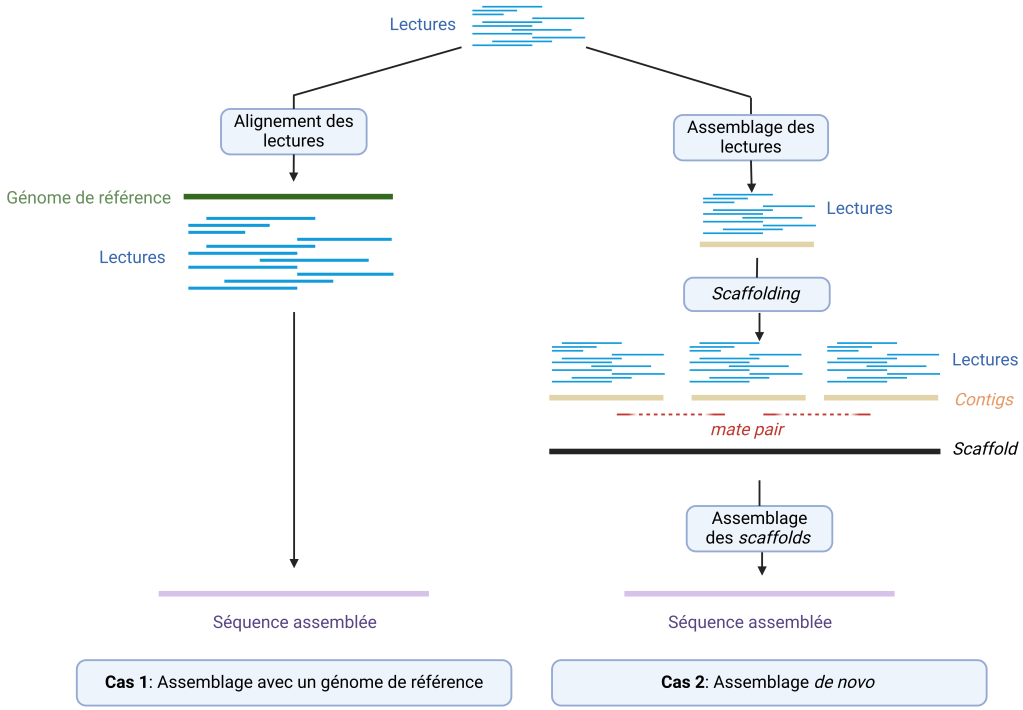

FIGURE A.5 – Schéma récapitulatif des méthodes d'assemblage. *Il importe de noter que le terme d'assemblage dans le cas d'un assemblage par mapping sur une référence est abusif. Il est utile à signaler parce que dans les premières années du séquençage massif à lectures courtes, le résultat d'assemblage par mapping a pu être déposé dans les bases de données comme étant un assemblage, sans que la méthodologie soit bien indiquée.* Créé avec BIORENDER.

Annexe B

Fichiers complémentaires - Article 3

B.1 Fichier complémentaire 1

Extraction protocol for samples B2F10, B2F12, B2F13 and B2F14

1. Immediately after collection, samples were filtered by gravity using fritted glass with pores going from 10 to 16 µm in diameter (category 4). This kind of filter allows the elimination of the larger environmental impurities, some eukaryotic cells, and the retention of small eukaryotic cells, viruses, bacterial cells, free DNA and free proteins. The filters are reusable after washing using successively bleach, water and 0.1M HCl. After filtration, samples were stored at 4°C until DNA extraction (occurring at most 10 days after sampling).
2. Prepare tubes for bead beating :
   - add 1 g of glass beads of 0.1 mm diameter (BIOSPEC PRODUCT 5562301) and 200 µl of water (Aqua B. Braun, Ecoteiner) containing 1 µg of tRNA (SIGMA R9001) to 2 ml screw capped tubes.
   - shake tubes for 20 sec at speed 6.5 (FastPrep-24, MP Biomedicals).
3. Add one volume of extraction buffer to the wastewater sample.
4. Transfer one ml to a glass beads tube (two tubes needed per ml of wastewater sample).
5. Shake the sample for one min at speed 4 using the FastPrep.
6. Recover sample fractions by punching a hole with a needle at the bottom of the tubes containing the samples and the beads, inserting these in 5 ml tubes and centrifuging at 1,500 rpm (centrifuge JOUAN GR412) for 5 min. The eluates from each tube (1.2 ml) are pooled if necessary.
7. Add lysozyme (SIGMA L-6876), zymolyase (ICN zymolyase-20T) and RNase A at 2 mg/ml, 50 µg/ml and 10 µg/ml final concentration, respectively.
8. Incubate the tubes at 37°C for 1 h.
9. Add SDS at 1% and Proteinase K (BIOPROBE systems PROK03) at 100 µg/ml (final concentrations).
10. Incubate for 2 h at 65°C in waterbath inverting the tube every 20 min.
11. Add 1 volume of chloroform (SIGMA-ALDRICH 32211-1L) and centrifuge at 10,000 rpm (centrifuge SIGMA 2K15) for 10 min at 4°C.
12. Transfer the supernatant to a clean polypropylene OakRidge tube or Eppendorf tube according to the volume of the sample.
13. Add one volume of phenol (ROTH, Roti-phenol 0038.2), centrifuge at 10,000 rpm (centrifuge SIGMA 2K15) for 10 min at 4°C.
14. Transfer the supernatant to clean tube.

15. Add one volume of chloroform (SIGMA-ALDRICH 32211-1L) and centrifuge at 10,000 rpm (centrifuge SIGMA 2K15) for 10 min at 4°C.

16. Transfer the supernatants to a clean OakRidge or Eppendorf tube and add glycogen (Sigma G8751) to a final concentration of 40 µg/ml.

17. Add 2 volumes of absolute ethanol (AnalaR NORMAPUR 28823.362) in order to precipitate the DNA.

18. Centrifuge at 10,000 rpm (centrifuge SIGMA 2K15) for 20 min and wash the pellet once with ethanol 70%.

19. Dry the pellet leaving the tube overnight on the bench and then re-suspend in an appropriate volume of water (Aqua B. Braun, Ecoteiner).

20. Perform an additional RNase treatment at 10 µg/ml incubating the sample at 37°C for 45 min.

21. Pre-hydrate a centricon-30 column adding 500 µl of water and centrifuging at 1000G for 30 min.

22. Bring the volume of the sample to 2 ml adding water.

23. Add the diluted sample to a centricon-30 column and centrifuge for 45 min at 1000 xg at 15°C.

24. Add 2 ml of water to the sample DNA retained by the column and centrifuge for 45 min at 1000 G at 15°C.

25. Repeat twice adding 2 ml of water.

B.2 Fichier complémentaire 2

Noms des souches	**Numéros d'accession**
Bacillus anthracis strain Ames	GCF_000007845.1
Bacillus anthracis strain Sterne	GCF_000008165.1
Bacillus anthracis strain Ames Ancestor	GCF_000008445.1
Bacillus anthracis strain CDC 684	GCF_000021445.1
Bacillus anthracis strain A0248	GCF_000022865.1
Bacillus anthracis strain CNEVA-9066	GCF_000167235.1
Bacillus anthracis strain A1055	GCF_000167255.1
Bacillus anthracis strain Vollum	GCF_000167275.1
Bacillus anthracis strain Kruger B	GCF_000167295.1
Bacillus anthracis strain Western North America USA6153	GCF_000167315.1
Bacillus anthracis strain Australia 94	GCF_000167335.1
Bacillus anthracis strain Tsiankovskii-I	GCF_000181675.2
Bacillus anthracis strain A0488	GCF_000181835.1
Bacillus anthracis strain A0193	GCF_000181915.1
Bacillus anthracis strain A0442	GCF_000181935.1
Bacillus anthracis strain A0465	GCF_000181995.1
Bacillus anthracis strain A0174	GCF_000182055.1
Bacillus anthracis strain A0389	GCF_000219895.1
Bacillus anthracis strain H9401	GCF_000258885.1
Bacillus anthracis strain Heroin Ba4599	GCF_000278385.1

Suite dans la page suivante

Noms des souches	**Numéros d'accession**
Bacillus anthracis strain UR-1	GCF_000292565.1
Bacillus anthracis strain BF1	GCF_000295695.1
Bacillus anthracis strain Carbosap	GCF_000310045.1
Bacillus anthracis strain 3166	GCF_000319715.1
Bacillus anthracis strain Sen2Col2	GCF_000359425.1
Bacillus anthracis strain Sen3	GCF_000359445.1
Bacillus anthracis strain Gmb1	GCF_000359465.1
Bacillus anthracis strain A16R	GCF_000512775.2
Bacillus anthracis strain A16	GCF_000512835.2
Bacillus anthracis strain V770-NP1-R	GCF_000521345.1
Bacillus anthracis strain CZC5	GCF_000534935.2
Bacillus anthracis strain 8903-G	GCF_000558965.1
Bacillus anthracis strain 9080-G	GCF_000558985.1
Bacillus anthracis strain 52-G	GCF_000559005.1
Bacillus anthracis strain SVA11	GCF_000583105.1
Bacillus anthracis strain 95014	GCF_000585275.1
Bacillus anthracis strain ANSES_08-8_20	GCF_000697515.2
Bacillus anthracis strain ANSES_99-100	GCF_000697535.2
Bacillus anthracis strain ANSES_00-82	GCF_000697555.1
Bacillus anthracis strain HYU01	GCF_000725325.1
Bacillus anthracis strain Carbosap	GCF_000732465.1
Bacillus anthracis strain STI-1	GCF_000740925.2
Bacillus anthracis strain Smith 1013	GCF_000742315.1
Bacillus anthracis strain 2000031021	GCF_000742655.1
Bacillus anthracis strain delta Sterne	GCF_000742695.1
Bacillus anthracis strain BFV	GCF_000742875.1
Bacillus anthracis strain Vollum	GCF_000742895.1
Bacillus anthracis strain Zimbabwe 89	GCF_000743805.1
Bacillus anthracis strain A.Br.003	GCF_000743825.1
Bacillus anthracis strain Cvac02	GCF_000747335.1
Bacillus anthracis strain Han	GCF_000747375.1
Bacillus anthracis strain 2000031006	GCF_000782875.1
Bacillus anthracis strain 2000031023	GCF_000782885.1
Bacillus anthracis strain 2000031027	GCF_000782895.1
Bacillus anthracis strain 2000032892	GCF_000782905.1
Bacillus anthracis strain 2000031031	GCF_000782955.1
Bacillus anthracis strain 2000031039	GCF_000782975.1
Bacillus anthracis strain 2000031052	GCF_000782995.1
Bacillus anthracis strain 2000031075	GCF_000783015.1
Bacillus anthracis strain 2000031709	GCF_000783035.1
Bacillus anthracis strain 2000031765	GCF_000783055.1
Bacillus anthracis strain 2000032819	GCF_000783075.1
Bacillus anthracis strain 2000032879	GCF_000783095.1
Bacillus anthracis strain 2000032951	GCF_000783115.1
Bacillus anthracis strain 2000032967	GCF_000783135.1
Bacillus anthracis strain 2000032968	GCF_000783155.1
Bacillus anthracis strain 2000032975	GCF_000783165.1

