



Développement d'outils de typage moléculaire de haute résolution pour la détection et la différenciation de *Bacillus anthracis*

Guillaume Girault

► To cite this version:

Guillaume Girault. Développement d'outils de typage moléculaire de haute résolution pour la détection et la différenciation de *Bacillus anthracis*. Biochimie, Biologie Moléculaire. AgroParisTech, 2015. Français. NNT : 2015AGPT0007 . tel-03081379

HAL Id: tel-03081379

<https://pastel.hal.science/tel-03081379>

Submitted on 18 Dec 2020

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

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



Doctorat ParisTech

THÈSE

pour obtenir le grade de docteur délivré par

L'Institut des Sciences et Industries du Vivant et de l'Environnement (AgroParisTech)

Spécialité : Sciences du Vivant

présentée et soutenue publiquement par

Guillaume GIRAULT

le 21 janvier 2015

Développement d'outils de typage moléculaire de haute résolution pour la détection et la différenciation de *Bacillus anthracis*

Directrice de thèse : **Dr Sylviane DERZELLE**

Jury

M. Pierre GOOSSENS, Directeur de Laboratoire, Institut Pasteur
M. Christophe NGUYEN-THE, Directeur de recherche, INRA
M. Marc ARTOIS, Professeur, VetAgro sup
M. Joachim FREY, Professeur, Institute for Veterinary Bacteriology, Université de Berne
Mme Emmanuelle GUILLOT-COMBE, Responsable Biologie et Biotechnologies, MRIS, DGA
Mme Nadia HADDAD, Professeur, UMR Bipar, ENVA
Mme Nalini RAMA RAO, Chercheur, MICALIS, UMR 1319, INRA
Mme Sylviane DERZELLE, Chargée de projet, UZB, ANSES

Rapporteur
Rapporteur
Examinateur
Examinateur
Examinateur
Examinateur
Examinateur
Directrice

REMERCIEMENTS

Ce travail a été réalisé au sein de l'Unité des Zoonoses Bactériennes de l'ANSES, sous la direction du Docteur Sylviane Derzelle, et co-financé par une bourse de l'ANSES et de la Direction Générale de l'Armement (DGA).

Je souhaite remercier Sylviane Derzelle pour son dévouement à mon apprentissage, ses conseils et sa disponibilité durant ces 3 années de thèse.

Je remercie le directeur du Laboratoire de Santé Animale, Pascal Boireau, de m'avoir accueilli au sein de son établissement et pour son implication dans la fin de ma thèse.

Je remercie particulièrement Claire Ponsart de m'avoir accueilli au sein de l'Unité des Zoonoses Bactériennes ainsi que pour son aide précieuse dans la fin de ma thèse.

Je souhaite vivement remercier Pierre Goossens et Christophe Nguyen-Thé pour avoir accepté de juger ce travail en tant que rapporteurs. Mes remerciements vont également à Nadia Haddad, Nalini Rama Rao, Emmanuelle Guillot-Combe, Joachim Frey et Marc Artois pour m'avoir fait l'honneur de participer à mon jury de thèse en tant qu'examinateurs.

Je remercie particulièrement François Thibault pour sa gentillesse, son soutien lors des congrès ainsi que pour ses précieux conseils tout au long de mon doctorat. Je remercie Yannick Morel et Guilhem Larigauderie pour les conseils prodigués lors de mes comités de thèse.

Je tiens à remercier Gilles Vergnaud de m'avoir accueilli au sein de son équipe à l'IGM d'Orsay pour me former à l'analyse bioinformatique des données de séquençage.

Un immense merci à Yann Blouin qui m'a formé à l'utilisation de BioNumerics, à la programmation en langage Python et pour tout ce qu'il m'a apporté : tu es mon maître Jedi !

Je remercie également Catherine Eng et Philippe Le Flèche de la DGA pour leur aide lors de mon projet ainsi que pour leur bonne humeur lors de mes visites au centre MNRBC de Vert-le-Petit.

Je remercie toute l'équipe de la plateforme de séquençage IMAGIF pour leur apport dans mon travail et leur disponibilité.

Tous mes remerciements à Simon Thierry, Fabien Vorimore, Nicolas Versillé, Aleksandar Zocovic, Yannick Corde, Ali Cherfa, Antoine Drapeau, Krystel De Cruz, Amandine Hauer, Tabatha Bulach, Virginie Hulin, Gilles Le Carrou, Vincent Sester, Julien Portier et Célia Rodriguez Nevado pour leur bonne humeur et tous les moments passés ensemble. J'ai passé trois années merveilleuses à vos côtés : merci pour tout du fond du cœur.

Un merci particulier à Virginie Mick pour tout ce qu'elle a fait pour moi, que ce soit en congrès, au labo ou pour la fin de ma thèse : MERCI !

Je remercie Nora Madani, Christiane Mendy, Emeline Cherchame et Cécile Colaneri de l'équipe Charbon : merci de votre accueil, de vos conseils et de votre bonne humeur.

Tous mes remerciements à Karine Laroucau et Maryne Jay, toujours là pour moi.

Merci à Rachid Aaziz, Claire Bertin, Madeleine Carle, Cécile Marzet, Rebiha Mesbah, Gaby Vecchio, Laura Boschioli, Sylvie Henault, Claudine Karoui et Lorraine Michelet pour votre gentillesse et votre disponibilité.

Merci à l'équipe de foot du COSMOS Rocquencourt : vous me permettez de me défouler tous les lundis, et votre bonne humeur me donne de l'énergie pour toute la semaine.

Tous mes remerciements et bien plus à mes amis Max, Adrien, Arnaud, Benji, Maou, Chris (et ses enfants), Marion, Charlène, Soline, Julie, Axel, Christelle et Linda. Merci pour tous ces moments passés ensemble (et ceux à venir). Merci également à Michael, Elie, LeBron et Ezio de m'avoir accompagné durant ces 3 années.

Un immense merci à mes parents, pour tout ce que vous avez fait pour moi depuis le début. J'en suis aujourd'hui là grâce à vous. Vous m'avez toujours soutenu et je ne pourrai jamais assez vous remercier. Merci à mon frère, tu es le meilleur et je suis fier de toi !!

Un grand merci à toute ma famille, mes grands-mères, mes grands-pères (j'espère que vous êtes fiers de moi), mes tantes, mes oncles, mon cousin et mes cousines.

A mon petit Phénix qui me réveille tous les matins et me donne le sourire dès que je le vois. Mon petit chat, tu es mon plus fidèle compagnon.

Enfin, je remercie tout particulièrement Laetitia qui partage ma vie depuis 9 ans. Tu es mon soutien le plus solide, tu es toujours de bons conseils, et je suis l'homme le plus heureux à tes côtés. Puissent les années à venir nous apporter autant de bonheur (et bien plus encore).

RESUME

Bacillus anthracis est une bactérie pathogène de répartition mondiale qui est responsable d'une zoonose, le charbon bactérien. Appartenant au genre *Bacillus*, elle possède la capacité de sporuler : cette particularité lui permet de rester en dormance dans les sols et de résister à de nombreux stress (UV, chaleur...). Infectant principalement les mammifères, cette bactérie est responsable de foyers animaux et humains recensés annuellement à travers le monde. Certaines régions sont endémiques, d'autres, comme la France, ne connaissent que quelques cas sporadiques par an. Malgré l'incidence décroissante de la maladie à travers le monde, *B. anthracis* reste néanmoins d'un intérêt majeur pour de nombreux pays étant donné son usage potentiel en tant qu'arme biologique. L'étude de cette bactérie a une finalité duale, d'une part pour la mortalité qu'elle cause au sein des cheptels bovins et de la faune sauvage, et d'autre part, pour son potentiel en tant qu'arme de destruction massive.

Cette espèce bactérienne est considérée comme très clonale : toutes les souches sont extrêmement proches d'un point de vue génétique. Afin de pouvoir identifier précisément les souches et remonter à la source d'infection, diverses méthodes de diagnostic et de typage moléculaire sont disponibles. Cependant, toutes ne possèdent pas le pouvoir de résolution souhaité, la robustesse ou la simplicité d'utilisation.

Mon travail a consisté à étudier la diversité des souches de *B. anthracis* disponibles en Europe. Pour cela, un séquençage du génome complet a été réalisé sur un panel d'environ 250 souches (France, Europe) afin de disposer d'une très grande diversité. Une analyse bioinformatique comparative a permis de reconstruire les génomes des isolats européens et l'identification de polymorphismes de séquence spécifiques à ces souches (ici, les Single Nucleotide Polymorphisms = SNP). Ces marqueurs ont conduit à établir avec précision la phylogénie de 292 souches de *B. anthracis* au niveau mondial. Plusieurs hypothèses concernant les origines et l'évolution de ce pathogène ont été proposées. Une proposition d'une nouvelle sous-lignée a également été faite. Enfin, un panel de soixante nouveaux marqueurs de type SNP a été identifié et validé pour génotyper les groupes et lignées phylogénétiquement apparentés présents en Europe.

Deux méthodes de biologie moléculaire basées sur l'amplification PCR de ces marqueurs SNP ont été mises au point. La première permet une analyse rapide à faible coût (PCR HRM) alors que la seconde permet un multiplexage des tests (Luminex).

Mon travail a permis un accroissement significatif de la connaissance de ce pathogène qu'est *B. anthracis*. Les outils de typage mis en place permettront à la France de disposer d'un outil de diagnostic performant, assurant la traçabilité et l'identification rapide des souches impliquées lors de foyers de fièvre charbonneuse ou lors de cas importés.

ABSTRACT

Bacillus anthracis is a pathogenic bacterium with a worldwide repartition. It is the causative agent of a zoonosis named Anthrax. Belonging to the *Bacillus* genus, *B. anthracis* has the ability to sporulate: this particularity allows the bacterium to stay as quiescent spores into the soils and to resist against different kind of stresses (UV, heat treatment...). Mammals are principally infected and human or animal outbreaks are reported annually in the world. Several regions are endemic while some others, like France, report more sporadic cases. Even if Anthrax incidence is in constant decrease all over the world, *B. anthracis* is still a pathogen of interest for many countries because of its potential use as a biological weapon. The study of this bacterium has a two-tier purpose: first, it is the cause of a relative mortality in livestock and wildlife, and second it is potentially used as a weapon of mass destruction.

This bacterial species is considered highly monomorphic and all strains are extremely closed genetically. In order to precisely identify strains during outbreaks or bioterrorism attempt and track the source of infection, several diagnosis and typing methods are available. However, not all of these methods have the discrimination power, the robustness or the ease of use that is required in the laboratory.

During this work, I have studied the diversity of European isolates of *B. anthracis*. A whole genome sequencing approach for approximately 250 strains has been done with a great diversity into strains (France, Europe). A comparative bioinformatic analysis allowed genomes reconstruction and polymorphisms identification among European isolates (Single Nucleotide Polymorphism = SNP). These markers leaded to establish a precise phylogeny among 292 *B. anthracis* isolates at a world level. Several hypotheses concerning the origins and the evolution of this pathogen have been proposed. A new sublineage has been potentially discovered. A panel of sixty SNPs has been identified and confirmed to genotype the major groups and lineages phylogenetically related in Europe.

Two molecular typing approaches based on PCR amplification have been developed. Using the identified SNP, they can be used to discriminate strains between each others. The first one is a quick and low cost approach (PCR HRM) whereas the second can be used to multiplex a lot of analyses (Luminex).

My work allowed a significative increase into *B. anthracis* knowledge. The typing tools developed will allow traceability and quick identification of the strains involved in an Anthrax outbreak or in a suspected bioterrorist attempt in France.

TABLE DES MATIERES

RESUME	1
ABSTRACT	3
AVANT-PROPOS	9
LISTE DES ABREVIATIONS	11
ANALYSE BIBLIOGRAPHIQUE	15
Chapitre 1 : <i>Bacillus anthracis</i> et la fièvre charbonneuse	15
1. <i>Bacillus anthracis</i>	17
1.1. La Fièvre charbonneuse à travers les époques	17
1.2. Taxonomie	20
1.3. Caractéristiques bactériologiques	22
1.3.1. La bactérie	22
1.3.2. La spore	24
1.4. Structure du génome	26
1.5. Facteurs de virulence	29
1.5.1. Capsule polypeptidique	29
1.5.2. Toxines protéiques	29
1.6. Cycle d'infection	32
2. La fièvre charbonneuse	32
2.1. Situation de la maladie au niveau mondial	32
2.2. Epidémiologie	34
2.3. Signes cliniques	35
2.3.1. Formes cliniques selon l'espèce animale	35
2.3.2. Chez l'Homme	39
2.4. Diagnostic de la maladie	43
2.4.1. Chez l'animal	43
2.4.2. Chez l'Homme	43
2.5. Traitements	44
2.6. Prévention	45
3. Arme biologique potentielle	46

Chapitre 2 : Détection et typage de <i>Bacillus anthracis</i>	51
1. Méthodes de détection et de typage moléculaire de <i>B. anthracis</i>	53
1.1. Détection	53
1.1.1. Phages appliqués au diagnostic	53
1.1.2. Réaction en Chaîne par Polymérase (PCR)	53
2. Typage moléculaire	54
2.1. Polymorphisme de Longueurs des Fragments Amplifiés (AFLP)	54
2.2. Multiple Locus VNTR Analysis (MLVA)	56
2.3. Single Nucleotide Repeats (SNR)	59
2.4. Single Nucleotide Polymorphisms (SNP)	59
2.5. Essais progressifs résolutifs hiérarchiques basés sur les acides nucléiques (PHRANA)	62
3. Phylogénie des souches de <i>B. anthracis</i> à travers le monde	63
Chapitre 3 : Analyse du génome	67
1. Séquençage du génome complet	69
1.1. Méthodes NGS	69
1.1.1. Séquençage 454®	69
1.1.2. Séquençage Illumina®	71
1.1.3. Séquençage SOLID®	71
2. Applications des NGS	74
3. Analyse bioinformatique	74
3.1. Assemblage de génome <i>de novo</i>	74
3.2. Assemblage de génome par homologie	75
CONTEXTE ET OBJECTIFS	77
ETUDE EXPERIMENTALE	81
Chapitre 1 : Etude de la diversité de <i>Bacillus anthracis</i> en France	81
1. Utilisation des SNP sur génome complet pour étudier la diversité des souches françaises de <i>B. anthracis</i>	83
2. Assemblage <i>de novo</i> de souches de <i>B. anthracis</i> isolées en France	97
Chapitre 2 : Etude de la diversité de <i>Bacillus anthracis</i> en Europe	101
1. Génotypage des souches italiennes de <i>B. anthracis</i>	103
2. Cas importés dans la filière du textile	123

3. Génomique des souches européennes de <i>B. anthracis</i>	145
Chapitre 3 : Développement de méthodes de typage de <i>Bacillus anthracis</i>	169
1. Application du séquençage à haut débit: identification de marqueurs SNP informatifs permettant le génomique de <i>B. anthracis</i>	171
2. Génomique de <i>B. anthracis</i> : utilisation des SNP via une approche par multiplexage	183
DISCUSSION GENERALE	195
CONCLUSIONS ET PERSPECTIVES	217
REFERENCES	221
ANNEXES.....	235

AVANT-PROPOS

La fièvre charbonneuse est une maladie zoonotique aiguë potentiellement mortelle causée par une bactérie sporulée, *Bacillus anthracis*. Cette maladie peut atteindre tous les mammifères, y compris les humains, ainsi que certaines espèces d'oiseaux. *B. anthracis* constitue un germe redoutable du fait du caractère extrêmement résistant de ses spores et de son pouvoir pathogène. Les spores peuvent persister dans un sol infecté durant de longues périodes assurant de ce fait la persistance de l'espèce. D'autre part, celles-ci sont hautement résistantes à de nombreux stress tels que la température, les rayons ultraviolets ou la dessiccation, même en cas d'exposition prolongée.

En France, la fièvre charbonneuse apparaît de façon sporadique dans certains départements français ayant connu par le passé des épisodes charbonneux, essentiellement dans le Massif central, les Alpes, la Bourgogne et la Franche-Comté. Ces épisodes sont généralement de faible importance (un ou deux foyers, et un ou deux animaux morts par foyer) mais il n'est pas exceptionnel que des épisodes plus conséquents se produisent, comme ce fut le cas en 2008 dans le département du Doubs et en 2009 dans les départements de Haute Savoie et de l'Isère.

Ces caractéristiques ainsi que la relative facilité d'accès à ce germe de répartition mondiale font de *B. anthracis* l'espèce bactérienne la plus susceptible d'être utilisée pour une attaque bioterroriste. En 2001, des envois délibérés de lettres contenant des spores de *B. anthracis* ont semé la panique aux Etats-Unis d'Amérique, démontrant le potentiel d'utilisation en tant qu'arme biologique aux yeux du monde. Cet épisode est à l'origine d'un regain d'intérêt majeur pour cet organisme.

Le typage moléculaire permet d'identifier rapidement l'origine probable d'une contamination et fournit de précieuses informations pour orienter les investigations. Les besoins d'un tel système, à la fois robuste et précis, sont similaires, que la thématique soit à visée militaire ou civile. Dans le cas d'un foyer de charbon naturel, le typage des souches peut fournir des informations intéressantes quant à l'introduction et la dispersion de différents profils dans un environnement naturel et peut permettre la comparaison entre les souches isolées chez l'animal et chez l'Homme. Dans le cas d'un acte terroriste, le typage permet de déterminer l'origine de la souche utilisée et, éventuellement, de remonter aux personnes l'ayant perpétré.

Un des exemples les plus représentatifs observés ces dernières années concerne les héroïnomanes. Au début des années 2000 et durant une dizaine d'années, plusieurs cas inhabituels de fièvre charbonneuse ont été diagnostiqués au sein de cette population en Europe, et une nouvelle forme de la maladie a été décrite, le charbon d'injection. Afin d'identifier la source de contamination des spores de l'héroïne, une enquête internationale fut entreprise. Au final, l'origine présumée de contamination fut identifiée grâce à l'utilisation de marqueurs spécifiques et via la combinaison de différentes méthodes (séquençage, analyse de SNP). Ces événements récents ont mis en évidence l'importance d'améliorer les moyens d'investigation des foyers de charbon, qu'ils soient naturels ou malveillants pour pouvoir caractériser finement les souches, retracer l'origine et l'historique de la contamination, voire relier entre eux les cas humains et animaux qui pourraient se déclarer dans une même région ou un même laps de temps.

LISTE DES ABREVIATIONS

ADN: Acide Désoxyribonucléique

AFLP: Amplified Fragment Length Polymorphism

AMP: Adénosine monophosphate

ANSES : Agence Nationale de Sécurité Sanitaire, Environnement et Travail

ARN: Acide Ribonucléique

ARNr: ARN ribosomal

ATP: Adénosine triphosphate

AVA: Anthrax vaccine adsorbed

canSNP: Canonical SNP

CGH: Hybridation génomique comparative

CIP : Collection Institut Pasteur

FC: Fièvre charbonneuse

HRM: High Resolution Melting

IRBA: Institut de Recherche Biomédicale des Armées

Kb: Kilo bases

Mb: Mega bases

MLST: Multilocus Sequence Typing

MLVA: Multiple Locus VNTR Analysis

NGS: Next Generation Sequencing

OIE: Organisation mondiale de la santé animale

ORF: Open Reading Frame

Pb: Paire de bases

PCR: Polymerase Chain Reaction

SEM: Scanning Electron Micrography

SNP: Single Nucleotide Polymorphism

SNR: Single Nucleotide Repeat

TEA: Trans-Eurasienne

URSS: Union des Républiques Socialistes Soviétiques

USA: United States of America – Etats-Unis d’Amérique

VNTR: Variable Number Tandem Repeat

WGS: Whole Genome Shotgun

« Tout obstacle renforce la détermination. Celui qui s'est fixé un but n'en change pas. »

(Léonard de Vinci)

ANALYSE BIBLIOGRAPHIQUE

Chapitre 1 :

***Bacillus anthracis* et la fièvre charbonneuse**

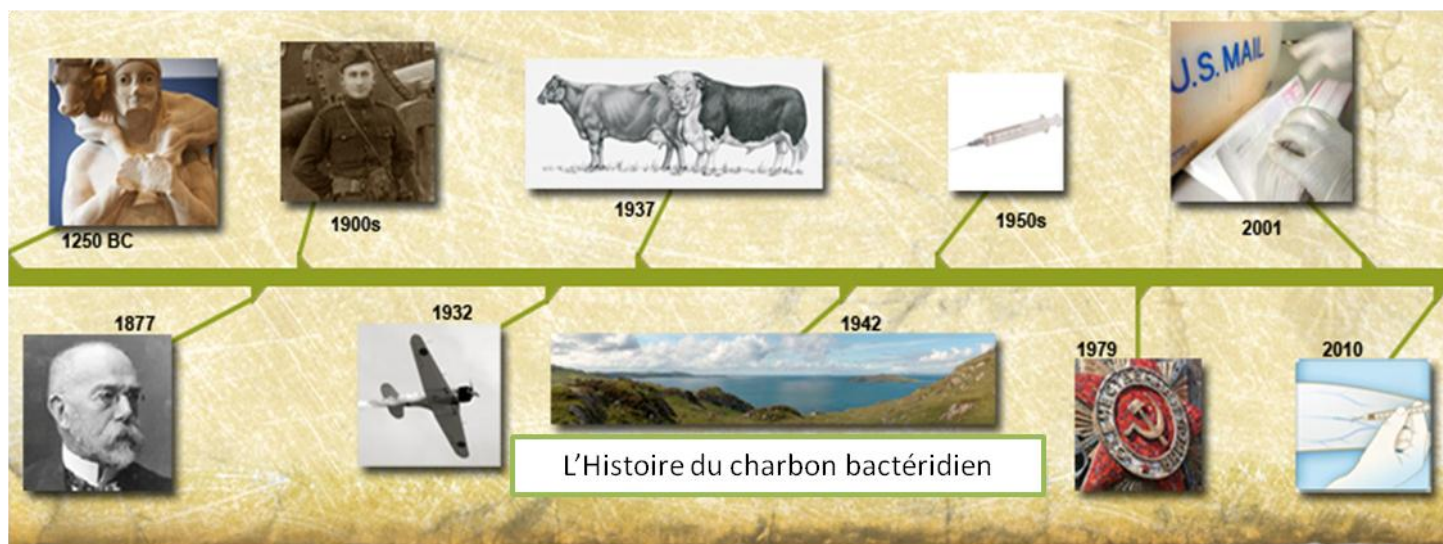


Figure 1 : le charbon bactérien à travers les époques
(Adapté d'après <http://www.cdc.gov/anthrax/history/index.html>).

- Antiquité: Description de maladies aux symptômes évocateurs de fièvre charbonneuse durant les 5^{èmes} et 6^{ème} plaies d'Egypte.
- 1877: Robert Koch isole le bacille en culture pure et l'injecte à des animaux. Ces études seront à l'origine d'un de ses postulats : « Un micro-organisme cultivé doit entraîner l'apparition de la maladie lorsqu'il est introduit dans un organisme sain ».
- 1881: Louis Pasteur développe le premier vaccin animal contre la fièvre charbonneuse.
- Années 1900: Premier usage « militaire » de la bactérie en tant qu'arme biologique durant la 1^{ère} guerre mondiale.
- 1932: Le Japon fait des expériences sur l'utilisation de la bactérie en tant qu'arme contre la Chine.
- 1937: Développement de la souche vaccinale par Sterne et première campagne « massive » de vaccination animale à travers le monde, ce qui permettra de réduire grandement l'incidence des cas humains.
- 1942: Test d'une forme de souche militarisée sur l'île de Gruinard en Ecosse par l'armée britannique. L'île restera inhabitable jusqu'au début des années 1990 où une décontamination chimique massive de l'île sera nécessaire.
- Années 1950: Développement du premier vaccin humain aux USA.
- 1979: Incident mortel de charbon à Sverdlovsk en ex-URSS. Une quantité inférieure à 1 g de spore « militarisées » est relâchée accidentellement par un site militaire de l'armée soviétique. De nombreux cas de charbon d'inhalation se déclarent chez les habitants de la ville la plus proche, conduisant à une soixantaine de décès.
- 2001: Envoi de lettres contaminées par le bacille du charbon aux USA (Amerithrax). Cinq personnes décéderont de charbon d'inhalation. La souche utilisée proviendrait d'un laboratoire de l'armée Américaine.
- 2010: Description d'une nouvelle forme de la maladie, le charbon d'injection, observé parmi les consommateurs d'héroïne en Europe occidentale (Allemagne, Danemark, France, Norvège et Royaume-Uni).

Tout au long de son Histoire, l'Homme a été confronté au monde microbien. Cette cohabitation peut être bénéfique pour les deux protagonistes, on parle alors de symbiose (certaines bactéries du microbiote intestinal humain), parfois elle reste neutre (certaines bactéries du microbiote intestinal humain), et quelque fois, elle est néfaste pour l'un ou l'autre (*Helicobacter pylori*, bactérie responsable d'ulcères gastriques). Depuis les temps anciens, l'Homme a connu de grandes pandémies liées à des pathogènes bactériens : la plus célèbre de mémoire humaine, la Peste Noire qui frappa l'Europe au 14^{ème} siècle, est due à la bactérie *Yersinia pestis*. Le choléra peut également être cité, de même que plusieurs virus, tels que la Variole, la Grippe ou le Virus Ebola plus récemment.

Ce n'est qu'à la fin du 19^{ème} siècle que certaines bactéries ont été identifiées comme étant l'agent infectieux de maladies. *Bacillus anthracis*, agent responsable de la fièvre charbonneuse, fut la première d'entre elles, une longue histoire qui continue à s'écrire aujourd'hui. Les lignes qui vont suivre racontent cette histoire à travers les époques, de l'Antiquité aux temps modernes.

1. *Bacillus anthracis*

1.1. La Fièvre charbonneuse à travers les époques

Quelques rares textes historiques relatent les épisodes tragiques dont ont été frappés les Hommes au cours de l'Histoire. Le récit de l'Exode d'Égypte mentionne les « plaies » qui frappèrent le peuple Égyptien. La Cinquième plaie frappa les troupeaux se trouvant dans les champs (chevaux, ânes, chameaux, bœufs et petit bétail) par une « Peste » très grave. La Sixième plaie, quant à elle, concerne l'apparition d'ulcères et de pustules suite à une exposition à de la suie de fourneau changée en fine poussière (Livre de l'Exode 9 :3) (**Figure 1**). Dans la Grèce et la Rome antique, la maladie est observée et Hippocrate décrit la maladie (4^{ème} siècle avant J-C). En 37 av. J-C, dans les « Géorgiques », le poète et scientifique grec Virgile fit référence à la maladie chez le bétail et les hommes manipulant la laine des moutons infectés (Dirckx, 1981; Sternbach, 2003). Il rapporta plusieurs observations intéressantes : « *les personnes emportant un vêtement fait de laine contaminée, leurs genoux étaient rapidement couverts par des papules enflammées et un exsudat répugnant ; s'ils tardaient trop longtemps à enlever le matériel, une violente inflammation recouvrait les parties du corps touchées* » (Dirckx, 1981).

Des descriptions de la maladie ont été faites au Moyen âge et à la Renaissance où elle était nommée « Poison noir », relatant la transmission de la maladie du bétail vers l'homme (Ambroise Paré en 1568). À l'époque, les auteurs attribuaient la maladie à un « virus » ou à une substance toxique. Au 18^{ème} siècle, une épidémie décima près de la moitié des moutons en Europe (Sternbach, 2003). Au sein de l'Angleterre Victorienne, la fièvre charbonneuse était connue comme la maladie des travailleurs de la laine : la fréquence d'infection chez ces travailleurs exposés aux fibres animales contaminées par la bactérie était importante (Brachman, 1980; Sternbach, 2003).

L'agent bactérien responsable de la fièvre charbonneuse (FC) n'a été identifié qu'au milieu du 19^{ème} siècle (**Figure 1**). Une relation entre le charbon et les sols était déjà connue: de manière cyclique, dans les mêmes parcelles, des morts de bovins étaient observées, sans signes annonciateurs. Les éleveurs avaient ainsi employé le terme de « champ maudit » pour caractériser ces pâtures. Plusieurs médecins et vétérinaires mirent en évidence le bacille du charbon comme agent pathogène de la maladie : ils parlaient alors de bactériidies. Pierre Rayet et Casimir Joseph Davaine découvrirent la bactérie responsable de la maladie en 1850 : ils parlaient de la présence de « bâtonnets » immobiles dans le sang des animaux malades (Carter, 1988; Sternbach, 2003). Mais ils ne réalisèrent pas l'importance de leurs observations immédiatement...

Robert Koch réalisa les premières cultures de la bactérie *in vitro* en 1876 (Koch, 1876). Il identifia le cycle de vie complet de la bactérie pour la première fois : il mit en évidence le phénomène de sporulation, permettant à la bactérie de survivre dans des environnements hostiles (Sternbach, 2003). Il dénomma la bactérie *Bacillus anthracis*. Ce fut la première maladie pour laquelle une origine microbienne a été clairement identifiée (Purcell *et al.*, 2006). Il affirma que la FC pouvait être seulement transmise d'un hôte à un autre en transférant le bacille (Sternbach, 2003), ce qui donna naissance à l'un de ses célèbres postulats: « *un microbe spécifique est responsable d'une maladie spécifique* ».

La fièvre charbonneuse fut également la première maladie pour laquelle un vaccin fut développé. En 1880, Henry Toussaint élabore le premier vaccin charbonneux à partir de sang de mouton contaminé chauffé à 55°C pendant 10 minutes. Puis, Louis Pasteur développa un vaccin vivant atténué (réalisé à partir d'une suspension de bactéries atténuées, obtenue après succession de cultures à 42-43°C) en 1881 et démontra son utilité lors d'une expérience à Pouilly le Fort (Pasteur *et al.*, 1881). Le vaccin fut administré à 25 moutons. On inocula à ces moutons une culture virulente de charbon (25 autres moutons non vaccinés constituèrent le groupe témoin et reçurent la suspension virulente). Les animaux vaccinés survécurent, ce qui ne fut pas le cas des 25 animaux du groupe témoin (Sternbach, 2003).

Au cours du 20^{ème} siècle, plusieurs cas de charbon d'inhalation furent observés en Europe et aux USA parmi les travailleurs du textile utilisant la laine ou la peau de mouton (Carter, 2004; Jernigan *et al.*, 2001; Kissling *et al.*, 2012; Wattiau *et al.*, 2008) (**Figure 2**). Puis, la maladie connut un déclin progressif dans le monde grâce à la prophylaxie par la vaccination des cheptels et à l'amélioration de l'hygiène sur les lieux de travail. Cette maladie sombra peu à peu dans l'oubli (INRS, 2001) (**Figure 1**). Cependant, cette maladie reste d'actualité. Certains pays déclarent des foyers de manière sporadique (France, Italie, USA...) mais la maladie reste endémique dans d'autres régions du globe (Europe de l'Est, Asie du Sud-Est, Afrique et Amérique du Sud).



Figure 2 : Dessin faisant part du danger observé chez les travailleurs de la laine
(Source : <http://100objectsbradford.wordpress.com/2011/07/13/21-death-and-the-woolorter-bradford-doctors-against-anthrax/>)

1.2. Taxonomie

Bacillus anthracis est une bactérie à Gram +, appartenant à l'embranchement des *Firmicutes*, à la classe des *Bacilli*, à l'ordre des *Bacillales*, à la famille des *Bacillaceae* et au genre *Bacillus*. Ce germe fait partie du groupe *Bacillus cereus sensu lato* qui inclut *Bacillus cereus sensu stricto*, *Bacillus mycoides*, *Bacillus pseudomycoides*, *Bacillus thuringiensis*, *Bacillus weihenstephanensis* et *Bacillus cytotoxicus* (Guinebretiere *et al.*, 2013; Rasko *et al.*, 2005). Ces sept espèces présentent de fortes similitudes génétiques (homologies ADN - ADN, ARNr 16S et 23S,...) (Helgason *et al.*, 2000; Kolstø *et al.*, 2009; Rasko *et al.*, 2005; Tourasse *et al.*, 2006) et elles sont souvent réunies au sein d'un unique groupe communément appelé le "groupe *Bacillus cereus*".

Les bactéries de ce groupe sont ubiquitaires dans l'environnement et retrouvées dans des habitats divers et variés tels que les sols, l'intestin des invertébrés, les plantes et les produits issus de l'industrie humaine (Tourasse *et al.*, 2011).

B. cereus est une cause d'infections d'origine alimentaire et de contamination dans les milieux hospitaliers. Il peut également être la cause d'infections plus sévères en tant que pathogène opportuniste chez l'homme (Bottone, 2010; Drobniewski, 1993; Ekman *et al.*, 2009; Kotiranta *et al.*, 2000).

B. thuringiensis est bien connu pour être un pathogène d'insectes utilisé en tant que pesticide d'origine biologique à travers le monde (Bravo *et al.*, 2007; Schnepf *et al.*, 1998; Soberon *et al.*, 2009; Soberon *et al.*, 2007). Il peut dans certains cas devenir un pathogène opportuniste chez l'homme et causer des nécroses, des infections pulmonaires ou des infections alimentaires (Ghelardi *et al.*, 2007; Hernandez *et al.*, 1998; McIntyre *et al.*, 2008).

B. weihenstephanensis est une bactérie tolérante au froid qui est retrouvée fréquemment en tant que contaminant des produits laitiers (Lechner *et al.*, 1998). Généralement, cette dernière ainsi que *B. mycoides* et *B. pseudomycoides* sont considérées comme étant non-pathogènes, bien que dans certains cas, elles peuvent se montrer toxiques (Beattie et Williams, 1999; Prüss *et al.*, 1999; Thorsen *et al.*, 2009; Thorsen *et al.*, 2006).

B. cytotoxicus a été décrite récemment (isolée lors d'une infection alimentaire sévère) et constitue le cluster le plus distant du groupe *B. cereus* (Guinebretiere *et al.*, 2013).

B. anthracis est quant à lui un pathogène animal responsable de la fièvre charbonneuse. Il est reconnu comme ayant récemment divergé de *Bacillus cereus* (il y a quelques dizaines de milliers d'années) (Keim *et al.*, 2009). (**Figure 3**).

Toutes les espèces du groupe *B. cereus sensu lato* sont donc considérées comme pathogènes ou ont le potentiel de l'être. En effet, la présence de gènes codant des entérotoxines et des enzymes dégradantes responsables de diarrhées ou de maladies intestinales a été identifiée au sein de ces espèces (Hendriksen *et al.*, 2006; Stenfors Arnesen *et al.*, 2008; Stenfors *et al.*, 2002).

Bien que toutes ces espèces soient extrêmement proches au niveau chromosomique, leurs spectres d'action et d'hôte montrent des divergences : les principales différences de pathogénicité sont dues à la présence de plasmides codant pour les principales protéines impliquées dans la virulence de chaque espèce, comme les toxines et la capsule chez *B. anthracis*, les toxines insecticides chez *B. thuringiensis* ou la toxine cereulide de certaines souches de *B. cereus* (Ehling-Schulz *et al.*, 2006; Hoton *et al.*, 2005).

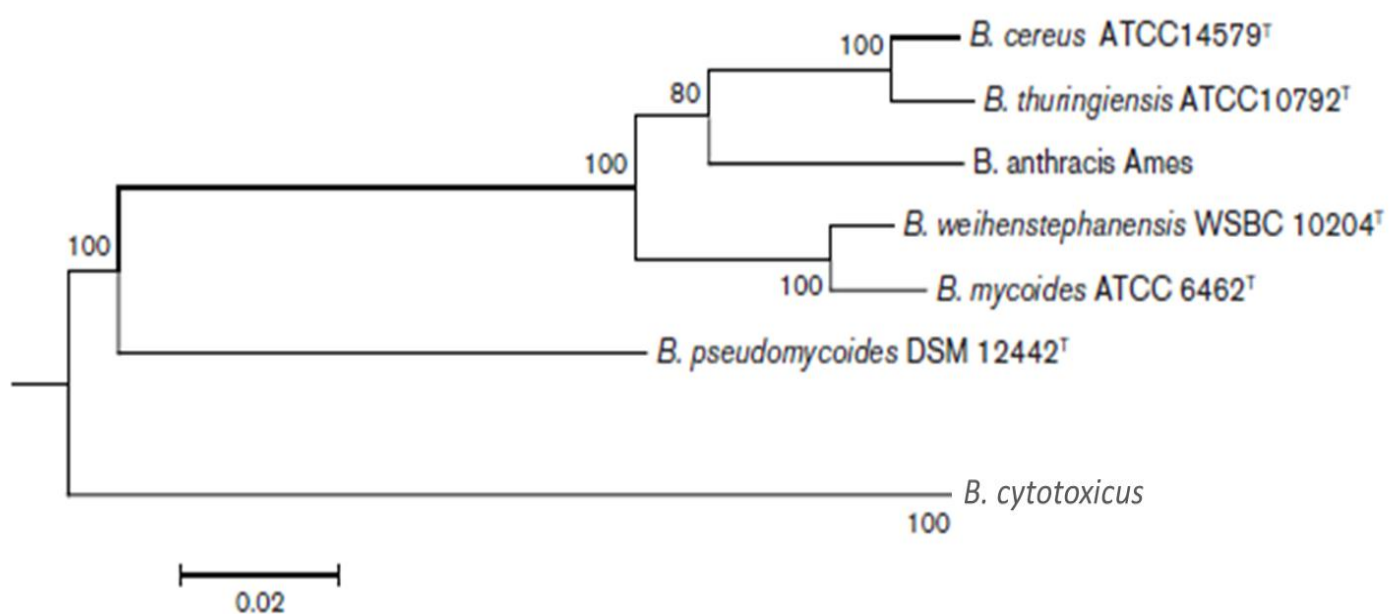


Figure 3 : Représentation du groupe *Bacillus cereus sensu lato* (Maximum-likelihood tree) obtenue à partir de séquences concaténées de gènes (schéma MLST de Tourasse *et al.*, 2006) (adapté d'après Guinebretière *et al.*, 2013)

1.3. Caractéristiques bactériologiques

1.3.1. La bactérie

B. anthracis est un bacille à Gram + d'une taille d'environ 1-1.5 µm de diamètre sur 3-10 µm de longueur (Purcell *et al.*, 2006). Cette bactérie tellurique est immobile, possède la capacité de sporuler et synthétise une capsule (Purcell *et al.*, 2006). Au cours de son cycle naturel, *B. anthracis* peut se présenter sous deux formes distinctes : la forme végétative, qui se multiplie dans l'hôte infecté, et la spore, forme infectante et de résistance.

Lorsqu'il est isolé de prélèvements provenant de cas pathologiques, *B. anthracis* se présente le plus souvent sous forme isolée ou en courtes chaînes. Cependant, en culture, il forme fréquemment des chaînes plus longues (**Figure 4a**) (Koehler, 2009).