Suite dans la page suivante

Noms des souches	Numéros d'accession
Bacillus anthracis strain 2000032979	GCF_000783195.1
Bacillus anthracis strain 2000032989	GCF_000783215.1
Bacillus anthracis strain 2002013094	GCF_000783235.1
Bacillus anthracis strain 2000031008	GCF_000793525.1
Bacillus anthracis strain 2000032832	GCF_000793545.1
Bacillus anthracis strain 2000031038	GCF_000793565.1
Bacillus anthracis strain Pollino	GCF_000831505.1
Bacillus anthracis strain PAK-1	GCF_000832425.1
Bacillus anthracis strain Vollum 1B	GCF_000832445.1
Bacillus anthracis strain K3	GCF_000832465.1
Bacillus anthracis strain Ohio ACB	GCF_000832505.1
Bacillus anthracis strain SK-102	GCF_000832565.1
Bacillus anthracis strain Pasteur	GCF_000832585.1
Bacillus anthracis strain Sterne	GCF_000832635.1
Bacillus anthracis strain BA1015	GCF_000832665.1
Bacillus anthracis strain BA1035	GCF_000832725.1
Bacillus anthracis strain RA3	GCF_000832745.1
Bacillus anthracis strain V770-NP-1R	GCF_000832785.1
Bacillus anthracis strain 2002013094	GCF_000832965.1
Bacillus anthracis strain Ames	GCF_000833065.1
Bacillus anthracis strain Canadian_bison	GCF_000833125.1
Bacillus anthracis strain Turkey32	GCF_000833275.1
Bacillus anthracis strain A1144	GCF_000875715.1
Bacillus anthracis strain A1039	GCF_000986915.1
Bacillus anthracis strain A1075	GCF_000986935.1
Bacillus anthracis strain 44-NIAH	GCF_001015005.1
Bacillus anthracis strain 52-40-NIAH	GCF_001015025.1
Bacillus anthracis strain K1285	GCF_001272985.1
Bacillus anthracis strain K2129	GCF_001273005.1
Bacillus anthracis strain K2883	GCF_001273025.1
Bacillus anthracis strain K0021	GCF_001273045.1
Bacillus anthracis strain K4834	GCF_001273065.1
Bacillus anthracis strain K3974	GCF_001273085.1
Bacillus anthracis strain K1129	GCF_001273105.1
Bacillus anthracis strain K8215	GCF_001273125.1
Bacillus anthracis strain PAK-1	GCF_001273145.1
Bacillus anthracis strain A0389	GCF_001276995.1
Bacillus anthracis strain A0021	GCF_001277085.1
Bacillus anthracis strain Larissa	GCF_001277955.1
Bacillus anthracis strain Stendal	GCF_001543225.1
Bacillus anthracis strain Tangail-1	GCF_001654475.1
Bacillus anthracis strain K2	GCF_001677295.1
Bacillus anthracis strain K1	GCF_001677305.1
Bacillus anthracis strain Parent2	GCF_001683065.1
Bacillus anthracis strain Parent1	GCF_001683095.1
Bacillus anthracis strain PR01	GCF_001683135.1
Bacillus anthracis strain PR02	GCF_001683155.1

Suite dans la page suivante

Noms des souches	**Numéros d'accession**
Bacillus anthracis strain PR05	GCF_001683175.1
Bacillus anthracis strain PR06	GCF_001683195.1
Bacillus anthracis strain PR07	GCF_001683215.1
Bacillus anthracis strain PR08	GCF_001683235.1
Bacillus anthracis strain PR09-1	GCF_001683255.1
Bacillus anthracis strain PR09-4	GCF_001683275.1
Bacillus anthracis strain PR10-4	GCF_001683295.1
Bacillus anthracis strain 55-VNIIVViM	GCF_001835485.1
Bacillus anthracis strain Tyrol 4675	GCF_001936375.1
Bacillus anthracis strain Sterne 34F2	GCF_002005265.1
Bacillus anthracis strain Ba-8884/94-Geo	GCF_002019425.1
Bacillus anthracis strain Ba-8785/92-Geo	GCF_002025265.1
Bacillus anthracis strain Ba-8784/92-Geo	GCF_002025275.1
Bacillus anthracis strain Ba-7673/89-Geo	GCF_002025335.1
Bacillus anthracis strain Ba-9065/08-Geo	GCF_002025345.1
Bacillus anthracis strain Ba-1897/12-Geo	GCF_002025375.1
Bacillus anthracis strain Ba-8782/92-Geo	GCF_002025395.1
Bacillus anthracis strain Ba-9108/08-Geo	GCF_002025415.1
Bacillus anthracis strain Ba-1802/12-Geo	GCF_002025435.1
Bacillus anthracis strain Ba-8776/92-Geo	GCF_002025455.1
Bacillus anthracis strain FDAARGOS_341	GCF_002208785.1
Bacillus anthracis strain DS201579	GCF_002233635.1
Bacillus anthracis strain 14RA5914	GCF_002277915.1
Bacillus anthracis strain Shikan-NIID	GCF_002356575.1
Bacillus anthracis strain BA663	GCF_002525685.1
Bacillus anthracis strain BA781	GCF_002525695.1
Bacillus anthracis strain BA500	GCF_002525705.1
Bacillus anthracis strain 34F2_Sterne	GCF_002525715.1
Bacillus anthracis strain BAP417	GCF_002525765.1
Bacillus anthracis strain BAP482	GCF_002525775.1
Bacillus anthracis strain BA721	GCF_002525785.1
Bacillus anthracis strain Kafkas-100	GCF_002896575.1
Bacillus anthracis strain Kafkas-149	GCF_002896585.1
Bacillus anthracis strain Kafkas-78	GCF_002896595.1
Bacillus anthracis strain Kafkas-86	GCF_002896635.1
Bacillus anthracis strain Kafkas-215	GCF_002896655.1
Bacillus anthracis strain Kafkas-68	GCF_002896665.1
Bacillus anthracis strain Kafkas-60	GCF_002896695.1
Bacillus anthracis strain STI1	GCF_002980615.2
Bacillus anthracis strain COVASU	GCF_003045745.1
Bacillus anthracis strain BA2968	GCF_003063925.1
Bacillus anthracis strain HKI4363/88	GCF_003063945.1
Bacillus anthracis strain A142	GCF_003063965.1
Bacillus anthracis strain A155	GCF_003063985.1
Bacillus anthracis strain SA047	GCF_003064005.1
Bacillus anthracis strain A16/LA896	GCF_003064015.1
Bacillus anthracis strain A24/TN_Bovine_Sokol	GCF_003064045.1

Suite dans la page suivante

Noms des souches	Numéros d'accession
Bacillus anthracis strain A46	GCF_003064085.1
Bacillus anthracis strain London_499	GCF_003227955.1
Bacillus anthracis strain Sterne 09RA8929	GCF_003335125.1
Bacillus anthracis strain UT308	GCF_003345105.1
Bacillus anthracis strain Strain 32	GCF_003367985.1
Bacillus anthracis strain SK57	GCF_003368005.1
Bacillus anthracis strain 81/1	GCF_003860145.1
Bacillus anthracis strain 17OD930	GCF_006088855.1
Bacillus anthracis strain PCr	GCF_006742565.1
Bacillus anthracis strain KC2011	GCF_008087155.1
Bacillus anthracis strain ANSES_94	GCF_900011745.1
Bacillus anthracis strain ANSES_86	GCF_900012555.1
Bacillus anthracis strain ANSES_101	GCF_900013465.1
Bacillus anthracis strain ANSES_89	GCF_900013475.1
Bacillus anthracis strain ANSES_92	GCF_900013485.1
Bacillus anthracis strain ANSES_118	GCF_900014335.1
Bacillus anthracis strain ANSES_122	GCF_900014345.1
Bacillus anthracis strain ANSES_32	GCF_900014355.1
Bacillus anthracis strain ANSES_88	GCF_900014365.1
Bacillus anthracis strain ANSES_90	GCF_900014375.1

TABLE B.1 – Supplementary file 2 : List of *Bacillus anthracis* strains for KRAKEN2 custom database

B.3 Fichier complémentaire 3

ID	Description	Type	Localisation	Spiked	Spiking*
B2F05	May 14th, 2017	Wastewater	France	Yes	BT-BC-FH
B2F10	May 11th, 2017	Wastewater	Germany	Yes	BT-BC-FH
B2F12	May 11th, 2017	Wastewater	Germany	No	/
B2F13	October 30th, 2017	Wastewater	Germany	No	/
B2F14	October 30th, 2017	Wastewater	Germany	Yes	BC-BA-FT-BM-YP
B2F15	October 30th, 2017	Wastewater	Germany	No	/
B2F16	October 30th, 2017	Wastewater	Germany	Yes	BC-BA-FT-BM-YP

TABLE B.2 – Summary of the samples included in this study

*Spiking
BT : *Bacillus thuringiensis*
BC : *Bacillus cereus*
FH : *Francisella hispaniensis*
BA : *Bacillus anthracis*
FT : *Francisella tularensis*

YP : *Yersinia pestis*

Bibliographie

ABADIA, G. et al. (2005). *Recommandations pour la surveillance et la lutte contre le charbon animal et humain*. Guide méthodologique. Santé publique France, p. 31. URL : https://www.santepubliquefrance.fr.

ABDEL-GLIL, Mostafa Y et al. (2021). « A whole-genome-based gene-by-gene typing system for standardized high-resolution strain typing of Bacillus anthracis ». In : *Journal of Clinical Microbiology* 59.7, p. 10-1128.

ABDELLI, Mehdi et al. (2023). « Get to Know Your Neighbors : Characterization of Close Bacillus anthracis Isolates and Toxin Profile Diversity in the Bacillus cereus Group ». In : *Microorganisms* 11.11, p. 2721.

ABO-ABA, SEM et al. (2015). « Draft genome sequence of Bacillus species from the rhizosphere of the desert plant Rhazya stricta ». In : *Genome Announcements* 3.6, p. 10-1128.

ABRAMOVA, Faina A et al. (1993). « Pathology of inhalational anthrax in 42 cases from the Sverdlovsk outbreak of 1979. » In : *Proceedings of the national academy of sciences* 90.6, p. 2291-2294.

ACHTMAN, Mark (2008). « Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens ». In : *Annu. Rev. Microbiol.* 62, p. 53-70.

ALIBEK, Kenneth, Catherine LOBANOVA et Serguei POPOV (2009). « Anthrax : a disease and a weapon ». In : *Bioterrorism and infectious agents : A new dilemma for the 21st century*, p. 1-35.

ALIKHAN, Nabil-Fareed et al. (2011). « BLAST Ring Image Generator (BRIG) : simple prokaryote genome comparisons ». In : *BMC genomics* 12.1, p. 1-10.

ALTSCHUL, Stephen F et al. (1990). « Basic local alignment search tool ». In : *Journal of molecular biology* 215.3, p. 403-410.

ANTONATION, Kym S et al. (2016). « Bacillus cereus biovar anthracis causing anthrax in sub-Saharan Africa—chromosomal monophyly and broad geographic distribution ». In : *PLoS neglected tropical diseases* 10.9, e0004923.

ANTONINI, James M (2003). « Health effects of welding ». In : *Critical reviews in toxicology* 33.1, p. 61-103.

ANTWERPEN, Markus H et al. (2008). « Real-time PCR system targeting a chromosomal marker specific for Bacillus anthracis ». In : *Molecular and cellular probes* 22.5-6, p. 313-315.

AUWERA, Géraldine Van der et Jacques MAHILLON (2008). « Transcriptional analysis of the conjugative plasmid pAW63 from Bacillus thuringiensis ». In : *Plasmid* 60.3, p. 190-199.

AUWERA, Géraldine A Van der, Lars ANDRUP et Jacques MAHILLON (2005). « Conjugative plasmid pAW63 brings new insights into the genesis of the Bacillus anthracis virulence plasmid pXO2 and of the Bacillus thuringiensis plasmid pBT9727 ». In : *BMC genomics* 6, p. 1-14.

AVASHIA, Swati B et al. (2007). « Fatal pneumonia among metalworkers due to inhalation exposure to Bacillus cereus containing Bacillus anthracis toxin genes ». In : *Clinical infectious diseases* 44.3, p. 414-416.

Avis de l'ANSES relatif à la saisine n° 2016-SA-0286 (2016). Rapp. tech. Avis de l'Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail relatif à l'évaluation de la présence de spores de Bacillus anthracis dans différents milieux (eau, sol, aliments) et l'évaluation du risque pour la santé humaine lié à différentes voies d'exposition (voie respiratoire, cutanée, digestive). 14 rue Pierre et Marie Curie, 94701 Maisons-Alfort Cedex, France : Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail (ANSES). URL : https://www.anses.fr/fr/system/files/BIORISK2016SA0286.pdf.

BAĞCI, Caner et al. (2019). « Introduction to the analysis of environmental sequences : metagenomics with MEGAN ». In : *Evolutionary genomics : statistical and computational methods*, p. 591-604.

BAKKEN, Lars R (1985). « Separation and purification of bacteria from soil ». In : *Applied and Environmental Microbiology* 49.6, p. 1482-1487.

BALDWIN, Victoria M (2020). « You can't B. cereus–a review of Bacillus cereus strains that cause anthrax-like disease ». In : *Frontiers in microbiology* 11, p. 1731.

BARRAS, Vincent et Gilbert GREUB (2014). « History of biological warfare and bioterrorism ». In : *Clinical Microbiology and infection* 20.6, p. 497-502.

BASSY, Olga et al. (2023). « Spanish outbreak isolates bridge phylogenies of European and American Bacillus anthracis ». In : *Microorganisms* 11.4, p. 889.

BE, Nicholas A et al. (2013). « Detection of Bacillus anthracis DNA in complex soil and air samples using next-generation sequencing ». In : *PLOS one* 8.9, e73455.

BEALL, Francis A, Martha J TAYLOR et Curtis B THORNE (1962). « Rapid lethal effect in rats of a third component found upon fractionating the toxin of Bacillus anthracis ». In : *Journal of Bacteriology* 83.6, p. 1274-1280.

BEATTY, Mark E et al. (2003). « Gastrointestinal anthrax : review of the literature ». In : *Archives of Internal Medicine* 163.20, p. 2527-2531.

BEKEMEYER, William B et Guy A ZIMMERMAN (1985). « Life-threatening complications associated with Bacillus cereus pneumonia ». In : *American Review of Respiratory Disease* 131.3, p. 466-469.

BERGER, T, M KASSIRER et AA ARAN (2014). « Injectional anthrax-new presentation of an old disease ». In : *Eurosurveillance* 19.32, p. 20877.

BEYER, Wolfgang et al. (2012). « Distribution and molecular evolution of Bacillus anthracis genotypes in Namibia ». In : *PLoS neglected tropical diseases* 6.3, e1534.

BHATTACHARJEE, Yudhijit (2009). *Paul Keim on His Life With the FBI During the Anthrax Investigation*.

BIORENDER (2020). *Nanopore Sequencing*. https://app.biorender.com/biorender-templates/figures/all/t-5f8717bcfb2c3900a82de0ae-nanopore-sequencing.

BLANCO-MÍGUEZ, Aitor et al. (2023). « Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4 ». In : *Nature Biotechnology*, p. 1-12.

BLANCOU, Jean (2000). *Histoire de la surveillance et du contrôle des maladies animales transmissibles*. Office international des épizooties.

BONIS, Mathilde et al. (2021). « Comparative phenotypic, genotypic and genomic analyses of Bacillus thuringiensis associated with foodborne outbreaks in France ». In : *PloS one* 16.2, e0246885.

BRADY, Arthur et Steven SALZBERG (2011). « PhymmBL expanded : confidence scores, custom databases, parallelization and more ». In : *Nature methods* 8.5, p. 367-368.

BRAUN, Peter et al. (2022). « Reoccurring bovine anthrax in Germany on the same pasture after 12 years ». In : *Journal of clinical microbiology* 60.3, e02291-21.

BREITWIESER, Florian P, Daniel N BAKER et Steven L SALZBERG (2018). « KrakenUniq : confident and fast metagenomics classification using unique k-mer counts ». In : *Genome biology* 19.1, p. 1-10.

BREITWIESER, Florian P, Jennifer LU et Steven L SALZBERG (2019). « A review of methods and databases for metagenomic classification and assembly ». In : *Briefings in bioinformatics* 20.4, p. 1125-1136.

BRETSCHNEIDER, Anne, David G HECKEL et Yannick PAUCHET (2016). « Three toxins, two receptors, one mechanism : Mode of action of Cry1A toxins from Bacillus thuringiensis in Heliothis virescens ». In : *Insect biochemistry and molecular biology* 76, p. 109-117.

BRÉZILLON, Christophe et al. (2015). « Capsules, toxins and AtxA as virulence factors of emerging Bacillus cereus biovar anthracis ». In : *PLoS neglected tropical diseases* 9.4, e0003455.

BRILLARD, Julien et Didier LERECLUS (2004). « Comparison of cytotoxin cytK promoters from Bacillus cereus strain ATCC 14579 and from a B. cereus food-poisoning strain ». In : *Microbiology* 150.8, p. 2699-2705.

BROSSIER, F et M MOCK (2001). « Toxins of Bacillus anthracis ». In : *Toxicon* 39.11, p. 1747-1755.

BROWN, Eric R et William B CHERRY (1955). « Specific identification of Bacillus anthracis by means of a variant bacteriophage ». In : *The Journal of Infectious Diseases*, p. 34-39.

BRUCE, Spencer A et al. (2020). « A classification framework for Bacillus anthracis defined by global genomic structure ». In : *Evolutionary Applications* 13.5, p. 935-944.

BRUCKNER, V., J. KOVÁCS et G. DÉNES (1953). « Structure of Poly-D-glutamic Acid isolated from Capsulated Strains of B. anthracis ». In : *Nature* 172, p. 508.

BURTON, Joshua N et al. (2014). « Species-level deconvolution of metagenome assemblies with Hi-C–based contact probability maps ». In : *G3 : Genes, Genomes, Genetics* 4.7, p. 1339-1346.

CANDELA, Thomas, Michele MOCK et Agnes FOUET (2005). « CapE, a 47-amino-acid peptide, is necessary for Bacillus anthracis polyglutamate capsule synthesis ». In : *Journal of Bacteriology* 187.22, p. 7765-7772.

CARLSON, Colin J et al. (2018). « Spores and soil from six sides : interdisciplinarity and the environmental biology of anthrax (Bacillus anthracis) ». In : *Biological Reviews* 93.4, p. 1813-1831.

CARLSON, Colin J et al. (2019). « The global distribution of Bacillus anthracis and associated anthrax risk to humans, livestock and wildlife ». In : *Nature microbiology* 4.8, p. 1337-1343.

CARROLL, Laura M, Rachel A CHENG et Jasna KOVAC (2020). « No assembly required : using BTyper3 to assess the congruency of a proposed taxonomic framework for the Bacillus cereus group with historical typing methods ». In : *Frontiers in Microbiology* 11, p. 580691.

CARROLL, Laura M, Martin WIEDMANN et Jasna KOVAC (2020). « Proposal of a taxonomic nomenclature for the Bacillus cereus group which reconciles genomic definitions of bacterial species with clinical and industrial phenotypes ». In : *MBio* 11.1, p. 10-1128.

CARROLL, Laura M et al. (2022a). « Keeping up with the Bacillus cereus group : taxonomy through the genomics era and beyond ». In : *Critical reviews in food science and nutrition* 62.28, p. 7677-7702.

CARROLL, Laura M et al. (2022b). « Strains Associated with Two 2020 Welder Anthrax Cases in the United States Belong to Separate Lineages within Bacillus cereus sensu lato ». In : *Pathogens* 11.8, p. 856.

CENTERS FOR DISEASE CONTROL AND PREVENTION (2018). *Bioterrorism Agents/Diseases*. Consulté le 27 septembre 2023. URL : https://emergency.cdc.gov/agent/agentlist-category.asp.