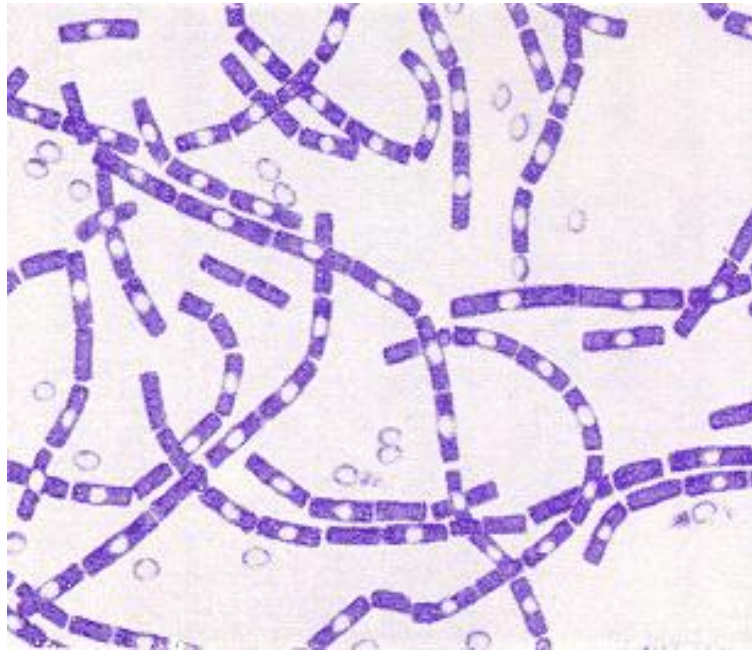
Sur le plan métabolique, *B. anthracis* se caractérise par des réactions catalase positive, nitrate réductase positive, uréase négative, et positive à la réaction de Voges-Proskauer (mise en évidence de la voie fermentaire du butan-2,3-diol lors de l'identification biochimique des entérobactéries) (OMS, 2008). Concernant le métabolisme des sucres, il acidifie le glucose (mais pas le mannitol ni le xylose).

Certaines caractéristiques phénotypiques permettent de différencier *B. anthracis* des autres membres du groupe *B. cereus sensu lato* : la sensibilité aux β-lactamases, le déficit en activité hémolytique ou la sensibilité au phage gamma (Økstad et Kolstø, 2011). Son immobilité est une des caractéristiques le différenciant des autres membres du groupe. Ceci pourrait être dû à des délétions qui ont été mises en évidence dans les gènes de l'opéron du flagelle, impactant 4 protéines essentielles.

B. anthracis est une bactérie aérobie-anaérobie facultative qui pousse en plus ou moins 24 heures sous atmosphère normale. Les bactéries se développent entre 14 et 45°C avec un optimum situé entre 35 et 38°C. Le pH est un autre facteur important concernant la croissance, avec une tolérance entre 6 et 8,5, et un optimum entre 7 et 7,4. Sur géloses nutritives, les colonies sont larges (3 à 5 mm de diamètre) et ont un aspect rugueux en « tête de méduse » (**Figure 4b**). Elles sont blanches ou grisâtres le plus souvent. Sur gélose au sang, le germe apparaît non hémolytique, bien que de rares souches aient présenté un caractère hémolytique (Gierczyński *et al.*, 2004).

La cellule végétative contient une couche de peptidoglycane importante, des acides lipotéichoïques, ainsi qu'une couche S (S-layer). Contrairement aux autres bactéries à Gram+, *B. anthracis* ne contient pas d'acides téichoïques et les protéines de la couche S-layer ne sont pas glycosylées.

a.



b.



Figure 4 : Culture de *Bacillus anthracis*

a. Vue microscopique de *B. anthracis* observé en coloration de Gram au microscope (grossissement 1500X). On y observe l'aspect caractéristique du bacille aux extrémités angulaires. De même, la spore est bien visible en position centrale (elle n'est pas colorée).

(<http://textbookofbacteriology.net/Anthrax.html>)

b. Aspect morphologique de *B. anthracis* sur gélose au sang après 16 heures d'incubation à 37°C (ANSES, UZB).

1.3.2. La spore

En présence de conditions appropriées (atmosphère humide, présence d'oxygène, température comprise entre 15 et 41°C avec un optimum compris entre 30 et 35°C), la bactérie peut sporuler, et ce, en moins de 24h. La spore, ovoïde et non déformante, occupe une position centrale (ou para-centrale) et a une taille d'approximativement 1 µm sur 2 µm (**Figure 5**). Elle possède un aspect réfringent lorsqu'elle est observée sous microscope à contraste de phase.

La structure de la spore peut être décrite comme une apposition séquentielle d'enveloppes autour du « cœur » qui contient le génome bactérien (**Figure 6**). Plusieurs couches successives entourent le cœur : membrane interne, cortex et paroi primordiale, membrane externe et manteau ainsi que l'exosporium (Gimenez, 2006). Sous cette forme, elle peut survivre dans le sol pendant plusieurs décennies. La persistance dans le sol est favorisée par différents facteurs, parmi lesquels peuvent être cités le pH (neutre ou légèrement alcalin), la température, l'humidité, l'activité hydrique, la disponibilité en oxygène ou la présence de certains ions (principalement certains cations et notamment l'ion manganèse Mn^{2+}).

Ces spores sont extrêmement résistantes à l'exposition de nombreux agents physiques ou chimiques. Certaines souches se montrent plus résistantes que d'autres aux diverses agressions extérieures (thermique, chimique, physique,...). Pour leur destruction, diverses méthodes sont préconisées. On peut citer la chaleur sèche (120-140 °C pendant 3 heures), la chaleur humide (121 °C pendant 10 minutes), le formol à 5 pour cent pendant 4 heures, le glutaraldéhyde à 2 pour cent pendant 2 heures, l'eau oxygénée à 3 pour cent pendant 1 heure ou l'acide peracétique à 0,6 pour cent pendant 1 heure.

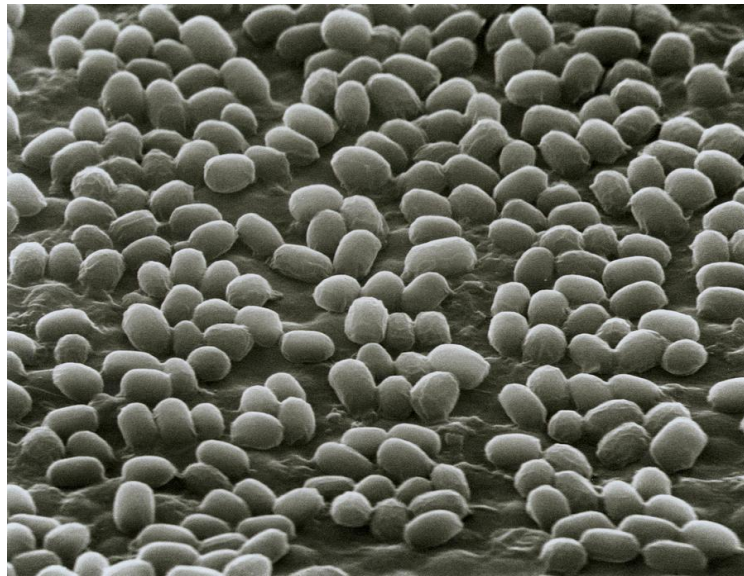


Figure 5 : Spores de *Bacillus anthracis* observées par microscopie électronique (SEM)
(<http://fineartamerica.com/featured/spores-of-bacillus-anthraxis-bacteria-nibsc.html>)

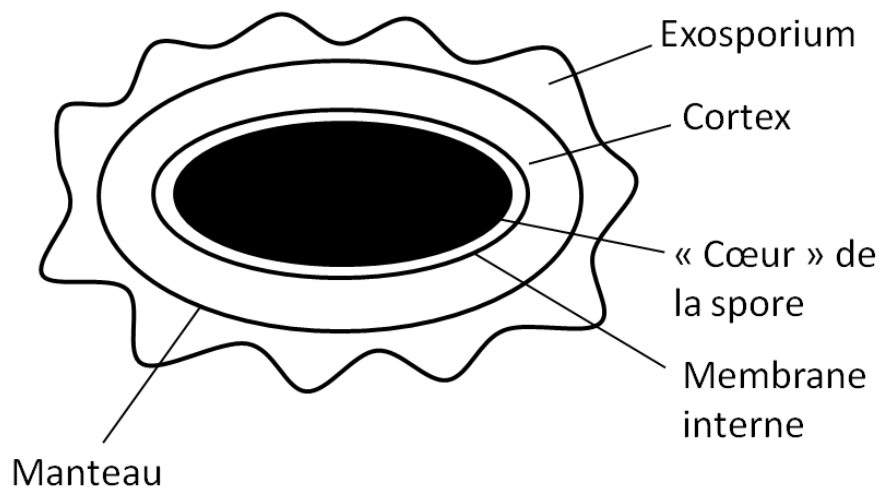


Figure 6 : Spore de *Bacillus anthracis*. Les différentes couches de la spore sont représentées (Cœur, membrane, manteau, cortex et exosporium)
(D'après <http://pubs.acs.org/subscribe/archive/mdd/v04/i06/html/06disease.html>)

1.4. Structure du génome

B. anthracis possède un chromosome circulaire d'environ 5.2 Mb avec un contenu en GC d'approximativement 35 % (valeurs du génome Ames Ancestor : 5.227.293 pb, 5508 ORFs, 35.4% de GC) et 11 opérons ribosomiques (Read *et al.*, 2003) (**Figure 7**).

Le génome de *B. anthracis* possède également deux plasmides circulaires et extra-chromosomiques:

- pXO1, décrit pour la première fois en 1983 (Mikesell *et al.*, 1983), a une taille d'environ 181 kb et un contenu en GC de 32% (valeurs du génome Ames Ancestor : 181677 pb, 217 gènes, 32.5% de GC) (Read *et al.*, 2003) (**Figure 7**).

- pXO2, décrit pour la première fois en 1985 (Uchida *et al.*, 1985), a une taille d'environ 94 kb et a un contenu en GC de 33% (valeurs du génome Ames Ancestor : 94829 pb, 113 gènes, 33% de GC) (Read *et al.*, 2003) (**Figure 7**).

Ces deux plasmides contiennent les gènes codant pour les principaux facteurs de virulence de la bactérie. Le plasmide pXO1 contient les gènes codant les toxines alors que le plasmide pXO2 contient les gènes impliqués dans la synthèse de la capsule.

En comparant plusieurs génomes (*B. cereus*, *B. thuringiensis* et *B. anthracis*) par hybridation génomique comparative (CGH), seules quatre régions sur le chromosome sont spécifiques de *B. anthracis* et correspondent à quatre prophages (Ivanova *et al.*, 2003; Read *et al.*, 2003).

Une autre différence majeure entre *B. anthracis* et les autres membres du groupe *B. cereus* concerne la mutation « non-sens » dans un régulateur de la transcription des facteurs de virulence extracellulaires : le gène *PlcR*. Cette mutation cause un arrêt prématuré de la traduction, rendant la protéine non fonctionnelle (Kolstø *et al.*, 2009). Dans la plupart des isolats de *B. cereus* et *B. thuringiensis*, *PlcR* est active et joue un rôle dans l'expression d'un certain nombre de facteurs de virulence (codés par le chromosome), parmi lesquels des phospholipases, des protéases, des hémolysines ainsi que des entérotoxines (Økstad et Kolstø, 2011).

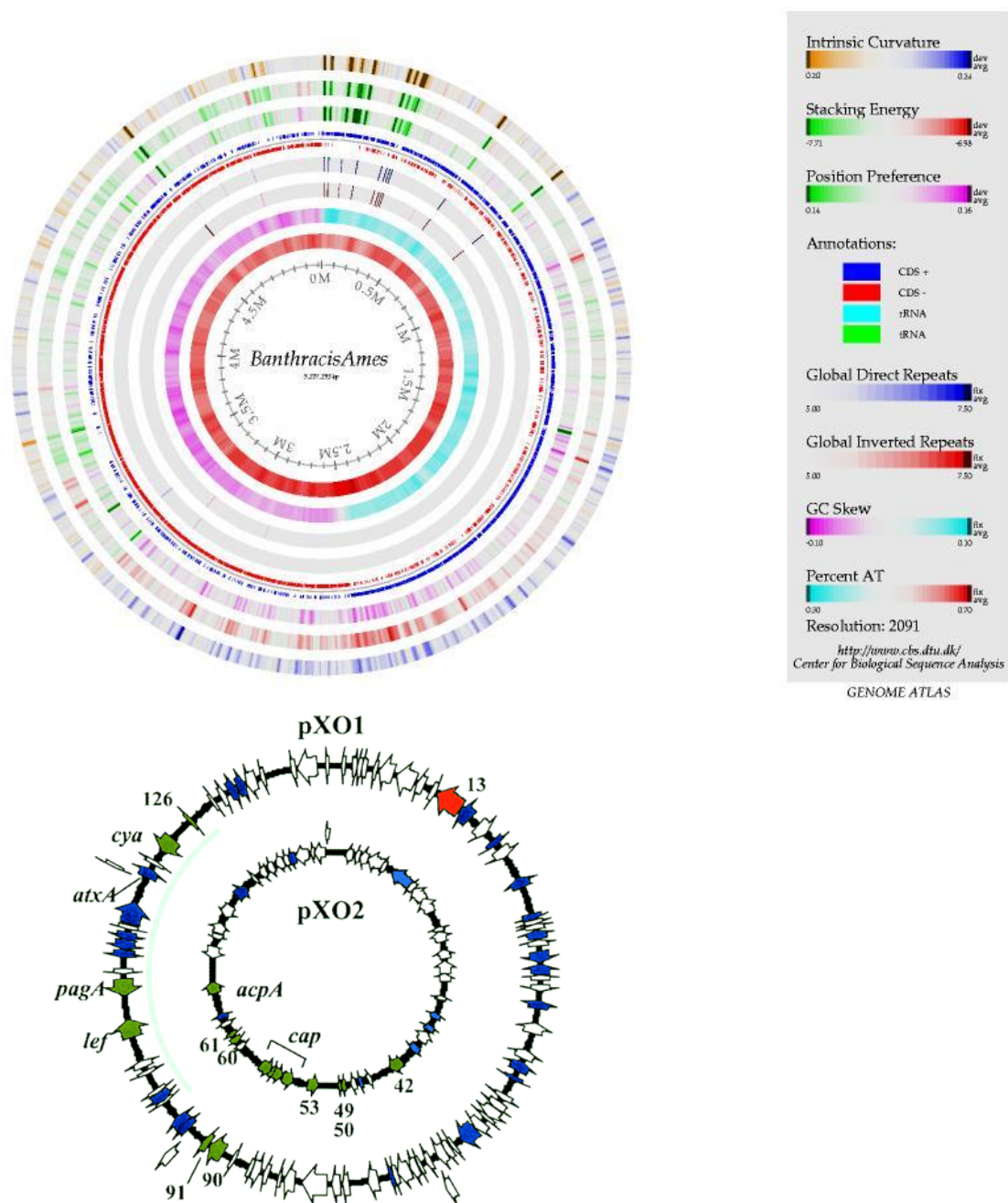
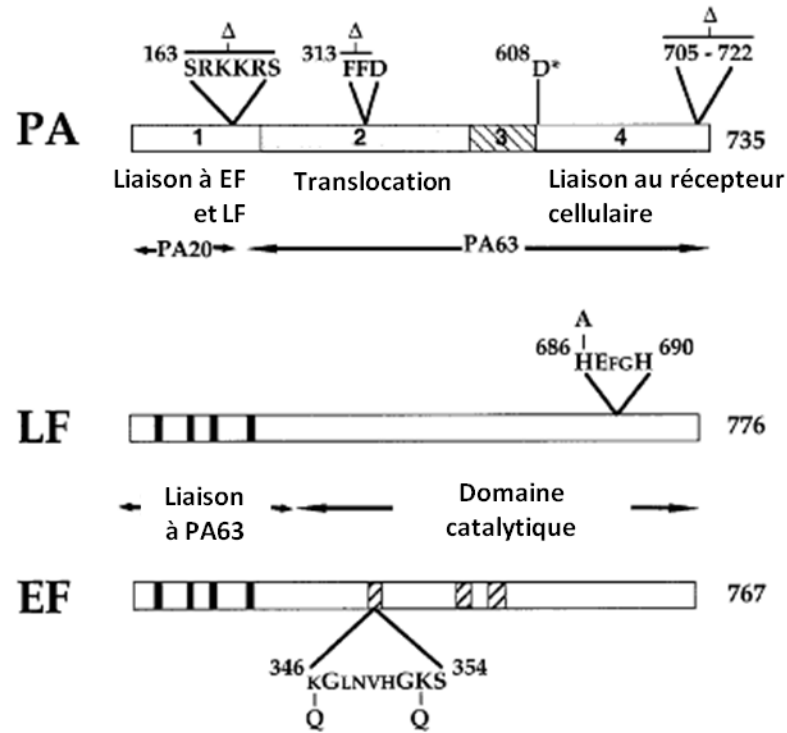


Figure 7: Représentations circulaires du chromosome et des plasmides de *B. anthracis*
 (Source : http://genome.cbs.dtu.dk/staff/dave/MScourse/Lekt_29Feb2003a.html pour le chromosome et (Bourgogne *et al.*, 2003) pour les plasmides)



Les domaines fonctionnels de PA sont numérotés

|| Régions similaires entre EF et LF

▨ Régions similaires avec l'adénylate cyclase de *B. pertussis*

Figure 8: Organisation structurale des toxines protéiques de *B. anthracis* (Brossier et Mock, 2001)

PA : Antigène Protecteur

LF : Facteur Létal

EF : Facteur Oedématogène

1.5. Facteurs de virulence

B. anthracis est une bactérie virulente. Les morts imputées à la fièvre charbonneuse sont principalement dues à une défaillance multi-organes (OMS, 2008). La mort survient par toxoinfection. La capsule et les toxines sont les facteurs majeurs de virulence de la bactérie.

D'autres facteurs de virulence sont supposés exister chez *B. anthracis*. En effet, le séquençage du génome de la bactérie a permis de mettre en évidence des gènes codant pour des protéines homologues à des facteurs de virulence identifiés chez d'autres bactéries pathogènes comme *Listeria monocytogenes* : des hémolysines et des phospholipases sont concernées (Read *et al.*, 2003).

1.5.1. Capsule polypeptidique

Le premier facteur de virulence, la capsule, est formée d'un polymère d'acide poly- γ -D-glutamique (Goodman et Nitecki, 1966). Elle se forme en présence de CO₂ et de bicarbonate à 37°C. En culture *in vitro*, sa synthèse est obtenue en couplant une atmosphère élevée en CO₂ à la présence de bicarbonate (OMS, 2008). Elle constitue le premier facteur conduisant à la maladie, en s'opposant à la phagocytose des cellules végétatives une fois dans l'organisme infecté. Les gènes codant pour sa synthèse (gènes *cap*) sont portés par le plasmide pXO2. Cinq gènes sont impliqués dans sa biosynthèse : *capA*, *capB*, *capC*, *capD* et *capE* (Candela *et al.*, 2005) et peuvent servir de marqueurs diagnostics. Structure peu immunogène, elle permet aux bacilles de se multiplier chez l'hôte.

1.5.2. Toxines protéiques

Le charbon bactérien est causé en partie par l'action combinée de toxines. Les gènes codant pour ces toxines sont portés par le plasmide pXO1. Trois protéines vont agir en synergie : l'antigène protecteur (PA, codé par le gène *pagA*), le facteur létal (LF, codé par le gène *lef*) et le facteur œdémogène (EF, codé par le gène *cya*). Ce complexe toxinique est synthétisé au cours de la phase de croissance exponentielle de la bactérie (Brossier et Mock, 2001; OMS, 2008).

Le facteur LF s'associe à l'antigène protecteur PA pour former la toxine létale, alors que le facteur EF forme la toxine œdémogène en s'associant avec l'antigène protecteur PA. Elles font partie des toxines de type A-B (A=activité enzymatique et B=domaine de fixation) (Brossier et Mock, 2001; OMS, 2008).

L'antigène protecteur (PA) sert de domaine de liaison aux facteurs EF et LF. Sous sa forme mature, cette protéine est dénommée PA83. Elle est constituée de 4 domaines fonctionnels (**Figure 8**). Le domaine 1 contient le site d'activation protéolytique. Ce domaine est essentiel, car il permet d'exposer le site de fixation aux facteurs EF et LF. Le domaine 2 est un tonneau Béta qui est impliqué dans la formation de pores. Le domaine 3 est le plus petit des quatre domaines et possède un patch hydrophobe qui est supposé être impliqué dans une interaction protéine-protéine. Le domaine 4 se trouve à l'extrémité C-terminale de la molécule et est requis pour la fixation au récepteur (Brossier et Mock, 2001).

Le facteur EF, caractérisé pour la première fois en 1982, est une adénylate cyclase calmoduline-dépendante, qui altère les mouvements d'eau et d'ions dans la cellule. Il conduit à l'œdème caractéristique de la fièvre charbonneuse en catalysant une production anormale d'AMP cyclique à partir d'ATP intracellulaire (la forte concentration en AMPc n'est pas létale pour les cellules hôtes). Le facteur EF aurait un rôle dans le retard de l'activation du processus inflammatoire. Différentes études ont montré que la toxine œdémogène (EF associé à PA) interrompait le réseau de cytokines

pendant l'infection. Cette toxine, lorsqu'elle est administrée, cause des dysfonctionnements du système cardio-vasculaire, des dommages au niveau tissulaire et des hémorragies qui conduisent à la mort (Brossier et Mock, 2001; OMS, 2008).

Le facteur LF est une protéase zinc-dépendante qui interrompt les voies de signalisations au niveau des cascades de phosphorylation. Il a été montré qu'elle clive les parties N-terminales de la plupart des MAPKK (Mitogen-Activated Protein Kinases) (Brossier et Mock, 2001; OMS, 2008).

Les macrophages sont le principal type de cellules affectées par le charbon bactérien. Ces derniers semblent être les médiateurs clés dans le choc induit par les toxines. Les voies de signalisation impliquant les MAPKKs jouent un rôle crucial dans l'activation des macrophages et sont directement impliquées dans la production de cytokines telles que le TNF α , l'IL-1 ou l'IL-6 (Brossier et Mock, 2001). De fortes concentrations en toxine létale conduisent à la lyse des macrophages et à une libération massive de TNF α et d'IL-1 par ces derniers. De plus, cette toxine perturbe la réponse immune des lymphocytes B et T par son action sur les cellules dendritiques (OMS, 2008).

Les effets dysrégulateurs des toxines ont été montrés sur l'ensemble des cellules du système immunitaire et notamment sur les cellules dendritiques (Valade *et al.*, 2009).

Lors d'une infection, la toxémie associée augmente avec le nombre de bacilles jusqu'à un certain seuil à partir duquel les toxines vont causer des dommages au niveau des parois cellulaires, ce qui a pour conséquence la formation de thromboses et d'hémorragies, conduisant à la mort de l'hôte infecté (OMS, 2008).

Le mode d'action des toxines protéiques est illustré en **Figure 9**.

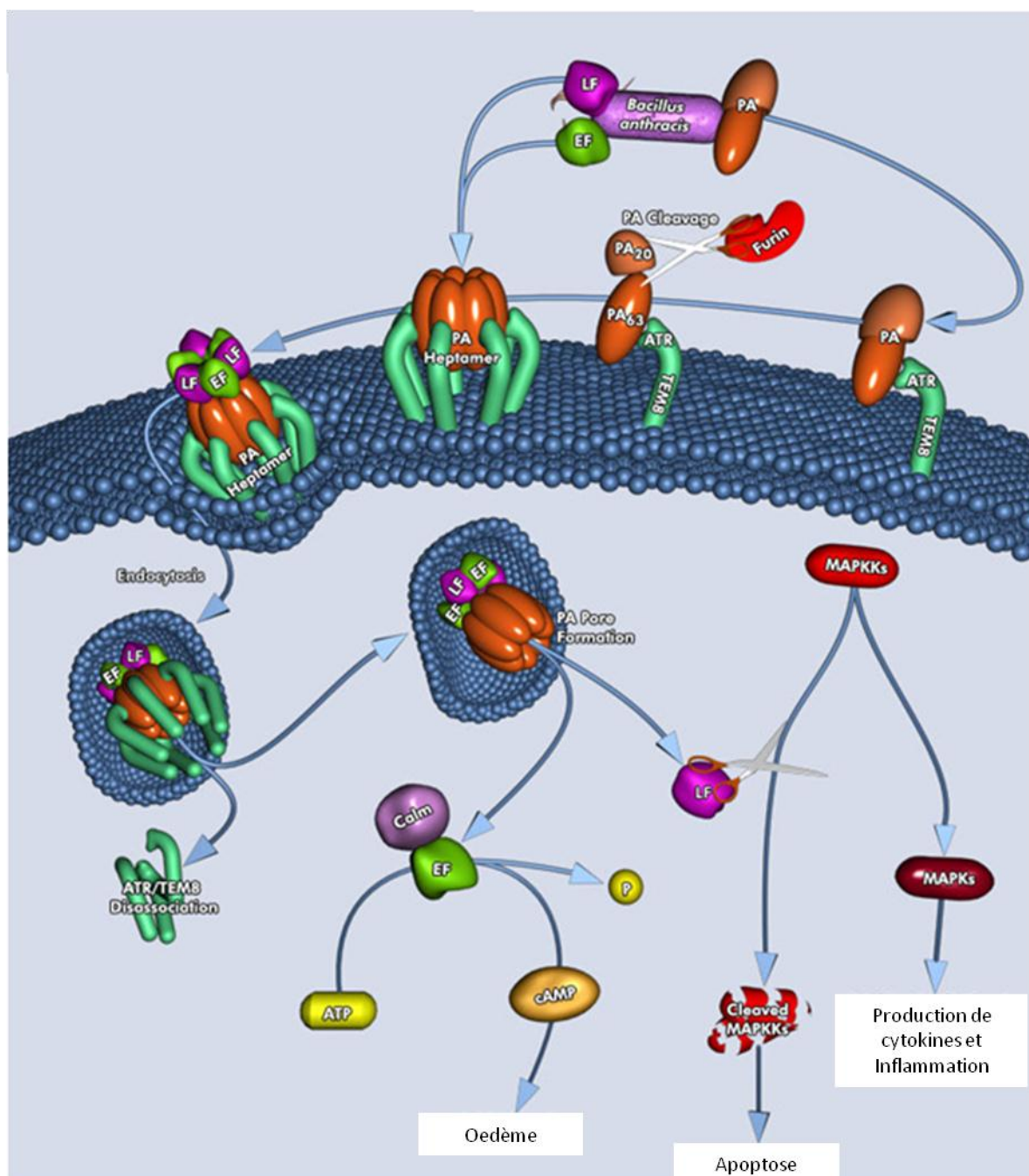


Figure 9: Mode d'action des toxines de *Bacillus anthracis*

(Source : adapté à partir de

<http://www.qiagen.com/fr/products/genes%20and%20pathways/pathway%20details.aspx?pwid=284>)

1.6. Cycle d'infection

B. anthracis est une bactérie tellurique. Sous sa forme sporulée, elle persiste dans le sol durant de longues périodes. De ce fait, les animaux peuvent ingérer les spores durant leur pâture (ingestion d'herbe contaminée) ou lors de « bain de poussière ». Une fois à l'intérieur de l'organisme, les spores vont atteindre les ganglions lymphatiques régionaux et la rate. Elles entrent ensuite en germination et la forme végétative se développe et se multiplie. La production de toxines associée à cette multiplication de cellules végétatives conduit à la mort de l'animal. L'animal malade est capable de disperser une grande quantité de bactéries via les effusions hémorragiques excrétées par les orifices naturels. En cas d'exposition à l'oxygène (ouverture du cadavre, fluide sortant des orifices), les bacilles vont entrer en sporulation très rapidement et se retrouver dans le sol, entretenant ainsi le cycle d'infection (**Figure 10**). Les insectes volants, notamment les insectes piqueurs, seraient susceptibles de transmettre la maladie et de jouer un rôle dans les régions où la maladie est endémique (Fasanella *et al.*, 2010).

Concernant l'Homme, la contamination est principalement accidentelle et concerne majoritairement les personnes manipulant les carcasses d'animaux morts du charbon (vétérinaires, équarisseurs, laborantins, travailleurs de la laine). Plusieurs voies de contamination sont observées (cutanée, injection, gastro-intestinal et inhalation). Les cas de charbon humain sont assez récurrents dans le monde et concernent la forme cutanée le plus souvent (95% des cas).

2. La fièvre charbonneuse

2.1. Situation de la maladie au niveau mondial

La fièvre charbonneuse est une zoonose qui touche une grande majorité des animaux à sang chaud, mais plus principalement les herbivores. Les amphibiens et les reptiles sont naturellement résistants (OMS, 2008). Jusqu'au développement des vaccins et à l'administration d'antibiotiques, le charbon bactérien était l'une des principales causes de mortalité du bétail (bovins, ovins, chevaux, porcs) dans le monde. Cette maladie de répartition mondiale est endémique dans la plupart des pays d'Afrique et d'Asie, dans un certain nombre de pays européens ainsi que dans certains pays d'Amérique du Nord et du Sud (**Figure 11**). Le charbon est recensé plus fréquemment dans les régions tropicales et subtropicales. Des milliers d'animaux en meurent chaque année.

Grâce à la mise en place de certains programmes de lutte au niveau national, il existe une décroissance du nombre de foyers dans certains pays (Turnbull *et al.*, 1999). Cependant, la maladie reste mal diagnostiquée et pas toujours déclarée dans certains pays où la maladie est endémique (Hugh-Jones, 1999).

B. anthracis a toujours été considéré comme l'un des principaux agents du bioterrorisme, du fait de son pouvoir pathogène. Les cas d'utilisation du bacille du charbon en tant qu'arme biologique sont heureusement très rares mais ont tout de même été observés dans le passé (aux USA en septembre 2001 (Imperiale et Casadevall, 2011; Jernigan *et al.*, 2001; Rasko *et al.*, 2011)).

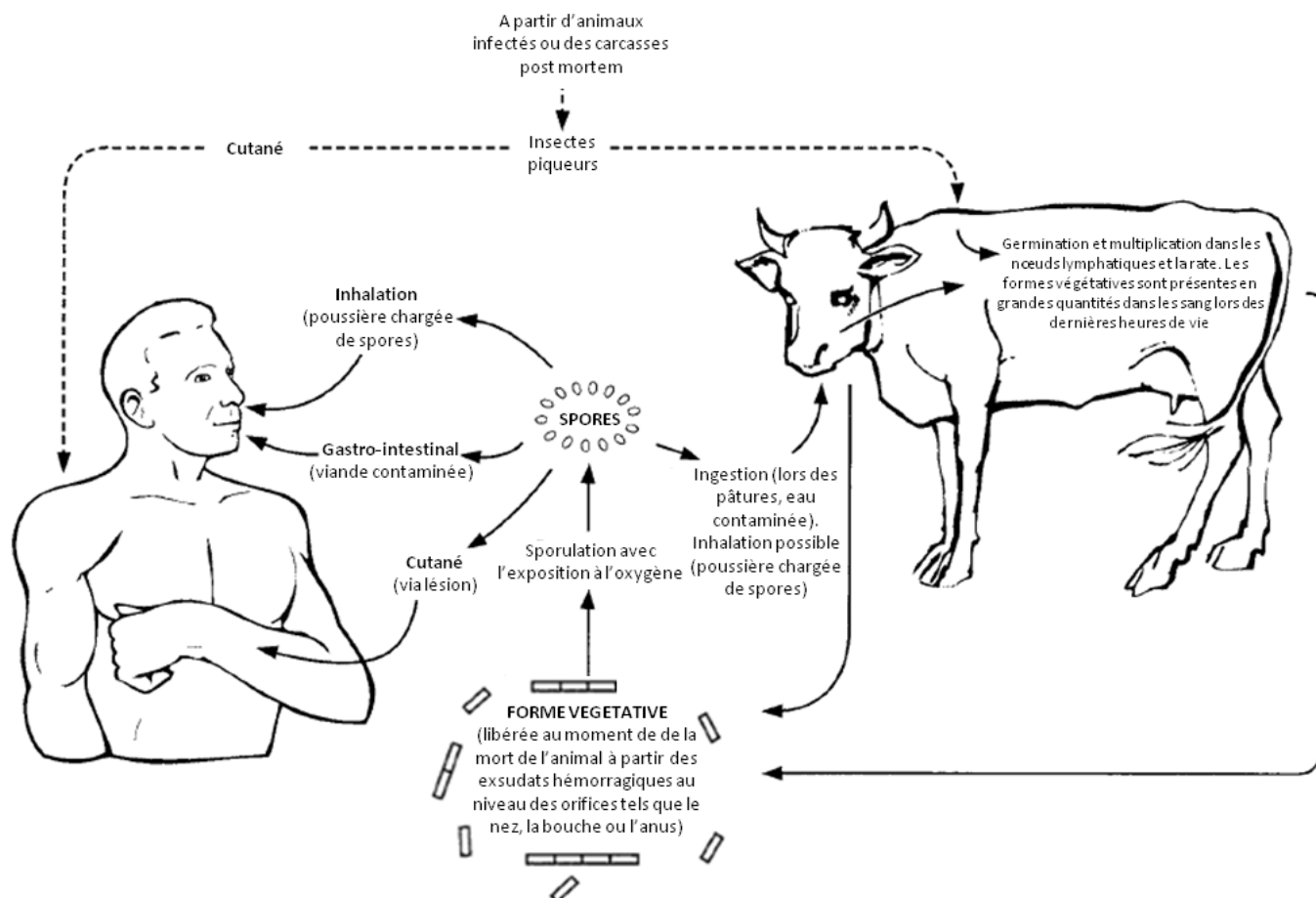


Figure 10: Cycle d'infection de *Bacillus anthracis* (adapté d'après (OMS, 2008))

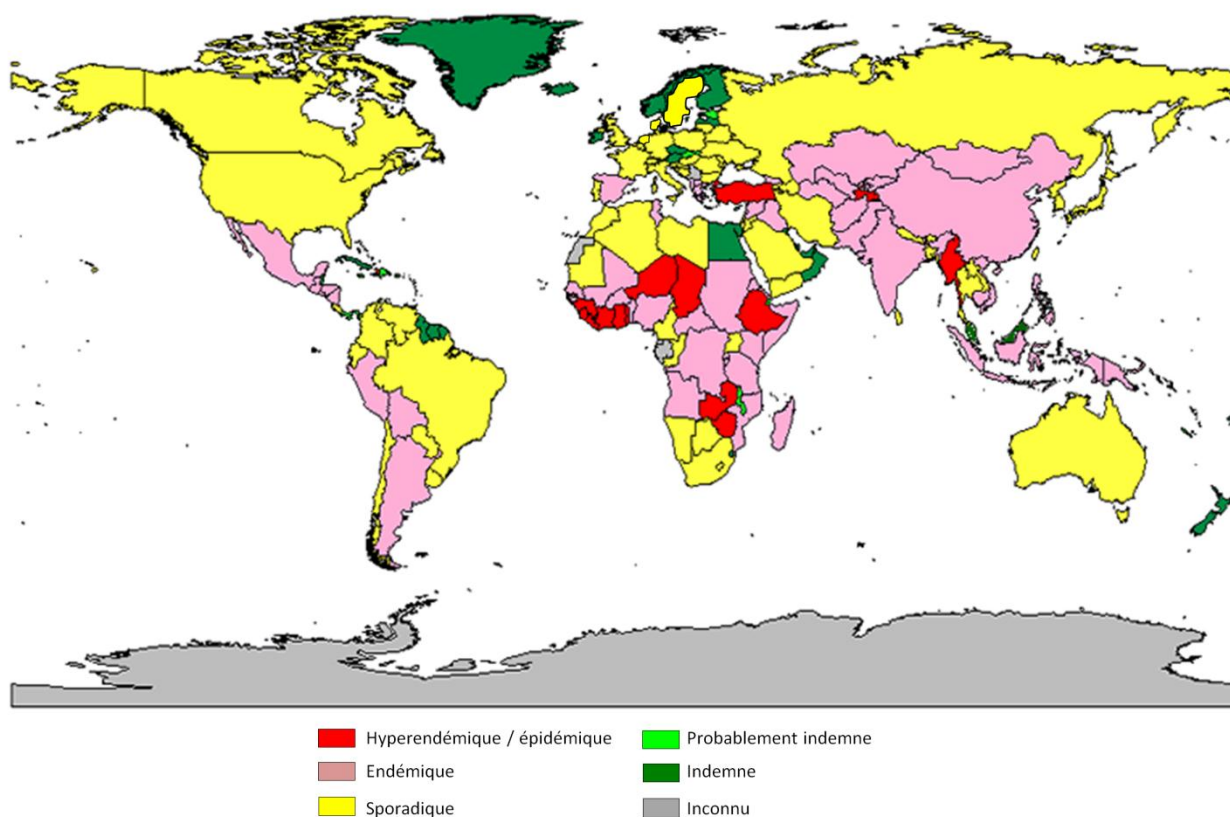


Figure 11 : Répartition de *Bacillus anthracis* à travers le monde. La représentation a été adaptée à partir des données obtenues à l'adresse suivante : http://www.vetmed.lsu.edu/whocc/mp_world.htm (données de 2003 pour la plupart des pays)

2.2. Epidémiologie

La forme sporulée de la bactérie constitue probablement la source d'infection majeure. Le réservoir permanent de la bactérie est donc le sol. Les facteurs favorisant cette persistance ne sont pas complètement élucidés même si un sol riche en matières organiques ainsi qu'un pH alcalin semblent être favorables à la survie des spores. Il est couramment admis que l'ingestion de spores par les animaux durant leur pâture n'est en soit pas suffisante pour le développement de la maladie. Il semblerait que la présence de lésions soit nécessaire pour initier l'infection (INRS, 2001; OMS, 2008). Les pics de contamination animale observés lors de saisons sèches (herbe rase pâturée avec de la terre) vont dans ce sens (INRS, 2001).

Les conditions météorologiques semblent avoir un impact sur la survenue de la maladie. Les cas de fièvre charbonneuse surviennent généralement après des saisons pluvieuses (hiver/printemps) suivies de périodes de sécheresse plus ou moins longues, puis lors de chaleurs accompagnées ou non d'orages (Vaissaire *et al.*, 2001). La pluie joue également un rôle dans la dissémination de la maladie en concentrant les spores sous terre (INVS, 2005). Ces conditions réunies rendent propices l'apparition de la maladie chez les ruminants (INVS, 2005). Les carnivores peuvent quant à eux être contaminés en ingérant de la viande infectée.

La dissémination des spores peut se faire via différentes espèces de charognards (vautours,...) (Lindeque et Turnbull, 1994). Ceci explique que des foyers peuvent apparaître dans des zones identifiées comme indemnes auparavant (INVS, 2005). Il a également été observé que les vers de terre peuvent jouer un rôle dans l'apparition de la maladie en dispersant les spores et en les ramenant à la surface (INVS, 2005). Enfin, l'Homme peut favoriser la dissémination de spores via de grands travaux tels que les drainages ou la construction de routes.

Le sang et les tissus sont considérés comme contaminants, de même que les excréments, le plus gros danger venant du cadavre lui-même (viscères, viande, os, peaux et phanères, ...) (INVS, 2005). Les aliments et sous-produits animaux, tels que les poudres d'os ou farines, peuvent constituer une source de contamination à distance du foyer d'origine (INVS, 2005). La transmission directe par contamination d'une plaie existante a été observée, de même que l'inhalation de spores (Bisons au Canada (Salb *et al.*, 2014)) ou les piqûres d'insecte (Braack et De Vos, 1990; Turell et Knudson, 1987).

2.3. Signes Cliniques

Après avoir pénétré au sein de l'organisme hôte, les spores entrent en germination et la forme végétative se multiplie rapidement. Les toxines sont alors libérées et les symptômes apparaissent. Chez les animaux sensibles, la fièvre charbonneuse provoque généralement des septicémies, le plus souvent mortelles (INVS, 2005).

2.3.1. Formes cliniques selon l'espèce animale

La période d'incubation de la maladie est assez variable (en moyenne de 36 à 72 h, voire plus) (OMS, 2008). Chez le bétail, le délai d'incubation varie entre 1 et 14 jours (OMS, 2008). Les formes de la maladie diffèrent légèrement selon l'espèce animale touchée.

Chez les bovins, plusieurs formes de la maladie sont observées :

- La forme aigüe (charbon septicémique) se manifeste par une atteinte brusque de l'état général de l'animal (frissons, hyperthermie entre 41 et 42°C). Entre 12 et 24 h, des troubles respiratoires (dyspnée) et circulatoires (tachycardie, cyanose des muqueuses,...) sont observés. Des symptômes digestifs et urinaires sont occasionnellement observés. La mort intervient alors en moins de 3 jours (**Figure 12**).

- Une forme suraigüe de la maladie existe. La mort intervient alors en 6 à 12 heures (INVS, 2005; OMS, 2008).

- La forme subaigüe (charbon « externe » ou « à tumeurs ») est caractérisée par une réaction œdémateuse importante (20 à 30 cm de diamètre en quelques heures). Un développement rapide de symptômes identiques à ceux observés lors de la forme aigüe intervient. La mort a alors lieu dans un délai de 4 à 5 jours (INVS, 2005; OMS, 2008).

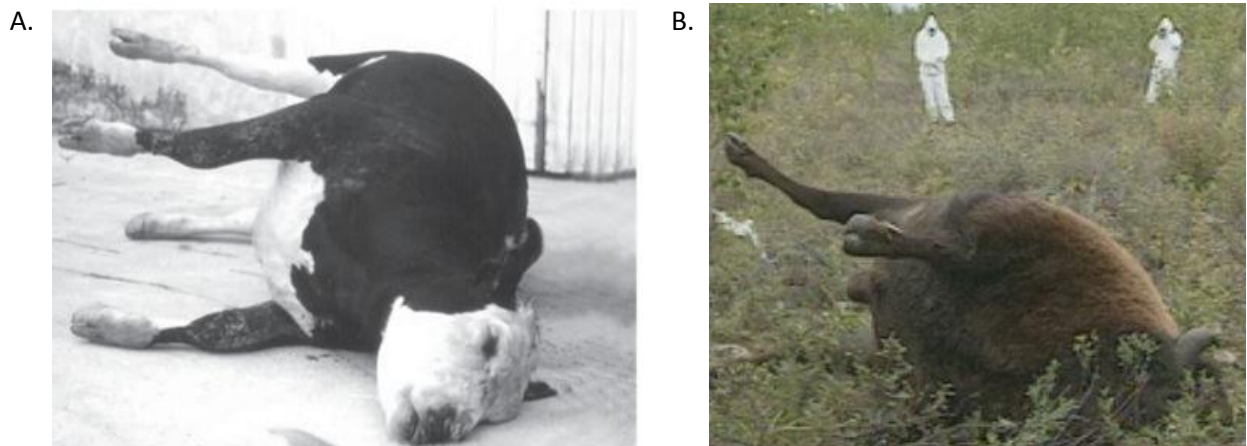


Figure 12 : Animaux morts du charbon bactérien.

A. Vache morte depuis 8 heures : le corps est en posture caractéristique (Schild *et al.*, 2006).

B. Cadavre d'un bison mort du charbon au Canada.

(<http://iamisatthedoors.wordpress.com/2012/07/12/outbreak-of-anthrax-killing-canadian-bison/>).

Chez les petits ruminants, les symptômes sont identiques à ceux observés chez les bovins. Certains symptômes neurologiques aigus et brefs sont observés. Les formes suraigües sont plus fréquentes et les symptômes urinaires sont plus précoces. La mort intervient alors entre 12 et 24h (INVS, 2005).

Les équidés ont quant à eux des symptômes distinctifs : les signes digestifs sont plus importants (coliques, entérite hémorragique), et la maladie évolue plus lentement (mort entre 2 et 6 jours) (INVS, 2005; OMS, 2008).

Cependant, il est à noter qu'il semble y avoir une différence de sensibilité chez les herbivores domestiques et sauvages liée non seulement à l'espèce, mais également à l'individu. Des facteurs autres que les conditions environnementales, les conditions de l'hôte ou le type de souche semblent ainsi entrer en cause mais restent à ce jour assez mal expliqués.

Les suidés sont plus résistants à la maladie que les espèces précitées. Les formes septicémiques sont rarement observées. Usuellement, on observe la forme dite d'angine charbonneuse (tuméfaction œdémateuse de la gorge, puis fièvre, dyspnée, troubles circulatoires, diarrhée et lésions cutanées). Les porcins guérissent de manière spontanée assez fréquemment. Cependant, la mort peut tout de même intervenir en 2 à 4 jours (INVS, 2005).

Les oiseaux semblent assez résistants au charbon, bien que certains cas de morts dues à la fièvre charbonneuse aient été rapportés (OMS, 2008). La température corporelle élevée du corps des oiseaux est l'une des hypothèses émises pour expliquer ce phénomène (OMS, 2008).

Chez les carnivores et autres espèces animales, les symptômes observés sont ceux d'une septicémie, qui s'avère rapidement mortelle (Hugh-Jones et De Vos, 2002; INVS, 2005). Les carnivores sont, en général, plus résistants que les herbivores (Dragon et Rennie, 1995; Gleiser, 1967). Leur résistance pourrait provenir de l'immunité humorale acquise naturellement (exposition fréquente aux carcasses d'animaux morts du charbon). Les anticorps leur permettraient de survivre à la maladie (Turnbull *et al.*, 1992). Un cas à part peut être fait concernant les guépards qui sont très touchés par le charbon en Namibie (OMS, 2008). Ils ne sont pas considérés comme charognards et ne se nourrissent que très peu de charognes, ceci étant dû à leur gabarit qui ne leur permet pas d'entrer en compétition avec d'autres animaux pour une carcasse. De ce fait, ils ne peuvent pas acquérir, à l'inverse des lions par exemple, d'immunité via le repas de carcasses, ce qui explique leur sensibilité à la maladie (OMS, 2008).

Les lésions occasionnées par la fièvre charbonneuse sont caractéristiques, même si toutes les lésions ne sont pas systématiquement observées (INVS, 2005). On observe essentiellement:

- une carcasse foncée, sans rigidité cadavérique et congestionnée,
- un sang de couleur noirâtre, épais, poisseux et incoagulable,
- une rate hypertrophiée (le volume de la rate étant parfois multiplié par 5), globuleuse, noirâtre, flasque, fragile (pulpe de consistance boueuse allant jusqu'à la liquéfaction)
- urine sanguinolente dans la vessie avec congestion rénale intense,
- intestin congestif, hémorragique ou nécrotique (sur une portion limitée),
- « tumeur charbonneuse », œdème gélatineux et ambré entourant un groupe ganglionnaire interne ou externe hypertrophié, hémorragique et nécrosé,
- hémorragies de la graisse sous-cutanée,
- foyers hémorragiques disséminés dans le parenchyme pulmonaire sain.

Le **Tableau 1** récapitule les différents symptômes observés chez les espèces animales pré-citées.

Tableau 1 : Récapitulatif des formes cliniques observées chez les animaux
NR : Non rapporté

Espèce	Délai d'incubation avant la mort	Signes cliniques	Lésions	Sensibilité
Bovins	6 à 12h (forme suraigüe)	Frissons, hyperthermie, dyspnée, tachycardie, cyanose des muqueuses, symptômes digestifs et urinaires occasionnels. Œdème important, puis symptômes identiques à la forme aigüe.	• Carcasse foncée sans rigidité cadavérique et congestionnée • Sang noirâtre et incoagulable • Rate hypertrophiée, globuleuse, fragile, noirâtre et flasque • intestin congestif, hémorragique ou nécrotique • urine sanguinolente dans la vessie avec congestion rénale intense • hémorragies de la graisse sous-cutanée • hémorragies disséminées dans la parenchyme pulmonaire sain • « tumeur charbonneuse »	++
	3 jours (forme aigüe)			
	4 à 5 jours (forme subaigüe)			
Petits ruminants	Entre 12 et 24h (forme suraigüe fréquente)	Symptômes identiques à ceux observés chez les bovins, avec symptômes urinaires plus précoces.		+++
Equidés	Entre 2 et 6 jours	Coliques, entérite hémorragique		+
Suidés	Entre 2 et 4 jours	Tuméfaction œdémateuse de la gorge, fièvre, dyspnée, troubles circulatoires, diarrhée, lésions cutanées.		+/- (guérison spontanée fréquente)
Oiseaux	NR	NR		+/-
Carnivores	NR	Septicémie		+/-

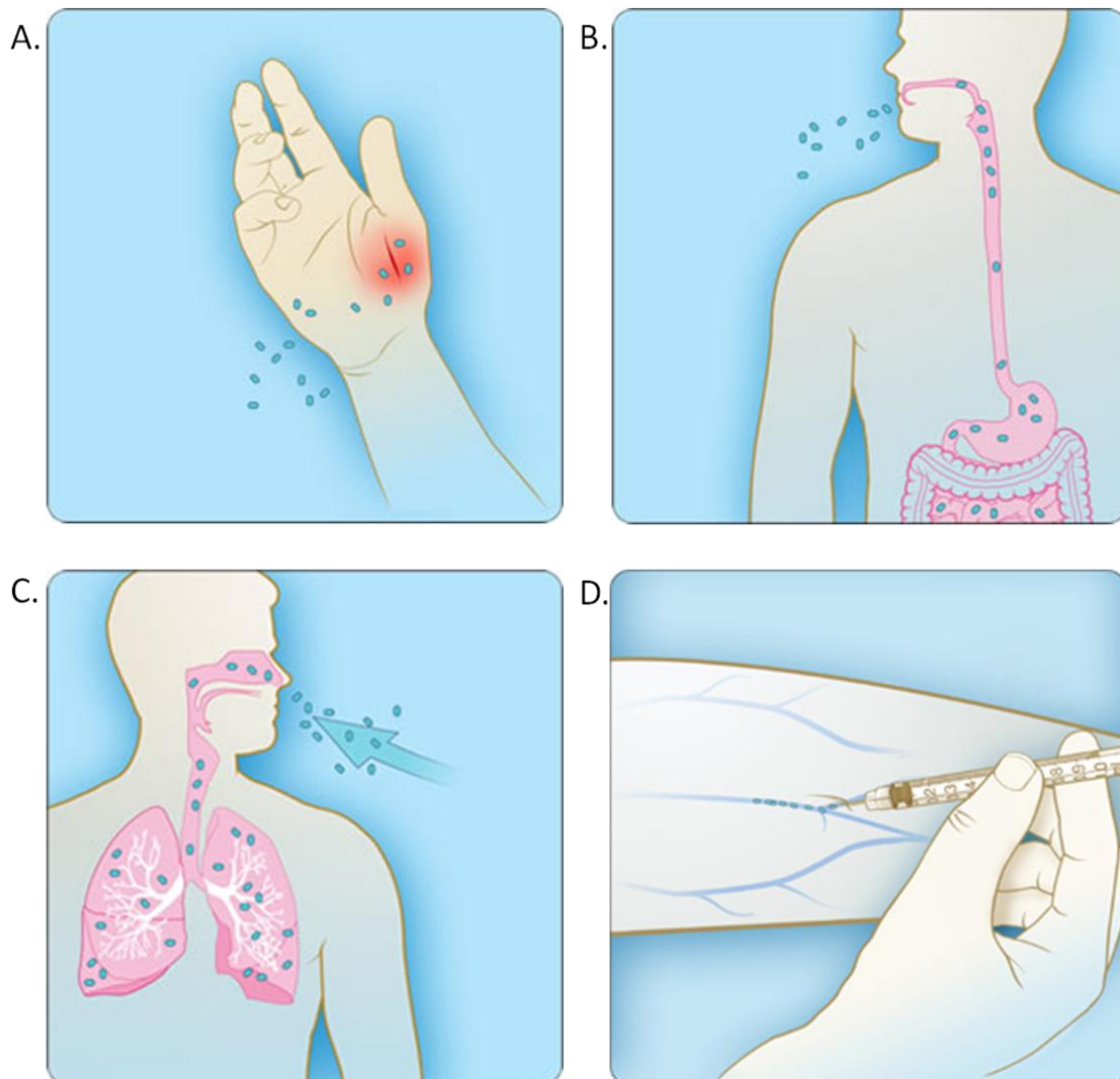


Figure 13 : Illustration des différents modes de contamination observés chez l'Homme

- A. Charbon cutané
- B. Charbon gastro-intestinal
- C. Charbon d'inhalation
- D. Charbon d'injection

(Source : <http://www.cdc.gov anthrax/types/>)

2.3.2. Chez l'Homme

Chez l'Homme, la maladie peut se présenter sous 3 ou 4 formes, selon le site d'entrée du pathogène. Les trois formes principalement observées sont le charbon cutané, le charbon gastro-intestinal et le charbon d'inhalation. La quatrième est une variante du charbon cutané (charbon d'injection).

Le charbon dit cutané est la forme la plus fréquemment observée (95% des cas) (INRS, 2001; Purcell *et al.*, 2006). Les spores pénètrent dans l'organisme via une blessure (plaie, piqûre d'insecte,...) (**Figure 13A**). Les bras, les mains, la face et le cou sont les zones les plus fréquemment touchées (INVS, 2005). L'incubation est dite silencieuse, puis une papule érythémateuse apparaît au bout de quelques jours après la contamination (1 à 5 jours d'incubation en moyenne) (INRS, 2001; Purcell *et al.*, 2006; Shafazand *et al.*, 1999). Cette papule évolue vers la lésion caractéristique du charbon cutané : l'escarre noirâtre (**Figure 14**). L'évolution est généralement favorable, l'escarre régressant après une dizaine de jours pour disparaître en 2 à 3 semaines (Purcell *et al.*, 2006). Le temps de disparition dépend de la taille, la localisation et la sévérité de la lésion (OMS, 2008). Il est parfois nécessaire de réaliser un traitement chirurgical pour exciser celle-ci (OMS, 2008). Cependant, dans certains cas, l'escarre peut progresser et l'œdème devient important (visage ou membre). On parle alors d'un « œdème malin » et les patients présentent fièvre et maux de tête, avec une sensation de malaise (OMS, 2008; Purcell *et al.*, 2006). Bien que l'antibiothérapie ne soit pas indispensable pour le traitement de l'escarre, elle permet néanmoins d'éviter l'évolution vers la forme systémique (INRS, 2001). En absence de traitement antibiotique (ou dans le cas d'un traitement trop tardif), 5 à 20 % des cas de charbon cutané peuvent évoluer vers une forme septicémique de la maladie (fièvre élevée, complications, état de choc et décès) (INRS, 2001). Les formes septicémiques de la maladie sont rares, et avec le traitement antibiotique approprié, la mortalité chute à moins de 1% (Purcell *et al.*, 2006).

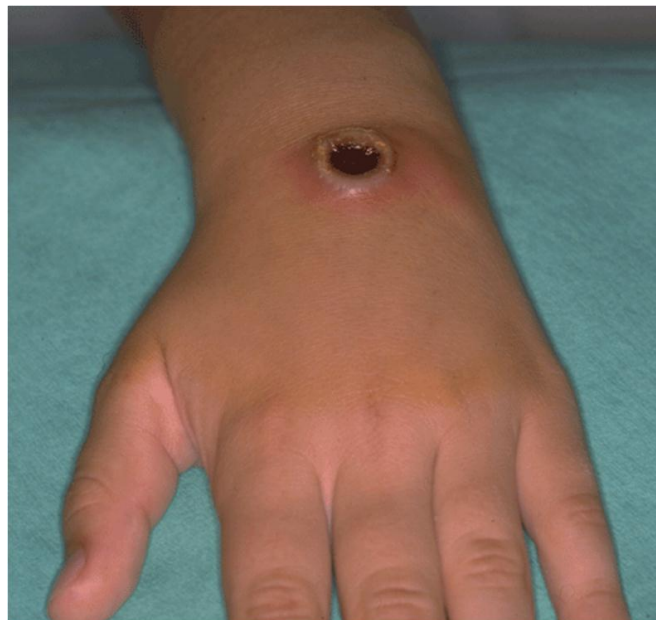


Figure 14 : Charbon cutané. L'escarre noirâtre caractéristique de cette forme de maladie est ici bien visible. (Source : http://textbookofbacteriology.net/pathogenesis_4.html)

Récemment, une nouvelle forme de charbon cutané a été décrite chez les héroïnomanes : elle est appelée charbon d'injection (Berger *et al.*, 2014; Booth *et al.*, 2014; Grunow *et al.*, 2013; Hanczaruk *et al.*, 2014; Price *et al.*, 2012) (**Figure 13D**). L'héroïne est contaminée par les spores : une fois injectées, les spores vont germer et un œdème important se développe à partir du point d'injection. Le diagnostic est souvent compliqué pour les praticiens qui ne pensent pas au charbon en premier lieu. Cependant, les cas de charbon d'injection ayant été fréquents en Europe ces dernières années, causant plusieurs morts (Grunow *et al.*, 2013; Hanczaruk *et al.*, 2014; Price *et al.*, 2012), les praticiens sont désormais plus à même de diagnostiquer les consommateurs d'héroïne et de leur prodiguer les soins leur permettant de guérir. Le taux de mortalité a été évalué à 37 % (Berger *et al.*, 2014).

Lors de la consommation d'aliments contaminés (lait cru ou viande insuffisamment cuite), une forme de charbon dite gastro-intestinale peut survenir (INRS, 2001; INVS, 2005; Purcell *et al.*, 2006) (**Figure 13B**). L'infection se développe à différents endroits du tractus gastro-intestinal, tels que l'œsophage, l'estomac, le duodénum, le jéjunum, l'iléon ou le cæcum (OMS, 2008). La plupart du temps, ce sont l'iléon et le cæcum qui sont touchés (OMS, 2008). Les symptômes apparaissent généralement 2 à 5 jours après l'ingestion des aliments contaminés. Ces symptômes initiaux non-spécifiques du charbon bactérien (OMS, 2008; Purcell *et al.*, 2006) évoquent un syndrome abdominal subaigu (nausées, vomissements, douleurs abdominales, asthénies,...). Très rapidement, une fièvre élevée apparaît, accompagnée de diarrhée sanglante, d'une sensation de malaise, puis d'un état de choc (INRS, 2001). Le décès survient dans 25 à 60% des cas (il peut parfois survenir en quelques heures), principalement dû à de multiples hémorragies entraînant un collapsus (INRS, 2001). Peu de cas sont décrits, ceci étant certainement dû à la mauvaise identification de cette forme de la maladie, qui est souvent diagnostiquée de manière incorrecte (Sirisanthana et Brown, 2002).

Souvent appelé charbon pulmonaire, et ce de manière impropre puisqu'il ne s'agit pas réellement d'une pneumopathie (INRS, 2001; INVS, 2005), le charbon d'inhalation est dû à la pénétration de spores par voie aérienne : un aérosol de poussières (entre 1 et 5 µm de diamètre) contaminé par des spores vient se déposer au niveau des alvéoles pulmonaires (Cote *et al.*, 2011; INRS, 2001) (**Figure 13C**).

Après avoir pénétré dans l'organisme, les spores sont phagocytées dans les poumons par les macrophages, puis transportées vers le système lymphatique où la germination a lieu (INRS, 2001; OMS, 2008). Cette dernière intervient jusqu'à 60 jours après inhalation des spores (INRS, 2001). Les cellules végétatives détruisent les macrophages, et sont alors libérées dans la circulation sanguine, ce qui conduit, à terme, à une septicémie (OMS, 2008).

Les symptômes apparaissent très rapidement une fois la germination effectuée. Dans les premiers stades, ils s'apparentent à un syndrome pseudo-grippal (INVS, 2005). La suspicion de charbon dépend de la connaissance de l'historique du patient (voyages, contact avec des animaux, joueurs de djembé...) (OMS, 2008), ce qui complique le diagnostic précoce de la maladie. Les symptômes sont les suivants : fièvre, dyspnée, toux sèche, céphalées, évoluant brutalement en quelques jours vers une détresse respiratoire avec dyspnée, cyanose et coma (INRS, 2001).

Lors de la phase aigüe de la maladie, les formes végétatives s'échappent des nœuds lymphatiques endommagés où elles se sont multipliées et s'accompagnent de la libération de toxines (OMS, 2008). A ce stade, les dommages au niveau des tissus pulmonaires, dus à l'action de la toxine létale sur l'endothélium des capillaires pulmonaires, deviennent apparents en radio pulmonaire (OMS, 2008). A l'examen clinique, un signe peut néanmoins orienter le diagnostic : il s'agit de l'élargissement du médiastin observé par radiologie (**Figure 15**), lié à des nécroses et hémorragies des nœuds

lymphatiques et des tissus (Bush *et al.*, 2001; INRS, 2001; INVS, 2005; OMS, 2008; Purcell *et al.*, 2006). Le décès est très fréquent (90% des cas), surtout en l'absence de traitement précoce aux antibiotiques (INVS, 2005). Il survient entre 1 et 4 jours après l'apparition des premiers symptômes, d'où l'importance d'une prise en charge rapide accompagnée d'un traitement adéquat (INRS, 2001; INVS, 2005). Il s'agit de la forme la plus sévère de la maladie, qui survient majoritairement lors de contact avec des peaux d'animaux contaminées (tanneurs et joueurs de Djembé) ou lors de cas de bioterrorisme.



Figure 15 : Radiographie pulmonaire montrant l'élargissement médiastinal caractéristique observé lors des cas de charbon d'inhalation chez l'Homme (Purcell *et al.*, 2006)

Il existe également une forme particulière de charbon, contractée par inhalation des spores au niveau ethmoïdal ou pouvant résulter de complications des autres formes : on parle alors de charbon méningé (méningite hémorragique grave, évoluant en coma et la mort) (INRS, 2001; INVS, 2005; Sejvar *et al.*, 2005; Shafazand *et al.*, 1999). Dans ce cas, même avec un traitement antibiotique adéquat précoce, le taux de mortalité est élevé (90%).

Le **Tableau 2** récapitule les principales caractéristiques des formes de charbon observées chez l'Homme.

Tableau 2 : Symptômes et lésions des formes cliniques observées chez l'Homme
(Adapté d'après la thèse (Gimenez, 2006))

	Charbon cutané	Charbon d'injection	Charbon gastro-intestinal	Charbon d'inhalation	Charbon méningé
Symptômes et signes	<i>Stades précoces:</i> - Anneaux de petites vésicules entourées de papules centrales érythémateuses prurigineuses mais non douloureuses. <i>Stades avancés:</i> - Escarres noires nécrotiques non douloureuses - Œdème important - Lymphadénopathie régionale	<i>Si injection de l'héroïne:</i> - Zone de tuméfaction (gonflement) et de rougeur au point d'injection, avec ou sans douleur associée - Abcès ou ulcère au point d'injection - Fièvre, maux de tête, nausées (occasionnel) <i>Si inhalation de l'héroïne:</i> - Syndrome grippal (fièvre, maux de tête, douleurs musculaires, toux)	- Nausées et vomissements - Mal de gorge - Dysphagie - Fièvre - Douleurs abdominales - Ascite et hématoméso - Elargissement du médiastin - Œdème et lymphadénopathie cervicale - Anorexie - Diarrhée, sanguinolente parfois	- Fièvre et/ou sueur, frissons - Fatigue sévère ou malaise - Toux minimale ou non productive - Gène dans la poitrine ou douleur pleurétique - Douleur abdominale, nausée, vomissements, - Détresse respiratoire - Mal de tête - Myalgies, mal de gorge - Elargissement médiastinal (radiographie) - Lymphadénopathie médiastinale - Effusions pleurales	- Brusque sensation de malaise - Intense douleur au niveau de la tête - Fièvre - Frissons - Nausées et/ou vomissements
Lésions	- Œdème - Nécrose - Dépôt de fibrine - Infiltration de leucocytes mononucléaires et polymorphonucléaires	- Œdème étendu au niveau du bras - Nécrose	- Lésions intestinales ulcéraives ou hémorragiques - Lymphadénite mésentérique hémorragique - Ascite profuse - Obstruction, perforation intestinale	- Lymphadénopathie médiastinale - Médiastinite hémorragique - Petites effusions pleurales bilatérales - Effusion péricardique - Elévation du pourcentage de neutrophiles	- Hémorragie au niveau de la matière grise, des ventricules et espace sub-arachnoïde - LCR hémorragique - Pléocytose
Mortalité	- <1 % avec traitement - 10 à 20 % sans traitement	37%	25 à 60 %	90 % (surtout en absence de traitement précoce)	90 %

2.4. Diagnostic de la maladie

2.4.1. Chez l'animal

Chez les herbivores, une mort brutale, des saignements au niveau des orifices naturels, des œdèmes observés à différents endroits (pelvien, sous-glossien) ainsi qu'un abdomen distendu, sont des signes permettant l'évocation du charbon bactérien (INVS, 2005). Si le cadavre n'est pas ouvert, la putréfaction est rapide et la forme végétative est détruite, car aucune sporulation n'est observée en anaérobiose (INVS, 2005). Bien qu'il soit recommandé de ne pas autopsier un animal lors d'une mort suspecte et possiblement imputable au charbon, il arrive que la carcasse soit ouverte ailleurs que dans des locaux spécifiques, conduisant à la sporulation et assurant la persistance de la maladie. Ainsi, lorsqu'une carcasse est vieille ou putréfiée, il est tout de même possible d'isoler *B. anthracis* à partir de fragments de peau résiduels ou à partir de matériaux souillés de sang (OMS, 2008).

Chez la plupart des espèces mortes du charbon, le sang est hautement contaminé par la bactérie (OMS, 2008). En effet, des concentrations de l'ordre de 10^6 à 10^9 ufc/ml sont observées chez différentes espèces, à l'exception du porc (OMS, 2008). Après coloration de Gram, la présence du bacille pourra être révélée. La mise en culture doit être rapide, la bactérie pouvant disparaître au bout de quelques heures (INVS, 2005). Si l'animal a été traité par antibiotiques avant sa mort (plus de 12 heures), il n'est en général plus possible de mettre en évidence les bacilles dans les prélèvements (INVS, 2005; OMS, 2008).

La culture de *B. anthracis* est aisée sur différents milieux de culture (gélose ordinaire, gélose au sang) à 37°C en atmosphère ordinaire (INVS, 2005). En 24 heures, des colonies rugueuses sont ainsi obtenues. L'antibiogramme permet le plus souvent d'indiquer une sensibilité à la pénicilline (INVS, 2005). Bactérie immobile, elle ne possède pas de capacité hémolytique, à l'exception de quelques rares cas (Gierczyński *et al.*, 2004), ce qui permet de la distinguer de *B. cereus*.

La technique d'amplification génique *in vitro* (PCR) permet d'identifier les gènes de virulence et de différencier les souches sauvages des souches vaccinales (INVS, 2005).

2.4.2. Chez l'Homme

Le diagnostic clinique de la maladie est simple concernant le charbon cutané : en présence de la lésion caractéristique du charbon et en fonction de l'historique du patient, il est assez aisé de poser un diagnostic concernant ce type de contamination (INVS, 2005). Le diagnostic est plus difficile pour le charbon d'inhalation. Les signes cliniques initiaux sont similaires à ceux observés dans la plupart des infections respiratoires virales ou bactériennes, y compris les plus banales (INRS, 2001). L'historique du patient est alors extrêmement important (exposition à une poudre, menace bioterroriste) (INRS, 2001). Il en va de même pour le charbon gastro-intestinal, où les symptômes précoces sont similaires à une gastroentérite ou à une appendicite (INRS, 2001). Dans ce cas-là, l'historique du patient est critique pour diagnostiquer la maladie (consommation de viande issue d'un abattage sans contrôle vétérinaire,...) (INRS, 2001).

Selon la forme clinique, la bactérie peut être isolée dans le liquide vésiculaire, sous l'escarre, dans le sang, les expectorations, les selles ou le liquide céphalorachidien (**Figure 16**) (INRS, 2001; INVS, 2005). De même que chez l'animal, une antibiothérapie précoce rend difficile la mise en évidence de la bactérie (INRS, 2001). Cependant, elle n'empêche pas l'action des toxines.

L'utilisation de la PCR permet, comme chez l'animal, d'identifier la bactérie, de même que les méthodes de typage bactérien. Un diagnostic sérologique peut également être réalisé, mais servira

uniquement de diagnostic rétrospectif, en cas de traitement antibiotique avant le prélèvement (INVS, 2005). L'antibiogramme est également utile pour le choix du médicament à administrer.

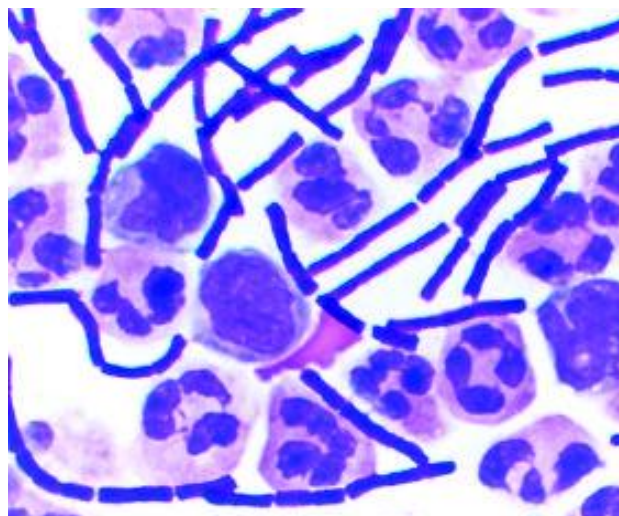


Figure 16 : Coloration de Gram à partir de liquide céphalorachidien montrant la présence caractéristique de bacilles en chaînette (Jernigan *et al.*, 2001)

2.5. Traitements

Le bacille du charbon est une bactérie sensible à de nombreux antibiotiques, parmi lesquels, la famille des bêta-lactamines (comme la pénicilline) et des cyclines (tétracycline) (Heine *et al.*, 2007; INRS, 2001; INVS, 2005). Chez l'Homme, il est désormais recommandé d'utiliser les fluoroquinolones (ciprofloxacine, ofloxacine, lévofloxacine) ou la doxycycline (INRS, 2001; INVS, 2005). Le traitement est à adapter en fonction des résultats de l'antibiogramme (INVS, 2005).

Le traitement de la fièvre charbonneuse repose essentiellement sur l'antibiothérapie, que la maladie soit observée chez l'animal (à l'exception du mouton chez lequel le traitement n'a pas le temps d'être efficace quand il est cliniquement atteint) ou chez l'Homme. Lors d'un cas de charbon cutané, l'antibiothérapie est nécessaire pour contrôler la diffusion bactérienne (traitement de 7 jours), mais n'influe pas sur l'évolution locale de la lésion (INVS, 2005). Le traitement doit être administré en urgence dans les cas de charbon autres que cutanés (INVS, 2005). Son efficacité dépendra grandement de la précocité du diagnostic : en effet, passé un certain délai, l'antibiothérapie devient inefficace. La mort sera inéluctable, ceci étant du soit à l'action des toxines, soit à la septicémie développée (INVS, 2005; OMS, 2008). Lors d'un cas de charbon d'inhalation, la durée du traitement par antibiotiques est bien plus longue : les spores peuvent persister durant plusieurs semaines au sein des alvéoles pulmonaires avant d'entamer leur germination. Il est donc indispensable de pratiquer une antibiothérapie longue (fluoroquinolones, cyclines) et ce, afin d'éviter toute résurgence de la maladie. Il est ainsi recommandé de pratiquer un traitement de 60 jours (Friedlander *et al.*, 1993; INVS, 2005).

Des cas de résistances contre la pénicilline ont été observés (par production de β -lactamases) (Bradaric et Punda-Polic, 1992; Patra *et al.*, 1998), lors de l'épisode des lettres contaminées en septembre 2001 aux USA notamment (OMS, 2008) ou lors d'un foyer animal en Suède en 2011 (Agren *et al.*, 2014). Une résistance à la ciprofloxacine a également été observée dans une étude menée au Japon (Serizawa *et al.*, 2010) ainsi qu'à d'autres antibiotiques (Athamna *et al.*, 2004; Vogler *et al.*, 2012).

2.6. Prévention

De nombreux pays ont réussi à contrôler la maladie en mettant en place des séries de mesures pour interrompre le cycle de la bactérie : traitement approprié des cadavres d'animaux (incinération, enterrement,...), désinfection des sites de morts des animaux (chaux, formaldéhyde,...), décontamination et élimination des matériels contaminés et vaccination des animaux en zone dite « à risque ». Lorsqu'un foyer de charbon est déclaré dans un troupeau, le reste du troupeau doit être mis à l'écart et placé sous surveillance. Lors d'une suspicion d'exposition au bacille du Charbon, une prophylaxie par antibiotiques est conseillée chez l'Homme (OMS, 2008; Purcell *et al.*, 2006). Chez l'animal, la principale action prophylactique est la vaccination chez les animaux apparemment sains associée à un traitement des animaux atteints.

La vaccination est le seul moyen de prévention de la maladie (OMS, 2008) (**Figure 17**). Les vaccins animaux sont disponibles dans la plupart des pays où la maladie est endémique ou sporadique (OMS, 2008). Ce sont des vaccins vivants de souches atténuées, sous forme de spores (OMS, 2008). En France, un seul vaccin est disponible (Carbovac®) (INVS, 2005).

La fièvre charbonneuse est la première maladie pour laquelle un vaccin fut développé (Brossier et Mock, 2001). La souche atténuée de *B. anthracis* isolée par Louis Pasteur en 1881 a été utilisée en tant que vaccin vivant (Pasteur *et al.*, 1881). En 1937, Max Sterne isola une souche toxigénique (possédant le plasmide pXO1) mais non capsulée (déplétion du plasmide pXO2). Elle fait désormais partie des souches utilisées mondialement pour le contrôle du charbon bactérien chez le bétail (souche 34 F/2), bien qu'elle possède une certaine virulence résiduelle (Brossier et Mock, 2001). Depuis, certains dérivés génétiques ont permis de modifier la souche Sterne et d'éviter cette virulence résiduelle.

Des vaccins analogues (vaccins vivants atténués) sont disponibles pour l'Homme en Russie (Cherkasskiy, 1999) et en Chine. Au Royaume-Uni ou aux Etats-Unis, ce sont des vaccins AVA (Anthrax vaccine adsorbed) acellulaires contenant l'antigène protecteur (PA) qui sont produits (Cybulski *et al.*, 2009; OMS, 2008). L'effet protecteur des anticorps anti-PA exerce un effet neutralisant sur les toxines (Brey, 2005). Ces vaccins sont réservés aux personnes présentant un risque très élevé d'exposition à la maladie (OMS, 2008). La vaccination prophylactique par vaccin AVA confère une immunité protectrice après 3 à 6 doses. Un rappel tous les 12 mois est nécessaire pour maintenir la protection (Turnbull, 2000). Il est à noter qu'à l'heure actuelle, aucun vaccin humain n'est disponible en France (en développement entre l'IRBA et l'Institut Pasteur).



Figure 17 : Exemple de vaccin humain contre le charbon (BioThrax®).

(Source : <http://www.pharmaceutical-drug-manufacturers.com/pharmaceutical-drugs/anthrax-vaccines.html>)

3. Arme biologique potentielle

Depuis des temps anciens, les hommes ont utilisé des germes pathogènes en tant qu'arme. La première mention documentée de cet emploi remonte au Moyen-Orient en -1350 av. JC quand la bactérie *Francisella tularensis* (agent de la tularémie) fut utilisée : les villageois laissèrent des béliers contaminés dans leurs villages pillés. En Chine, l'utilisation de cadavres pestiférés envoyés dans les villages assiégés fut un des premiers exemples d'arme bactériologique. Les origines modernes des armes biologiques remontent quant à elles à la première guerre mondiale lorsque l'armée allemande tenta d'utiliser ces agents pour des actes de sabotage (ONU). Par la suite, plusieurs pays créèrent des unités pour la production d'armes biologiques (France, Royaume-Uni, USA, Union Soviétique, Japon,...) (ONU).

Ces germes pathogènes sont utilisés délibérément pour tuer ou affaiblir les armées ou les populations ennemies (hommes ou animaux). Les bactéries peuvent être utilisées, tout comme les virus, les champignons, les rickettsies ou les toxines (ONU). Elles présentent un faible coût de production et sont relativement faciles à produire (contrairement aux armes chimiques ou nucléaires) (ONU). Malgré leurs avantages, elles sont cependant considérées comme peu fiables et militairement inférieures aux armes chimiques et nucléaires (ONU). Le processus de production est généralement propre à chaque agent, mais certaines étapes sont communes, telles que la sélection et la culture du micro-organisme, l'augmentation de la virulence à des fins militaires et la stabilisation de la culture de manière à éviter toute dégradation lors du stockage ou de l'utilisation (ONU).

Classées en tant qu'armes de destruction massive, tout comme les armes chimiques et nucléaires, leur utilisation est interdite depuis 1925 (Protocole de Genève), de même que leur mise au point et leur possession (United-Nations, 1972).

B. anthracis fait donc partie des germes utilisables en tant qu'arme biologique (Tasota *et al.*, 2002). En 1942, le Royaume-Uni réalisa un test en larguant plusieurs bombes remplies de spores du bacille du charbon (souche Vollum 14578) sur Gruinard Island (Ecosse) afin d'étudier la persistance des spores dans l'environnement (ONU). L'île n'était habitée que par des moutons (n=80), et quelques jours après le largage des bombes, la mort de plusieurs moutons fut observée, pour finalement aboutir à la mort de tout le troupeau quelques jours plus tard. La contamination des sols par les spores était très importante, montrant un niveau impropre à la vie sur l'île (Manchee *et al.*, 1981) (**Figure 18**). Elle fut ainsi placée en quarantaine et déclarée « incompatible » avec la vie animale et humaine. En 1986, la décontamination de l'île fut entreprise : 280 tonnes de solution de formaldéhyde et d'eau de mer ont été pulvérisées sur les 196 hectares de l'île. Un troupeau de moutons fut placé sur l'île et survécut. Ainsi, le 24 avril 1990, après 48 ans de quarantaine et 4 ans après l'application de la solution de décontamination, l'île fut déclarée saine pour être habitée.

Depuis ce test, plusieurs cas ont relaté son utilisation en tant qu'arme.



Figure 18 : Gruinard Island au Royaume-Uni (Ecosse).