— (2020). *History of Anthrax*. https://www.cdc.gov/anthrax/basics/anthrax-history.html. Consulté le 4 novembre 2023.

— (2023). *Types of Anthrax*. https://www.cdc.gov/anthrax/basics/types/index.html. Accessed : 2023-10-12.

CHEN, Jianshun et al. (2014). « First reported fatal Bacillus thuringiensis infections in Chinese soft-shelled turtles (Trionyx sinensis) ». In : *Aquaculture* 428, p. 16-20.

CHENG, Li-Wu et al. (2023). « Pathogenicity and Genomic Characterization of a Novel Genospecies, Bacillus shihchuchen, of the Bacillus cereus Group Isolated from Chinese Softshell Turtle (Pelodiscus sinensis) ». In : *International Journal of Molecular Sciences* 24.11, p. 9636.

COETZER, JAW, GR THOMSON et RC TUSTIN (1995). « Infectious diseases of livestock with special reference to Southern Africa ». In : *Journal of the South African Veterinary Association* 66.2, p. 106.

COHAN, Frederick M (2006). « Towards a conceptual and operational union of bacterial systematics, ecology, and evolution ». In : *Philosophical Transactions of the Royal Society B : Biological Sciences* 361.1475, p. 1985-1996.

CRICKMORE, N et al. (2020). « Bacterial pesticidal protein resource center ». In : *Data. Available online : https ://www. bpprc. org (accessed on 10 January 2024)*.

CROUCHER, Nicholas J et al. (2015). « Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins ». In : *Nucleic acids research* 43.3, e15-e15.

DELMONT, Tom O et al. (2011). « Accessing the soil metagenome for studies of microbial diversity ». In : *Applied and environmental microbiology* 77.4, p. 1315-1324.

DERZELLE, Sylviane et Simon THIERRY (2013). « Genetic diversity of Bacillus anthracis in Europe : genotyping methods in forensic and epidemiologic investigations ». In : *Biosecurity and Bioterrorism : Biodefense Strategy, Practice, and Science* 11.S1, S166-S176.

DESAI, Narayan et al. (2012). « From genomics to metagenomics ». In : *Current opinion in biotechnology* 23.1, p. 72-76.

DIETRICH, Richard et al. (2021). « The food poisoning toxins of Bacillus cereus ». In : *Toxins* 13.2, p. 98.

DIXON, T.C. et al. (1999). « Anthrax ». In : *New England Journal of Medicine* 341, p. 815-826.

DOGANAY, Mehmet et Hayati DEMIRASLAN (2015). « Human anthrax as a re-emerging disease ». In : *Recent Patents on Anti-Infective Drug Discovery* 10.1, p. 10-29.

DONG, Mei-Jing, Hao LUO et Feng GAO (2022). « Ori-Finder 2022 : a comprehensive web server for prediction and analysis of bacterial replication origins ». In : *Genomics, Proteomics & Bioinformatics* 20.6, p. 1207-1213.

DOYLE, Ronald J, Fariborz NEDJAT-HAIEM et Jyoti S SINGH (1984). « Hydrophobic characteristics of Bacillus spores ». In : *Current Microbiology* 10, p. 329-332.

DRAGON, DC et BT ELKIN (2001). « An overview of early anthrax outbreaks in northern Canada : Field reports of the Health of Animals Branch, Agriculture Canada, 1962–71 ». In : *Arctic*, p. 32-40.

DRAGON, DC, RP RENNIE et BT ELKIN (2001). « Detection of anthrax spores in endemic regions of northern Canada ». In : *Journal of applied microbiology* 91.3, p. 435-441.

DRAGON, DC et al. (1999). « A review of anthrax in Canada and implications for research on the disease in northern bison ». In : *Journal of Applied Microbiology* 87.2, p. 208-213.

DROMIGNY, Éric (2009). *Bacillus anthracis*. Paris : Tec & Doc. ISBN : 9782743011949.

DRYSDALE, Melissa et al. (2004). « atxA controls Bacillus anthracis capsule synthesis via acpA and a newly discovered regulator, acpB ». In : *Journal of bacteriology* 186.2, p. 307-315.

DUPKE, S et al. (2018). « Analysis of a newly discovered antigen of Bacillus cereus biovar anthracis for its suitability in specific serological antibody testing ». In : *Journal of applied microbiology* 126.1, p. 311-323.

DUPKE, Susann et al. (2020). « Serological evidence for human exposure to Bacillus cereus biovar anthracis in the villages around Taï National Park, Côte d'Ivoire ». In : *PLoS Neglected Tropical Diseases* 14.5, e0008292.

DURMAZ, Rıza et al. (2012). « Molecular epidemiology of the Bacillus anthracis isolates collected throughout Turkey from 1983 to 2011 ». In : *European journal of clinical microbiology & infectious diseases* 31, p. 2783-2790.

EASTERDAY, W Ryan et al. (2005). « Use of single nucleotide polymorphisms in the plcR gene for specific identification of Bacillus anthracis ». In : *Journal of clinical microbiology* 43.4, p. 1995-1997.

EHLING-SCHULZ, M, D LERECLUS et TM KOEHLER (2019). *The Bacillus cereus group : Bacillus species with pathogenic potential. Microbiol Spectr 7 : GPP3-0032-2018*.

EHLING-SCHULZ, Monika, Martina FRICKER et Siegfried SCHERER (2004). « Identification of emetic toxin producing Bacillus cereus strains by a novel molecular assay ». In : *FEMS microbiology letters* 232.2, p. 189-195.

EHLING-SCHULZ, Monika et al. (2006). « Cereulide synthetase gene cluster from emetic Bacillus cereus : structure and location on a mega virulence plasmid related to Bacillus anthracis toxin plasmid pXO1 ». In : *BMC microbiology* 6.1, p. 1-11.

ENDEN, Erwin Van den, Alphons VAN GOMPEL et Marjan VAN ESBROECK (2006). « Cutaneous anthrax, Belgian traveler ». In : *Emerging Infectious Diseases* 12.3, p. 523.

EREMENKO, Eugene et al. (2021). « Phylogenetics of Bacillus anthracis isolates from Russia and bordering countries ». In : *Infection, Genetics and Evolution* 92, p. 104890.

ETIENNE-TOUMELIN, Isabelle et al. (1995). « Characterization of the Bacillus anthracis S-layer : cloning and sequencing of the structural gene ». In : *Journal of Bacteriology* 177.3, p. 614-620.

EUZÉBY, Jean Paul (1997). « List of Bacterial Names with Standing in Nomenclature : a folder available on the Internet ». In : *International Journal of Systematic and Evolutionary Microbiology* 47.2, p. 590-592.

FEIL, Edward J et Mark C ENRIGHT (2004). « Analyses of clonality and the evolution of bacterial pathogens ». In : *Current opinion in microbiology* 7.3, p. 308-313.

FERRIÈRES, Madeleine (2002). « Chapitre 10. De l'épizootie à l'épidémie ». In : *Histoire des peurs alimentaires. Du Moyen Âge à l'aube du XXe siècle*. Sous la dir. de Madeleine FERRIÈRES. L'Univers historique. Le Seuil, p. 262-293. URL : https://www.cairn.info/histoire-des-peurs-alimentaires-du-moyen-age-a-l-a--9782020476614-page-262.htm.

FOURNIER, Nicolas (1769). *Observations et expériences sur le charbon malin : avec une méthode assurée de la guérir*. Paris : Defay.

FRAZIER, Aletta Ann, Teri J FRANKS et Jeffrey R GALVIN (2006). « Inhalational anthrax ». In : *Journal of thoracic imaging* 21.4, p. 252-258.

GAINER, Robert S, Gilles VERGNAUD et Martin E HUGH-JONES (2020). « A Review of Arguments for the Existence of Latent Infections of Bacillus anthracis, and Research Needed to Understand their Role in the Outbreaks of Anthrax ». In : *Microorganisms* 8.6, p. 800.

GIBBS, E Paul J (2014). « The evolution of One Health : a decade of progress and challenges for the future ». In : *Veterinary Record* 174.4, p. 85-91.

GIERCZYŃSKI, Rafał et al. (2004). « Intriguing diversity of Bacillus anthracis in eastern Poland–the molecular echoes of the past outbreaks ». In : *FEMS microbiology letters* 239.2, p. 235-240.

GIRAULT, Guillaume (2015). « Développement d'outils de typage moléculaire de haute résolution pour la détection et la différenciation de Bacillus anthracis ». Thèse de doct. AgroParisTech.

GOHAR, Michel et al. (2002). « Two-dimensional electrophoresis analysis of the extracellular proteome of Bacillus cereus reveals the importance of the PlcR regulon ». In : *Proteomics* 2.6, p. 784-791.

GOLDFEDER, Rachel L et al. (2017). « Human genome sequencing at the population scale : a primer on high-throughput DNA sequencing and analysis ». In : *American journal of epidemiology* 186.8, p. 1000-1009.

GOODING, J Justin (2006). « Biosensor technology for detecting biological warfare agents : Recent progress and future trends ». In : *Analytica chimica acta* 559.2, p. 137-151.

GRANUM, Per Einar, Kristin O'SULLIVAN et Terje LUND (1999). « The sequence of the non-haemolytic enterotoxin operon from Bacillus cereus ». In : *FEMS microbiology letters* 177.2, p. 225-229.

GUIDI-RONTANI, Chantal et al. (1999a). « Germination of Bacillus anthracis spores within alveolar macrophages ». In : *Molecular microbiology* 31.1, p. 9-17.

GUIDI-RONTANI, Chantal et al. (1999b). « Identification and characterization of a germination operon on the virulence plasmid pXOl of Bacillus anthracis ». In : *Molecular microbiology* 33.2, p. 407-414.

GUIGNOT, Julie, Michèle MOCK et Agnès FOUET (1997). « AtxA activates the transcription of genes harbored by both Bacillus anthracis virulence plasmids ». In : *FEMS microbiology letters* 147.2, p. 203-207.

GUINEBRETIÈRE, Marie-Hélène et al. (2008). « Ecological diversification in the Bacillus cereus group ». In : *Environmental Microbiology* 10.4, p. 851-865.

GUINEBRETIÈRE, Marie-Hélène et al. (2013). « Bacillus cytotoxicus sp. nov. is a novel thermotolerant species of the Bacillus cereus group occasionally associated with food poisoning ». In : *International journal of systematic and evolutionary microbiology* 63.Pt_1, p. 31-40.