(Source : <http://cs.stanford.edu/people/eroberts/courses/ww2/projects/chemical-biological-warfare/uk.htm>)

En avril et mai 1979, plusieurs cas de charbon bactérien furent observés à Sverdlovsk (ex-URSS). Ces cas touchèrent aussi bien le bétail que les humains (Meselson *et al.*, 1994). En tant que signataire du traité sur la non-utilisation et détention d'armes biologiques, l'ex-URSS déclara que l'origine de la contamination provenait de viande issue de bétail mort du charbon, avec, au total, 96 cas de charbon humains, dont 79 de type gastro-intestinal (64 morts) et 17 de type cutané (aucun mort) rapportés (Meselson *et al.*, 1994). Les cas furent observés dans une localité à proximité d'une installation militaire axant ses recherches sur le développement d'armes biologiques. Les conclusions d'une enquête menée *a posteriori* furent sans équivoque : les cas de fièvre charbonneuse observés ont en fait été causés par la libération accidentelle d'un aérosol d'environ 1 gramme de spores sèches de *B. anthracis*, causant la mort de 68 personnes (Meselson *et al.*, 1994). Cette épidémie reste à l'heure actuelle le foyer de charbon d'inhalation humain le plus important et le mieux documenté (Meselson *et al.*, 1994). Des animaux infectés par la maladie ont été retrouvés mort à une distance approximative de 50 km du site de libération de l'aérosol, alors que pour les cas humains, la distance était d'environ 4 km (Meselson *et al.*, 1994). Chez l'Homme, la période d'incubation a été observée avec une moyenne de 9 à 10 jours, avec une durée maximale d'incubation de 43 jours chez un patient (Meselson *et al.*, 1994). Cet épisode montra au monde que, malgré la signature de traités pour la non-utilisation et non-détention d'armes biologiques, une certaine expérimentation « clandestine » existait toujours.

Plus récemment, il a été découvert que la secte Aum Shinrikyo, tristement connue pour son attaque du métro de Tokyo au gaz sarin, avait tenté d'utiliser le bacille du charbon pour commettre un attentat. En juin 1993, une suspension liquide de *B. anthracis* fut utilisée sous forme de spray à proximité de Tokyo (Keim *et al.*, 2001). Les habitants de l'immeuble duquel l'aérosolisation a eu lieu se sont plaints d'odeurs désagréables. Des échantillons d'une solution liquide avaient alors été prélevés par la Police. Plusieurs années après les faits, une analyse rétrospective a été faite par P. Keim et collaborateurs. Après étude du liquide utilisé par la Secte, il s'est avéré que la souche utilisée était la souche Sterne 34F2 (souche vaccinale animale caractérisée par l'absence du plasmide pXO2

(Keim *et al.*, 2001)). Ces observations expliquent probablement l'absence de victimes (Keim *et al.*, 2001).

En septembre 2001, quelques jours seulement après les attentats du World Trade Center, une série d'attaques à l'aide d'enveloppes contaminées au bacille du charbon eut lieu (**Figure 19**). Les attaques débutèrent le 18 septembre 2001 : 5 bureaux de grands médias furent ciblés, ainsi que 2 sénateurs américains (FBI; Rasko *et al.*, 2011). La poudre utilisée pour cette première vague était assez grossière, de couleur brune, et causa principalement des cas de charbon cutané. Cependant, le 09 octobre 2001, une seconde vague d'enveloppes contaminées fut envoyée à 2 sénateurs démocrates : cette fois-ci, la poudre contenue dans les enveloppes était beaucoup plus fine et contenait beaucoup plus de spores que dans les précédentes lettres (Beecher, 2006; FBI). Ces spores étaient de qualité militaire (elles sont dites « militarisées ») et causèrent la forme la plus sévère de la maladie, le charbon d'inhalation. Les premières hypothèses se portèrent sur des laboratoires de type militaire, les connaissances et les équipements nécessaires à la production de spores de cette qualité ne pouvant provenir que de ce type de structures. Au total, 5 personnes sont mortes (charbon d'inhalation) et 17 autres survécurent lors de cet épisode appelé « Amerithrax » (Bush *et al.*, 2001; FBI; Jernigan *et al.*, 2001; Purcell *et al.*, 2006). L'enquête a conclu que la souche utilisée (RMR-1029) provenait du site de l'USAMRIID (U.S. Army Medical Research Institute for Infectious Diseases) à Fort Detrick dans le Maryland (FBI). Le principal suspect de l'affaire des lettres contaminées, Bruce Ivins, s'est suicidé en 2008.

Ces quelques événements récents ont mis en évidence l'importance de disposer d'outils d'investigation criminalistique pour conduire l'enquête qui suivrait une attaque avérée ou supposée. D'autre part, il est raisonnable de penser que, la démonstration par un pays de sa capacité à identifier l'éventuel agresseur et ses soutiens, aura un effet dissuasif sur les éventuels promoteurs de telles agressions.



Figure 19: Lettres contaminées par le bacille du charbon lors de l'épisode Amerithrax de septembre 2001 (Source : FBI)

ANALYSE BIBLIOGRAPHIQUE

Chapitre 2 :

Détection et typage de *Bacillus anthracis*

1. Méthodes de détection et de typage moléculaire de *B. anthracis*

B. anthracis forme une lignée très monomorphe au sein du groupe *B. cereus*. Il présente très peu de variation génétique dans sa population (Pearson *et al.*, 2004; Van Ert *et al.*, 2007a). Les souches les plus divergentes de *B. anthracis* présentent une homologie inter-souches de plus de 99.99 % au niveau nucléotidique (Pearson *et al.*, 2004; Rasko *et al.*, 2011; Van Ert *et al.*, 2007a). Cette homogénéité remarquable est principalement liée au cycle de vie particulier de ce pathogène : *B. anthracis* passe la majorité de sa vie sous forme de spores dormantes. Les échanges de matériel génétique ainsi que l'accumulation de mutations ponctuelles au sein de cette espèce sont donc extrêmement réduits. De ce fait, il est très difficile de différencier deux souches entre elles en utilisant des méthodes classiques de diagnostic telles que la bactériologie ou le typage par ARNr. Les méthodes sérologiques et biochimiques permettent d'identifier le bacille du charbon mais ne sont pas totalement fiables lorsqu'il s'agit de distinguer les différents membres du groupe *B. cereus* (Henderson *et al.*, 1994).

1.1. Détection

Traditionnellement, l'identification de *B. anthracis* est réalisée en utilisant les différences phénotypiques entre *B. anthracis* et les autres membres du groupe *B. cereus* : déficit de mobilité et d'hémolyse, sensibilité à la pénicilline, morphologie des colonies,... (Hoffmaster *et al.*, 2002). Cependant, toutes ces méthodes nécessitent au moins 24h pour obtenir la confirmation des résultats. Il est important de bénéficier d'outils rapide de détection lors de cas épidémiques ou importés (Hoffmaster *et al.*, 2002).

1.1.1. Phages appliqués au diagnostic

Les phages peuvent être utilisés pour l'identification de *B. anthracis* (Rao *et al.*, 2010). Le phage gamma, spécifique de *B. anthracis* a été découvert par Brown et Cherry en 1955 (OMS, 2008). Il a permis la lyse de 41 souches de *B. anthracis* (sur 41) mais d'aucune souche de *B. cereus* (89 souches testées) ou souches d'espèces différentes de *Bacillus* (134 souches testées) (OMS, 2008). On parle de phages diagnostics désormais, car le phage gamma fait partie d'une famille élargie de phages spécifiques du bacille du charbon (OMS, 2008). Cependant, cette méthode n'est pas assez rapide.

1.1.2. Réaction en Chaîne par Polymérase (PCR)

Afin de détecter rapidement la présence de *B. anthracis* dans un échantillon, la méthode utilisée est la PCR. Les études d'homologie ADN ont montré que *B. anthracis* est extrêmement proche de *B. cereus*, *B. thuringiensis* et *B. mycoides* que ce soit par son contenu en G-C (entre 31 et 34%) ou au niveau de l'ADN ribosomique. Les séquences n'ont montré que quelques différences mineures (Henderson *et al.*, 1994).

La PCR permet, via une réaction d'amplification génique, de confirmer la présence de la bactérie en ciblant un marqueur chromosomique, ainsi que les gènes codant pour les toxines et la capsule (sur pXO1 et pXO2, respectivement) (Hoffmaster *et al.*, 2002; OMS, 2008). Des amorces ciblant les gènes *pag* (PA)(Price *et al.*, 1999) et *lef* (LF) sont décrites. D'autres gènes peuvent être ciblés (gènes *capA*, *capB* et *capC*) (OMS, 2008) mais sont susceptibles d'induire de faux positifs (OMS, 2008).

Les résultats obtenus avec les gènes cibles suivants ont été décrits comme corrects: *pagA* sur pXO1, *capA* sur pXO2 et *saspB* sur le chromosome (Hoffmaster *et al.*, 2002; Marston *et al.*, 2006). Les gènes

sspE, *cya* et *capB* ont également montré de bons résultats (Janse *et al.*, 2010). Cependant, beaucoup de marqueurs chromosomiques décrits comme spécifiques de *B. anthracis* (*BA813*, *rpoB*, *gyrA*, *gyB*, *saspB*, *plcR*, *BA5345*) sont également retrouvés chez certaines souches de *B. cereus* (Antwerpen *et al.*, 2008; Hurtle *et al.*, 2004). Il est important de pouvoir cibler un marqueur chromosomique, car bien que les plasmides permettent de distinguer *B. anthracis* des autres espèces du groupe *B. cereus*, certaines souches peuvent avoir perdu un ou les deux plasmides. A l'inverse, certaines souches ayant acquis ces plasmides peuvent être détectées grâce aux marqueurs plasmidiques. La combinaison de 3 marqueurs (1 chromosomique et 2 plasmidiques) permet une bonne détection de *B. anthracis*.

2. Typage moléculaire

Le typage moléculaire permet d'identifier avec précision différents organismes au sein d'une espèce. Les méthodes employées permettent ainsi de caractériser les différentes souches d'une espèce et de les comparer entre elles. Ces méthodes sont des outils indispensables lors d'études épidémiologiques, permettant d'investiguer la source d'infection ou les voies de transmission impliquées.

Les techniques de typage basées sur une étude du phénotype (sérotypes, biotypes, phages, antibiotypes) sont utilisées depuis de longues années (Sabat *et al.*, 2013). Cependant, des méthodes plus récentes ont émergé et permettent d'étudier les relations entre isolats au niveau moléculaire (Marston *et al.*, 2006; Sabat *et al.*, 2013). Une méthode telle que l'électrophorèse en champs pulsés (PFGE) présente un pouvoir discriminant intéressant (Zhong *et al.*, 2007), mais n'est pas facilement reproductible et n'est donc pas une technique de choix pour le typage moléculaire à grande échelle. Une méthode doit être simple, robuste et facilement transférable et standardisable d'un laboratoire à un autre.

Les analyses de marqueurs moléculaires permettent d'identifier des variations génétiques informatives (Keim *et al.*, 1997). Plusieurs études de caractérisation moléculaire et d'empreintes ADN ont préalablement été effectuées chez *B. anthracis*, montrant une absence de diversité au sein de l'espèce (Keim *et al.*, 1997). Ainsi, le séquençage de gènes a révélé une identité exacte pour l'ARN 16S, la région intergénique 16S-23S ainsi que pour la région intergénique *gyrA-gyrB* (Keim *et al.*, 1997).

Plusieurs méthodes de typage, ayant pour but de différencier les souches entre elles, ont été testées sur ce pathogène très monomorphe.

2.1. Polymorphisme de longueur des fragments amplifiés (AFLP)

Les marqueurs AFLP, basés sur des polymorphismes de fragments de restriction, sont détectés par PCR sélective (Keim *et al.*, 1997; Sabat *et al.*, 2013; Vos *et al.*, 1995). Cette méthode permet de tester plusieurs loci polymorphes sur un même gel d'agarose.

Cette approche a permis l'apport d'un niveau de discrimination assez important comparé à ce qui était obtenu auparavant au sein de cette espèce clonale. Paul Keim et ses collaborateurs ont utilisé 31 régions polymorphes au sein du génome de *B. anthracis*, permettant de définir deux grands groupes au sein de l'espèce *B. anthracis* : cette sub-division majeure a permis de distinguer un groupe A plus important en nombre de souches et un groupe B, avec quelques subdivisions observées au sein de ces groupes (Keim *et al.*, 1997) (**Figure 20**). Cependant, sur un total de 1221 fragments observés, 1184 (97%) étaient monomorphes, alors que près de 60 % des fragments totaux différaient avec les espèces proches que sont *B. cereus* et *B. thuringiensis* (Keim *et al.*, 1997).

Avec seulement 30 marqueurs polymorphes sur plus de 1000 fragments d'ADN, le pouvoir discriminant de l'AFLP reste assez faible pour l'espèce *B. anthracis* (Keim *et al.*, 2000).

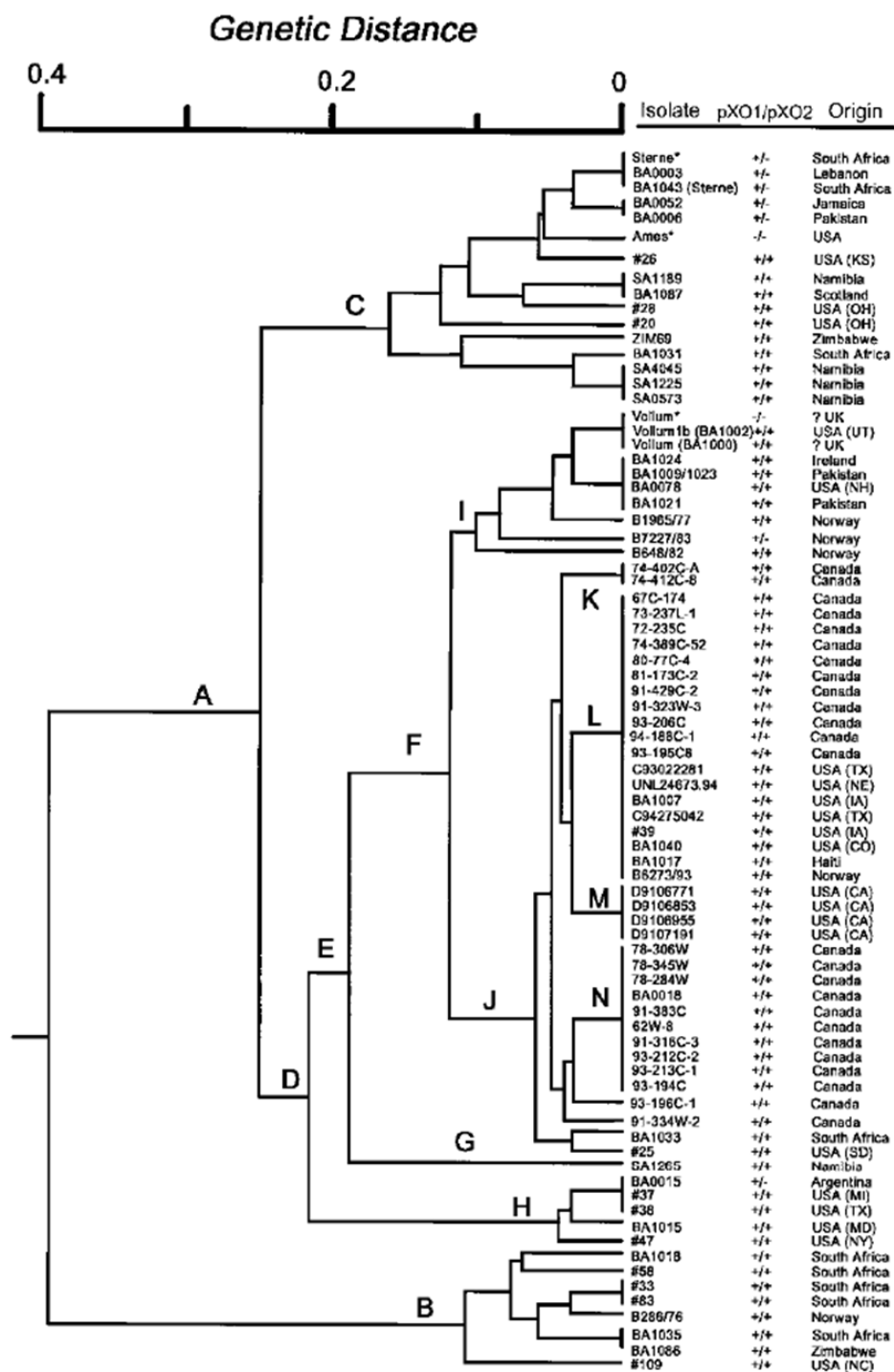


Figure 20 : Dendrogramme UPGMA des souches de *B. anthracis* obtenu par la méthode AFLP. Deux grandes lignées A et B sont discriminées (Keim *et al.*, 1997)

2.2. Multiple Locus Variable Number of Tandem Repeats Analysis (MLVA)

Une autre approche basée sur les séquences répétées dans le génome a vu le jour à la fin des années 2000. L'essor du séquençage du génome a montré l'existence de séquences répétées chez toutes les espèces bactériennes analysées (Vergnaud et Pourcel, 2009). La taille de ces répétitions peut varier d'un seul nucléotide à plusieurs centaines (Vergnaud et Pourcel, 2009). Ces régions répétées sont retrouvées dans des régions codantes ou intergéniques et pourraient jouer un rôle dans l'adaptation à l'environnement (Vergnaud et Pourcel, 2009). L'étude des séquences répétées en tandem, ou VNTR (Variable Number of Tandem Repeats) est appelée MLVA (Multiple Locus VNTR Analysis). Les VNTR ou minisatellites sont habituellement définis comme des séquences répétées de 9 à 100 pb représentant un motif allant d'une dizaine à plusieurs milliers de paires de bases. La diversité génétique observée au niveau des loci VNTR est, en général, liée à des erreurs de réplication qui résultent d'un dérapage (« slipped strand mispairing ») dans des zones instables du génome, provoquant une amplification ou une réduction de ces zones répétées. La fréquence d'occurrence est de l'ordre de 10^{-5} à 10^{-4} chez *B. anthracis* pour les loci les plus instables (Thierry *et al.*, 2014). Développée pour la première fois sur *Haemophilus influenzae* par Van Belkum et collaborateurs (Van Belkum *et al.*, 1997), la MLVA a depuis été développée pour de nombreuses espèces bactériennes (*Mycobacterium tuberculosis* (Frothingham et Meeker-O'Connell, 1998), *Yersinia pestis* (Adair *et al.*, 2000), ...). Pour la fièvre charbonneuse, la première étude eut lieu en 2000 par l'équipe de Paul Keim aux USA (Keim *et al.*, 2000).

La MLVA présente une plus grande capacité de discrimination que l'AFLP ainsi qu'une meilleure reproductibilité (Keim *et al.*, 2000). Une PCR avec 2 amorces dans les séquences flanquantes à la répétition permet d'amplifier la région répétée. Pour les séquences de plus petite taille, une électrophorèse capillaire est requise pour distinguer avec précision les différences de tailles des amplicons, la taille étant liée au nombre d'unités de répétitions. Le choix d'une méthode appropriée de détection doit donc être fait, à savoir, l'utilisation d'un gel d'agarose ou d'une électrophorèse capillaire.

La première étude MLVA chez *B. anthracis* fut réalisée à partir de 8 marqueurs VNTR, 6 chromosomiques et 2 plasmidiques (Keim *et al.*, 2000). Sur les 419 souches mondiales testées, 89 génotypes ont été identifiés (Keim *et al.*, 2000) (**Figure 21**). Les deux lignées clonales majeures identifiées par AFLP ont ainsi été retrouvées (lignées A et B), mais avec un bien meilleur pouvoir discriminant au sein de ces lignées qu'avec l'AFLP (Keim *et al.*, 2000). Ce panel a été largement utilisé pour génotyper différentes collections à travers le monde (Beyer *et al.*, 2012; Durmaz *et al.*, 2012; Fasanella *et al.*, 2005; Fouet *et al.*, 2002; Gierczyński *et al.*, 2004; Maho *et al.*, 2006; Merabishvili *et al.*, 2006; Patra *et al.*, 1998; Pilo *et al.*, 2008; Pilo *et al.*, 2011; Smith *et al.*, 2000; Sue *et al.*, 2007).

Dans le but d'améliorer le pouvoir discriminant de la méthode, de nouveaux VNTR ont été identifiés et ajoutés au set des 8 marqueurs initialement proposés. Un schéma à 20 VNTR a été proposé (Le Flèche *et al.*, 2001), puis un schéma à 15 VNTR (Keim *et al.*, 2004; Van Ert *et al.*, 2007a), utilisé notamment en Bulgarie (Antwerpen *et al.*, 2011). Enfin, un schéma à 25 marqueurs a été développé (Lista *et al.*, 2006). Il a été utilisé au Japon (Okutani *et al.*, 2010). Plus le nombre de marqueurs utilisés augmente, plus on obtient de génotypes : le pouvoir discriminant augmente.

Au total, un panel complet de 31 loci VNTR est disponible et a été utilisé pour tester la diversité de *B. anthracis* observée en Namibie et en France (Beyer *et al.*, 2012; Thierry *et al.*, 2014). Le **Tableau 3** récapitule les différents schémas MLVA utilisés ainsi que les VNTR concernés.

Cette méthode a longtemps été la référence pour le génotypage de *B. anthracis*. Elle est simple à utiliser (plusieurs panels selon la discrimination souhaitée (**Tableau 3**), la taille des loci VNTR à utiliser) et permet d'obtenir un pouvoir discriminant satisfaisant.

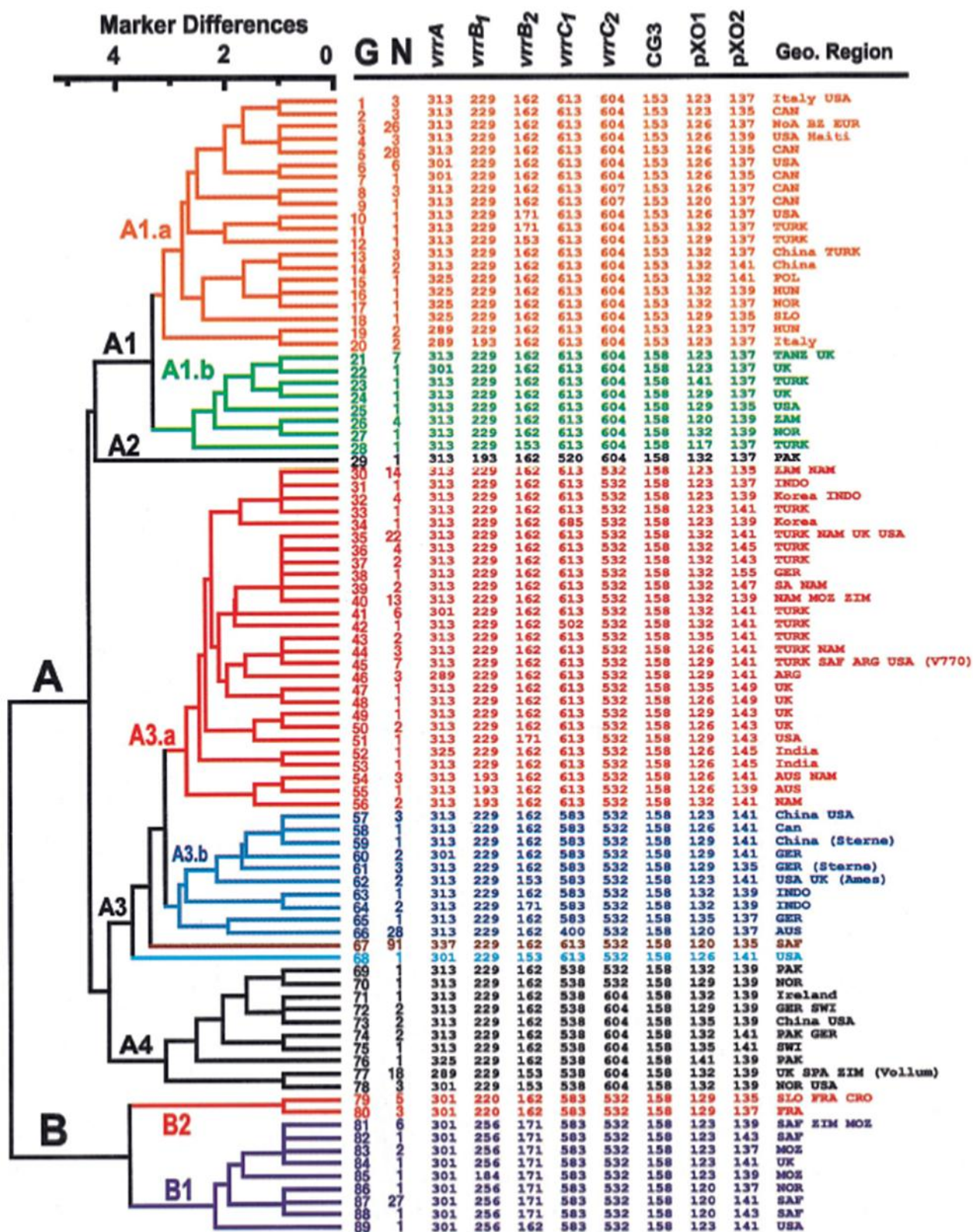


Figure 21: Dendrogramme UPGMA obtenu à partir de la MLVA8 sur 419 souches : 89 géotypes sont identifiés. Les 2 grandes lignées obtenues avec l'AFLP sont retrouvées ici, mais le niveau de discrimination obtenu est supérieur (Keim *et al.*, 2000)

Tableau 3 : Description des différents schémas MLVA utilisés chez *B. anthracis* ainsi que les VNTR concernés

Panel	MLVA8	MLVA15	MLVA20	MLVA25	MLVA31
Liste des marqueurs	pXO1	pXO1		pXO1	pXO1
	pXO2	pXO2		pXO2	pXO2
	<i>vrrA</i>	<i>vrrA</i>	<i>vrrA</i>	<i>vrrA</i>	<i>vrrA</i>
	<i>vrkB1</i>	<i>vrkB1</i>	<i>vrkB1</i>	<i>vrkB1</i>	<i>vrkB1</i>
	<i>vrkB2</i>	<i>vrkB2</i>	<i>vrkB2</i>	<i>vrkB2</i>	<i>vrkB2</i>
	<i>vrnC1</i>	<i>vrnC1</i>	<i>vrnC1</i>	<i>vrnC1</i>	<i>vrnC1</i>
	<i>vrnC2</i>	<i>vrnC2</i>	<i>vrnC2</i>	<i>vrnC2</i>	<i>vrnC2</i>
	CG3	CG3	CG3	CG3	CG3
		BaVNTR32 (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
Références	(Keim <i>et al.</i> , 2000)	(Keim <i>et al.</i> , 2004; Van Ert <i>et al.</i> , 2007a)	(Le Flèche <i>et al.</i> , 2001)	(Lista <i>et al.</i> , 2006)	(Beyer <i>et al.</i> , 2012; Thierry <i>et al.</i> , 2014)

2.3. Single Nucleotide Repeats (SNR)

Un type particulier de VNTR possédant un taux élevé de mutation permet de différencier des souches d'un même foyer présentant une diversité génétique extrêmement faible et un même génotype MLVA: ce sont les Single Nucleotide Repeats (SNR) (Garofolo *et al.*, 2010; Keim *et al.*, 2004; Kenefic *et al.*, 2008a; Kenefic *et al.*, 2008b; Stratilo *et al.*, 2006). Ces zones hyper-mutables ou « Hot spots » (taux de mutation estimé à 6×10^{-4} mutations par génération) (Keim *et al.*, 2004) sont constituées de répétitions d'un nucléotide, qui diffère en longueur entre les isolats étudiés (Stratilo *et al.*, 2006). Un panel de 22 SNR a été proposé (Stratilo *et al.*, 2006).

Ces marqueurs ne sont cependant pas utilisables pour des études de relations phylogénétiques entre isolats divers : les résultats seraient alors assimilables à de l'homoplasie (similitude d'un état de caractère qui ne provient pas d'un ancêtre commun, contrairement à l'homologie) (Keim *et al.*, 2004).

2.4. Single Nucleotide Polymorphisms (SNP)

Les Single Nucleotide Polymorphisms (SNP) sont générés par des substitutions de nucléotides, issues probablement d'erreurs de réplication de l'ADN qui ne sont pas réparées (Keim *et al.*, 2004). Ces événements sont assez rares chez *B. anthracis*. Ils sont observés à des taux approximatifs de 10^{-10} mutations par nucléotide et par génération (Keim *et al.*, 2004). Bien qu'ils n'apparaissent qu'à des fréquences faibles, les SNP peuvent être identifiés via des méthodes de recherche intensive à haut débit. Les nouvelles technologies de séquençage du génome complet permettent l'identification de milliers de SNP. Leur évolution suggère une origine unique, et ces marqueurs sont extrêmement stables (très faible taux de réversion) (Keim *et al.*, 2004), ce qui en fait des marqueurs diagnostics importants en phylogénie (Pilo *et al.*, 2011).

Un set de 13 SNP canoniques (canSNP) a été établi, permettant de distinguer 3 lignées majeures A, B et C au sein de l'espèce *B. anthracis*, ainsi que plusieurs sous-divisions en 7 sous-lignées (A.Br.Ames, A.Br.Aust94, A.Br.Vollum, A.Br.WNA, B.Br.KrugerB, B.Br.CNEVA et C.Br.A1055) et 5 groupes (A.Br.001/002, A.Br.003/004, A.Br.005/006, A.Br.008/009 et B.Br.001/002) (Van Ert *et al.*, 2007a) (**Figure 22**). Un 14^{ème} canSNP a intégré le panel précédent. Il permet de subdiviser la sous-lignée Transeurasienne (TEA, A.Br.008/009), en 2 groupes : A.Br.008/011 et A.Br.011/009 (Marston *et al.*, 2011) (**Figure 22**). Ces marqueurs canoniques permettent d'identifier un moment phylogénétique particulier dans l'histoire évolutive d'une espèce ou d'un organisme. Les études par typage moléculaire ont montré que certains génotypes particuliers sont retrouvés uniquement dans certaines régions géographiques. Cette relation génotype-localisation géographique peut s'expliquer par la forte clonalité et la récente apparition de l'espèce *B. anthracis* (Price *et al.*, 2012). En sélectionnant un nombre restreint de SNP représentatifs de *B. anthracis*, la position phylogénétique de n'importe quel isolat peut être identifiée (Van Ert *et al.*, 2007a).

Ces canSNP ont été utilisés par la suite pour génotyper un grand nombre de collections à travers le monde (Antwerpen *et al.*, 2011; Derzelle *et al.*, 2011a; Garofolo *et al.*, 2011; Khmaladze *et al.*, 2014; Kuroda *et al.*, 2010; Okinaka *et al.*, 2008; Simonson *et al.*, 2009).

L'étude détaillée de certaines sous-lignées a permis d'identifier de nouveaux canSNP. Pour la sous-lignée A.Br.WNA, sous-lignée majoritaire en Amérique du Nord, 10 SNP spécifiques ont été décrits, permettant de discriminer les souches présentes au Canada et aux USA en six sous-groupes (387 isolats testés) (Kenefic *et al.*, 2009). Le même type d'étude a été mené pour la sous-lignée A.Br.Ames en décrivant 6 SNP spécifiques (Van Ert *et al.*, 2007b), ainsi que pour d'autres sous-groupes (Birdsell *et al.*, 2012; Khmaladze *et al.*, 2014). La liste de ces SNP est présentée dans le **Tableau 4**.

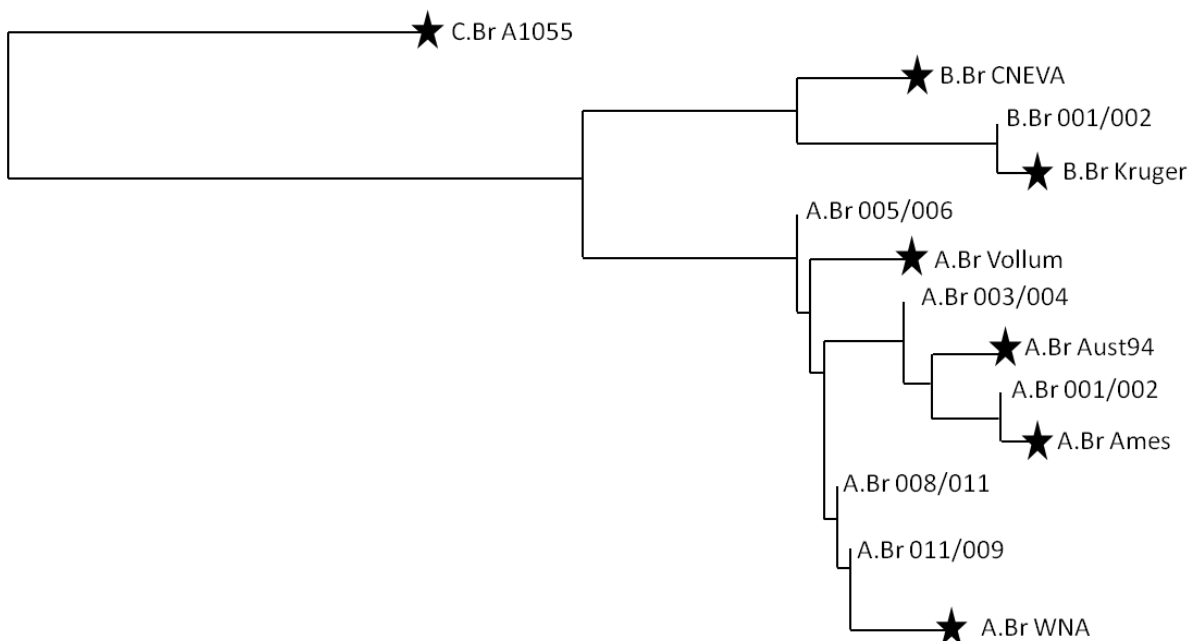


Figure 22 : Dendrogramme représentant la phylogénie de *B. anthracis* à partir de l'analyse de 14 canSNP. Adapté d'après (Marston *et al.*, 2011; Van Ert *et al.*, 2007a)

Les SNP représentent donc des marqueurs idéaux pour définir des groupes phylogénétiques. Cette approche est rapidement devenue la méthode de référence pour typer une souche de *B. anthracis* et la positionner au sein de l'espèce.

Beaucoup d'études ont utilisé ce type d'approches, y compris *a posteriori*, comme pour le cas de Sverdlovsk en ex-URSS (Okinaka *et al.*, 2008). Facilement identifiables *in silico* (Brodzik et Francoeur, 2011), les SNP peuvent être déterminés par différentes méthodes de biologie moléculaire basées sur la PCR. On peut citer la méthode Mismatch amplification mutation assay (MAMA) (Birdsell *et al.*, 2012; Easterday *et al.*, 2005; Li *et al.*, 2004; Price *et al.*, 2010) ou la méthode de Fusion à haute résolution (HRM) appliquée à la suite d'une étape de PCR en temps réel (Derzelle *et al.*, 2011b; Erali *et al.*, 2008; Fortini *et al.*, 2007).

Tableau 4 : Synthèse des canSNP décrits dans la littérature

Nom du canSNP	Position *	Groupe ciblé	Référence
A.Br.001	182106	A.Br.Ames	(Van Ert <i>et al.</i> , 2007a)
A.Br.002	947760	A.Br.Ames, A.Br.001/002	(Van Ert <i>et al.</i> , 2007a)
A.Br.003	1493280	A.Br.Ames, A.Br.001/002, A.Br.Aust94	(Van Ert <i>et al.</i> , 2007a)
A.Br.004	3600786	A.Br.Ames, A.Br.001/002, A.Br.Aust94, A.Br.003/004	(Van Ert <i>et al.</i> , 2007a)
A.Br.005	3842864	A.Br.005/006, B.Br.Kruger, B.Br.001/002, B.Br.CNEVA, C.Br.A1055	(Birdsell <i>et al.</i> , 2012)
A.Br.006	162509	Branche A	(Van Ert <i>et al.</i> , 2007a)
A.Br.007	266439	A.Br.Vollum	(Van Ert <i>et al.</i> , 2007a)
A.Br.008	3947375	A.Br.008/009, A.Br.WNA	(Van Ert <i>et al.</i> , 2007a)
A.Br.008/011 group	2034709	A.Br.008/011	(Birdsell <i>et al.</i> , 2012)
A.Br.008/011 group	2314544	A.Br.008/011	(Birdsell <i>et al.</i> , 2012)
A.Br.009	2589947	A.Br.WNA	(Van Ert <i>et al.</i> , 2007a)
A.Br.011	2552486	A.Br.011/009, A.Br.WNA	(Marston <i>et al.</i> , 2011)
A.Br.011	3692595	A.Br.011/009, A.Br.WNA	(Marston <i>et al.</i> , 2011)
A.Br.026	3640599	A.Br.Aust94	(Khmaladze <i>et al.</i> , 2014)
A.Br.027	4355524	A.Br.Aust94	(Khmaladze <i>et al.</i> , 2014)
A.Br.028	791256	A.Br.Aust94	(Khmaladze <i>et al.</i> , 2014)
A.Br.029	3960657	A.Br.Aust94	(Khmaladze <i>et al.</i> , 2014)
A.Br.030	3528668	A.Br.Aust94	(Khmaladze <i>et al.</i> , 2014)
A.Br.031	3018054	A.Br.Aust94	(Khmaladze <i>et al.</i> , 2014)
A.Br.032	3520170	A.Br.Aust94	(Khmaladze <i>et al.</i> , 2014)
A.Br.033	3610151	A.Br.Aust94	(Khmaladze <i>et al.</i> , 2014)
A/B.Br.001	3698013	C.Br.A1055	(Van Ert <i>et al.</i> , 2007a)
B.Br.001	1455402	B.Br.Kruger	(Van Ert <i>et al.</i> , 2007a)
B.Br.002	1056740	B.Br.Kruger, B.Br.001/002	(Van Ert <i>et al.</i> , 2007a)
B.Br.003	1494392	B.Br.Kruger, B.Br.001/002, B.Br.CNEVA	(Van Ert <i>et al.</i> , 2007a)
B.Br.004	69952	B.Br.CNEVA	(Van Ert <i>et al.</i> , 2007a)
Branch 1_26	4624132	A.Br.Ames	(Van Ert <i>et al.</i> , 2007b)
Branch 1_28	4929312	A.Br.Ames	(Van Ert <i>et al.</i> , 2007b)
Branch 1_31	2749671	A.Br.Ames	(Van Ert <i>et al.</i> , 2007b)
Branch 1_7	433277	A.Br.Ames	(Van Ert <i>et al.</i> , 2007b)
C.Br.001	2521201	C.Br.A1055	(Birdsell <i>et al.</i> , 2012)
pagA 981	2784	A.Br.008/009	(Okinaka <i>et al.</i> , 2008)
plcR	5081429	<i>B. anthracis</i>	(Birdsell <i>et al.</i> , 2012)
PS-1 (pXO1)	7452	A.Br.Ames	(Van Ert <i>et al.</i> , 2007b)
PS-52 (pXO2)	72924	A.Br.Ames	(Van Ert <i>et al.</i> , 2007b)
SNP1053700	1053700	A.Br.008/011 (souches Héroïne)	(Marston <i>et al.</i> , 2011)
SNP1118831	1118831	A.Br.008/011 (souches Héroïne)	(Marston <i>et al.</i> , 2011)
SNP1173928	1173928	A.Br.008/011 (souches Héroïne)	(Marston <i>et al.</i> , 2011)
SNP1530761	1530761	A.Br.008/011 (souches Héroïne)	(Marston <i>et al.</i> , 2011)
SNP1967560	1967560	A.Br.008/011 (souches Héroïne)	(Marston <i>et al.</i> , 2011)
SNP3287006	3287006	A.Br.008/011 (souches Héroïne)	(Marston <i>et al.</i> , 2011)
SNP3836105	3836105	A.Br.008/011 (souches Héroïne)	(Marston <i>et al.</i> , 2011)
SNP5013862	5013862	A.Br.008/011 (souches Héroïne)	(Marston <i>et al.</i> , 2011)
wna1141774	1141774	A.Br.WNA	(Kenefic <i>et al.</i> , 2009)
wna237471	237471	A.Br.WNA	(Kenefic <i>et al.</i> , 2009)
wna2994131	2994131	A.Br.WNA	(Kenefic <i>et al.</i> , 2009)
wna3368524	3368524	A.Br.WNA	(Kenefic <i>et al.</i> , 2009)
wna3631093	3631093	A.Br.WNA	(Kenefic <i>et al.</i> , 2009)
wna3682247	3682247	A.Br.WNA	(Kenefic <i>et al.</i> , 2009)
wna3732539	3732539	A.Br.WNA	(Kenefic <i>et al.</i> , 2009)
wna3774186	3774186	A.Br.WNA	(Kenefic <i>et al.</i> , 2009)
wna4461234	4461234	A.Br.WNA	(Kenefic <i>et al.</i> , 2009)
wna4718500	4718500	A.Br.WNA	(Kenefic <i>et al.</i> , 2009)

(* : Position des SNP sur le génome référence Ames Ancestor, pXO1 Ames Ancestor et pXO2 Ames Ancestor)

2.5. Essais progressifs résolutifs hiérarchiques basés sur les acides nucléiques (PHRANA)

Paul Keim et ses collaborateurs (2004) ont proposé une approche de typage basée sur l'utilisation successive de 3 types de marqueurs moléculaires : les canSNP, les VNTR avec la méthode MLVA et les SNR. Cette méthode est appelée PHRANA (Progressive Hierarchical Resolving Assays using Nucleic Acids).

Les marqueurs moléculaires tels que les VNTR, les SNR ou les SNP présentent des taux de mutations différents. Si le taux de mutation d'un marqueur est très faible, les polymorphismes au sein de ce marqueur ne sont identifiés que parmi des populations hautement diverses alors que les populations moins diverses apparaîtront de manière monomorphe (Keim *et al.*, 2004). A l'inverse, si les taux de mutation sont élevés, la diversité augmente rapidement sur les loci concernés, permettant de différencier les souches les plus proches : c'est le cas des SNR (Keim *et al.*, 2004). Cependant, un taux élevé de mutation est corrélé à une augmentation des similitudes de type homoplasie, qui sont retrouvées généralement dans des populations à diversité génétique importante. Ainsi, si des marqueurs hautement polymorphes sont utilisés pour l'analyse de populations diverses, de mauvais assignements de groupes sont inévitables.

Dans un premier temps, la collection de souches doit être testée via les canSNP. Les souches présentant le même profil sont ensuite typées via la MLVA. Enfin, les souches possédant le même génotype sont typées via les SNR, obtenant ainsi le plus haut niveau de discrimination possible avec ces 3 méthodes (**Figure 23**). Cette approche permet de distinguer beaucoup plus de génotypes qu'avec les méthodes précédemment décrites et prises séparément.

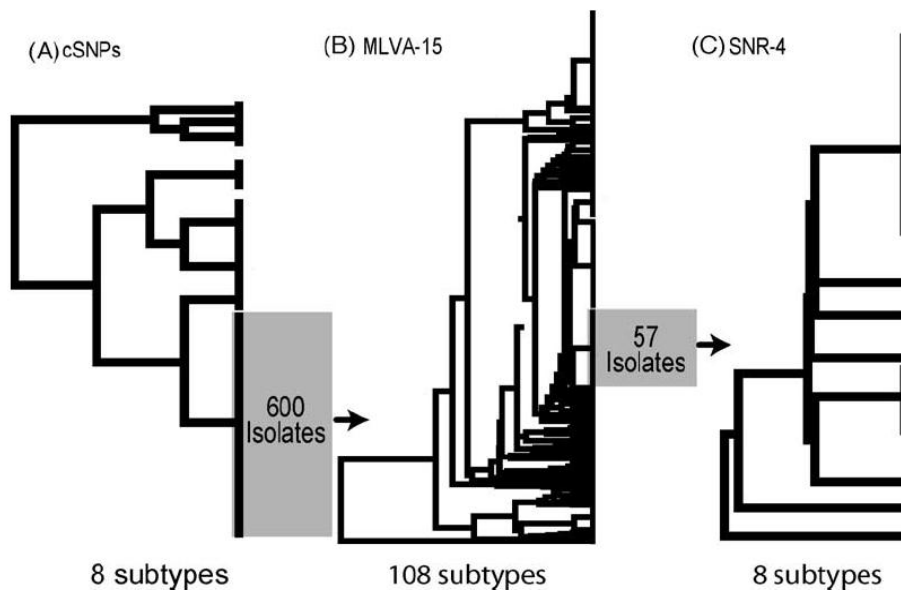


Figure 23 : Méthode PHRANA utilisée par Paul Keim et ses collaborateurs et permettant d'obtenir un très bon pouvoir discriminant en combinant 3 approches : les canSNP, la MLVA-15 et les SNR-4 (Keim *et al.*, 2004)

3. Phylogénie des souches de *B. anthracis* à travers le monde

L'utilisation du typage par canSNP ainsi que les comparaisons sur génomes complets de plusieurs souches d'origines variées ont permis la reconstruction de la phylogénie de *B. anthracis* de manière extrêmement précise (Pearson *et al.*, 2004; Van Ert *et al.*, 2007a). Trois lignées majeures (A, B et C) ainsi que 13 sous-lignées ou groupes ont été définis (**Figure 24**).

La lignée A comprend la majorité des cas de fièvre charbonneuse recensés à travers le monde (Keim *et al.*, 2000). Les 8 groupes de cette lignée sont répartis de manière globale dans le monde et il arrive qu'une coexistence dans une même région soit observée, bien que très souvent, un groupe reste majoritaire.

La sous-lignée TEA (A.Br.008/011-A.Br.011/009) est la plus représentée dans le monde, concernant la majorité des sous-lignées relevées en Europe, en Chine, en Russie, au Kazakhstan et en Afrique de l'Ouest (Derzelle et Thierry, 2013; Garofolo *et al.*, 2011; Price *et al.*, 2012; Van Ert *et al.*, 2007a).

La plupart des isolats retrouvés en Chine et en Asie correspondent au groupe A.Br.001/002. La sous-lignée A.Br.Ames, introduite au Texas (USA), a émergé de ce groupe issu de Mongolie (Simonson *et al.*, 2009).

L'Amérique abrite les groupes A.Br.WNA et A.Br.003/004 (Nord et Sud respectivement), alors que la sous-lignée A.Br.005/006 est retrouvée majoritairement en Afrique.

Le sud de l'Afrique ainsi que certains pays d'Asie (notamment le Pakistan et l'Afghanistan) comprennent majoritairement la sous-lignée A.Br.Vollum, alors que les collections de pays tels que la Turquie, l'Inde, la Namibie, la Géorgie ou l'Australie sont dominées par le groupe A.Br.Aust94 (Beyer *et al.*, 2012; Durmaz *et al.*, 2012; Kenefic *et al.*, 2009; Khmaladze *et al.*, 2014; Simonson *et al.*, 2009; Van Ert *et al.*, 2007a).

La lignée B est séparée en 2 clades distincts, et géographiquement implantés dans des régions spécifiques. Le clade B1 (incluant les groupes B.Br.001/002 et B.Br.Kruger) est retrouvé en Afrique du Sud, plus particulièrement dans le Parc National Kruger, où il coexiste avec certaines souches A.Br.Vollum (Smith *et al.*, 2000; Van Ert *et al.*, 2007a). La sous-lignée B.Br.001/002 a récemment été identifiée en Corée et au Bhutan (Thapa *et al.*, 2014).

Le clade B2 (B.Br.CNEVA) est exclusivement retrouvé en Europe (Antwerpen *et al.*, 2012; Derzelle *et al.*, 2011a; Garofolo *et al.*, 2011; Gierczyński *et al.*, 2004; Pilo *et al.*, 2008; Steiner *et al.*, 2013; Van Ert *et al.*, 2007a), à l'exception d'un cas recensé en Californie (Van Ert *et al.*, 2007a).

La lignée C (C.Br.A1055) est quant à elle beaucoup plus anecdotique puisqu'elle ne comporte à l'heure actuelle que 3 isolats d'origine inconnue (Sue *et al.*, 2007; Van Ert *et al.*, 2007a).

En Europe, la population de *B. anthracis* est très diversifiée, pouvant aller jusqu'à la coexistence de 6 groupes phylogénétiques au sein d'un même pays (**Figure 24**). Trois sous-lignées sont plus représentées que les autres : la TEA (A.Br.008/011-A.Br.011/009), A.Br.001/002 et B.Br.CNEVA. Cette dernière est implantée en Europe centrale, allant des Pyrénées (Sud de la France) à la Pologne (nord-est) en passant par les Alpes italiennes, la Suisse, l'Allemagne, la Croatie et la Slovaquie. Dans ces pays, une coexistence avec la sous-lignée TEA est généralement observée, même si, au sein d'un même pays, elles peuvent être localisées dans plusieurs régions différentes. La sous-lignée TEA a été isolée dans la plupart des pays européens, notamment dans le sud et l'est de l'Europe. Elle représente le groupe majoritaire en Italie, Bulgarie, Hongrie et Albanie (Antwerpen *et al.*, 2011; Fasanella *et al.*, 2005; Garofolo *et al.*, 2011; Van Ert *et al.*, 2007a). Le groupe A.Br.001/002 est retrouvé en Europe du Nord (Allemagne, Danemark, Belgique, Pays-Bas, Royaume-Uni) mais également en France (nord-est) (Calavas *et al.*, 2008).

Les autres groupes retrouvés en Europe incluent A.Br.005/006, A.Br.Aust94, A.Br.Vollum, B.Br.001/002, A.Br.Ames et A.Br.003/004 mais sont beaucoup plus minoritaires. Cette diversité

observée est due à l'activité des hommes telle que le commerce international ou les anciennes colonies.

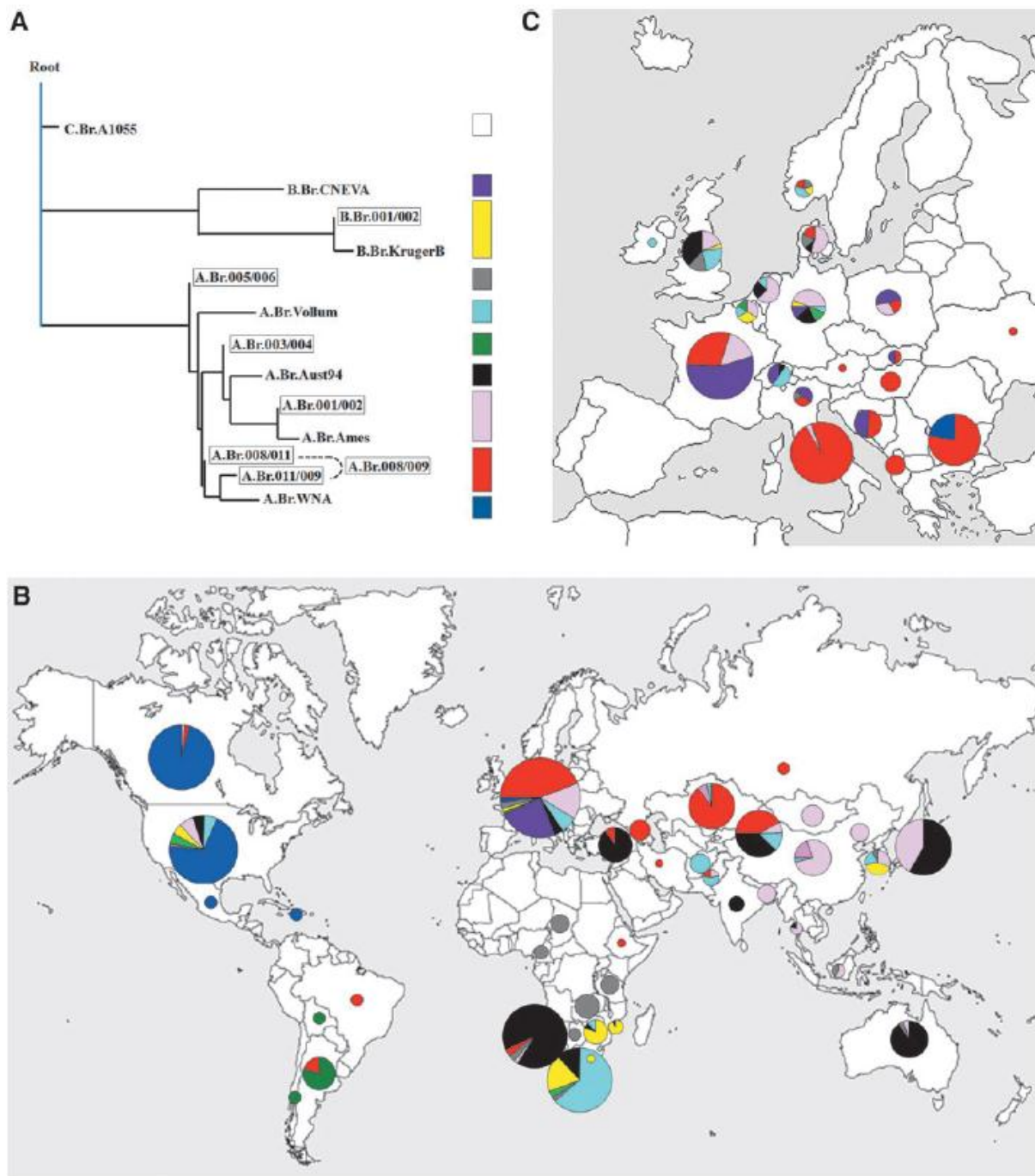


Figure 24: Répartition de *B. anthracis* dans le monde en fonction des génotypes identifiés par canSNP

- A. Phylogénie des groupes majeurs de *B. anthracis* (Marston *et al.*, 2011; Van Ert *et al.*, 2007a)
- B. Distribution mondiale des sous-lignées et groupes de *B. anthracis*
- C. Distribution européenne des sous-lignées et groupes de *B. anthracis*

(Figure extraite de (Derzelle et Thierry, 2013))

L'utilisation des méthodes de typage combinées à la phylogénie d'une espèce bactérienne peut s'avérer critique lorsqu'il s'agit de déterminer avec certitude qu'une souche particulière est à l'origine d'un foyer de FC ou d'un cas humain. Un des exemples les plus récents de cette utilisation concerne les cas de charbon observés chez les consommateurs d'héroïne (plusieurs cas recensés en Europe depuis le début des années 2000), où les souches impliquées appartiennent au groupe A.Br.008/011 (Price *et al.*, 2012). La principale préoccupation des enquêteurs était de définir l'origine de la contamination de l'héroïne. Cette dernière provenant d'Afghanistan, la première hypothèse fut de penser à une contamination lors de la production. Or, les souches afghanes appartiennent majoritairement à la sous-lignée A.Br.Vollum. Après différents tests sur une collection de la sous-lignée TEA (groupe A.Br.008/011), il s'est avéré que les souches les plus proches de celle ayant causé la mort d'un héroïnomanes en Ecosse provenaient de Turquie. L'hypothèse a donc été que la contamination de l'héroïne aurait eu lieu lors du transport de la drogue à travers les pays, plutôt que lors de sa production ou de sa consommation (Price *et al.*, 2012). De plus, 2 SNP spécifiques de la souche isolée en Ecosse (Ba4599) furent utilisés pour typer 92 souches de la sous-lignée TEA et 2 souches non-TEA, ainsi que 35 échantillons positifs à *B. anthracis* associés à l'épidémie chez les héroïnomanes : tous les échantillons de l'épidémie possédaient l'allèle identifié chez Ba4599, alors que les autres portaient l'autre allèle. Par ailleurs, un allemand a contracté la maladie après injection de drogue, mais n'avait pas voyagé au Royaume-Uni : ceci a permis de penser qu'un seul pool originel d'héroïne a été disséminé en Europe (Price *et al.*, 2012).

ANALYSE BIBLIOGRAPHIQUE

Chapitre 3 : Analyse du génome

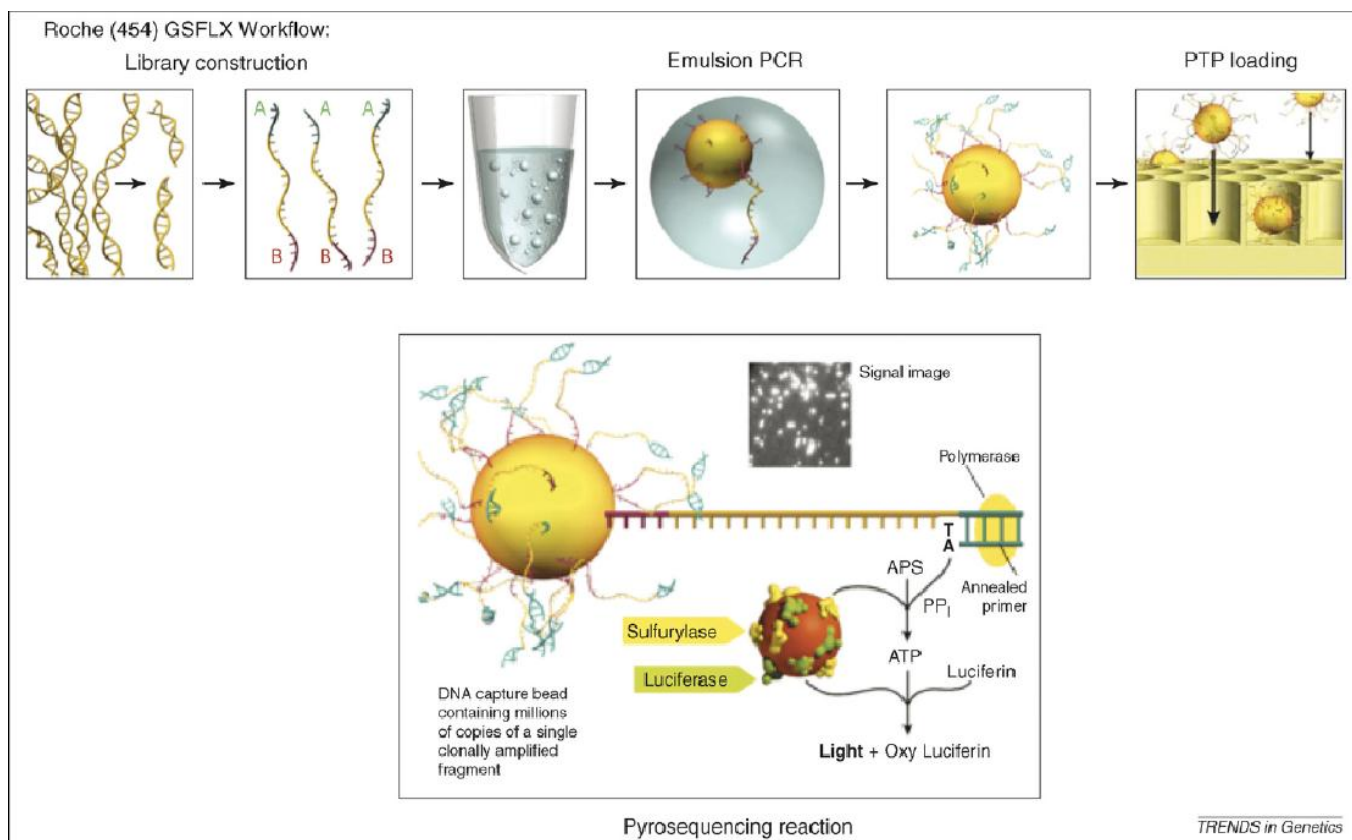


Figure 25: Principe de la méthode de séquençage 454 basée sur le pyroséquençage.
Figure extraite de (Gibrat, 2014)

1. Séquençage du génome complet

L'ADN est la molécule codant l'information génétique de la plupart des organismes vivants. Sa séquence est composée d'un « alphabet » à 4 bases : l'adénine, la thymine, la guanine et la cytosine. Ces nucléotides sont organisés en gènes et opérons. Le séquençage de l'ADN est le procédé qui consiste à déterminer l'ordre des nucléotides au sein d'une molécule d'ADN. Dans les années 1970, plusieurs avancées dans le séquençage de l'ADN ont été observées.

A la fin des années 1970, deux équipes de recherche, les groupes de Sanger et de Gilbert (Maxam et Gilbert, 1977; Sanger et Coulson, 1975), érigèrent les bases du séquençage : Sanger développa un procédé de séquençage rapide de l'ADN basé sur des « terminateurs de chaînes » (Sanger *et al.*, 1977). Cette technique, communément appelée séquençage Sanger, est devenue la méthode la plus utilisée. Le premier organisme totalement séquencé fut ainsi obtenu en 1977 (bactériophage phiX174). De nombreux organismes suivirent (*Escherichia coli*, *Saccharomyces cerevisiae*, etc.).

En 1995, Craig Venter et ses associés du TIGR (The Institute for Genomic Research) publièrent le génome complet de la bactérie *Haemophilus influenzae*. Ce fut le premier exemple de l'utilisation du séquençage global de manière aléatoire ou « whole genome shotgun » pour l'obtention d'un génome bactérien complet. Lorsque l'on séquence l'ADN d'un organisme, on ne peut pas encore obtenir la totalité de la molécule en une seule lecture. La technique "shotgun" repose sur un principe simple : découper un génome en un grand nombre de fragments de petite taille. Les extrémités d'une partie de ces fragments sont ensuite séquencées, puis ces séquences sont assemblées sur la base de leurs chevauchements grâce à des programmes informatiques pour essayer de produire une séquence complète. Les difficultés d'une telle technique se situent à deux niveaux : avoir assez de fragments pour couvrir le génome dans sa totalité, et réussir l'assemblage (Furelaud et Esnault).

Plus récemment, la méthode Sanger a été partiellement supplantée par ce que l'on appelle les NGS (Next Generation Sequencing) : les nouvelles technologies de séquençage (Morozova et Marra, 2008).

1.1. Méthodes NGS

Basées sur le séquençage massif en parallèle, trois méthodes principales ont été développées: la méthode 454®, la méthode Illumina/Solexa® et la méthode SOLID®. Ces méthodes diffèrent dans leurs caractéristiques (tailles des lectures, nombre de lectures générées, ...) et ont permis de diminuer les coûts de séquençage. Ces méthodes sont ainsi recommandées lorsque l'on souhaite caractériser une souche (Read *et al.*, 2002): c'est la méthode de typage la plus fine, car basée sur la lecture de l'information génétique.

1.1.1. Séquençage 454 ®

Une des principales limites de la méthode Sanger est le recours à l'amplification *in vivo* de fragments d'ADN à séquencer, clonés au préalable dans des bactéries hôtes (Morozova et Marra, 2008). Cette étape est susceptible d'introduire un biais lié à l'hôte, en plus d'être longue et fastidieuse. La technologie 454 (pyroséquençage), la première des NGS apparue sur le marché, contourne cette étape de clonage : elle utilise une méthode particulière d'amplification d'ADN, la PCR en émulsion. Le principe de la méthode est présenté en **Figure 25**.

Dans un premier temps, l'ADN génomique est fragmenté par sonication. Les brins sont ensuite séparés et un adaptateur complémentaire d'une amorce PCR est ajouté sur les brins. La réaction PCR se fait dans des micelles, d'où le terme de PCR en émulsion. Dans chaque micelle se trouve une micro-bille qui fixe un fragment d'ADN. Ce dernier est amplifié à l'intérieur de la

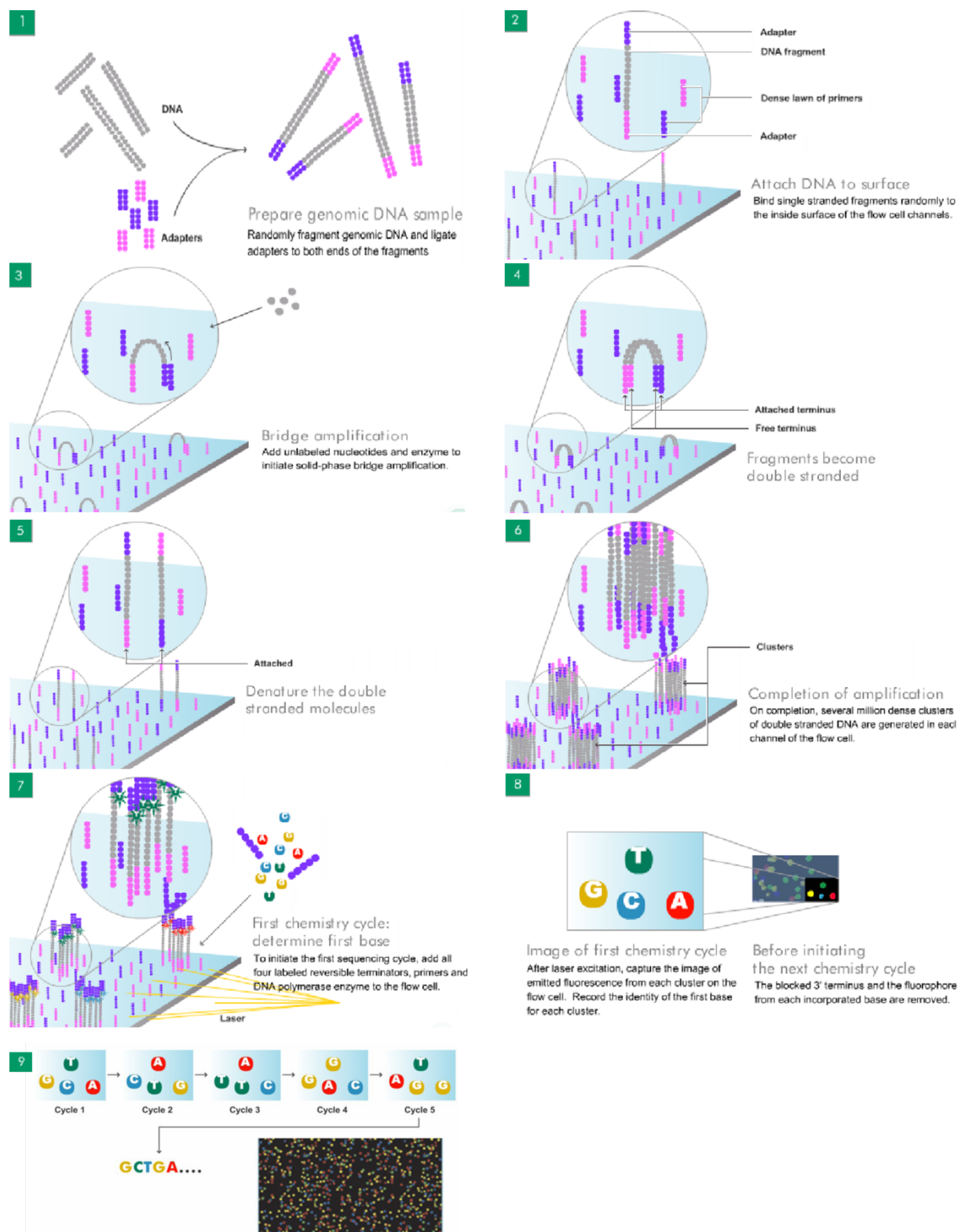


Figure 26: Principe de la méthode de séquençage Illumina.
Figures extraites de (Gibrat, 2014)

micelle (10^7 copies d'un ADN unique) (Morozova et Marra, 2008). Chaque bille est ensuite placée dans un micro-puits. Les nucléotides sont ajoutés tour à tour dans les puits et lorsque l'un d'eux est incorporé, un pyrophosphate inorganique (PPi) est libéré. Le signal relatif à cette émission est mesuré : il est proportionnel au nombre de photons émis, et donc au nombre de nucléotides incorporés. La séquence est ainsi reconstituée nucléotide par nucléotide.

L'un des principaux biais associé à cette méthode est lié au principe de mesure des photons : un effet de saturation du capteur est observé en présence d'un homopolymère. L'estimation est alors incorrecte (Gibrat, 2014). Les principales caractéristiques de cette méthode sont résumées dans la **Figure 28**.

1.1.2. Séquençage Illumina®

Une autre méthode de séquençage permet d'éviter l'étape de clonage limitante de la méthode Sanger : il s'agit du séquençage Illumina. Le principe de cette technique est illustré en **Figure 26**.

La première étape consiste à fragmenter l'ADN génomique (sonication, transposase,...). Des adaptateurs sont ajoutés aux extrémités des fragments d'ADN. Ces derniers sont fixés par une extrémité sur une surface solide où sont situées des amorces complémentaires aux adaptateurs : la « flow cell ». Les molécules se courbent ensuite pour s'hybrider aux adaptateurs complémentaires en formant un « pont ». L'amplification est réalisée à partir de cette structure en « pont » et permet d'obtenir près de 40 millions de clusters d'amplification (chaque cluster contient approximativement 1000 copies d'une même molécule) (Morozova et Marra, 2008). Le cycle de séquençage est initié en ajoutant les 4 nucléotides marqués de couleur différente, les amorces et l'ADN polymérase. La séquence ADN est déduite à la fin de chaque incorporation de nucléotides : un laser permet d'exciter les nucléotides et la fluorescence émise est enregistrée.

Le séquençage Illumina permet une meilleure reconstitution des homopolymères que la méthode 454. Cependant, les lectures de séquençage sont de taille inférieures, ce qui peut poser problème lors de la reconstruction de séquences répétées au sein du génome (Morozova et Marra, 2008). Les principales caractéristiques de cette méthode sont résumées dans la **Figure 28**.

1.1.3. Séquençage SOLID®

Cette méthode de séquençage est basée sur une approche par hybridation-ligation (SOLID pour « Supported Oligonucleotide Ligation and Detection system ») (Morozova et Marra, 2008). Le principe de cette méthode est illustré en **Figure 27**.

Les bibliothèques de séquençage sont constituées via une PCR en émulsion. Les produits PCR sont transférés sur une surface de verre où le séquençage a lieu, via un cycle d'hybridation-ligation. Une combinaison de 16 dinucléotides marqués par quatre fluorophores différents est utilisée. Ce système permet de marquer chaque position deux fois. Le nucléotide de la séquence est défini par l'analyse de la couleur de deux réactions de ligations successives (Morozova et Marra, 2008). Cela permet ainsi de faire la distinction entre une erreur de séquençage et un polymorphisme de séquence : l'erreur sera détectée dans seulement une réaction de ligation, alors que le polymorphisme le sera dans les deux. Les principales caractéristiques de cette méthode sont résumées dans la **Figure 28**.

Une troisième génération de séquenceurs arrive sur le marché depuis peu, avec des tailles de lectures de plus en plus importantes (Ion-Torrent, Pacific Bioscience, Minlon Oxford Nanopore). Ces technologies ne sont pas encore autant utilisées que celles citées précédemment (**Figure 28**).

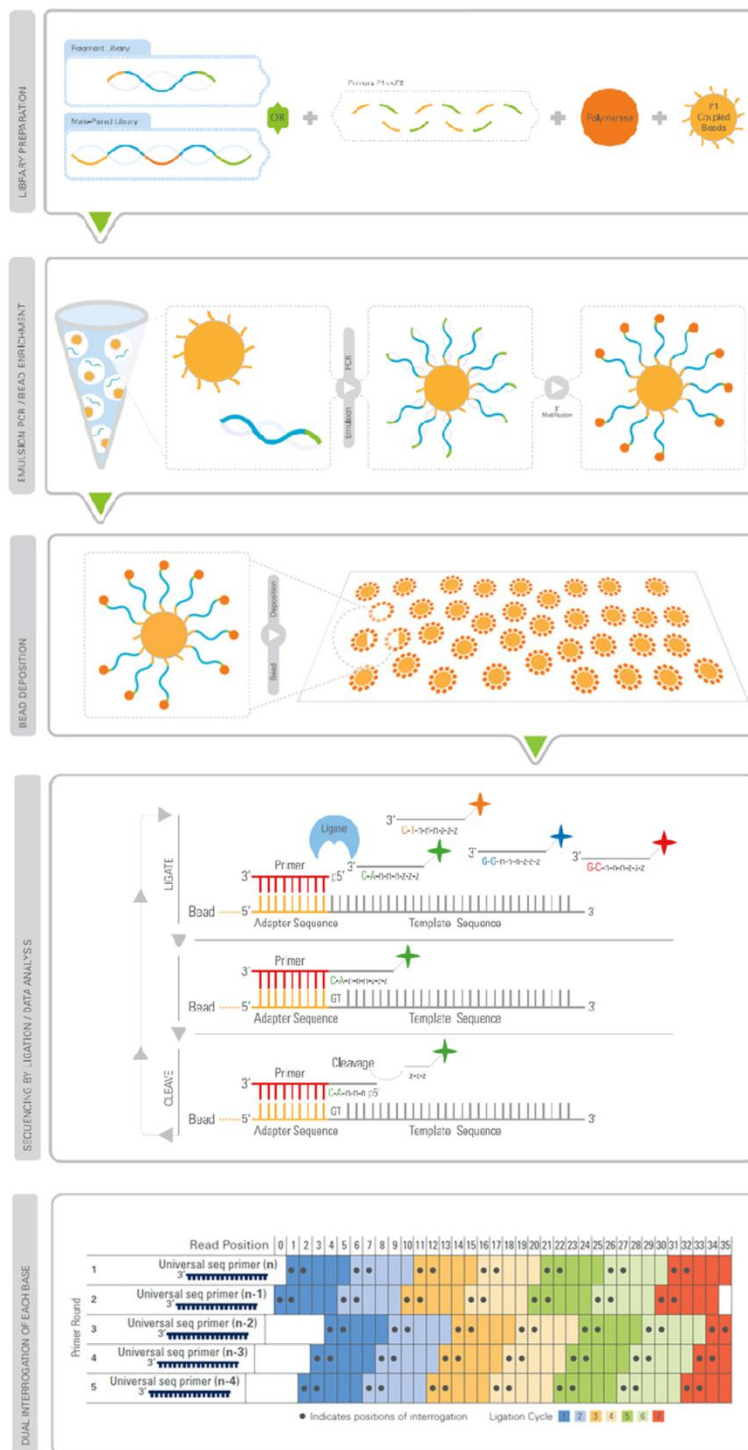


Figure 27: Principe de la méthode de séquençage SOLID

(Source : <http://www.appliedbiosystems.com/absite/us/en/home/applications-technologies/solid-next-generation-sequencing/next-generation-systems/solid-sequencing-chemistry.html>)

Différentes méthodes de séquençage

- **Première génération**
 - Méthode Sanger : **séquençage par synthèse** (clonage)
- **Deuxième génération**
 - Méthode 454 : **séquençage par synthèse** (amplification PCR)
 - Méthode Illumina/Solexa : **séquençage par synthèse** (amplification PCR)
 - Méthode SOLID : **séquençage par ligation** (amplification PCR)
- **Troisième génération**
 - Méthode Pacific Bioscience : **séquençage par synthèse** (molécule unique)
 - Méthode Ion Torrent : **séquençage par synthèse** (amplification PCR)
 - Méthode Nanopore : **séquençage « nanopore »** (molécule unique)

Instrument	Durée d'un run (heures)	Nombre de lectures (millions)	Taille lectures (bases)	Production (Mbase/run)	Coût (\$/Mbase)
Sanger 3730xl	2	0.000096	650	0.06	10000
454 FLX+	18-20	1	700	900	85
Illumina Hiseq2000 v3	240	<3000	100+100	<600000	0.03
SOLID 5500xl	192	1400	75+35	155000	0.04
Ion Torrent 318 Chip	2	4-8	200	800	5
PacBio RS	3	0.05	Moyenne : 8.5kb Top 5% : >18kb	375	100
Oxford Nanopore Minlon	??	??	Médiane : 5kb Plus de 20kb	??	??

Figure 28 : Caractéristiques des principales méthodes de séquençage passées et actuelles.
Figure extraite de (Gibrat, 2014)

2. Application des NGS

Les possibilités d'utilisation des NGS sont multiples. Ayant accès à la quasi-totalité de l'information génétique de l'organisme séquencé, les possibilités d'analyse sont immenses (re-séquençage de génomes, identification et expression de gènes, biologie comparative, épidémiologie, identification de facteurs de virulence...) (Grada et Weinbrecht, 2013). Cependant, elles nécessitent un traitement des données important. Bien que l'on parle de séquençage du génome complet, la totalité du génome n'est jamais reconstruite via l'utilisation d'assembleurs ou très difficilement. Il est souvent nécessaire de combiner différentes techniques de séquençage lorsque l'on souhaite obtenir un génome dans sa quasi-totalité. Cependant, une très grande partie du génome (>99%) est souvent reconstruite avec une seule méthode de séquençage, ce qui est très souvent suffisant. Tout dépend de la finalité du projet.

Grâce au séquençage du génome complet, il apparaît désormais possible d'identifier de nouveaux polymorphismes de séquence permettant de discriminer des isolats entre eux (Chen *et al.*, 2010; Cummings *et al.*, 2010). Plusieurs projets de séquençage de génome de *B. anthracis* ont été menés et ont permis l'obtention de séquences complètes ou de « draft » génomes selon les méthodes utilisées (Antwerpen *et al.*, 2012; Birdsell *et al.*, 2013; Chun *et al.*, 2012; Kim *et al.*, 2014; Ohnishi *et al.*, 2014; Ravel *et al.*, 2009; Read *et al.*, 2003).

Avec des coûts de moins en moins importants, les NGS sont de plus en plus utilisées au sein de plateformes ou de laboratoires, et pourraient devenir à terme la méthode de typage de référence.

3. Analyse bioinformatique

Bien que de plus en plus accessibles en raison de la diminution des coûts, les NGS restent une méthode complexe à analyser. Les données générées sont conséquentes (**Figure 28**) et leur analyse nécessite un savoir-faire particulier, ainsi qu'un matériel approprié. Après l'étape de séquençage proprement dite, les données brutes (lectures de séquençage) doivent être traitées afin de reconstituer le génome de l'organisme séquencé. Deux alternatives sont possibles : l'assemblage *de novo* ou l'assemblage par homologie.

3.1. Assemblage de génome *de novo*

Lorsqu'aucun génome de référence n'est connu ou que l'on souhaite étudier les remaniements chromosomiques et autres variations d'un génome, une approche sans *a priori* est à privilégier. On parle d'assemblage *de novo*.