HACHISUKA, Yoetsu, Satoshi KOZUKA et Masayuki TSUJIKAWA (1984). « Exosporia and appendages of spores of Bacillus species ». In : *Microbiology and immunology* 28.5, p. 619-624.

HAGENBOURGER, Martin (2003). « The French Post Office and anthrax : Key lessons and new questions ». In : *Journal of Contingencies and Crisis Management* 11.3, p. 124-128.

HAHN, Beth L, Sonia SHARMA et Peter G SOHNLE (2005). « Analysis of epidermal entry in experimental cutaneous Bacillus anthracis infections in mice ». In : *Journal of Laboratory and Clinical Medicine* 146.2, p. 95-102.

HAN, Cliff S et al. (2006). « Pathogenomic sequence analysis of Bacillus cereus and Bacillus thuringiensis isolates closely related to Bacillus anthracis ». In : *Journal of bacteriology* 188.9, p. 3382-3390.

HANDELSMAN, Jo et al. (1998). « Molecular biological access to the chemistry of unknown soil microbes : a new frontier for natural products ». In : *Chemistry & biology* 5.10, R245-R249.

HELGASON, Erlendur et al. (2004). « Multilocus sequence typing scheme for bacteria of the Bacillus cereus group ». In : *Applied and environmental microbiology* 70.1, p. 191-201.

HERNANDEZ, Eric et al. (1998). « Bacillus thuringiensis subsp. konkukian (serotype H34) superinfection : case report and experimental evidence of pathogenicity in immunosuppressed mice ». In : *Journal of clinical microbiology* 36.7, p. 2138-2139.

HILMAS, CJ et J ANDERSON (2015). « Chapter 29—Anthrax A2. Gupta RC, ed ». In : *Handbook of Toxicology of Chemical Warfare Agents*, p. 387-410.

HINNEKENS, Pauline et Jacques MAHILLON (2022). « Conjugation-mediated transfer of pXO16, a large plasmid from Bacillus thuringiensis sv. israelensis, across the Bacillus cereus group and its impact on host phenotype ». In : *Plasmid* 122, p. 102639.

HINNEKENS, Pauline et al. (2019). « pXO16, the large conjugative plasmid from Bacillus thuringiensis serovar israelensis displays an extended host spectrum ». In : *Plasmid* 102, p. 46-50.

HOFFMANN, Constanze et al. (2017). « Persistent anthrax as a major driver of wildlife mortality in a tropical rainforest ». In : *Nature* 548.7665, p. 82-86.

HOFFMASTER, Alex R et Theresa M KOEHLER (1999). « Autogenous regulation of the Bacillus anthracis pag operon ». In : *Journal of bacteriology* 181.15, p. 4485-4492.

HOFFMASTER, Alex R et al. (2004). « Identification of anthrax toxin genes in a Bacillus cereus associated with an illness resembling inhalation anthrax ». In : *Proceedings of the National Academy of Sciences* 101.22, p. 8449-8454.

HOFFMASTER, Alex R et al. (2006). « Characterization of Bacillus cereus isolates associated with fatal pneumonias : strains are closely related to Bacillus anthracis and harbor B. anthracis virulence genes ». In : *Journal of clinical microbiology* 44.9, p. 3352-3360.

HOLTY, Jon-Erik C, Rebecca Y KIM et Dena M BRAVATA (2006). « Anthrax : a systematic review of atypical presentations ». In : *Annals of emergency medicine* 48.2, p. 200-211.

HOMÈRE (s. d.). *L'Iliade*.

HU, Taishan et al. (2021). « Next-generation sequencing technologies : An overview ». In : *Human Immunology* 82.11, p. 801-811.

HU, Xiaomin et al. (2009). « Distribution, diversity, and potential mobility of extrachromosomal elements related to the Bacillus anthracis pXO1 and pXO2 virulence plasmids ». In : *Applied and environmental microbiology* 75.10, p. 3016-3028.

HUANG, E. (2023). *The History of DNA Sequencing*. https://app.biorender.com/biorender-templates/figures/all/t-64060407e9f63f5b95cb8019-the-history-of-dna-sequencing.

HUANG, Weichun et al. (2012). « ART : a next-generation sequencing read simulator ». In : *Bioinformatics* 28.4, p. 593-594.

HUGH-JONES, ME et V DE VOS (2002). « Anthrax and wildlife ». In : *Revue Scientifique et Technique-Office International des Epizooties* 21.1, p. 359-384.

HULTON, CSJ, CF HIGGINS et PM SHARP (1991). « ERIC sequences : a novel family of repetitive elements in the genomes of Escherichia coli, Salmonella typhimurium and other enterobacteria ». In : *Molecular microbiology* 5.4, p. 825-834.

INGLESBY, Thomas V et al. (2002). « Anthrax as a biological weapon, 2002 : updated recommendations for management ». In : *Jama* 287.17, p. 2236-2252.

IRENGE, Leonid M et al. (2020). « Complete Genome Sequence of an Environmental Bacillus cereus Isolate Belonging to the Bacillus anthracis Clade ». In : *Microbiology resource announcements* 9.47, p. 10-1128.

JAIN, Chirag et al. (2018). « High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries ». In : *Nature communications* 9.1, p. 5114.

JENSEN, GB et al. (2003). « The hidden lifestyles of Bacillus cereus and relatives ». In : *Environmental microbiology* 5.8, p. 631-640.

JENSEN, J, H KLEEMEYER et al. (1953). « The Bacteriological Differential Diagnosis of Anthrax by means of a Specific Test (" String of Pearls-Test"). » In : *Zentralblatt fur Bakteriologie, Parasitenkunde, Infektionskrankheiten und Hygiene* 159.8, p. 494-500.

JERNIGAN, Daniel B et al. (2002). « Investigation of bioterrorism-related anthrax, United States, 2001 : epidemiologic findings ». In : *Emerging infectious diseases* 8.10, p. 1019.

JIMÉNEZ, Guillermo et al. (2013). « Description of Bacillus toyonensis sp. nov., a novel species of the Bacillus cereus group, and pairwise genome comparisons of the species of the group by means of ANI calculations ». In : *Systematic and applied microbiology* 36.6, p. 383-391.

JOLLEY, Keith A, James E BRAY et Martin CJ MAIDEN (2018). « Open-access bacterial population genomics : BIGSdb software, the PubMLST. org website and their applications ». In : *Wellcome open research* 3.

JUERGENSMEYER, Margaret A et al. (2006). « A selective chromogenic agar that distinguishes Bacillus anthracis from Bacillus cereus and Bacillus thuringiensis ». In : *Journal of food protection* 69.8, p. 2002-2006.

JUNG, Kyoung Hwa et al. (2012). « Genetic populations of Bacillus anthracis isolates from Korea ». In : *Journal of Veterinary Science* 13.4, p. 385-393.

KAMAL, Nazia et al. (2017). « Structural and immunochemical relatedness suggests a conserved pathogenicity motif for secondary cell wall polysaccharides in Bacillus anthracis and infection-associated Bacillus cereus ». In : *Plos one* 12.8, e0183115.

KAMAL, SM et al. (2011). « Anthrax : an update ». In : *Asian Pacific journal of tropical biomedicine* 1.6, p. 496-501.

KASSEN, Rees, Martin LLEWELLYN et Paul B RAINEY (2004). « Ecological constraints on diversification in a model adaptive radiation ». In : *Nature* 431.7011, p. 984-988.

KEIM, P et al. (2000). « Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within Bacillus anthracis ». In : *Journal of bacteriology* 182.10, p. 2928-2936.

KEIM, Paul et al. (1997). « Molecular evolution and diversity in Bacillus anthracis as detected by amplified fragment length polymorphism markers ». In : *Journal of bacteriology* 179.3, p. 818-824.

KEIM, Paul et al. (2004). « Anthrax molecular epidemiology and forensics : using the appropriate marker for different evolutionary scales ». In : *Infection, Genetics and Evolution* 4.3, p. 205-213.

KEIM, Paul S et David M WAGNER (2009). « Humans and evolutionary and ecological forces shaped the phylogeography of recently emerged diseases ». In : *Nature Reviews Microbiology* 7.11, p. 813-821.

KENEFIC, Leo J et al. (2009). « Pre-columbian origins for north american anthrax ». In : *PLoS One* 4.3, e4813.

KHMALADZE, Ekaterine et al. (2014). « Phylogeography of Bacillus anthracis in the country of Georgia shows evidence of population structuring and is dissimilar to other regional genotypes ». In : *PLoS One* 9.7, e102651.

KIM, Daehwan et al. (2016). « Centrifuge : rapid and sensitive classification of metagenomic sequences ». In : *Genome research* 26.12, p. 1721-1729.

KLEE, Silke R et al. (2006). « Characterization of Bacillus anthracis-like bacteria isolated from wild great apes from Cote d'Ivoire and Cameroon ». In : *Journal of bacteriology* 188.15, p. 5333-5344.

KLEE, Silke R et al. (2010). « The genome of a Bacillus isolate causing anthrax in chimpanzees combines chromosomal properties of B. cereus with B. anthracis virulence plasmids ». In : *PloS one* 5.7, e10986.

KNISELY, Ralph F (1966). « Selective medium for Bacillus anthracis ». In : *Journal of Bacteriology* 92.3, p. 784-786.

KOCH, Robert (1876). « The etiology of anthrax, based on the life history of Bacillus anthracis ». In : *Beiträge zur Biologie der Pflanzen* 2.2, p. 277-310.

KOEHLER, Theresa M, Zhihao DAI et Mary KAUFMAN-YARBRAY (1994). « Regulation of the Bacillus anthracis protective antigen gene : CO2 and a trans-acting element activate transcription from one of two promoters ». In : *Journal of bacteriology* 176.3, p. 586-595.

KOEHLER, TM (2002). « Bacillus anthracis genetics and virulence gene regulation ». In : *Anthrax*, p. 143-164.

KOLSTØ, Anne-Brit, Nicolas J TOURASSE et Ole Andreas ØKSTAD (2009). « What sets Bacillus anthracis apart from other Bacillus species ? » In : *Annual review of microbiology* 63, p. 451-476.

KOLTON, Cari B et al. (2017). « Bacillus anthracis gamma phage lysis among soil bacteria : an update on test specificity ». In : *BMC Research Notes* 10.1, p. 1-6.

KORF, Ian, Mark YANDELL et Joseph BEDELL (2003). *Blast*. " O'Reilly Media, Inc."