Pour réaliser cela, divers programmes informatiques appelés assembleurs sont disponibles (Velvet, Ray, Celera, Mira, Abyss...), chacun possédant des spécificités propres. Les algorithmes informatiques qui permettent à ces programmes de reconstruire les génomes sont complexes et ne sont pas détaillés dans ce manuscrit.

Un pipeline d'assemblage est défini pour obtenir le résultat : dans un premier temps, les lectures sont importées, puis filtrées en fonction de leur qualité (étape de Trimming). L'assembleur récupère toutes les lectures de séquençage triées pour les regrouper en contigs lorsqu'elles se chevauchent, en fonction d'un paramètre variable permettant d'optimiser l'assemblage : le k-mer. Ainsi, à partir de millions de lectures courtes, on obtient une dizaine (plus ou moins selon l'assembleur, la qualité de séquençage,...) de contigs représentant le génome séquencé. Plusieurs paramètres sont disponibles, permettant d'évaluer la qualité de l'assemblage effectué : taille totale des contigs obtenue, pourcentage de lectures utilisées pour l'assemblage et la valeur N50 (plus petite taille de

contig/scaffold telle que 50% de l'assemblage est contenu dans des contigs/scaffolds plus grands que cette taille).

Si l'on souhaite combler les espaces (gaps) entre les contigs, on peut par exemple coupler la première technologie de séquençage utilisée (ex : Illumina, lectures courtes) avec une autre (ex : 454, lectures longues). On peut ainsi parvenir à reconstruire la « totalité » du génome de l'organisme séquencé (**Figure 29**).

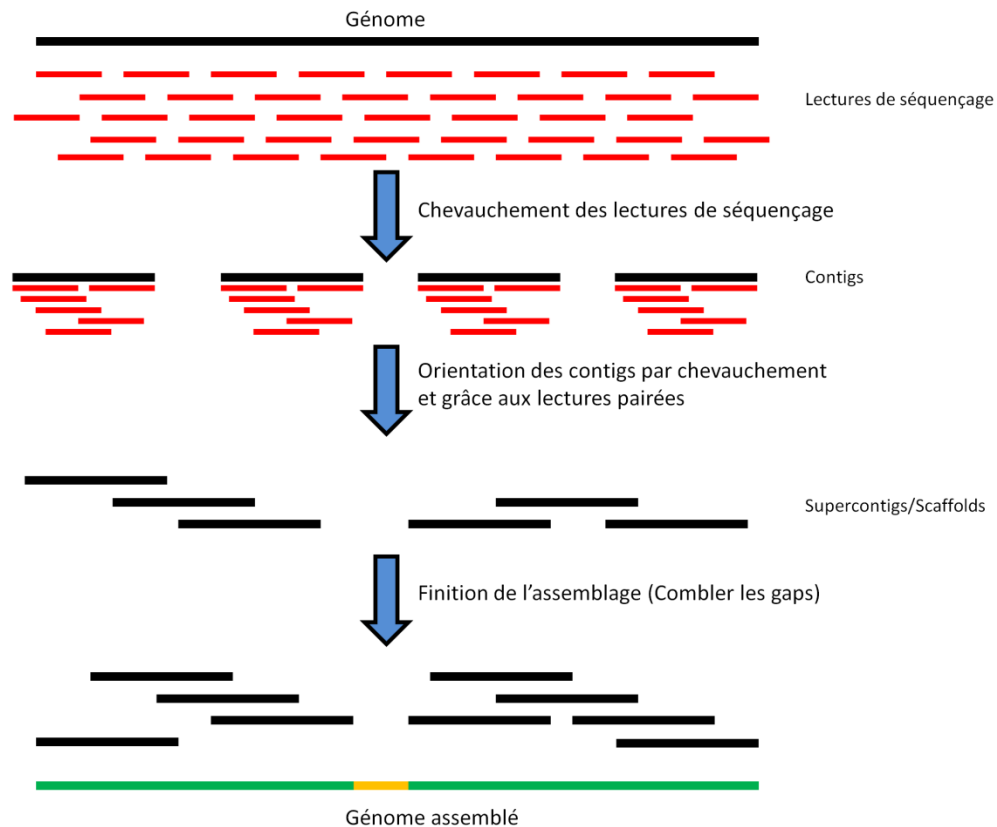


Figure 29: Principe de l'assemblage *de novo* (adapté d'après (Chen, 2011))

3.2. Assemblage de génome par homologie

Lorsqu'un génome de référence est disponible et que l'on recherche des polymorphismes de séquence tels que les SNP, une standardisation des génomes peut s'avérer utile. Un assemblage par homologie est alors conseillé. On parle de « mapping » des lectures de séquençage sur le génome de référence : elles sont alignées sur le génome de référence en fonction d'un pourcentage d'identité. Le génome final obtenu a une taille similaire au génome de référence. Avec cette approche, un certain nombre d'informations seront perdues, telles que certaines insertions-délétions, ce qui implique de bien choisir la stratégie à adopter en rapport avec la finalité du projet.

CONTEXTE ET OBJECTIFS

Zoonose aiguë potentiellement mortelle, la fièvre charbonneuse peut infecter tous les mammifères, humains compris, ainsi que quelques espèces d'oiseaux. Les herbivores sont les plus touchés. Ils contractent généralement la maladie en ingérant des spores, qui constituent la forme infectante de la bactérie, au cours de leur pâture. Chez l'Homme, le contact d'animaux infectés, l'ingestion d'aliments contaminés ou la respiration de poussières contenant des spores, sont responsables de la maladie. Le réservoir naturel est le sol. Sous sa forme sporulée, *B. anthracis* peut survivre durant de longues périodes dans les sols où il se trouve. De par ses caractéristiques et son pouvoir pathogène, le bacille du Charbon a toujours été considéré comme une arme biologique potentielle.

De répartition mondiale, la maladie reste endémique dans certaines régions du globe alors qu'elle est peu prévalente dans d'autres. Classée dans la liste de l'Organisation mondiale de la santé animale (OIE), la fièvre charbonneuse fait partie en France des maladies réglementées de catégorie 1 (articles L.223-2 et suivants, et D.223-2 du Code rural) et donne lieu à l'application de mesures de police sanitaire. Elle est reconnue comme maladie professionnelle et est redevenue une maladie à déclaration obligatoire chez l'Homme depuis 2001.

En Europe, on retrouve la maladie dans la plupart des pays. En France, elle apparaît de façon sporadique selon les années en fonction de la météorologie, de l'arrêt des vaccinations ou de l'activité humaine impactant le sol. Certaines années ont été particulièrement propices au développement de la maladie. Ce fut le cas en 2008 dans le Doubs (Calavas *et al.*, 2008) ainsi qu'en 2009 (Madani *et al.*, 2009).

B. anthracis, espèce jeune sur le plan phylogénétique, est considéré comme un des pathogènes les plus clonaux, faisant preuve d'une très grande homogénéité sur le plan génomique. Les souches sont extrêmement proches les unes des autres et il n'est pas aisé de les différencier en utilisant les méthodes classiques. Identifier clairement l'origine d'une contamination ou la souche impliquée dans un foyer de fièvre charbonneuse est une donnée importante, qui facilite l'analyse des causes et les décisions de gestion sanitaire. Ainsi, la stratégie thérapeutique, en particulier l'antibiothérapie, ou la traçabilité d'une tentative malveillante seront plus facilement mises en place. Plusieurs méthodes de différenciation de la bactérie ont été testées au cours des 20 dernières années, certaines avec plus de succès que d'autres. Parmi les méthodes récentes les plus prometteuses, on peut citer la MLVA, les SNP et les SNR, méthodes robustes, précises et présentant un pouvoir discriminant important. Toutefois, c'est désormais le séquençage à haut débit du génome complet qui permet d'étudier le plus précisément les relations phylogénétiques d'une espèce bactérienne tout en réalisant le génotypage de cette espèce.

L'objectif de cette thèse était de développer une méthode de typage rapide, robuste et à haut pouvoir discriminant, utilisable pour le diagnostic et l'épidémiologie des foyers de charbon en France et en Europe. Un tel outil représente un intérêt évident pour la sécurité intérieure afin de détecter rapidement toute diffusion accidentelle ou malveillante de ce germe et de réagir vite et de manière adaptée. Le but ultime de ce travail est d'acquérir la capacité de distinguer suffisamment de différences entre deux souches de façon à déterminer avec certitude qu'une souche particulière est à l'origine d'une infection naturelle ou d'un acte de malveillance, ainsi que l'acquisition de suffisamment d'informations pour identifier l'origine géographique de toute souche de *B. anthracis* présente sur Terre.

Pour ce faire, deux axes de travail ont été développés : dans un premier temps, étudier la diversité des souches de *B. anthracis* isolées en Europe via une approche de séquençage à haut débit (à partir de la collection de souches du LNR (124 souches) ainsi que sur une centaine de souches d'origines diverses (issues de collaborations inter-laboratoires)), puis, dans un second temps, identifier des marqueurs génétiques rares et discriminants. L'identification de marqueurs polymorphes « évolutivement » stables à l'origine de cette diversité était un préalable au développement d'un outil de typage moléculaire précis, rapide et peu coûteux, adapté à la caractérisation des souches naturellement présentes en France et dans le monde. Le principal avantage est de disposer d'un outil discriminant sans avoir à recourir au séquençage.

Plusieurs objectifs spécifiques ont été définis à partir du séquençage de près de 250 isolats et l'obtention du génome complet des souches séquencées :

- 1) Etudier la biodiversité des isolats français de *B. anthracis* et proposer un premier panel de marqueurs polymorphes visant à développer un outil de typage moléculaire rapide et précis
- 2) Investiguer les relations phylogénétiques existantes entre les isolats français et mondiaux et identification de panels complémentaires de marqueurs polymorphes destinés à être utilisés en typage moléculaire pour identifier rapidement toute souche
- 3) Développement de deux méthodes de biologie moléculaire capables de discriminer des mutations ponctuelles de type SNP : la PCR-HRM et le Luminex.

ETUDE EXPERIMENTALE

Chapitre 1:

Etude de la diversité de *Bacillus anthracis* en France

1. Utilisation des SNP sur génome complet pour étudier la diversité des souches françaises de *B. anthracis*

1.1 Présentation de l'étude

Afin d'étudier la diversité des isolats de *B. anthracis* en France, une approche par séquençage à haut débit couplée à une recherche des polymorphismes de séquence (SNP) sur génome complet a été effectuée. Dans un premier temps, un protocole d'extraction d'ADN génomique (chromosome et plasmides) a été mis au point. La quasi-totalité de la collection de souches du LNR Charbon a pu être séquencée (124 souches). Une fois le séquençage effectué, les données ont été traitées informatiquement à l'aide du logiciel BioNumerics (Applied Maths). Les génomes ont été assemblés par homologie par rapport au génome de référence (Ames Ancestor). Les séquences de chaque souche ont ensuite été alignées sur le génome Ames Ancestor afin d'identifier tous les polymorphismes de séquences de type SNP (Single Nucleotide Polymorphism). Toute donnée pouvant introduire un biais dans l'analyse est écartée (VNTR, opérons ribosomiques, données incomplètes, SNP en clusters...).

Afin de positionner les souches françaises par rapport aux souches disponibles dans les bases de données, 4 génomes d'intérêt ont été ajoutés à l'étude (Ames Ancestor, Sterne, A1055 et une souche africaine de la collection ANSES). Près de 4000 SNP ont ainsi été identifiés et ont permis l'étude des relations existantes entre ces souches via une approche par parcimonie avec génération d'arbres phylogénétiques (Minimum Spanning Trees). Les trois sous-lignées françaises sont bien retrouvées et bien discriminées. La résolution obtenue est telle que l'on est en mesure de définir des SNP spécifiques d'une souche. Chaque sous-lignée a été caractérisée en détail.

1.2. Développement d'un outil de typage : la PCR HRM

L'outil de typage développé au sein du LNR Charbon est basé sur une PCR en temps réel suivie d'une analyse par courbe de fusion à haute résolution (PCR HRM). Ce traitement post-PCR permet la détection de polymorphismes génétiques. Rapide et simple à mettre en œuvre, elle est également peu onéreuse.

L'analyse HRM est réalisée à partir d'ADN double brin. Une taille d'amplicon courte (inférieure à 100 pb) est conseillée. Une fois l'étape de PCR terminée, les amplicons subissent une élévation progressive de la température. Une fois la température de fusion de l'amplicon atteinte, l'ADN sous forme double brin se sépare : on appelle cela la fusion de l'ADN. La distinction entre deux populations d'amplicons ne différant que par un seul nucléotide s'effectue en observant les différences de températures de fusion en fonction du nucléotide observé. La **Figure 30** présente le type de courbe obtenue classiquement lors de l'utilisation de cette méthode.

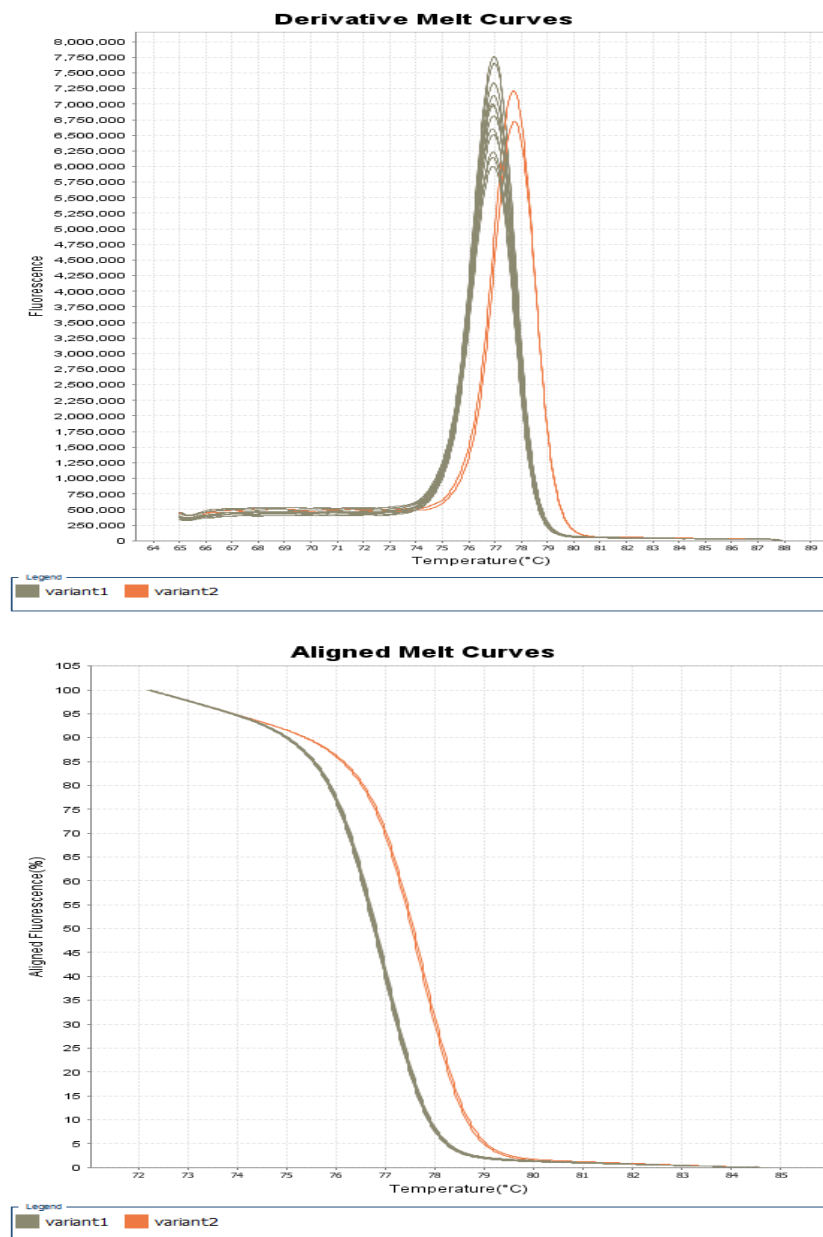


Figure 30: Courbes de fusion obtenues lors d’une PCR HRM

Article 1:

High-throughput sequencing of *Bacillus anthracis* in France: investigating genome diversity and population structure using whole-genome SNP discovery

Publié le 16 avril 2014 dans *BMC Genomics*

RESEARCH ARTICLE

Open Access

High-throughput sequencing of *Bacillus anthracis* in France: investigating genome diversity and population structure using whole-genome SNP discovery

Guillaume Girault¹, Yann Blouin², Gilles Vergnaud² and Sylviane Derzelle^{1*}

Abstract

Background: Single nucleotide polymorphisms (SNPs) are ideal signatures for subtyping monomorphic pathogens such as *Bacillus anthracis*. Here we report the use of next-generation sequencing technology to investigate the historical, geographic and genetic diversity of *Bacillus anthracis* in France. 122 strains isolated over a 60-years period throughout the country were whole-genome sequenced and comparative analyses were carried out with a focus on SNPs discovery to discriminate regional sub-groups of strains.

Results: A total of 1581 chromosomal SNPs precisely establish the phylogenetic relationships existing between the French strains. Phylogeography patterns within the three canSNP sub-lineages present in France (i.e. B.Br.CNEVA, A.Br.011/009 and A.Br.001/002) were observed. One of the more remarkable findings was the identification of a variety of genotypes within the A.Br.011/009 sub-group that are persisting in the different regions of France. The 560 SNPs defining the A.Br.011/009- affiliated French strains split the Trans-Eurasian sub-group into six distinct branches without any intermediate nodes. Distinct sub-branches, with some geographic clustering, were resolved. The 345 SNPs defining the major B.Br CNEVA sub-lineage clustered three main phylogeographic clades, the Alps, the Pyrenees, and the Massif Central, with a small Saône-et-Loire sub-cluster nested within the latter group. The French strains affiliated to the minor A.Br.001/002 group were characterized by 226 SNPs. All recent isolates collected from the Doubs *department* were closely related. Identification of SNPs from whole-genome sequences facilitates high-resolution strain tracking and provides the level of discrimination required for outbreak investigations. Eight diagnostic SNPs, representative of the main French-specific phylogeographic clusters, were therefore selected and developed into high-resolution melting SNP discriminative assays.

Conclusions: This work has established one of the most accurate phylogenetic reconstruction of *B. anthracis* population structure in a country. An extensive next-generation sequencing (NGS) dataset of 122 French strains have been created that allowed the identification of novel diagnostic SNPs useful to rapidly determine the geographic origin of any strain found in France.

Keywords: *Bacillus anthracis*, Whole genome sequencing, Single nucleotide polymorphism, Molecular typing, Comparative genomics, HRM

* Correspondence: sylviane.derzelle@anses.fr

¹University Paris-Est, Anses, Animal Health Laboratory, Bacterial Zoonoses Unit, Maisons-Alfort 94706, France

Full list of author information is available at the end of the article



© 2014 Girault et al.; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

Background

Bacillus anthracis, the causative agent of anthrax, is a spore-forming bacterium found throughout the world. The bacterium mainly affects wild and domesticated herbivores, causing serious, often fatal disease. Spores which are the infectant form of the bacteria can remain in soils for decades before being ingested by grazing animals and later on contaminated browse by blow flies [1]. Consequently, opportunity for accumulating DNA mutations is limited, resulting in relatively little genetic variation within the species.

Single nucleotide polymorphisms (SNPs) are the most common genetic variation found in genomes of all species. It also represents important biologically informative DNA markers extensively used to elucidate deep phylogenetic relationships among worldwide strains, due to their evolutionary stability [2]. By querying a large number of them against collections of diverse strains, a set of canonical SNPs (canSNPs) that define major clades within the *B. anthracis* species has been identified and used for subdividing all isolates into three major lineages (A, B and C) and 13 major sub-lineages or sub-groups [2-5]. Other SNPs that define the A.Br. Ames and A.Br.WNA lineages [6,7] or that lie on various branches of the *B. anthracis* SNP tree [8] have also been published.

As a consequence of the rapid development of technology in the area of high throughput sequencing, extensive genomic sampling is now possible at a reasonable cost. By interrogating nearly every base of the genome, whole genome sequencing is the genotyping tool giving the highest possible resolution, with comparison information spanning from distant phylogenetic relationships to the highest level of subtyping. Identification of thousands of SNPs retrieved from compiled NGS sequences facilitates high-resolution strain tracking and provides the level of discrimination required for microbial forensics. Canonical SNP typing and whole genome comparisons of multiple strains are now becoming the reference method in *B. anthracis* genotyping.

In France, anthrax was a widespread cause of disease in livestock and humans until the beginning of the 20th century. The development of live vaccines for animals by Louis Pasteur (1880–1881), followed by widespread vaccination, had then allowed the numbers of animals affected, and indirectly human cases, to steadily decline. The prevention of new cases and proper carcass disposal with the implementation of knackeries prevented the replenishment of spores in soil in many regions of France and concentrations progressively declined [9]. Today, anthrax is considered as a sporadic disease that continues to cause occasionally outbreaks in livestock. A few cases with one to several dead animals may be recorded annually. However two large outbreaks with several

dozens of animals affected still occurred over the past five years in areas where outbreaks had been reported in the past [10].

We have previously reported the canSNP characterization and MLVA typing of a hundreds of strains of *B. anthracis* isolated in France [11]. French strains were assigned into the three B.Br.CNEVA, A.Br.011/009 and A.Br.001/002 canSNP sub-groups and further resolved into about fifty MLVA genotypes (unpublished data). To gain further resolution into the phylogenetic relationships among French isolates, 122 strains have been sequenced using parallel sequencing technology. Comparative genomics, with a focus on SNPs discovery, were then conducted to identify diagnostic SNPs that specifically discriminate regional sub-groups of strains.

Results

canSNP genotyping

One hundred and thirty-six “field strains” representative of anthrax activity in France since about half a century (1953–2011) were selected for this study. All strains were first subjected to canSNP typing [11] to place them into a broader phylogenetic context [3]. An additional canSNP assay designed around the A.Br.011 SNP, that further divides the Trans-Eurasian A.Br.008/009 group into two sub-groups (A.Br.008/011 and A.Br.011/009) [4], was also run against the whole collection (data not shown). All French A.Br.008/009 samples (n = 39) were defined as belonging to the A.Br.011/009 sub-group.

Whole genome sequencing

In order to accurately determine the evolutionary relationships among these strains, 122 isolates were characterized by paired-end whole genome sequencing. We sequenced 67 B.Br.CNEVA-affiliated strains collected in the four areas where this sub-lineage is prevalent (i.e. the Alps (n = 34/38), the Pyrenees (n = 9), the Massif Central region (n = 18/20) and the Saône-et-Loire *département* (n = 6)), 31 A.Br.011/009-affiliated strains isolated throughout the country (n = 31/36) and 24 of the 26 A.Br.001/002-affiliated strains collected from the Doubs *département* (n = 21/22) and the North-East of France (n = 3). An African strain (IEMVT 89) affiliated to the A.Br.005/006 sub-group positioned at the basis of the main A radiation was also whole-genome sequenced for comparison.

The Illumina sequencers produced 4 to 25 million reads per strain after applying the quality filter of the Illumina base-calling pipeline. The filter-passed reads were aligned to the Ames Ancestor reference genome, resulting in more than a 56-fold sequencing depth on average and a genome coverage ranging from 98.9 to 99.9% (Additional file 1: Table S1).

Extraction of whole strain-specific SNPs among *B. anthracis* strains

Comparative analysis of the genomic sequences was next carried out. Available data for the *B. anthracis* Ames Ancestor, Sterne and A1055 strains were also included. The *B. cereus* AH820 genome was used as outgroup. Figure 1 illustrates the minimum spanning tree (MST) generated by the chromosomal whole-genome SNPs data. A total of 3987 chromosomal SNPs were discovered that precisely positioned French strains within the *B. anthracis* SNP tree. Additional SNPs (128 in pXO1 and 92 in pXO2) were also identified in both plasmids (Analysis of the 122 French strains, including the African strains IEMVT 89, data not shown). Consistent with previous MLVA and canSNP analyses [11], all French isolates were clustered into three distinct groups, i.e. B.Br.CNEVA, A.Br.001/002 and A.Br.011/009. The major B.Br.CNEVA group was separated by a long phylogenetic branch (438 SNPs) to the A and C branches radiation. A total of 1581 SNPs resolved the French population structure at an unprecedented resolution. Strain-specific SNPs can be defined for nearly all particular strains. Moreover, particular genotypes corresponding with geographic regions were observed.

B.Br.CNEVA phylogenetic analysis

More than half of the *B. anthracis* isolates from France clustered within the B.Br.CNEVA lineage. We therefore focused on further resolving the positions of its members. The B.Br.CNEVA sub-lineage was characterized by 779 unique chromosomal SNPs, 49 pXO1 SNPs and 49 pXO2 SNPs. Our analysis identified 345 diagnostic SNPs that discriminate between French strains, as well as 29 plasmidic SNPs (Figure 2). Positions of the 67 whole-genome sequenced strains relative to each others are illustrated on Figure 2. A geographic clustering into four sub-groups (Alps, Pyrenees, Massif Central and, nested in this latter, the Saône-et-Loire sub-cluster) was observed using both chromosomal and plasmidic SNPs analysis. Four markers that specifically define one of the newly described geographic sub-groups, as well as one additional SNP shared by all French B.Br.CNEVA genomes, were selected and incorporated into SNP discrimination HRM assays (Table 1).

A.Br.011/009 phylogenetic analysis

Comparison of A.Br.011/009 genomes yielded a total of 574 unique chromosomal SNPs, including 560 SNPs that resolve the sub-group into six distinct sub-branches without any intermediate nodes (Figure 3A). This sub-structure was poorly reflected in both pXO1 and pXO2 SNPs profiles. The 38 Plasmidic SNPs resolved the 31 samples in a higher number of putative branches, without formal clustering (Figure 3B and C). In contrast to the B.Br.CNEVA lineage, a less apparent geographic

clustering was also observed within each identified sub-branches (Figure 3A). Branches 1 and 2 include strains isolated in the north-east quarter of France, the outermost samples belonging to branch 2. Branches 3 and 6 are composed of strains isolated around the center (close to Paris) and in the South-East quarter of France, respectively. While the only strain of branch 5 was isolated close to the Pyrenees (South-West), the two samples from branch 4 are from remote geographical regions (West and South). Specific SNPs for the two more relevant branches (1 and 2) were developed into real-time HRM assays (Table 1).

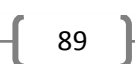
A.Br.001/002 phylogenetic analysis

Concerning the minor A.Br.001/002 subgroup, sequence analysis identified 226 chromosomal SNPs that differentiate the 24 French strains. Notably, all isolates from the Doubs *department* (21 out 24 samples) formed a single clonal cluster characterized by 44 SNPs (Figure 1). All strains were closely related, differing by a maximum of 2 chromosomal SNPs. Similar SNP patterns were found using both virulence plasmids data (data not shown). Most of them ($n = 20$) were collected during 17 clustered animal outbreaks occurring in the summer of 2008. An additional strain was isolated in the same area in 2011 and presented the dominant 2008's SNP genotype, suggesting a common point source for these repeated outbreaks. One SNP marker was selected for further testing based upon its specificity for the Doubs strains (Table 1). The three other strains are older "outbreak" samples (1953, 1954 and 1981) isolated in other *departments* of France. They were found to be unrelated to the Doubs recent outbreak isolates. Several specific SNPs differentiating these strains were also found on both plasmids.

French specific SNP discrimination genotyping assays

We designed eight novel canSNPs discrimination assays based on High-Resolution Melting (HRM)-PCR technology (Table 1). Assays targeted five SNPs identified among the French B.Br.CNEVA genomes, one SNP specific to the A.Br.001/002 strains isolated in Doubs, and two SNPs defining the two large sub-branches 1 and 2 of the A.Br.011/009 sub-group (Figure 1). These diagnostic assays were then successfully validated across the 138 *B. anthracis* strains of our collection. The two expected alternate alleles exhibited distinct melting curves and melting temperatures (T_m s), allowing unambiguous grouping of each allele. On average, differences in T_m ranging from 0.5 to 0.8°C were observed between the both allelic states (Table 2). The seventeen non-sequenced genomes of the collection fall on the correct geographical groups (data not shown).

The eight novel canSNPs were further validated *in silico* against all *B. anthracis* genomes ($n = 30$) available in public NCBI database and Sequence Read Archive (SRA) [12]. The new markers were also screened across



(See figure on previous page.)

Figure 1 Phylogeny of 126 *B. anthracis* strains based on whole-genome SNP analysis. **A.** Minimum spanning tree based on 3987 chromosomal SNPs. The 3 canSNP groups present in France are color-coded: B.Br CNEVA in light blue, A.Br 011/009 in purple and A.Br 001/002 in green. The African lineage A.Br 005/006 is indicated in red. Positions of the *B. anthracis* Sterne (in green), Ames ancestor (in yellow) and A1055 (in black) strains are also marked. Each circle represents a unique SNP genotype. The diameter of each circle varies according to the number of isolates having the same genotype. The length of each branch is proportional (logarithmic scale) to the number of SNPs identified between strains. Indicated in red are the position and name of the new identified French canSNPs. The star marks the approximate branching point of the *B. anthracis* lineage within the *B. cereus* group. Based on a parsimony approach, the tree size is 4018, i.e. it contains approximately 0.77 % of homoplasia. **B.** Linear phylogenetic tree rooted with the *B. cereus* AH820 strain as outgroup. This figure illustrates the relationship between French and globally diverse *B. anthracis* strains.

a hundred of non-French *B. anthracis* DNAs of diverse origins (described by 10 canSNP typing, unpublished data) to confirm their specificity. The eight assays accurately separated the French strains from the remaining globally diverse genotypes (data not shown).

Discussion

Comparative whole-genome sequencing offers a powerful way for in-depth characterization of any bacterial

pathogen. It also provides an unbiased approach for informative SNPs discovery [13]. In this study, an ultimate picture of the genetic diversity found within the *B. anthracis* population in France has been established. An extensive NGS dataset of 122 autochthonous strains have been created that allowed the phylogenetic linkage of geographically diverse isolates and the identification of novel SNPs signatures useful to rapidly determine the geographic origin of any strain ecologically established in France.

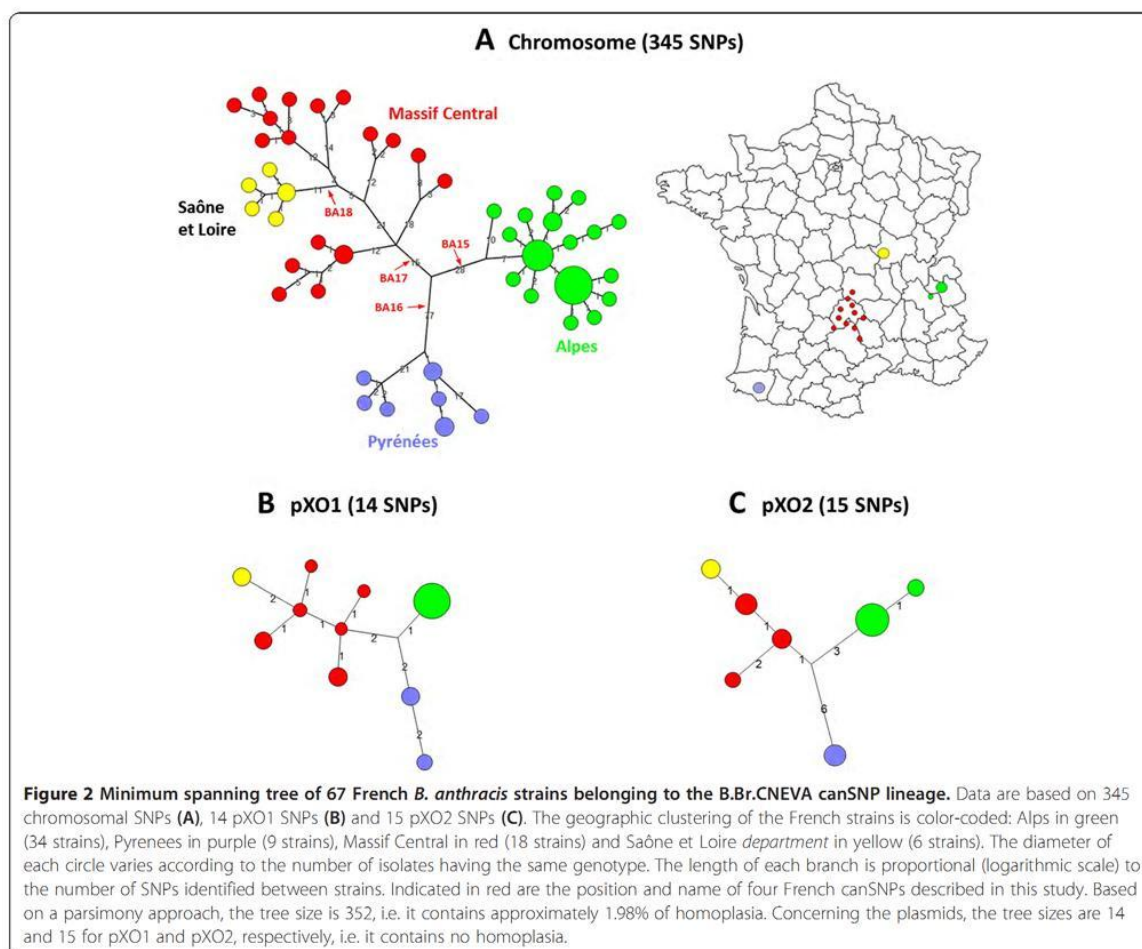


Table 1 French specific canSNPs and primer sequences used for HRM analysis

canSNP	Position*	Target	SNP	Forward primer (5'-3')	Reverse primer (5'-3')	Product size (bp)
BA15	126639	Alps (B.CNEVA)	A to G	CCACAAGGTGGAATTACTAAAGA	GGTTCACCTGTTTCGGATCT	80
BA16	3765357	Pyrenees (B.CNEVA)	A to G	GGTGGTTTCGGATATGCACT	AAAGGTGCTGGGTAGTAAGG	68
BA17	4719494	MC + SL (B.CNEVA)	G to A	TTAGATCTCGTTTCGGTTCC	CAATGAGTGACGGCTCCAA	79
BA18	2390832	SL (B.CNEVA)	T to C	CCAGGCAAATACATTGTGGA	TTACAGTCTGTGTGCCGTTG	73
BA19	2573536	French B.CNEVA	A to G	CATATATTTACCTCTTTATGAACA	GATAAAGGCTGTCGGATGG	90
BA20	3434997	Doubs (A.001/002)	A to C	AGCGAGCCAATTTGGAACCGA	AGGCGGGATTGTGTGGATGT	63
BA21	3562427	Central NE (A.011/009)	A to C	AGCAAAAAGTCGGCAAAGAA	ACAGAGCTTCTCCGAACATG	82
BA22	820195	NE (A.011/009)	T to C	AGTGTGTCAATCCCAATTC	CGCAGCAATATTCGCTATCA	99

*localisation on the Ames Ancestor chromosome (GenBank accession no. AE017334.2). MC: Massif Central; SL: Saône et Loire; NE: North-Eastern.

Within the B.Br. CNEVA lineage, which is the most successful group in frequency found in France, a clustering concordant with the geographical origin of each strain was highlighted. This phylogeography pattern is consistent with a single introduction of this lineage in

the country, followed by dispersal of the ancestral population toward West and South, with progressively derived local populations of *B. anthracis* strains. Ecological establishment and *in situ* differentiation within three regions with environmental conditions particularly favorable

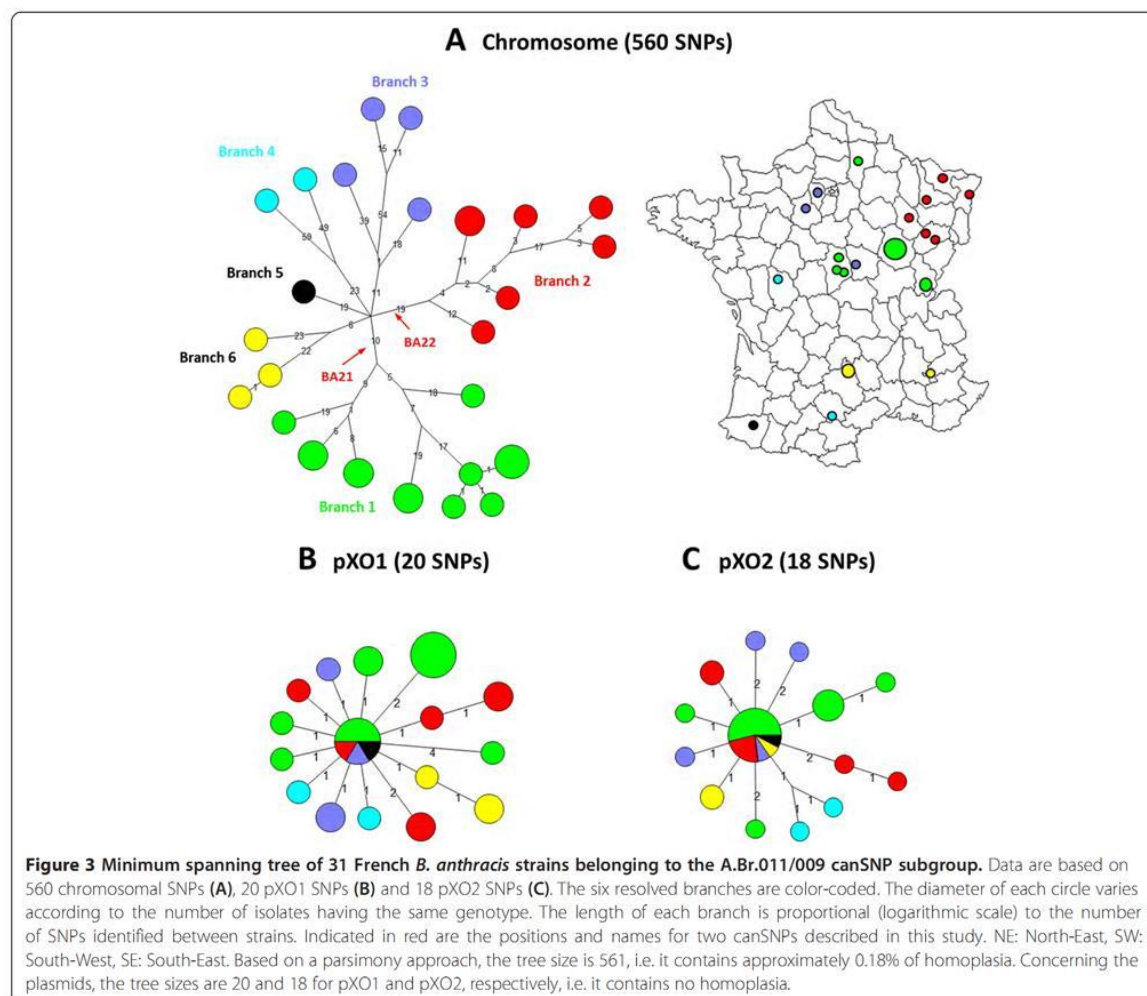


Table 2 Melting temperatures (T_m) determined for French-specific canSNP by HRM

canSNP	Region	Allele	T _m values (°C)	Nb strains
BA15	Alps	G	75.16 ± 0.05	38
	Others	A	74.61 ± 0.07	100
BA16	Pyr	G	79.93 ± 0.04	9
	Others	A	79.39 ± 0.08	129
BA17	MCSL	A	77.27 ± 0.08	26
	Others	G	77.87 ± 0.13	112
BA18	SL	C	79.83 ± 0.10	6
	Others	T	79.13 ± 0.10	132
BA19	France	G	75.84 ± 0.07	73
	Others	A	75.22 ± 0.07	65
BA20	Doubs	C	80.36 ± 0.05	21
	Others	A	79.83 ± 0.06	117
BA21	CNE	C	77.90 ± 0.05	15
	Others	A	77.08 ± 0.08	123
BA22	NE	C	77.50 ± 0.08	10
	Others	T	76.93 ± 0.08	128

T_m values are means ± standard deviation calculated from 138 strains. MCSL: Massif Central and Saône-et-Loire; SL: Saône-et-Loire; Pyr: Pyrenees; CNE: Central North-Eastern; NE: North-Eastern.

for the survival of the spores (i.e. grasslands made of pastoral valleys) are observed: The Alps, The Pyrenees and the Massif Central, with further derived sub-clade in Saône-et-Loire *department* (discriminative canSNPs BA15 to BA18). The French B.Br.CNEVA lineage forms a large cluster distinct from similar strains isolated elsewhere in Europe (canSNP BA19). It is worthwhile to mention that A strains are adapted to more diverse environments than B strains, which are restricted to more narrow environmental conditions [14]. This trend is reflected on a global scale. A strains have experienced a global dissemination and population explosion, while both B lineages are uncommon in much of the world and appear to be only successful locally or regionally [3].

In contrast to the B.Br.CNEVA lineage, French strains affiliated to both A.Br.001/002 and A.Br.011/009 groups appeared to be more disparate and less related to one another. Several distinct introductions and subsequent *in situ* differentiation of both canSNP groups in France could therefore not be excluded. One of the more remarkable findings from the whole genome SNP analysis was the discovery of a complex pattern of six phylogenetic sub-branches (composed of up to fifteen strains) within the A.Br.011/009 isolates collected in France. The TEA group (A.Br.008/009) is a cluster that predominates throughout Europe, the Middle East and the Western most Chinese province of Xinjiang [3,15]. The identification of six sub-branches, with some geographic clustering

(canSNPs BA21 and BA22), indicates an extensive history for the particular A.Br.011/009 TEA sub-group in France that could be linked to an expansion into many novel environments throughout the country or a combination of possibly repeated introduction and infections. Similar pattern of loosely related strains (≈100 SNPs) was also observed for the A.Br.001/002 canSNP group, except for strains isolated from the Doubs *department* (canSNP BA20). All but one of these Doubs specimens were isolated from a single episode (involving 17 clustered outbreaks) associated with the death of 39 animals in 2008 [16]. Only a short evolutionary time period, i.e. very few SNPs, separate all Doubs isolates, reflecting recurring outbreaks in this specific area. Further comparisons with European strains affiliated to the same canSNP groups are, however, need for drawing sound conclusions on the evolution and natural history of both A.Br.011/009 and A.Br.001/002 canSNP groups in France.

Although other relationships are topologically similar to MLVA predictions, some discrepancies exist between whole-genome SNP- and MLVA-based phylogenetic reconstructions (data not shown). There result from homoplasmy coupled with weaker character support in the VNTR system. SNPs are more stable and definitive markers than VNTRs. In the MLVA-31 minimum spanning trees, for example, the French B.Br.CNEVA strains are grouped into three clonal complexes (CCs) and a few singletons unrelated to others strains from close geographical origins (Thierry et al., unpublished data). Most of the strains from The Alps, The Massif Central and part of the Saône-et-Loire samples were clustered in a single, large clonal complex (CC1). Strains from the Pyrenees formed a distinct clade (CC2), as well as half of the Saône-et-Loire strains (CC3). The Saône-et-Loire isolates were erroneously separated within two different clusters based on difference in up to four VNTR loci between CC1 and CC3. The discriminative power of the 31 VNTR markers is also more limited to resolve strains from a specific region or outbreak, as illustrated by the A.Br.001/002 strains collected in the Doubs 2008 and 2011. While MLVA31 differentiated five genotypes, nine distinct SNPs profiles could be resolved by whole-genome SNPs scanning. Furthermore, the latter clearly demonstrated that all Doubs strains are genetically related and differs by only one to two SNPs throughout the whole genomes. SNP-based phylogenetic reconstruction of *B. anthracis* population structure is more reliable. Whole genome comparison and canSNPs genotyping represent the future of molecular epidemiology analysis.

Conclusions

In this study, an ultimate picture of the genetic diversity found within the *B. anthracis* population in France has been established. The knowledge derived from this work

is extremely useful for future outbreak investigations, not only due to the specific genomic information that facilitates detailed comparisons between different bacterial isolates, but also because the data can be used to add unprecedented resolving power to current molecular typing tools. Using HRM, eight novel canSNP assays that narrowly defined the three *B. anthracis* genetic sub-lineages found in France were developed. As a consequence, whether an index strain is representative of ecologically established strains from France or have been imported from other areas can be rapidly determined.

Methods

Bacterial strains and biosafety procedures

A total of 136 *B. anthracis* strains belonging to B.Br CNEVA (n = 73), A.Br 011/009 (n = 36), A.Br 001/002 (n = 26) and A.Br 005/006 (n = 1) were used in this study. All but one strain were isolated in France. It included nine old reference strains from the Pasteur Institute's CIP collection (e.g. 17JB, CIP 53.169, CIP 74.12, CIP 77.02, CIP 81.89, CIP A204, CIP A205, CIP A206 and CIP A211), 126 strains collected during animal or human anthrax outbreaks (mostly from bovine origin) that have occurred in France over the past 31 years (1982–2013), among which the previously sequenced CNEVA 9066 and A0465, and one African strain (IEMVT 89–1620) affiliated to the A.Br.005/006 sub-group. *B. anthracis* species was confirmed by bacteriology and PCR. All *B. anthracis* manipulations were performed in a biosafety level 3 laboratory using class II type A2 biosafety cabinet.

DNA extraction

Genomic DNAs were obtained from vegetative cells grown at 37°C on 5% horse blood agar plates. DNA was purified using the QIAGEN® Genomic-tip 100/G columns and QIAGEN® Genomic DNA Buffer Set. Briefly, bacterial colonies were harvested by scraping the agar surfaces from 16 to 18 h-old Petri dishes. Cell pellet was resuspended in 3.3 ml of Buffer B1 containing 7 µl of RNase A (100 mg/ml, QIAGEN) and incubated at 37°C for 45 min with 350 µl of lysozyme (100 mg/ml, Roche) and 100 µl of proteinase K stock solution (QIAGEN). Following addition of 1.2 ml of Buffer B2, DNA lysate was further incubated at 50°C for 1 h. Particle-free sample was next applied to the equilibrated QIAGEN Genomic-tip and DNA purification on anion-exchange resin processed according to the manufacturer's recommendations. After isopropanol precipitation, genomic DNA was resuspended in 400 µl of 10 mM Tris HCl (pH 8) for at least 2 h at 50°C.

DNA solutions were transferred to a 0.22 µm sterile Ultrafree-MC spin filter (Millipore) and centrifuged for 2 min at a maximal speed of 8000 × g to ensure the complete removal of live forms of *B. anthracis* from DNA.

Viability testing was systematically performed before DNA was taken out of the BSL-3 facility. An aliquot of each DNA preparation (a quarter) was spread on Petri dishes and grown at 37°C for 24 h.

Draft whole genome sequencing (WGS) and data analysis

In this study, 122 French strains belonging to 3 sub-lineages (A.Br 001/002 (n = 24), A.Br 011/009 (n = 31) and B.Br CNEVA (n = 67)) were sequenced at the IMAGIF sequencing platform (Imagif, Gif sur Yvette, France) [17]. Isolates were subjected to paired-end whole genome sequencing on either the Illumina Genome Analyzer IIx instrument (16 samples, paired-end data of 2x75pb) or the Illumina HiSeq2000 platform (107 samples, paired-end data of 2x100pb) (Illumina Inc., San Diego, CA, USA). Genome coverage of at least 57× was obtained (Additional file 1: Table S1, average depth ranging from 57× to 433×). The number of reads that passed Illumina quality filters varied from 4.5 to 26.8 million.

Ames Ancestor [GenBank:AE017334.2] was used as the reference genome for assembly. Ames Ancestor plasmid pXO1 [GenBank:AE017336.2] and pXO2 [GenBank:AE017335.3] were used as references for plasmids assembly. Short reads data sets were exported on the FastQ format and mapped to the Ames Ancestor genome and both pXO1 and pXO2 plasmidic sequences using BioNumerics version 6.6 (Applied Maths, Belgium) and Power assembler module asking for a similarity of at least 90%. A set of SNPs was deduced for each genome sequence data using BioNumerics Chromosome Comparisons module. Individual lists were compiled by using an in-house Python script, and data filtered to remove SNP positions at which one or more isolate displayed an ambiguous residue call or missing data. Ribosomal operons and VNTR loci were also excluded from the analysis, so as contiguous SNPs (in a 10pb window). The list of canSNP positions is provided in Table 1.

The draft genome sequences of one representative French strains affiliated to the A.Br.001/002 (08-8_20, Doubs) [NCBI: JHCB000000000], A.Br.011/009 (99–100, Branch 3) [NCBI: JHDR000000000] and B.Br.CNEVA (00–82, Alps) [NCBI: JHDS000000000] have been deposited in the NCBI database. The whole genome sequences for *B. anthracis* strains CNEVA-9066 (B.Br.CNEVA, Massif Central) [NCBI: NZ_AAEN000000000.1], A0465 (B.Br.CNEVA, Pyrenees) [NCBI: NZ_ABLH000000000.1], Ames ancestor (A.Br.Ames) [NCBI: NC_007530.2], Sterne (A.Br.001/002) [NCBI: NC_005945.1] and *B. cereus* AH820 (outgroup) [NCBI: NC_011773.1] can be found in the NCBI microbial genome website at <http://www.ncbi.nlm.nih.gov/>.

Whole genome phylogenetic analysis

A minimum spanning tree was drawn in BioNumerics by using the filtered whole genome sequencing SNP data

as input. The tree was rooted by using the *B. cereus* AH820 strain as outgroup. *B. cereus* AH820 resides very close to the *B. anthracis* cluster using multilocus sequence typing [18]. Nodes were numbered by BioNumerics. The canSNPs along the three branches leading to the French strains were identified from whole genome sequencing data by searching for SNPs with allelic states shared only by these different subgroups.

SNP discrimination assays by HRM

We designed High Resolution Melting (HRM) assays for eight French-specific SNPs using Primer 3⁺ software [19]. The positions of these SNPs in the Ames Ancestor genome [GenBank: AE017334.2] and the primers sequences used are listed in Table 1.

Amplification was performed on the ViiA7[™] Real-Time PCR System (Life Technologies) using the LightCycler[®] 480 High Resolution Melting Master Mix (Roche Diagnostics). The reaction mixture consisted of 0.2 μM of each primer, 1 × LightCycler[®] 480 HRM master mix and 2.5 mM MgCl₂ in a 10-μl final volume. The following parameters were used: 10 min at 95°C were followed by 40 cycles consisting of 10 s at 95°C, 10 s at 58°C and 20 s at 72°C. Samples were next heated to 95°C for 30 s, cooled down to 65°C for 1 min and heated from 65°C to 88°C at a rate of 1°C/s with 25 acquisitions/°C. HRM data were analyzed by the ViiA7[™] Software (version 1.2.1).

Availability of supporting data

The ANSES Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the BioProject number PRJNA242332, with accession numbers JHCB00000000 (A.Br 001/002, 08-8_20, Doubs), JHDR00000000 (A.Br 011/009, 99–100, Branch 3) and JHDS00000000 (B.Br. CNEVA, 00–82, Alps).

Phylogenetic trees data are available from the Dryad Digital Repository (<http://datadryad.org/>) with the following identifier: <http://doi.org/10.5061/dryad.rc6m9>.

Additional file

Additional file 1: Table S1. Strains used in this study and NGS data.

Abbreviations

DNA: Deoxyribonucleic acid; SNP: Single nucleotide polymorphism; canSNP: Canonical single nucleotide polymorphism; VNTR: Variable-number tandem repeat; MLVA: Multiple-locus variable-number tandem repeat analysis; WGS: Whole-genome sequencing; NGS: Next generation sequencing; PCR: Polymerase chain reaction; HRM: High resolution melting; CC: Clonal complex; NCBI: National center for biotechnology information; TEA: Trans-Eurasian.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

GV and SD conceived the study. GG and SD performed the experiments, analyzed the data and wrote the paper. YB developed scripts used for whole-genome SNP analysis. All authors read and approved the final manuscript.

Acknowledgements

We wish to acknowledge Nora Madani (Anses, France) for access to the strains collection of the national anthrax reference. This work was supported by the French National Research Agency (ANR) and Direction Générale de l'Armement (DGA) (project number ANR11ASTR0007). GG is a PhD student co-supported by Anses and DGA grants. This work has benefited from the facilities and expertise of the high throughput sequencing platform of IMAGIF (Centre de Recherche de Gif [<http://www.imagif.cnrs.fr>]).

Author details

¹University Paris-Est, Anses, Animal Health Laboratory, Bacterial Zoonoses Unit, Maisons-Alfort 94706, France. ²University Paris-Sud, Institut de Génétique et Microbiologie, UMR 8621, Orsay 91405, France.

Received: 21 November 2013 Accepted: 9 April 2014

Published: 16 April 2014

References

- Hugh-Jones M, Blackburn J: **The ecology of *Bacillus anthracis*.** *Mol Aspects Med* 2009, **30**:356–367.
- Pearson T, Busch JD, Ravel J, Read TD, Rhoton SD, U'Ren JM, Simonson TS, Kachur SM, Leadem RR, Cardon ML, Van Ert MN, Huynh LY, Fraser CM, Keim P: **Phylogenetic discovery bias in *Bacillus anthracis* using single-nucleotide polymorphisms from whole-genome sequencing.** *Proc Natl Acad Sci U S A* 2004, **101**:13536–13541.
- Van Ert MN, Easterday WR, Huynh LY, Okinaka RT, Hugh-Jones ME, Ravel J, Zanecki SR, Pearson T, Simonson TS, U'Ren JM, Kachur SM, Leadem-Dougherty RR, Rhoton SD, Guenevier Z, Farlow J, Coker PR, Smith KL, Wang B, Kenefic LJ, Fraser-Liggett CM, Wagner DM, Keim P: **Global genetic population structure of *Bacillus anthracis*.** *PLoS One* 2007, **2**:e461.
- Marston CK, Allen CA, Beaudry J, Price EP, Wolken SR, Pearson T, Keim P, Hoffmaster AR: **Molecular epidemiology of anthrax cases associated with recreational use of animal hides and yarn in the United States.** *PLoS One* 2011, **6**:e28274.
- Price EP, Seymour ML, Sarovich DS, Latham J, Wolken SR, Mason J, Vincent G, Drees KP, Beckstrom-Sternberg SM, Phillippy AM, Koren S, Okinaka RT, Chung WK, Schupp JM, Wagner DM, Vipond R, Foster JT, Bergman NH, Burans J, Pearson T, Brooks T, Keim P: **Molecular epidemiologic investigation of an anthrax outbreak among heroin users, Europe.** *Emerg Infect Dis* 2012, **18**:1307–1313.
- Van Ert MN, Easterday WR, Simonson TS, U'Ren JM, Pearson T, Kenefic LJ, Busch JD, Huynh LY, Dukerich M, Trim CB, Beaudry J, Welty-Bernard A, Read T, Fraser CM, Ravel J, Keim P: **Strain-specific single-nucleotide polymorphism assays for the *Bacillus anthracis* Ames strain.** *J Clin Microbiol* 2007, **45**:47–53.
- Kenefic LJ, Pearson T, Okinaka RT, Schupp JM, Wagner DM, Ravel J, Hoffmaster AR, Trim CP, Chung WK, Beaudry JA, Foster JT, Mead JI, Keim P: **Pre-Columbian origins for North American anthrax.** *PLoS One* 2009, **4**:e4813.
- Birdsell DN, Pearson T, Price EP, Hornstra HM, Nera RD, Stone N, Gruendike J, Kaufman EL, Pettus AH, Hurlon AN, Buchhagen JL, Harms NJ, Chanturia G, Gyuranecz M, Wagner DM, Keim PS: **Melt analysis of mismatch amplification mutation assays (Melt-MAMA): a functional study of a cost-effective SNP genotyping assay in bacterial models.** *PLoS One* 2012, **7**:e32866.
- Vaissaire J, Mock M, Le Doujet C, Levy M: **Le charbon bactérien. Épidémiologie de la maladie en France.** *Méd Mal Infect* 2001, **31**:257–271.
- Madani N, Mendy C, Moutou F, Garin-Bastuji B: **La fièvre charbonneuse en France. Episodes de l'été 2009 et foyers enregistrés sur la dernière décennie (1999–2009).** *Bulletin épidémiologique* 2010, **38**:17–19.
- Derzelle S, Laroche S, Le Fleche P, Hauck Y, Thierry S, Vergnaud G, Madani N: **Characterization of genetic diversity of *Bacillus anthracis* in France by using high-resolution melting assays and multilocus variable-number tandem-repeat analysis.** *J Clin Microbiol* 2011, **49**:4286–4292.
- NCBI. <http://www.ncbi.nlm.nih.gov/>.

13. Pearson T, Okinaka RT, Foster JT, Keim P: **Phylogenetic understanding of clonal populations in an era of whole genome sequencing.** *Infect Genet Evol* 2009, **9**:1010–1019.
14. Smith KL, DeVos V, Bryden H, Price LB, Hugh-Jones ME, Keim P: ***Bacillus anthracis* diversity in Kruger National Park.** *J Clin Microbiol* 2000, **38**:3780–3784.
15. Simonson TS, Okinaka RT, Wang B, Easterday WR, Huynh L, U'Ren JM, Dukerich M, Zanecki SR, Kenefic LJ, Beaudry J, Schupp JM, Pearson T, Wagner DM, Hoffmaster A, Ravel J, Keim P: ***Bacillus anthracis* in China and its relationship to worldwide lineages.** *BMC Microbiol* 2009, **9**:71.
16. Calavas D, Sala C, Vaissaire J, Condé J, Thien-Aubert H, Hesseemann M, Woronoff-Rhen N: **Retour d'expérience sur un épisode de fièvre charbonneuse chez les bovins dans le Doubs au cours de l'été 2008.** *Bulletin épidémiologique* 2009, **32**:1–7.
17. IMAGIF: <http://www.imagif.cnrs.fr>.
18. Kolsto AB, Tourasse NJ, Okstad OA: **What sets *Bacillus anthracis* apart from other *Bacillus* species?** *Annu Rev Microbiol* 2009, **63**:451–476.
19. **Primer 3+ software.** <http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/>.

doi:10.1186/1471-2164-15-288

Cite this article as: Girault *et al.*: High-throughput sequencing of *Bacillus anthracis* in France: investigating genome diversity and population structure using whole-genome SNP discovery. *BMC Genomics* 2014 **15**:288.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit



2. Assemblage *de novo* de souches de *B. anthracis* isolées en France

Article 2:

Draft Genomes of Three Strains Representative of the *Bacillus anthracis* diversity found in France

Publié le 31 Juillet 2014 dans *Genome Announcements*

Draft Genomes of Three Strains Representative of the *Bacillus anthracis* Diversity Found in France

Guillaume Girault,^a Nicolas Parisot,^{b,c} Eric Peyretailade,^b Pierre Peyret,^b Sylviane Derzelle^a

University Paris-Est, ANSES, Animal Health Laboratory, Bacterial Zoonoses Unit, Maisons-Alfort, France^a; Clermont Université, Université d'Auvergne, EA 4678 CIDAM, BP 10448, Clermont-Ferrand, France^b; UMR CNRS 6023, Université Blaise Pascal, Clermont-Ferrand, France^c

We report here the draft genomes of three *Bacillus anthracis* strains isolated in France: 08-8_20 (A.Br.001/002), 99-100 (A.Br.011/009), and 00-82 (B.Br CNEVA). The total lengths of assemblies are 5,440,708 bp, 5,446,472 bp, and 5,436,014 bp for 08-8_20, 99-100, and 00-82, respectively.

Received 2 July 2014 Accepted 16 July 2014 Published 31 July 2014

Citation Girault G, Parisot N, Peyretailade E, Peyret P, Derzelle S. 2014. Draft genomes of three strains representative of the *Bacillus anthracis* diversity found in France. *Genome Announc* 2(4):e00736-14. doi:10.1128/genomeA.00736-14.

Copyright © 2014 Girault et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](http://creativecommons.org/licenses/by/3.0/).

Address correspondence to Sylviane Derzelle, sylviane.derzelle@anses.fr.

Anthrax is an acute to peracute infectious disease of mammals (1). *Bacillus anthracis*, its causal agent, is a Gram-positive endospore-forming bacterium that primarily affects grazing herbivores but can be used as an agent of bioterrorism (2). Although the incidence of anthrax is declining in many parts of the world, spore-contaminated areas are still widely distributed and present a persistent risk for infection. Spores of *B. anthracis* are extremely resistant to several stresses and can remain dormant in soils for several decades before given the chance to infect another host (1).

Considered as an evolutionary young species that recently emerged from *Bacillus cereus* (3), *B. anthracis* has evolved as a highly fit clonal pathogen that has spread throughout the world and has become ecologically established in local genetically distinct populations (4). In France, anthrax is considered to be a sporadic disease that continues to cause occasional outbreaks in livestock. A few cases resulting in one to several dead animals may be recorded annually, mainly in cattle. In our ongoing efforts to elucidate the worldwide phylogenetic relationships between *B. anthracis* isolates, a canonical single-nucleotide polymorphism (canSNP) characterization and next-generation sequencing (5) of hundreds of French strains isolated over a 60-year period throughout the country were recently performed. We report here the draft genomes of three of these strains, which are representative of the different *B. anthracis* canSNP sublineages present in France, i.e., A.Br.001/002 (strain 08-8_20), A.Br.011/009 (strain 99-100), and B.Br CNEVA (strain 00-82) (6).

Whole-genome shotgun (WGS) sequencing of the *B. anthracis* 08-8_20, 99-100, and 00-82 strains was performed with the Illumina sequencing technology using the Genome Analyzer (GA)IIx, HiSeq 2000, and MiSeq instruments (Illumina, Inc.). Paired-end reads of 75 bp (08-8_20), 100 bp (08-8_20, 99-100, and 00-82), 150 bp (08-8_20 and 99-100), and 250 bp (08-8_20 and 99-100) in length were generated. The number of reads that passed the Illumina quality filters were, respectively, 0.297 to 9 million for strain 08-8_20, 0.497 to 8.7 million for strain 99-100, and 14.9 million for strain 00-82. A total of 18.6 million reads (08-8_20), 15.1 million reads (99-100), and 13.9 million reads (00-82) were *de novo*

assembled using IDBA-UD version 1.0.9 (7), with a *k*-mer size from 20 to 100 nucleotides.

The *N*₅₀ contig sizes of the assembly were 440.3 kbp, 868.5 kbp, and 868.4 kbp for strains 08-8_20, 99-100, and 00-82, respectively, and 10, 9, and 9 contigs >100 kbp accounting for 83.8%, 85.7%, and 86.6% of the sequenced nucleotides were obtained for the three strains, respectively. The total lengths of the three draft genomes were 5,440,708 bp (31 contigs), with an average G+C content of 35.65% (strain 08-8_20); 5,446,472 bp (30 contigs), with an average G+C content of 35.67% (strain 99-100); and 5,436,014 bp (28 contigs), with an average G+C content of 36.41% (strain 00-82).

A detailed comparative genomic analysis of these French strains with more *B. anthracis* strains from Western Europe will contribute to improve our ability to react rapidly when the identity and real origin of a strain need to be established (8). Extensive WGS data about autochthonous but also worldwide isolates are required so that accurate hypotheses can be made about isolate origins.

Nucleotide sequence accession numbers. The draft genome sequences for strains 08-8_20, 99-100, and 00-82 have been deposited at DDBJ/EMBL/GenBank under BioProject no. PRJNA242332, with accession no. [JHCB000000000](https://www.ncbi.nlm.nih.gov/assembly/JHCB000000000/) (08-8_20; A.Br 001/002), [JHDR000000000](https://www.ncbi.nlm.nih.gov/assembly/JHDR000000000/) (99-100; A.Br 011/009), and [JHDS000000000](https://www.ncbi.nlm.nih.gov/assembly/JHDS000000000/) (00-82; B.Br CNEVA).

ACKNOWLEDGMENTS

This work was supported by the French National Research Agency (A.N.R.) and Direction Générale de l'Armement (DGA) (project ANR11ASTR0007). G.G. is a Ph.D. student cosupported by ANSES and DGA grants.

This work has benefited from the facilities and expertise of the high-throughput sequencing platform of IMAGIF (Centre de Recherche de Gif [<http://www.imagif.cnrs.fr/>]).

REFERENCES

- OIE. 2008. Anthrax in humans and animals, 4th ed. World Organisation for Animal Health, Paris, France. http://www.who.int/csr/resources/publications/anthrax_webs.pdf.
- Jernigan JA, Stephens DS, Ashford DA, Omenaca C, Topiel MS, Gal-

- braith M, Tapper M, Fisk TL, Zaki S, Popovic T, Meyer RF, Quinn CP, Harper SA, Fridkin SK, Sejvar JJ, Shepard CW, McConnell M, Guarner J, Shieh W-J, Malecki JM, Gerberding JL, Hughes JM, Perkins BA, Members of the Anthrax Bioterrorism Investigation Team. 2001. Bioterrorism-related inhalational anthrax: the first 10 cases reported in the United States. *Emerg. Infect. Dis.* 7:933–944. http://wwwnc.cdc.gov/eid/article/7/6/01-0604_article.
3. Pearson T, Busch JD, Ravel J, Read TD, Rhoton SD, U'Ren JM, Simonson TS, Kachur SM, Leadem RR, Cardon ML, Van Ert MN, Huynh LY, Fraser CM, Keim P. 2004. Phylogenetic discovery bias in *Bacillus anthracis* using single-nucleotide polymorphisms from whole-genome sequencing. *Proc. Natl. Acad. Sci. U. S. A.* 101:13536–13541. <http://dx.doi.org/10.1073/pnas.0403844101>.
 4. Van Ert MN, Easterday WR, Huynh LY, Okinaka RT, Hugh-Jones ME, Ravel J, Zanecki SR, Pearson T, Simonson TS, U'Ren JM, Kachur SM, Leadem-Dougherty RR, Rhoton SD, Zinser G, Farlow J, Coker PR, Smith KL, Wang B, Kenefic LJ, Fraser-Liggett CM, Wagner DM, Keim P. 2007. Global genetic population structure of *Bacillus anthracis*. *PLoS One* 2:e461. <http://dx.doi.org/10.1371/journal.pone.0000461>.
 5. Girault G, Blouin Y, Vergnaud G, Derzelle S. 2014. High-throughput sequencing of *Bacillus anthracis* in France: investigating genome diversity and population structure using whole-genome SNP discovery. *BMC Genomics* 15:288. <http://dx.doi.org/10.1186/1471-2164-15-288>.
 6. Derzelle S, Laroche S, Le Flèche P, Hauck Y, Thierry S, Vergnaud G, Madani N. 2011. Characterization of genetic diversity of *Bacillus anthracis* in France by using high-resolution melting assays and multilocus variable-number tandem-repeat analysis. *J. Clin. Microbiol.* 49:4286–4292. <http://dx.doi.org/10.1128/JCM.05439-11>.
 7. Peng Y, Leung H, Yiu SM, Chin FYL. 2010. IDBA—a practical iterative de Bruijn graph de novo assembler, p 426–440. In Berger B (ed), *Research in computational molecular biology: 14th Annual International Conference, RECOMB 2010, Lisbon, Portugal, April 25–28, 2010, proceedings*. Springer, New York, NY.
 8. Derzelle S, Thierry S. 2013. Genetic diversity of *Bacillus anthracis* in Europe: genotyping methods in forensic and epidemiologic investigations. *Biosecur. Bioterror.* 11(Suppl 1):S166–S176. <http://dx.doi.org/10.1089/bsp.2013.0003>.

ETUDE EXPERIMENTALE

Chapitre 2:

Etude de la diversité de *Bacillus anthracis* en Europe

L'étude menée en France a permis de développer une approche à haut débit pour identifier des polymorphismes de séquence sur génome complet. Ces SNP ont permis de développer un outil de typage à haut pouvoir résolutif à l'échelle française au sein de l'espèce *B. anthracis*. Grâce à cela, une cartographie à haute résolution de la population de cette espèce sur le territoire français a été obtenue.

Afin d'améliorer l'outil de détection, l'étude a été étendue aux souches européennes et hors UE. Grâce à plusieurs collaborations entre laboratoires (Allemagne, Pologne, Danemark, Italie, Belgique, Pays-Bas, Hongrie, Suède et Suisse), l'ADN (ou les souches) provenant de ces pays ont pu être intégrés à l'étude de la diversité de la bactérie responsable de la fièvre charbonneuse. Nous avons ainsi pu obtenir des données sur 10 des 13 sous-lignées de *B. anthracis*. Dans l'objectif de développer un outil de typage exhaustif, ces collaborations se sont avérées extrêmement précieuses.

Différentes approches ont été testées : une globale à l'échelle européenne, intégrant les souches disponibles dans les bases de données, et plusieurs à l'échelle nationale (Pays) en fonction de la diversité présentée.

1. Génomotypage des souches italiennes de *B. anthracis*

Article 3:

Anthrax phylogenetic structure in Italy: investigating genome diversity using whole genome SNP discovery

Article en préparation

Anthrax phylogenetic structure in Italy: investigating genome diversity using whole-genome SNP discovery

Antonio FASANELLA^{1*}, Guillaume GIRAULT², Angela ACETI¹, Alessia AFFUSO¹ and Sylviane DERZELLE²

¹ Istituto Zooprofilattico Sperimentale of Puglia and Basilicata; Anthrax Reference Institute of Italy; Foggia, Italy.

² University Paris-Est, Anses, Animal Health Laboratory, Bacterial Zoonoses Unit, 94706 Maisons-Alfort, France

*corresponding author (do you prefer exchanging position with me?)

a.fasanella@izsfg.it

guillaume.girault@anses.fr

alessia.affu@hotmail.it

acetiangel@hotmai.com

sylviane.derzelle@anses.fr

Abstract:

Background *Bacillus anthracis*, the etiological agent of anthrax, is an evolutionarily young species that presents an extremely low genetic diversity. We used next-generation sequencing technology to investigate the phylogenetic relationships existing between strains of *Bacillus anthracis* in Italy. Twenty-one strains isolated over a 20-years period throughout the country were sequenced and comparative analyses were carried out with a focus on single nucleotide polymorphisms (SNPs) discovery.

Results Whole-genome SNPs screening resolved six phylogenetic clusters among the Italian strains belonging to the Trans-Eurasian group, including both A.Br.008/011 and A.Br.011/009 canSNP sub-groups, which are predominant in Italy. A.Br.011/009-affiliated strains were subdivided into four phylogeographic clades: Basilicata-Campania, Puglia-Lazio-Toscana, Sardegna and Carbosap-like strains from Sicilia. Two minor but distinct A.Br.008/011-affiliated clades, with some geographic clustering in Sicily or Umbria, were also identified. Isolates from the B.Br.CNEVA lineage that were previously reported in the Northeast of Italy were demonstrated to be only weakly related to French strains from the same lineage. Seven diagnostic SNPs, defining the newly described clusters, were next selected and developed into discriminative assays using a melting temperature-shift method.

Conclusions Those specific canSNPs accurately resolved the Italian population structure of *B. anthracis* into seven phylogenetic clusters.

Keywords: *Bacillus anthracis*, Whole genome sequencing, Single Nucleotide Polymorphism, genotyping, comparative genomics, Melting temperature-shift method

Introduction

Anthrax is a non-contagious infectious disease that affects a wide range of animal species. All mammals are known to be susceptible to anthrax, including humans, but domesticated and wild ruminants are the most at risk. *Bacillus anthracis*, the bacterial agent of anthrax is a gram-positive bacterium capable of forming endospores that can survive in adverse environmental conditions for several decades. In herbivores, the disease is usually acute or peracute, with a fatal outcome in most cases. Sudden death, accompanied sometimes by oozing of blood from the natural body openings, is a consistent feature of animal anthrax (Hugh-Jones et De Vos, 2002). In humans, the disease manifests itself in four distinct forms depending on the route of entry of the bacterium: cutaneous anthrax, with case-fatality rate considered low; injectional anthrax associated with drug users who consumed spore-contaminated heroin; gastrointestinal, and inhalation anthrax for which treatment must be given as soon as possible to reduce mortality. Due to its high rate of infectivity and lethality, its global dispersal pattern and the extreme stability of its spores, *B. anthracis* is considered as a major threat in what concerns bioterrorism.

Italy is probably one of the few countries in Europe, along with Greece, where anthrax is still relatively common, despite vaccination campaigns carried out since the late 1950s. Anthrax is hypo-enzootic in Italy but almost exclusively affects animals at pasture (Fasanella *et al.*, 2005; Garofolo *et al.*, 2011). Over the last twenty years, 121 outbreaks of animal anthrax have been reported across the country (Figure 1). The most affected regions are in the South and the major islands, especially in Basilicata, Puglia, Campania and Sicily. Outbreaks data from Calabria are very likely underestimated.

Most of the animal anthrax outbreaks recorded in Italy are characterized by sporadic episodes taking place during the summer in known enzootic area (including old knackeries, former tanneries or burial sites). It typically involves a very low number of animals (one to three on average) grazing on contaminated pastures (Brachman *et al.*, 1966; Schmid et Kaufmann, 2002). The occurrence of new cases is generally linked to climate changes (with incidents of warmer and extreme weather) or human activities that interfere with soil surface (works of drainage or excavation). Animals develop a hyperacute form of the disease from which they die without showing any symptoms. In rare cases, these sporadic outbreaks progress into an epidemic-like episode, leading to many clustered outbreaks involving a very limited area and a large number of animals. Epidemic-like anthrax outbreaks were so recorded in the Madonie Natural Park (Sicily, 1997) (Di Marco, 2004), the Pollino Natural Park (Basilicata, 2004) (Fasanella *et al.*, 2007; Garofolo *et al.*, 2010), the Mugello Park (Tuscany, 2008) (Garofolo *et al.*, 2011) and the Regional Park of Gallipoli Cognato and Small Dolomites of Lucania (Campania and Basilicata, 2011) (Palazzo *et al.*, 2012). These outbreaks were found in some animals which developed a fatal, sub-acute form of anthrax characterized by the presence of extensive edema due to an injection-source infection. A possible role for bloodsucking insects, and especially tabanids, as vectors in the transmission of *B. anthracis* in these epidemic-like infections has therefore been suggested. Atypical outbreaks, occurring in areas considered to be free of anthrax and in atypical seasons or climatic conditions, may also be observed occasionally in Italy. These unexpected infections are linked to the consumption by livestock of cut forage harvested from contaminated soils, or feed products such as meat or bone meal supplements derived from processing of

infected carcasses. Three such cases were reported during the last decade in Italy, e.g. in Parcines (Trentino Alto Adige, December 2005) (Kreidl, 2006), Foligno (Umbria, March 2009) and Firenzuola (Tuscany, October 2012).

There have been continuous technological improvements for microbial genomic characterization in the past decade, moving to biologically informative assays such as MLVA and MLST to whole-genome SNP analysis based on next generation sequencing. SNPs have become important genomic markers used for the detection and differentiation of *Bacillus anthracis* (Girault *et al.*, 2014a; Marston *et al.*, 2011; Van Ert *et al.*, 2007a). Because of their low mutation rates, SNPs also represent ideal targets for phylogenetic analyses. Approaches based on genome SNPs discovery are now seen as the ultimate resolution tool for epidemiological typing of bacteria as they permits detailed, targeted analysis of variations within even monomorphic pathogens such as *B. anthracis*.

We have previously reported the investigation of the molecular diversity found among Italian strains of *B. anthracis* by MLVA-8 (Fasanella *et al.*, 2005). Canonical SNP genotyping, MLVA-15 and single nucleotide repeat (SNR) analyses were further performed to improve resolution of the population structure (Garofolo *et al.*, 2010). Three canSNP sublineages were identified in Italy. A.Br.008/009 is the dominant canSNP genotype in the country (Garofolo *et al.*, 2011). B.Br.CNEVA and A.Br.005/006 are rare lineages exclusively found in some areas in the Northeast region. In this study, we used a whole-genome sequencing approach which focuses on identifying novel, strain-specific informative SNPs to gain further resolution into the phylogenetic relationships among Italian strains. Twenty-one isolates representative of the major MLVA-genotypes identified within the Italian *B. anthracis* population were sequenced and diagnostic SNPs that discriminate regional sub-groups were identified.

Results

MLVA-15 and canSNP typing

Thirty-one different genotypes circulating in Italy were resolved (Table 1), fourteen contain more than one strain. The geographical repartition of the different genotypes is presented on the map of Italy (Figure 1). The two hundreds of strains of *Bacillus anthracis* (n =220) collected in Italy during the last three decades fall into 31 genotypes distributed in two clusters and one singleton (Figure 2). Most genotypes (28 out of 31) were part of a large, unique clonal complex with very small genetic variations (1 to 2 VNTR loci), suggesting that most strains of *B. anthracis* circulating in Italy are the result of local evolution of a common ancestral strain. Strains belonging to this cluster were found all over the country, but primarily in the Southern and central regions with two-third of recent outbreaks (90 out of 121) occurring in Basilicata, Puglia and Campania (Figures 1 and 2). GT25 formed a singleton loosely connected to the clonal complex. It included one strain restricted to a single area from Veneto region. A second small cluster was constituted by GT14 and GT24. It grouped three genetically distinct strains originated from Veneto and Trentino Regions in the North of Italy.

All strains were subsequently subjected to canSNP typing to place them into a broader phylogenetic context. Strains belonging to CC1 (216 out of 220) were affiliated to the Trans-Eurasian (TEA) A.Br.008/009 canSNP group. An additional canSNP assay designed around the A.Br.011 SNP [4] that further divides the TEA group into two sub-groups (A.Br.008/011 and A.Br.011/009) was also run against the whole collection by HRM (Derzelle, 2014). Most of the TEA Italian strains were classified as A.Br.011/009, seventeen as A.Br.008/011. Only a few isolates were associated with the B.Br.CNEVA and A.Br.005/006 canSNP groups. GT14 and GT24 were composed of three B.Br.CNEVA-affiliated strains; GT25 of one A.Br.005/006 strain.