KOZLOV, Alexey M et al. (2019). « RAxML-NG : a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference ». In : *Bioinformatics* 35.21, p. 4453-4455.

KUEHN, Andrea et al. (2009). « Development of antibodies against anthrose tetrasaccharide for specific detection of Bacillus anthracis spores ». In : *Clinical and Vaccine Immunology* 16.12, p. 1728-1737.

LE FLÈCHE, Philippe et al. (2001). « A tandem repeats database for bacterial genomes : application to the genotyping of Yersinia pestis and Bacillus anthracis ». In : *BMC microbiology* 1.1, p. 1-14.

LECHNER, Sabine et al. (1998). « Bacillus weihenstephanensis sp. nov. is a new psychrotolerant species of the Bacillus cereus group ». In : *International journal of systematic and evolutionary microbiology* 48.4, p. 1373-1382.

LEENDERTZ, Fabian H et al. (2004). « Anthrax kills wild chimpanzees in a tropical rainforest ». In : *Nature* 430.6998, p. 451-452.

LEENDERTZ, Fabian H et al. (2006). « Anthrax in Western and Central African great apes ». In : *American journal of primatology* 68.9, p. 928-933.

LEPPLA, SH, N ARORA et M VARUGHESE (1999). « Anthrax toxin fusion proteins for intracellular delivery of macromolecules ». In : *Journal of applied microbiology* 87.2, p. 284-284.

LEPPLA, Stephen H (1982). « Anthrax toxin edema factor : a bacterial adenylate cyclase that increases cyclic AMP concentrations of eukaryotic cells. » In : *Proceedings of the National Academy of Sciences* 79.10, p. 3162-3166.

— (1988). « [16] Production and purification of anthrax toxin ». In : *Methods in enzymology*. T. 165. Elsevier, p. 103-116.

LETUNIC, Ivica et Peer BORK (2021). « Interactive Tree Of Life (iTOL) v5 : an online tool for phylogenetic tree display and annotation ». In : *Nucleic acids research* 49.W1, W293-W296.

LI, Yan-qiu et al. (2017). « Complete genome sequence of Bacillus thuringiensis Bt185, a potential soil insect biocontrol agent ». In : *Journal of integrative agriculture* 16.3, p. 749-751.

LIENEMANN, Taru et al. (2018). « Genotyping and phylogenetic placement of Bacillus anthracis isolates from Finland, a country with rare anthrax cases ». In : *BMC microbiology* 18, p. 1-7.

LINDNER, Martin S et Bernhard Y RENARD (2013). « Metagenomic abundance estimation and diagnostic testing on species level ». In : *Nucleic acids research* 41.1, e10-e10.

LISTA, Florigio et al. (2006). « Genotyping of Bacillus anthracis strains based on automated capillary 25-loci multiple locus variable-number tandem repeats analysis ». In : *BMC microbiology* 6, p. 1-15.

LIU, Junfeng et al. (2023). « Centrifuge+ : improving metagenomic analysis upon Centrifuge ». In : *bioRxiv*, p. 2023-02.

LIU, Yang et al. (2017). « Proposal of nine novel species of the Bacillus cereus group ». In : *International journal of systematic and evolutionary microbiology* 67.8, p. 2499-2508.

LOGSDON, Glennis A, Mitchell R VOLLGER et Evan E EICHLER (2020). « Long-read human genome sequencing and its applications ». In : *Nature Reviews Genetics* 21.10, p. 597-614.

LU, Jennifer et al. (2017). « Bracken : estimating species abundance in metagenomics data ». In : *PeerJ Computer Science* 3, e104.

LUND, Terje, Marie-Laure DE BUYSER et Per Einar GRANUM (2000). « A new cytotoxin from Bacillus cereus that may cause necrotic enteritis ». In : *Molecular microbiology* 38.2, p. 254-261.

LUO, Chengwei et al. (2015). « ConStrains identifies microbial strains in metagenomic datasets ». In : *Nature biotechnology* 33.10, p. 1045-1052.

MAKART, Lionel et al. (2018). « A novel T4SS-mediated DNA transfer used by pXO16, a conjugative plasmid from Bacillus thuringiensis serovar israelensis ». In : *Environmental microbiology* 20.4, p. 1550-1561.

MAKINO, Sl et al. (1989). « Molecular characterization and protein analysis of the cap region, which is essential for encapsulation in Bacillus anthracis ». In : *Journal of bacteriology* 171.2, p. 722-730.

MAKINO, Sou-ichi et al. (2002). « Effect of the lower molecular capsule released from the cell surface of Bacillus anthracis on the pathogenesis of anthrax ». In : *The Journal of infectious diseases* 186.2, p. 227-233.

MANCHEE, Richard J et al. (1983). « Decontamination of Bacillus anthracis on Gruinard Island ? » In : *Nature* 303.5914, p. 239-240.

MANCHEE, Richard J et al. (1994). « Formaldehyde solution effectively inactivates spores of Bacillus anthracis on the Scottish island of Gruinard ». In : *Applied and environmental microbiology* 60.11, p. 4167-4171.

MANCHEE, RJ et al. (1981). « Bacillus anthracis on Gruinard island ». In : *Nature* 294.5838, p. 254-255.

MARÇAIS, Guillaume et Carl KINGSFORD (2011). « A fast, lock-free approach for efficient parallel counting of occurrences of k-mers ». In : *Bioinformatics* 27.6, p. 764-770.

MARSTON, Chung K et al. (2011). « Molecular epidemiology of anthrax cases associated with recreational use of animal hides and yarn in the United States ». In : *PLoS One* 6.12, e28274.

MARSTON, Chung K et al. (2016). « Anthrax toxin-expressing Bacillus cereus isolated from an anthrax-like eschar ». In : *PLoS One* 11.6, e0156987.

MARSTON, CK et al. (2008). « Evaluation of two selective media for the isolation of Bacillus anthracis ». In : *Letters in applied microbiology* 47.1, p. 25-30.

MARTIN, Bernard et al. (1992). « A highly conserved repeated DNA element located in the chromosome of Streptococcus pneumoniae ». In : *Nucleic acids research* 20.13, p. 3479-3483.

MCINTYRE, Alexa BR et al. (2017). « Comprehensive benchmarking and ensemble approaches for metagenomic classifiers ». In : *Genome biology* 18.1, p. 1-19.

MENZEL, Peter, Kim Lee NG et Anders KROGH (2016). « Fast and sensitive taxonomic classification for metagenomics with Kaiju ». In : *Nature communications* 7.1, p. 11257.

MESELSON, Matthew et al. (1994). « The Sverdlovsk anthrax outbreak of 1979 ». In : *Science* 266.5188, p. 1202-1208.

MESNAGE, Stéphane, Michel HAUSTANT et Agnès FOUET (2001). « A general strategy for identification of S-layer genes in the Bacillus cereus group : molecular characterization of such a gene in Bacillus thuringiensis subsp. galleriae NRRL 4045 ». In : *Microbiology* 147.5, p. 1343-1351.

MESNAGE, Stéphane et al. (1997). « Molecular characterization of the Bacillus anthracis main S-layer component : evidence that it is the major cell-associated antigen ». In : *Molecular microbiology* 23.6, p. 1147-1155.

MESNAGE, Stephane et al. (1998). « The capsule and S-layer : two independent and yet compatible macromolecular structures in Bacillus anthracis ». In : *Journal of bacteriology* 180.1, p. 52-58.

MIKESELL, Perry et al. (1983). « Evidence for plasmid-mediated toxin production in Bacillus anthracis ». In : *Infection and immunity* 39.1, p. 371-376.

MILLER, J Michael et al. (1997). « Fulminating bacteremia and pneumonia due to Bacillus cereus ». In : *Journal of clinical microbiology* 35.2, p. 504-507.

MOCK, Michèle et Agnès FOUET (2001). « Anthrax ». In : *Annual Review of Microbiology* 55, p. 647-671.

MORAVEK, Maximilian et al. (2004). « Colony immunoblot assay for the detection of hemolysin BL enterotoxin producing Bacillus cereus ». In : *FEMS microbiology letters* 238.1, p. 107-113.

MORENS, David M (2003). « Characterizing a "new" disease : epizootic and epidemic anthrax, 1769–1780 ». In : *American journal of public health* 93.6, p. 886-893.

MUKARATI, Norman L et al. (2020). « The pattern of anthrax at the wildlife-livestock-human interface in Zimbabwe ». In : *PLoS Neglected Tropical Diseases* 14.10, e0008800.

NAKAMURA, LK (1998). « Bacillus pseudomycoides sp. nov. » In : *International Journal of Systematic and Evolutionary Microbiology* 48.3, p. 1031-1035.

NAKAMURA, LK et MA JACKSON (1995). « Clarification of the taxonomy of Bacillus mycoides ». In : *International Journal of Systematic and Evolutionary Microbiology* 45.1, p. 46-49.

NATIONAL INSTITUTE OF ALLERGY AND INFECTIOUS DISEASES (2018). *NIAID Emerging Infectious Diseases/Pathogens*. Consulté le 27 septembre 2023. URL : https://www.niaid.nih.gov/research/emerging-infectious-diseases-pathogens.

NORRIS, Michael H et al. (2023). « Genomic and Phylogenetic Analysis of Bacillus cereus Biovar anthracis Isolated from Archival Bone Samples Reveals Earlier Natural History of the Pathogen ». In : *Pathogens* 12.8, p. 1065.

OKINAKA, Richard, Talima PEARSON et Paul KEIM (2006). « Anthrax, but not Bacillus anthracis ? » In : *PLoS Pathogens* 2.11, e122.

ØKSTAD, Ole Andreas et Anne-Brit KOLSTØ (2010). « Genomics of Bacillus species ». In : *Genomics of foodborne bacterial pathogens*. Springer, p. 29-53.

OLSON, Kyle B (1999). « Aum Shinrikyo : once and future threat ? » In : *Emerging infectious diseases* 5.4, p. 513.

ONA, S. (2020a). *Next Generation Sequencing (Illumina)*. https://app.biorender.com/biorender-templates/figures/all/t-5ef134a11c72b100ad8d13ac-next-generation-sequencing-illumina.

— (2020b). *Sanger Sequencing*. https://app.biorender.com/biorender-templates/figures/all/t-5ef132f6c7bcd500b388a9c3-sanger-sequencing.

ONDOV, Brian D, Nicholas H BERGMAN et Adam M PHILLIPPY (2011). « Interactive metagenomic visualization in a Web browser ». In : *BMC bioinformatics* 12, p. 1-10.

ORR, Russell JS et al. (2023). « Reference genome assembly and annotation of two Bacillus cereus sensu lato strains from Etosha National Park, Namibia ». In : *Microbiology Resource Announcements* 12.11, e00544-23.

PALMA, Leopoldo et al. (2014). « Bacillus thuringiensis toxins : an overview of their biocidal activity ». In : *Toxins* 6.12, p. 3296-3325.

PARÉ, Ambroise (1568). *Traicté de la peste : de la petite verolle & rougeolle, avec une brefve description de la lepre..*

PATIN, Etienne et al. (2017). « Dispersals and genetic adaptation of Bantu-speaking populations in Africa and North America ». In : *Science* 356.6337, p. 543-546.

PATRA, Guy et al. (1998). « Molecular characterization of Bacillus strains involved in outbreaks of anthrax in France in 1997 ». In : *Journal of clinical microbiology* 36.11, p. 3412-3414.

PBS FRONTLINE (2011). *Paul Keim : We Were Surprised It Was the Ames Strain*. https://www.pbs.org/wgbh/frontline/article/paul-keim-we-were-surprised-it-was-the-ames-strain/. Accessed : 2023-11-02.

PEARSON, Talima et al. (2004). « Phylogenetic discovery bias in Bacillus anthracis using single-nucleotide polymorphisms from whole-genome sequencing ». In : *Proceedings of the National Academy of Sciences* 101.37, p. 13536-13541.

PEARSON, William R (1991). « Searching protein sequence libraries : comparison of the sensitivity and selectivity of the Smith-Waterman and FASTA algorithms ». In : *Genomics* 11.3, p. 635-650.

PENA-GONZALEZ, Angela et al. (2017). « Draft genome sequence of Bacillus cereus LA2007, a human-pathogenic isolate harboring anthrax-like plasmids ». In : *Genome Announcements* 5.16, p. 10-1128.

PENA-GONZALEZ, Angela et al. (2018). « Genomic characterization and copy number variation of Bacillus anthracis plasmids pXO1 and pXO2 in a historical collection of 412 strains ». In : *Msystems* 3.4, e00065-18.

PETIT III, Robert A et al. (2018). « Fine-scale differentiation between Bacillus anthracis and Bacillus cereus group signatures in metagenome shotgun data ». In : *PeerJ* 6, e5515.

PILO, Paola et al. (2011). « Bovine Bacillus anthracis in Cameroon ». In : *Applied and environmental microbiology* 77.16, p. 5818-5821.

PISARENKO, Sergey V et al. (2019). « Genotyping and phylogenetic location of one clinical isolate of Bacillus anthracis isolated from a human in Russia ». In : *BMC microbiology* 19.1, p. 1-9.

PISARENKO, Sergey V et al. (2021). « Molecular genotyping of 15 B. anthracis strains isolated in Eastern Siberia and Far East ». In : *Molecular Phylogenetics and Evolution* 159, p. 107116.

PRIEST, Fergus G et al. (2004). « Population structure and evolution of the Bacillus cereus group ». In : *Journal of bacteriology* 186.23, p. 7959-7970.

PRINCE, Alice S et al. (2003). « The host response to anthrax lethal toxin : unexpected observations ». In : *The Journal of clinical investigation* 112.5, p. 656-658.

PURCELL, B.K., P.L. WORSHAM et A.M. FRIEDLANDER (2006). « Anthrax ». In : *Medical Aspects of Biological Warfare*. TMM Publications, p. 69-90.

RAPPÉ, Michael S et Stephen J GIOVANNONI (2003). « The uncultured microbial majority ». In : *Annual Reviews in Microbiology* 57.1, p. 369-394.

Rapport d'information déposé en application de l'article 145 du règlement, par la commission de la défense nationale et des forces armées, en conclusion des travaux d'une mission d'information sur la défense NRBC (fév. 2022). 5112. Commission de la Défense Nationale et des Forces Armées.

RASKO, David A et al. (2011). « Bacillus anthracis comparative genome analysis in support of the Amerithrax investigation ». In : *Proceedings of the National Academy of Sciences* 108.12, p. 5027-5032.

READ, Timothy D et al. (2003). « The genome sequence of Bacillus anthracis Ames and comparison to closely related bacteria ». In : *Nature* 423.6935, p. 81-86.

REDDY, A, L BATTISTI et CB THORNE (1987). « Identification of self-transmissible plasmids in four Bacillus thuringiensis subspecies ». In : *Journal of bacteriology* 169.11, p. 5263-5270.

« Report of the Secretary-General on the status of the implementation of the Special Commission's plan for the ongoing monitoring and verification of Iraq's compliance with relevant parts of section C of Security Council resolution 687 (1991). » (1994). In : PUB, 38 p. URL : http://sanctionsplatform.ohchr.org/record/8688.

RICHARDSON, Lorna et al. (2023). « MGnify : the microbiome sequence data analysis resource in 2023 ». In : *Nucleic Acids Research* 51.D1, p. D753-D759.

ROBERTS, Michael et al. (2004). « Reducing storage requirements for biological sequence comparison ». In : *Bioinformatics* 20.18, p. 3363-3369.

ROBERTSON, Donald L et Stephen H LEPPLA (1986). « Molecular cloning and expression in Escherichia coli of the lethal factor gene of Bacillus anthracis ». In : *Gene* 44.1, p. 71-78.

ROLAND, Alex (2005). « Greek Fire, Poison Arrows, and Scorpion Bombs : Biological and Chemical Warfare in the Ancient World ». In : *Technology and Culture* 46.4, p. 878-879.

ROMERO-ALVAREZ, Daniel et al. (2020). « Potential distributions of Bacillus anthracis and Bacillus cereus biovar anthracis causing anthrax in Africa ». In : *PLoS neglected tropical diseases* 14.3, e0008131.

SAHL, Jason W et al. (2015). « Phylogenetically typing bacterial strains from partial SNP genotypes observed from direct sequencing of clinical specimen metagenomic data ». In : *Genome medicine* 7, p. 1-13.

SAHL, Jason W et al. (2016). « A Bacillus anthracis genome sequence from the Sverdlovsk 1979 autopsy specimens ». In : *MBio* 7.5, p. 10-1128.

SAILE, Elke et Theresa M KOEHLER (2006). « Bacillus anthracis multiplication, persistence, and genetic exchange in the rhizosphere of grass plants ». In : *Applied and environmental microbiology* 72.5, p. 3168-3174.

SANTOS, FDS et al. (2018). « Bacillus toyonensis improves immune response in the mice vaccinated with recombinant antigen of bovine herpesvirus type 5 ». In : *Beneficial Microbes* 9.1, p. 133-142.

SCARFF, Jennifer M et al. (2018). « Expression and contribution to virulence of each polysaccharide capsule of Bacillus cereus strain G9241 ». In : *PLoS One* 13.8, e0202701.

SCHWARTZ, Maxime (2009). « Dr. Jekyll and Mr. Hyde : a short history of anthrax ». In : *Molecular aspects of medicine* 30.6, p. 347-355.

SEELOS, Christian (1999). « Lessons from Iraq on bioweapons ». In : *Nature* 398.6724, p. 187-188.

SEGATA, Nicola et al. (2012). « Metagenomic microbial community profiling using unique clade-specific marker genes ». In : *Nature methods* 9.8, p. 811-814.

SEJVAR, James J, Fred C TENOVER et David S STEPHENS (2005). « Management of anthrax meningitis ». In : *The Lancet infectious diseases* 5.5, p. 287-295.

SHEVTSOV, Alexandr et al. (2021). « Bacillus anthracis phylogeography : new clues from Kazakhstan, Central Asia ». In : *Frontiers in Microbiology* 12, p. 778225.

SICHTIG, Heike et al. (2019). « FDA-ARGOS is a database with public quality-controlled reference genomes for diagnostic use and regulatory science ». In : *Nature communications* 10.1, p. 3313.

SIMON, H Ye et al. (2019). « Benchmarking metagenomics tools for taxonomic classification ». In : *Cell* 178.4, p. 779-794.

SIMONSON, Tatum S et al. (2009). « Bacillus anthracis in China and its relationship to worldwide lineages ». In : *BMC microbiology* 9, p. 1-11.

SLAMTI, Leyla et al. (2004). « Distinct mutations in PlcR explain why some strains of the Bacillus cereus group are nonhemolytic ». In : *Journal of bacteriology* 186.11, p. 3531-3538.

SMITH, KL et al. (1999). « Meso-scale ecology of anthrax in southern Africa : a pilot study of diversity and clustering ». In : *Journal of Applied Microbiology* 87.2, p. 204-207.

SMITH, KL et al. (2000). « Bacillus anthracis diversity in kruger national park ». In : *Journal of clinical microbiology* 38.10, p. 3780-3784.

SMITH, Martin D et al. (2015). « Less is more : an adaptive branch-site random effects model for efficient detection of episodic diversifying selection ». In : *Molecular biology and evolution* 32.5, p. 1342-1353.

SOBRAL, Daniel (2012). « De l'usage du polymorphisme de répétitions en tandem pour l'étude des populations bactériennes : mise au point et validation d'un système de génotypage automatisé utilisant la technique de MLVA ». Thèse de doct. Université Paris Sud-Paris XI.

SOHN, Michael B et al. (2014). « Accurate genome relative abundance estimation for closely related species in a metagenomic sample ». In : *BMC bioinformatics* 15.1, p. 1-13.

STENFORS ARNESEN, Lotte P, Annette FAGERLUND et Per Einar GRANUM (2008). « From soil to gut : Bacillus cereus and its food poisoning toxins ». In : *FEMS microbiology reviews* 32.4, p. 579-606.

STERN, MJ, E PROSSNITZ et G Ferro-Luzzi AMES (1988). « Role of the intercistronic region in post-transcriptional control of gene expression in the histidine transport operon of Salmonella typhimurium : involvement of REP sequences ». In : *Molecular microbiology* 2.1, p. 141-152.

STERNBACH, George (2003). « The history of anthrax ». In : *The Journal of emergency medicine* 24.4, p. 463-467.

STOPLER, T, V CAMUESCU et M VOICULESCU (1965). « Bronchopneumonia with lethal evolution determined by a microorganism of the genus Bacillus (B. cereus) ». In : *Rumanian medical review* 19, p. 7-9.

SUE, David et al. (2006). « Capsule production in Bacillus cereus strains associated with severe pneumonia ». In : *Journal of clinical microbiology* 44.9, p. 3426-3428.

SWEENEY, Daniel A et al. (2011). « Anthrax infection ». In : *American journal of respiratory and critical care medicine* 184.12, p. 1333-1341.

Systématique en microbiologie. Notion d'espèce. Classification universelle mixte consensuelle et phylogénétique en bactériologie (2022). http://www.perrin33.com/microbiologie/systematique_1.php. Accessed : 2024-01-24.

TAKAHASHI, H (2004). « Bacillus anthracis incident, Kameido, Tokyo, 1993 (vol 10, pg 119, 1993) ». In : *EMERGING INFECTIOUS DISEASES* 10.2, p. 385-385.

TAMBORRINI, Marco et al. (2010). « Anthrax spore detection by a luminex assay based on monoclonal antibodies that recognize anthrose-containing oligosaccharides ». In : *Clinical and Vaccine Immunology* 17.9, p. 1446-1451.

TAN, AP et al. (2011). « Isolation and identification of Bacillus cereus from Trionyx sinensis ». In : *Guangdong Agric. Sci* 20, p. 115-119.

TESSIER, Emilie (2022). « Endocytose du facteur oedémateux de Bacillus anthracis et des Bacillus cereus anthracis-like ». Thèse de doct. Université Paris-Saclay.

THAPA, Nirmal K et al. (2014). « Investigation and control of anthrax outbreak at the human–animal interface, Bhutan, 2010 ». In : *Emerging infectious diseases* 20.9, p. 1524.

THE WHITE HOUSE (2009). *National Strategy for Countering BioThreats*. https://obamawhitehouse.archives.gov/sites/default/files/National_Strategy_for_Countering_BioThreats.pdf. Accessed : 2024-02-11.

THIERRY, Simon et al. (2014). « Genotyping of French Bacillus anthracis strains based on 31-loci multi locus VNTR analysis : epidemiology, marker evaluation, and update of the internet genotype database ». In : *PLoS One* 9.6, e95131.

THORNE, Curtis B, Dorothy M MOLNAR et Richard E STRANGE (1960). « Production of toxin in vitro by Bacillus anthracis and its separation into two components ». In : *Journal of Bacteriology* 79.3, p. 450-455.

TIMOFEEV, Vitalii et al. (2019). « Insights from Bacillus anthracis strains isolated from permafrost in the tundra zone of Russia ». In : *PloS one* 14.5, e0209140.

TIMOFEEV, Vitalii et al. (2023). « New Research on the Bacillus anthracis Genetic Diversity in Siberia ». In : *Pathogens* 12.10, p. 1257.

TIPPETTS, M TODD et DONALD L ROBERTSON (1988). « Molecular cloning and expression of the Bacillus anthracis edema factor toxin gene : a calmodulin-dependent adenylate cyclase ». In : *Journal of bacteriology* 170.5, p. 2263-2266.

TOMASO, Herbert et al. (2006). « Growth characteristics of Bacillus anthracis compared to other Bacillus spp. on the selective nutrient media Anthrax Blood Agar® and Cereus Ident Agar® ». In : *Systematic and applied microbiology* 29.1, p. 24-28.

TOVO, Anna et al. (2020). « Taxonomic classification method for metagenomics based on core protein families with Core-Kaiju ». In : *Nucleic acids research* 48.16, e93-e93.

TSAI, Jia-Ming, Hsin-Wei KUO et Winton CHENG (2023). « Retrospective Screening of Anthrax-like Disease Induced by Bacillus tropicus str. JMT from Chinese Soft-Shell Turtles in Taiwan ». In : *Pathogens* 12.5, p. 693.

TURNBAUGH, Peter J et al. (2007). « The human microbiome project ». In : *Nature* 449.7164, p. 804-810.

TURNBULL, PCB (1996). « Guidance on environments known to be or suspected of being contaminated with anthrax spore ». In : *Land contamination and reclamation* 4, p. 37-45.

— (1999). « Definitive identification of Bacillus anthracis—a review ». In : *Journal of applied microbiology* 87.2, p. 237-240.

TURNBULL, PCB et al. (1992). « Serology and anthrax in humans, livestock and Etosha National Park wildlife ». In : *Epidemiology & Infection* 108.2, p. 299-313.

TURNBULL, Peter C. B. et Sean V. SHADOMY (2010). « Anthrax from 5000 BC to AD 2010 ». In : *Bacillus anthracis and anthrax*. Sous la dir. de Nicholas H. BERGMAN. Wiley-Blackwell, p. 3-9. DOI : 10.1002/9780470891193.ch1. URL : https://doi.org/10.1002/9780470891193.ch1.

UCHIDA, I et al. (1993). « Identification of a novel gene, dep, associated with depolymerization of the capsular polymer in Bacillus anthracis ». In : *Molecular microbiology* 9.3, p. 487-496.

UCHIDA, Ikuo et al. (1985). « Association of the encapsulation of Bacillus anthracis with a 60 megadalton plasmid ». In : *Microbiology* 131.2, p. 363-367.

VAN ERT, Matthew N et al. (2007). « Global genetic population structure of Bacillus anthracis ». In : *PloS one* 2.5, e461.

VAN ROSSUM, Thea et al. (2020). « Diversity within species : interpreting strains in microbiomes ». In : *Nature Reviews Microbiology* 18.9, p. 491-506.

VENKATESWARAN, Kasthuri et al. (2017). « Draft genome sequences from a novel clade of Bacillus cereus sensu lato strains, isolated from the International Space Station ». In : *Genome announcements* 5.32, p. 10-1128.

VENTER, J Craig et al. (2001). « The sequence of the human genome ». In : *science* 291.5507, p. 1304-1351.

VERGNAUD, Gilles (2020). « Bacillus anthracis Evolution : Taking Advantage of the Topology of the Phylogenetic Tree and Human History to Propose Dating Points. » In : *Erciyes Medical Journal/Erciyes Tip Dergisi* 42.4.

VERGNAUD, Gilles et al. (2016). « Comparison of French and worldwide Bacillus anthracis strains favors a recent, post-Columbian origin of the predominant North-American clade ». In : *PLoS One* 11.2, e0146216.

VIRGILE (s. d.). *Les Géorgiques de Virgile*.

VODKIN, Michael H et Stephen H LEPPLA (1983). « Cloning of the protective antigen gene of Bacillus anthracis ». In : *Cell* 34.2, p. 693-697.

VOGLER, Amy J et al. (2002). « Molecular analysis of rifampin resistance in Bacillus anthracis and Bacillus cereus ». In : *Antimicrobial agents and chemotherapy* 46.2, p. 511-513.

WANG, Dian-Bing et al. (2021). « Biosensors for the Detection of Bacillus anthracis ». In : *Accounts of chemical research* 54.24, p. 4451-4461.

WILSON, James M et al. (2016). « Reanalysis of the anthrax epidemic in Rhodesia, 1978–1984 ». In : *PeerJ* 4, e2686.

WILSON, Melissa K et al. (2011). « Bacillus cereus G9241 makes anthrax toxin and capsule like highly virulent B. anthracis Ames but behaves like attenuated toxigenic nonencapsulated B. anthracis Sterne in rabbits and mice ». In : *Infection and immunity* 79.8, p. 3012-3019.

WOESE, Carl R et George E FOX (1977). « Phylogenetic structure of the prokaryotic domain : the primary kingdoms ». In : *Proceedings of the National Academy of Sciences* 74.11, p. 5088-5090.

WOOD, Derrick E, Jennifer LU et Ben LANGMEAD (2019). « Improved metagenomic analysis with Kraken 2 ». In : *Genome biology* 20.1, p. 257.

WOOD, Derrick E et Steven L SALZBERG (2014). « Kraken : ultrafast metagenomic sequence classification using exact alignments ». In : *Genome biology* 15.3, p. 1-12.

WORLD HEALTH ORGANIZATION, International Office of Epizootics (2008). *Anthrax in humans and animals*. World Health Organization.

WRIGHT, Angela M et al. (2011). « Rapidly progressive, fatal, inhalation anthrax-like infection in a human : case report, pathogen genome sequencing, pathology, and coordinated response ». In : *Archives of pathology & laboratory medicine* 135.11, p. 1447-1459.

XIA, Li C et al. (2011). « Accurate genome relative abundance estimation based on shotgun metagenomic reads ». In : *PloS one* 6.12, e27992.

YUAN, Xuemei et al. (2020). « Complete genome sequence of novel isolate SYJ15 of Bacillus cereus group, a highly lethal pathogen isolated from Chinese soft shell turtle (Pelodiscus Sinensis) ». In : *Archives of microbiology* 202, p. 85-92.

ZAHAVY, E et al. (2003). « Detection of frequency resonance energy transfer pair on double-labeled microsphere and Bacillus anthracis spores by flow cytometry ». In : *Applied and Environmental Microbiology* 69.4, p. 2330-2339.

ZASADA, Aleksandra A (2018). « Injectional anthrax in human : A new face of the old disease ». In : *Adv Clin Exp Med* 27.4, p. 553-558.

— (2020). « Detection and identification of Bacillus anthracis : From conventional to molecular microbiology methods ». In : *Microorganisms* 8.1, p. 125.

ZHANG, Yong et al. (2022). « Enhancing the Phytoremediation of Heavy Metals by Combining Hyperaccumulator and Heavy Metal-Resistant Plant Growth-Promoting Bacteria ». In : *Frontiers in Plant Science* 13, p. 912350.

ZHENG, Jinshui et al. (2013). « Evolution and dynamics of megaplasmids with genome sizes larger than 100 kb in the Bacillus cereus group ». In : *BMC evolutionary biology* 13, p. 1-11.

ZHOU, Zhemin et al. (2018). « GrapeTree : visualization of core genomic relationships among 100,000 bacterial pathogens ». In : *Genome research* 28.9, p. 1395-1404.

ZILINSKAS, Raymond A (1997). « Iraq's biological weapons : the past as future ? » In : *Jama* 278.5, p. 418-424.

ZIMMERMANN, Fee et al. (2017). « Low antibody prevalence against Bacillus cereus biovar anthracis in Taï National Park, Côte d'Ivoire, indicates high rate of lethal infections in wildlife ». In : *PLOS Neglected Tropical Diseases* 11.9, e0005960.

ZWICK, Michael E et al. (2012). « Genomic characterization of the Bacillus cereus sensu lato species : backdrop to the evolution of Bacillus anthracis ». In : *Genome research* 22.8, p. 1512-1524.