Whole genome sequencing

Twenty-one “field strains”, representative of MLVA genotypes diversity and anthrax activity in Italy since 20 years, were selected for whole genome sequencing (Table 2). Fifteen of these strains were classified as A.Br.011/009. The panel was also composed of two B.Br.CNEVA-, one A.Br.005/006- and three A.Br.008/011-affiliated strains. All strains were characterized by paired-end whole genome sequencing using Illumina sequencers. After applying the quality filter of the Illumina base-calling pipeline, filter-passed reads were aligned to the Ames Ancestor reference genome, resulting in more than a 115-fold sequencing depth on average and a genome coverage ranging from 98.9 to 99.9% (Table 2).

Whole genome SNP discovery and phylogenetic clustering

Comparative analysis of the Italian genomic sequences was next carried out. Available data for the *B. anthracis* Ames Ancestor (A.Br.001/002), A1055 (C.Br.A1055), KrugerB (B.Br.KrugerB), Carbosap (A.Br.011/009) and USA6153 (A.Br.WNA) strains were also included, as well as thirty French strains belonging to either the B.Br.CNEVA (n = 10) or A.Br.011/009 (n = 20) sub-groups (Girault *et al.*, 2014a). Figure 3 illustrates the minimum spanning tree (MST) generated by the chromosomal whole-genome SNPs data. The two B.Br.CNEVA strains isolated in the Italian Alps (Garofolo *et al.*, 2011) were clearly separated from the French B.Br.CNEVA strains. They formed a distinct branch that split at a distance of 19 SNPs from the basis of the French radiation (BA19, canSNP 2573536). The unique A.Br.005/006 isolate was distantly linked (hundreds of SNPs) to the others canSNP groups within the *B. anthracis* SNP tree. The Italian population structure of Trans-Eurasian strains was resolved into six clusters of distinct geographic distribution (Figure 4). Four ones were part of the A.Br.011/009 sub-group. The two others belong to A.Br.008/011. Strain-specific SNPs could be identified for all particular strains sequenced.

A.Br.008/009 phylogenetic analysis

The Italian A.Br.011/009-affiliated strains were classified in four phylogeographic clades: Basilicata-Campania, Puglia-Lazio-Toscana, Sardegna and Carbosap-like strains. Their precise positioning within the *B. anthracis* SNP tree with respect to French TEA strains is shown in Figures 2 and 3. The four clusters fit into three of the six sub-branches previously described for the French A.Br.011/009 population (Girault *et al.*, 2014a) (Figure 4). Strains isolated from Puglia, Lazio and Toscana belong to the first branch (Branch 1), together with French isolates collected in the Center-North-East quarter of France (CNE clade). The split between Italian and French clusters occurs very shortly afterwards the A.Br.011/009 branching point, one SNP later (SNP 3952578). Branch 2 includes two Italian clusters, Basilicata-Campania and Sardegna, in addition to the French outermost north-eastern strains (NE clade). It is also

composed of two main sub-branches weakly connected to each other (SNP 658847). One corresponds to the Basilicata-Campania clade, the other one includes the Sardegna isolate and the French NE-clade and lead to the A.Br.WNA lineage found in North America. The flast Italian A.Br.011/009 cluster is composed of one Carbosap-like strain isolated from a donkey in Sicilia. The cluster fit into the Branch 3 defined according to the French *B. anthracis* population, together with a few strains isolated close to Paris, two atypical strains of Pasteur vaccine retaining both the pXO1 and pXO2 plasmids (i.e. CIP 74.12 and n.2-17JB (Uchida *et al.*, 1985)) and the Italian Carbosap vaccine (Harrington *et al.*, 2013).

Concerning the two A.Br.008/011 clusters involving strains from Umbria (Branch A) or Sicilia (Branch B), the corresponding branches were observed to radiate at a distance of two SNPs (SNPs 2552486 (A.Br.011) and 1612053) from the TEA branch leading to the A.Br.011/009 branching point. Another information obtained with the genome of the Umbria's strain shows the derived state for three previously published SNPs (15393,3242118 and 5182989) identified in the TEA branch leading to the Hungarian A0417 strain (localized between A.Br.008 and A.Br.A0417) (Birdsell *et al.*, 2012).

Italian cluster-specific SNP discrimination assays

Diagnostic canSNPs specific to each Italian cluster except the A.Br.005/006-affiliated strain were selected and developed into real-time PCR assays (Table 3). Seven SNPs discrimination assays based on Melting Temperature (*T_m*)-shift method or high-resolution melting (unpublished data) were designed to target the six TEA clades and the B.Br.CNEVA strains (Figure 4). Two allele-specific (AS) forward primers containing a 3'-terminal base matching only one of the two expected SNP alleles, a destabilizing mismatch at the penultimate position and a 5' GC-rich tail of unequal length, were designed for each *T_m*-shift assays (Table 3).

The novel canSNPs were validated *in silico* against all *B. anthracis* genomes (n = 32) available in public NCBI database and Sequence Read Archive (SRA). The Italian-specific SNP assays were then successfully used to screen the global collection of *B. anthracis* strains of the Italian Anthrax Reference Center. As expected, the resulting PCR products had a distinct melting temperatures (*T_m*) that depends on the AS primers that was responsible for the amplification, allowing unambiguous identification of each novel canSNP (Table 4). The assays accurately separated the Italian strains of various geographic origins from other world-wide isolates, including the French ones (data not shown).

Discussion

The Trans-Eurasian canSNP group (A.Br.008/009) is a successful subpopulation of *B. anthracis* that predominates throughout Europe, Russia, Kazakhstan, the Caucasus and the Western most Chinese province of Xinjiang (Aikembayev *et al.*, 2010; Merabishvili *et al.*, 2006; Simonson *et al.*, 2009; Van Ert *et al.*, 2007a). Its dispersal is thought to have been mediated by ancient east-west human trade routes, such as the "silk Road" (Simonson *et al.*, 2009; Van Ert *et al.*, 2007a). The origin of the group (e.g. Asia versus Europe) and precise distribution of its two subgroups (A.Br.008/011 and A.Br.011/009) on the Eurasian landmass are still difficult to determine. To solve this issue, it would be necessary to acquire and genotype more Eurasian isolates. The A.Br.008/011 has been reported to comprise isolates

from China, Pakistan, Iran, Turkey, Republic of Georgia, Ukraine, Russia, Hungary, Poland, Albania, Slovakia, Austria and Italy (Price 2012). The A.Br.011/009 sub-group - that led to the Western North American sublineage (A.Br.WNA) - is part of the French *B. anthracis* population (Girault *et al.*, 2014a) and represents about 90% of the strains isolated in Italy (this study).

Ancient introductions of members of the TEA group probably seeded the initial *B. anthracis* population in Italy. Six different phylo-geographic clusters established in Italy have been revealed by the present whole-genome SNP analysis, including strains belonging either to A.Br.008/011 or A.Br.011/009 sub-groups. These genetic data indicate a process of differentiation on a regional scale leading to spatially structured genetic variation in several regions of Italy (Basilicata/Campania, Puglia/Lazio/Toscana, Sardinia, Umbria or Sicily). In contrast to France, two phylogenetic clades affiliated to A.Br.008/011 (termed here Branches A and B) are present in Italy. Branch B includes several Sicilian strains. Branch A is composed of a single isolate from Umbria (Ba152) collected in 2009 during an atypical anthrax outbreak. This peculiar genome shows a derived state of three previously published SNPs specific to the Hungarian strain A0417 (Birdsell *et al.*, 2012). Both “Italian” A.Br.008/011 branches radiate from the TEA branch leading to the A.Br.011/009 sub-group near the A.Br.011/009 branching point (SNPs 2552486 (A.Br.011) and SNP 1612053) (Figure 3). Taking into account that the A.Br.008/011 sub-group appears to be basal to the A.Br.011/009 sub-lineage (Marston *et al.*, 2011), this may reflect an introduction of the TEA group into Italy earlier than in France.

Interestingly, French and Italian strains affiliated to A.Br.011/009 seem to be related to each other, suggesting that they might share some common ancestors. The four Italian clusters fit into three of the six sub-branches previously described for the French A.Br.011/009 population (Girault *et al.*, 2014a). Strains isolated from Puglia, Lazio and Toscana belong to Branch 1, together with isolates collected in the Center-North-East quarter of France (CNE clade). The bifurcation point between Italian and French strains occurs very shortly after the A.Br.011/009 branching point, i.e. one SNP later (SNP 3952578). Branch 2 includes two Italian clades, the Basilicata/Campania and the Sardinia, in addition to the French North-East clade (NE clade). This second branch splits up into two related sub-branches sharing a single SNP near the base of Branch 2 (SNP 658847). The Basilicata-Campania clade forms one of these sub-branches, the other is composed of three sub-groups: the French NE-clade, the Sardinia Ba048 isolate and the A.Br.WNA lineage. Trade exchange between northern Sardinia and France through Corsica has likely occurred in the past, possibly explaining the close genetic relationship observed. The last Italian A.Br.011/009 strain (Ba010) fits into Branch 3 that includes, among others, atypical vaccine strains retaining both the pXO1 and pXO2 plasmids such as Pasteur II strains (CIP 74.12 and n.2-17JB (Uchida *et al.*, 1985)) and the Italian Carbosap vaccine (Harrington *et al.*, 2013). This field strain (Ba010) was isolated in Sicily from a donkey vaccinated during the epidemic that hit the Madonie Park in 1996-1997. The Carbosap vaccine strain was used in Italy from 1949 to 2006. Although such attenuated live vaccines are effective to protect cattle and sheep, it is not unexpected for an occasional livestock animal to die from the vaccine in return for herd immunity (Scorpio *et al.*, 2006) and donkeys are known to be very sensitive to *B. anthracis*.

Another finding revealed by this study is the phylogenetic linkage existing between Branch 2 of A.Br.011/009 (defined according to French *B. anthracis* population (Girault *et al.*, 2014a)), in particular the Ba048 strain isolated in Sardinia, and the derived North American sub-lineage A.Br.WNA that is dominant today in central Canada and much of the western United States. The origin of the Old World ancestor from which derived the A.Br.WNA lineage is a controversial issue. The more ancestral isolates of the A.Br.WNA canSNP group are found in the upper reaches of central Canada and genetic data portrays a migration pattern from North to South (Kenefic *et al.*, 2009). Our results are consistent with a possible introduction to North America from Europe, possibly via French colonization of Canada 500 years ago (Van Ert *et al.*, 2007a), in opposition to the pre-Columbian hypothesis dating back the introduction of anthrax in North America during the late Pleistocene epoch (13,000 ybp) through the Bering Strait (Kenefic *et al.*, 2009). More complete genomic data are clearly required to substantiate any of both hypothesis (e.g. Far East versus Europe, ancient versus recent). Phylogenetic reconstructions by incomplete taxon sampling inherently bias any analysis (Pearson *et al.*, 2004; Pearson *et al.*, 2009). Nevertheless, recent MLVA data on *B. anthracis* strains circulating in Siberia and Far East have highlighted a considerable degree of genetic divergence between those TEA strains and isolates from Europe and Western North America (Afanasev *et al.*, 2013).

Strains belonging to the B.Br.CNEVA lineage are recovered in a relatively small area in the North-East of Italy (Garofolo *et al.*, 2011). The relatively low diversity pinpointed between the two sequenced samples in the study is consistent with a single introduction of the B lineage into the country, followed by ecological establishment and progressive *in situ* differentiation around the Italian Alps area. Consistent with this hypothesis, the Italian strains form a distinct cluster (canSNP BA30) relative to French B.Br.CNEVA-affiliated isolates, including those collected in the Alps. The bifurcation point leading to the Italian strains lies upstream (18 SNPs) of the French radiation. This may reflect an earlier introduction of this lineage into Italy, reinforcing the hypothesis of a dispersal of the ancestral B.Br.CNEVA population further westwards in Europe, with progressively derived local populations of *B. anthracis* strains (Girault *et al.*, 2014a).

The presence of an A.Br.005/006 strain in the Northeastern territories of Italy has been suggested to be related to trade exchanges dating back to the Maritime Republics period (Middle Ages), when city states competed for trade and commerce throughout the Mediterranean (Garofolo *et al.*, 2011). This sub-group is well represented in Africa but rare in Europe (Van Ert *et al.*, 2007a) and could have been imported through contaminated animal products (e.g. wool, hides, or bone meal) from Africa to Northern Italy in the past. In industrialized regions, and particularly in Europe, the population structure of *B. anthracis* is usually largely diversified, resulting from both ancient introductions and more recent importations of distinct genotypes via transport and trade of contaminated commodities across large distances. Italy has always been a land of conquest, where several invaders have come and gone over thousands of years, and an important trading center in the Mediterranean Sea. It is therefore quite surprising that past importations of ill or dying animals or spore-infected items from Africa, the Middle East or Asia did not impacted more the local *B. anthracis* population structure. Exotic genotypes might not have found suitable soils to establish themselves due to the prior presence of strains of the very successful TEA group in geographic areas favorable for *B. anthracis* persistence in much of the country.

Conclusions

NGS and whole-genome SNP discovery allowed the identification of hundreds of SNPs specific to Italian strains. Compared to MLVA, SNP analysis provides further resolution and more reliable phylogenetic reconstruction of the Italian *B. anthracis* population structure. Seven novel assays based on a simple and cost-effective technique (Papp *et al.*, 2003; Wang *et al.*, 2005) have been designed for interrogating a few canonical SNPs. They will be useful to rapidly confirm the geographic origin of any *B. anthracis* strain causing outbreak in Italy.

Materials and Methods

Bacterial strains and biosafety procedures

A total of 220 *B. anthracis* strains belonging to B.Br CNEVA (n = 3), A.Br 011/009 (n = 199), A.Br 008/011 (n = 17) and A.Br 005/006 (n = 1) were used in this study. It included isolates collected during animal or human anthrax outbreaks that have occurred in Italy over the past 30 years (1984- 2012). All *B. anthracis* manipulations were performed in a biosafety level 3 laboratory using class II type A2 biosafety cabinet.

DNA extraction

Genomic DNAs for sequencing were obtained from vegetative cells grown at 37°C on 5% horse blood agar plates. After heat inactivation (98°C for 20 min), DNA was purified using the DNAeasy Blood and Tissue kit (Qiagen), following the protocol for Gram positive bacteria. DNA solutions were transferred to a 0.22 µm sterile Ultrafree-MC spin filter (Millipore) and centrifuged for 2 min at a maximal speed of 8000 × g to ensure the complete removal of live forms of *B. anthracis* from DNA. Viability testing by plate spreading was systematically performed before DNA was taken out of the BSL-3 facility.

Draft whole genome sequencing (WGS) and data analysis

In this study, 21 Italian strains belonging to 4 sublineages (A.Br 008/011 (n = 3), A.Br 011/009 (n = 15), A.Br.005/006 (n = 1) and B.Br CNEVA (n = 2)) were whole genome sequenced. Isolates were subjected to paired-end whole genome sequencing on either the Illumina HiSeq2000 platform (18 samples, paired-end data of 2 x 100 pb) (Illumina Inc., San Diego, CA, USA) or the Illumina MiSeq platform (3 samples, paired-end data of 2 x 250 pb) (Illumina Inc., San Diego, CA, USA). Genome coverage of at least 78 × was obtained (average depth ranging from 78× to 357×).

Ames Ancestor [GenBank: AE017334.2] was used as the reference genome for assembly. Short reads data sets were exported on the FastQ format and mapped to the Ames Ancestor genome and both pXO1 and pXO2 plasmidic sequences using BioNumerics version 6.6 (Applied Maths, Belgium) and Power assembler module asking for a similarity of at least 90%. A set of SNPs was deduced for each genome sequence data using BioNumerics Chromosome Comparisons module. Individual lists were compiled and data filtered to remove SNP positions at which one or more isolate displayed an ambiguous residue call or missing data. Ribosomal operons and VNTR loci were also excluded from the analysis, as well as contiguous SNPs (using a window-frame of 10 pb). The list of canSNP positions specific to Italian clusters is provided in Table 2.

The whole genome sequences for *B. anthracis* strains CNEVA-9066 (B.Br.CNEVA) [NZ_AAEN000000000.1], A0465 (B.Br.CNEVA) [NZ_ABLH000000000.1], Ames ancestor

(A.Br.Ames) [NC_ 007530.2], Tsiankovskii (A.Br.008/011)[ABDN000000000.2], WNA USA6153 (A.Br.WNA) [AAER000000000.1], Carbosap (A.Br.011/009)[ANAO000000000.1], Ames ancestor and A1055 (C.Br.A1055) [NZ_AAE000000000.1] strains can be found in the NCBI microbial genome website at <http://www.ncbi.nlm.nih.gov>.

Whole Genome Phylogenetic Analysis

A minimum spanning tree was drawn in BioNumerics by using the filtered whole genome sequencing SNP data as input. Nodes were numbered by BioNumerics. CanSNPs along the seven branches leading to the Italian strains were identified from whole genome sequencing data by searching for SNPs with allelic states shared only by these different subgroups.

SNP discrimination assays by *T_m*-shift allele-specific (AS) PCR assays

We designed allelic-discrimination assays for seven Italian-specific SNPs based on *T_m*-shift method. Two forward AS primers, with the 3' base matching only one of the SNP allele bases, and a common reverse primer were designed for each SNP using Primer 3⁺ software. In order to enhance allele specificity, a destabilizing mismatch at the 3' penultimate position of the AS primers was introduced following the guidelines for mismatch amplification mutation assay (MAMA) presented by Li et al. (Li *et al.*, 2004). A longer GC-rich tail was attached to the AS primer which have the higher *T_m* base at its 3' end (G or C allele) (Table 2). The positions of these SNPs in the Ames Ancestor genome [GenBank: AE017334.2] and the primers sequences used are listed in Table 2.

Amplification and melting analysis were performed on the ViiA™ 7 real-time PCR system (Life Technologies). The 10 µl-reaction mixture consisted of 0.2 µM of each primer and 1x Power SYBR® Green PCR Master Mix (Applied Biosystems). Reactions were raised to 50°C for 2 min, 95°C for 10 min and then cycled at 95°C for 15 s and 60°C for 1 min for 40 cycles. Melting curves were obtained using the ViiA™ 7 onboard software (version 1.2.1) and adding a linear denaturation ramp from 60°C to 90°C at the end of the PCR program, with 25 acquisitions/°C. *T_m* values were used for allelic discrimination purposes. PCR amplification and melting analysis were also conducted with the xxxx® real-time PCR system using the same conditions. All primers were purchased from MWG (MWG Biotech, Germany).

Competing interests

The authors declare that they have no competing interests

Authors' contributions

AF and SD conceived the experiments and wrote the paper. GG and SD analyzed the NGS data. AA and AA performed the SNPs discrimination assays experiments. All authors read and approved the final manuscript.

Acknowledgements

This work was supported by the French National Research Agency (ANR) and Direction Générale de l'Armement (DGA) (project number ANR11ASTR0007). GG is a PhD student co-supported by Anses and DGA grants.

This work has benefited from the facilities and expertise of the high throughput sequencing platform of IMAGIF (Centre de Recherche de Gif [<http://www.imagif.cnrs.fr>]).

References

- Afanasev, M., Kravets, E., Kugorjapova, Z., Balakhonov, S. 2013. The first molecular snapshot of the *Bacillus anthracis* strains circulating in Siberia and Far East. (Medical Biodefense Conference 2013, Munich, Germany).
- Aikembayev, A.M., Lukhnova, L., Temiraliyeva, G., Meka-Mechenko, T., Pazylov, Y., Zakaryan, S., Denissov, G., Easterday, W.R., Van Ert, M.N., Keim, P., Francesconi, S.C., Blackburn, J.K., Hugh-Jones, M., Hadfield, T., 2010. Historical distribution and molecular diversity of *Bacillus anthracis*, Kazakhstan. *Emerg Infect Dis* 16, 789-796.
- Birdsell, D.N., Pearson, T., Price, E.P., Hornstra, H.M., Nera, R.D., Stone, N., Gruendike, J., Kaufman, E.L., Pettus, A.H., Hurbon, A.N., Buchhagen, J.L., Harms, N.J., Chanturia, G., Gyuranecz, M., Wagner, D.M., Keim, P.S., 2012. Melt analysis of mismatch amplification mutation assays (Melt-MAMA): a functional study of a cost-effective SNP genotyping assay in bacterial models. *PLoS ONE* 7, e32866.
- Brachman, P.S., Kaufman, A.F., Dalldorf, F.G., 1966. Industrial Inhalation Anthrax. *Bacteriological reviews* 30.
- Derzelle, S. 2014. Single nucleotide polymorphisms discrimination using high-resolution melting analysis for the genotyping of *Bacillus anthracis*, In: Inácio, M.C.a.J. (Ed.) *Molecular Diagnosis in Veterinary Laboratory Practice* Humana Press, New York.
- Fasanella, A., Palazzo, L., Petrella, A., Quaranta, V., Romanelli, B., Garofolo, G., 2007. Anthrax in Red Deer (*Cervus elaphus*), Italy. *Emerging Infectious Diseases* 13, 1118-1119.
- Fasanella, A., Van Ert, M., Altamura, S.A., Garofolo, G., Buonavoglia, C., Leori, G., Huynh, L., Zanecki, S., Keim, P., 2005. Molecular diversity of *Bacillus anthracis* in Italy. *Journal of clinical microbiology* 43, 3398-3401.
- Garofolo, G., Ciammaruconi, A., Fasanella, A., Scasciacchia, S., Adone, R., Pittiglio, V., Lista, F., 2010. SNR analysis: molecular investigation of an anthrax epidemic. *BMC veterinary research* 6, 11.
- Garofolo, G., Serrecchia, L., Corro, M., Fasanella, A., 2011. Anthrax phylogenetic structure in Northern Italy. *BMC research notes* 4, 273.
- Girault, G., Blouin, Y., Vergnaud, G., Derzelle, S., 2014. High-throughput sequencing of *Bacillus anthracis* in France: investigating genome diversity and population structure using whole-genome SNP discovery. *BMC genomics* 15.
- Harrington, R., Ondov, B.D., Radune, D., Friss, M.B., Klubnik, J., Diviak, L., Hnath, J., Cendrowski, S.R., Blank, T.E., Karaolis, D., Friedlander, A.M., Burans, J.P., Rosovitz, M.J., Treangen, T., Phillippy, A.M., Bergman, N.H., 2013. Genome Sequence of the Attenuated Carbosap Vaccine Strain of *Bacillus anthracis*. *Genome announcements* 1, 1-2.
- Hugh-Jones, M.E., De Vos, V., 2002. Anthrax and wildlife. *Rev. sci. tech. Off. int. Epiz.* 21, 359-383.
- Kenefic, L.J., Pearson, T., Okinaka, R.T., Schupp, J.M., Wagner, D.M., Hoffmaster, A.R., Trim, C.B., Chung, W.K., Beaudry, J.A., Jiang, L., Gajer, P., Foster, J.T., Mead, J.I., Ravel, J., Keim, P., 2009. Pre-Columbian origins for North American anthrax. *PLoS ONE* 4, e4813.
- Li, B., Kadura, I., Fu, D.-J., Watson, D.E., 2004. Genotyping with TaqMAMA. *Genomics* 83, 311-320.
- Marston, C.K., Allen, C.A., Beaudry, J., Price, E.P., Wolken, S.R., Pearson, T., Keim, P., Hoffmaster, A.R., 2011. Molecular epidemiology of anthrax cases associated with recreational use of animal hides and yarn in the United States. *PLoS ONE* 6, e28274.
- Merabishvili, M., Natidze, M., Rigvava, S., Brusetti, L., Raddadi, N., Borin, S., Chanishvili, N., Tediashvili, M., Sharp, R., Barbeschi, M., Visca, P., Daffonchio, D., 2006. Diversity of *Bacillus anthracis* strains in Georgia and of vaccine strains from the former Soviet Union. *Applied and environmental microbiology* 72, 5631-5636.
- Palazzo, L., De Carlo, E., Santagada, G., Serrecchia, L., Aceti, A., Guarino, A., Adone, R., Fasanella, A., 2012. Recent Epidemic-Like Anthrax Outbreaks in Italy: What Are the Probable Causes? *Open Journal of Veterinary Medicine* 2, 74-76.

- Papp, A.C., Pinsonneault, J.K., Cooke, G., Sadée, W., 2003. Single nucleotide polymorphism genotyping using allele-specific PCR and fluorescence melting curves. *Biotechniques* 34, 1068-1072.
- Pearson, T., Busch, J.D., Ravel, J., Read, T.D., Rhoton, S.D., U'Ren, J.M., Simonson, T.S., Kachur, S.M., Leadem, R.R., Cardon, M.L., Van Ert, M.N., Huynh, L.Y., Fraser, C.M., Keim, P., 2004. Phylogenetic discovery bias in *Bacillus anthracis* using single-nucleotide polymorphisms from whole-genome sequencing. *Proceedings of the National Academy of Sciences of the United States of America* 101, 13536-13541.
- Pearson, T., Okinaka, R.T., Foster, J.T., Keim, P., 2009. Phylogenetic understanding of clonal populations in an era of whole genome sequencing. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* 9, 1010-1019.
- Schmid, G., Kaufmann, A., 2002. Anthrax in Europe: its epidemiology, clinical characteristics, and role in bioterrorism. *Clin Microbiol Infect* 8, 479-488.
- Simonson, T.S., Okinaka, R.T., Wang, B., Easterday, W.R., Huynh, L., U'Ren, J.M., Dukerich, M., Zanecki, S.R., Kenefic, L.J., Beaudry, J., Schupp, J.M., Pearson, T., Wagner, D.M., Hoffmaster, A., Ravel, J., Keim, P., 2009. *Bacillus anthracis* in China and its relationship to worldwide lineages. *BMC microbiology* 9, 71.
- Uchida, I., Sekizaki, T., Hashimoto, K., Terakado, N., 1985. Association of the Encapsulation of *Bacillus anthracis* with a 60 Megadalton Plasmid. *Journal of General Microbiology* 131.
- Van Ert, M.N., Easterday, W.R., Huynh, L.Y., Okinaka, R.T., Hugh-Jones, M.E., Ravel, J., Zanecki, S.R., Pearson, T., Simonson, T.S., U'Ren, J.M., Kachur, S.M., Leadem-Dougherty, R.R., Rhoton, S.D., Zinser, G., Forlow, J., Coker, P.R., Smith, K.L., Wang, B., Kenefic, L.J., Fraser-Liggett, C.M., Wagner, D.M., Keim, P., 2007. Global Genetic Population Structure of *Bacillus anthracis*. *Plos ONE*.
- Wang, J., Chuang, K., Ahluwalia, M., Patel, S., Umblas, N., Mirel, D., Higuchi, R., Germe, S., 2005. High-throughput SNP genotyping by single-tube PCR with Tm-shift primers. *Biotechniques* 39, 885-893.

Figure legends

Figure 1. Map of Italy. Geographical distribution of the anthrax outbreaks (n=121) occurred in the last 20 years and associated MLVA-15 genotypes (GT). The most affected region is the region of Basilicata with 62 outbreaks reported, followed by Puglia and Sicily with 17 each and Campania with 11 outbreaks.

Geographical localization and SNP typing of the 21 strains of *Bacillus anthracis* whole-genome sequenced are also indicated by colored circles (red, B.Br.CNEVA; grey: A.Br.005/006); yellow-orange, A.Br.008/011; Blue-gree and violet: A.Br.011/009).

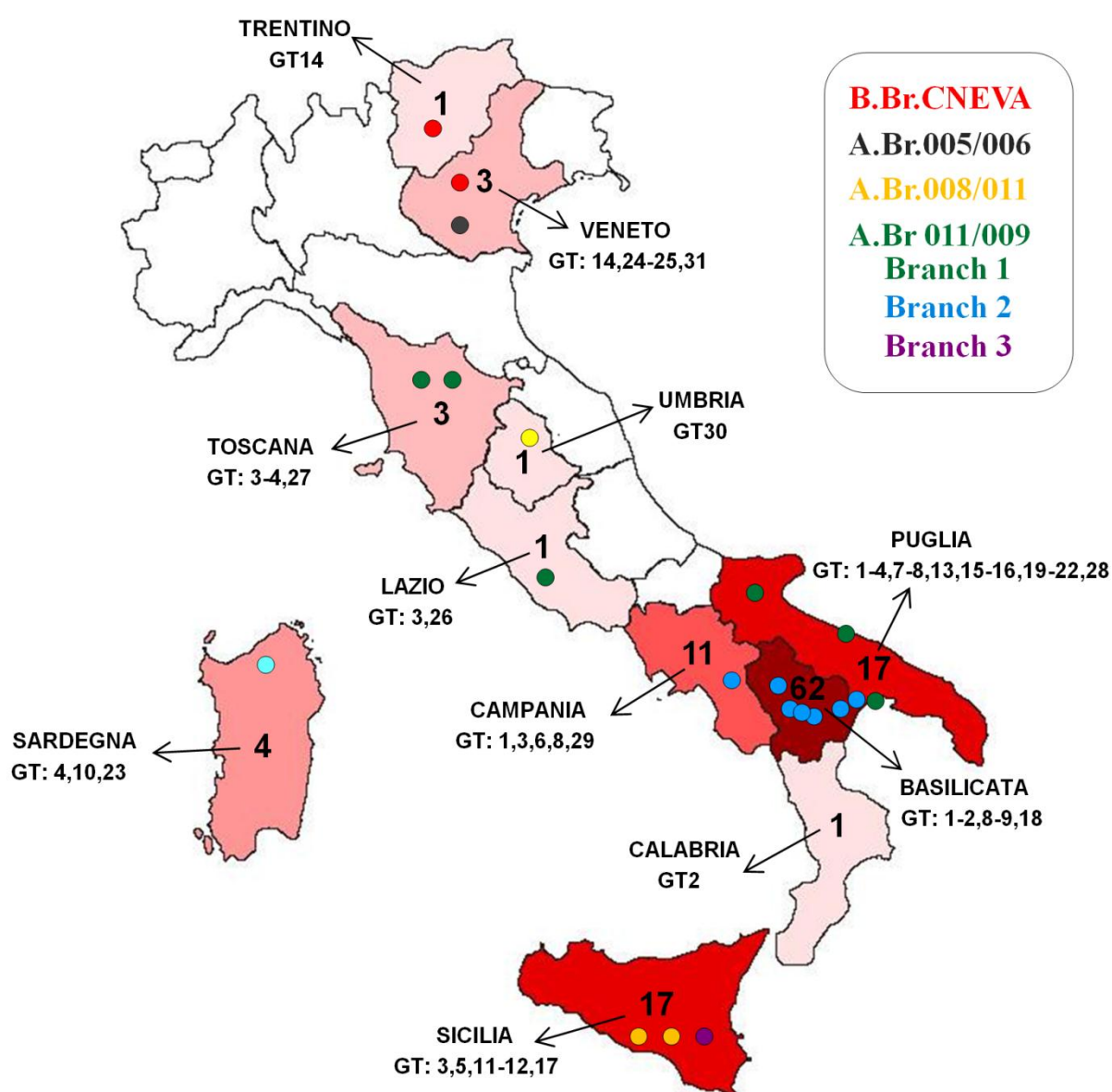


Figure 2. Minimum spanning tree of 220 Italian *B. anthracis* strains based on categorical analysis of 15 VNTRs. Each circle represents a unique genotype. The diameter of each circle varies according to the number of isolates having the same genotype. Genotypes connected by a shaded background differ by a maximum of 2 of the 15 VNTR markers and could be considered as a “clonal complex”. Thick and regular connecting lines represent a difference of one and two markers, respectively; thin interrupted lines represent a difference of three markers; thin dashed lines represent four or more differences. The length of each branch is also proportional to the number of differences.

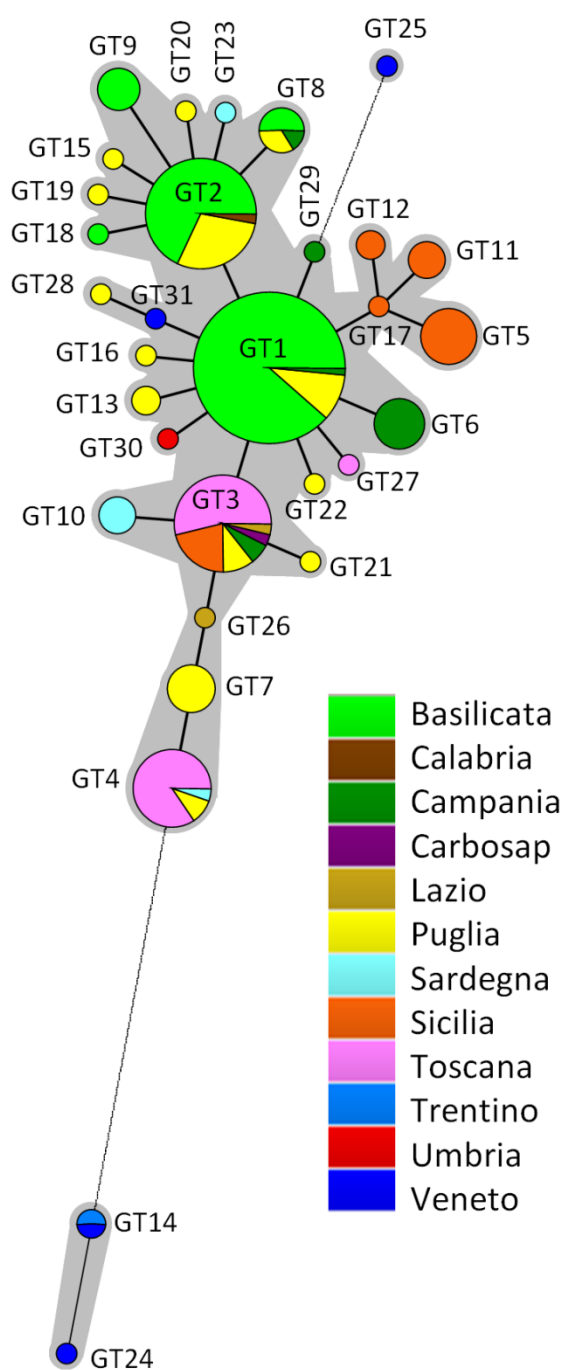


Figure 3. Minimum spanning tree of Italian and French *B. anthracis* strains based on whole-genome SNP analysis. Data are based on 5317 chromosomal SNPs. Red and blue circles represent Italian and French strains, respectively. Positions of the *B. anthracis* Tsiankovskii, Ba4599, UR-1, Ba3166, Ba3154, USA6153, A0174, A0193, Carbosap, Sen2Col2, Sen3, Gmb1, Ames Ancestor, CZC5, KrugerB, A0442, SVA11 and A1055 strains are also marked (black circles). The different canSNP groups are color-coded. The length of each branch is proportional (logarithmic scale) to the number of SNPs identified between strains. Indicated in red or blue are the position and name of the selected canSNPs specific to Italy or France, respectively.

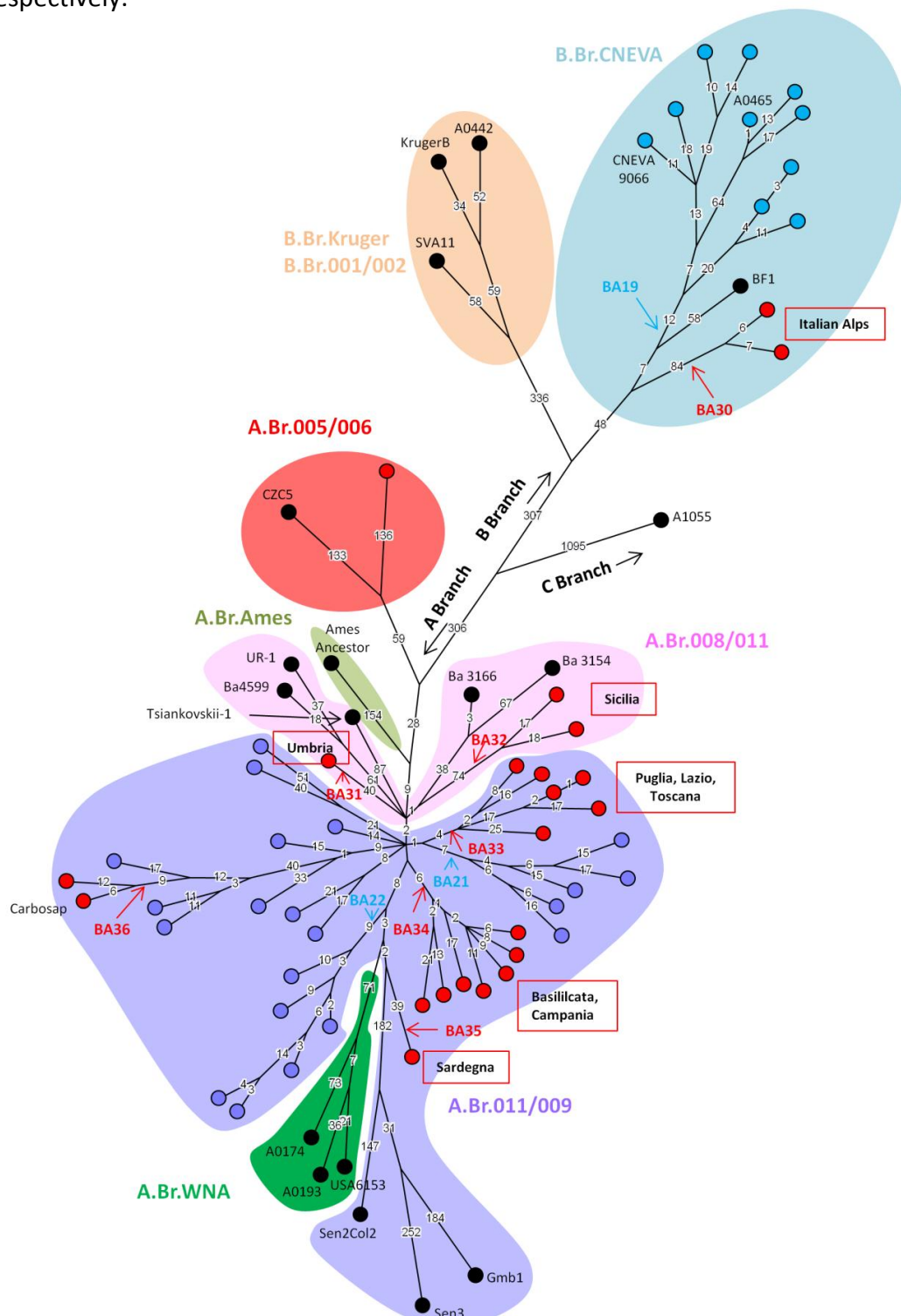


Figure 4. Minimum spanning tree based on whole-genome SNP analysis of Italian and French *B. anthracis* strains affiliated to the TransEurasian canSNP group. The six A.Br.011/009 branches are color-coded: Branch 1 in green, Branch 2 in red, Branch 3 in blue, Branch 4 in light blue, Branch 5 in pink and Branch 6 in yellow. The A.Br.008/011 and A.Br.WNA canSNP groups are indicated in purple and brown, respectively. Positions of the *B. anthracis* Tsiankovskii, WNA USA6153, Carbosap, CIP74.12 and n.2-17JB strains are also marked. Italian strains are represented by triangles and French strains by circles. Each circle or triangle represents a unique SNP genotype. The length of each branch is proportional (logarithmic scale) to the number of SNPs identified between strains.

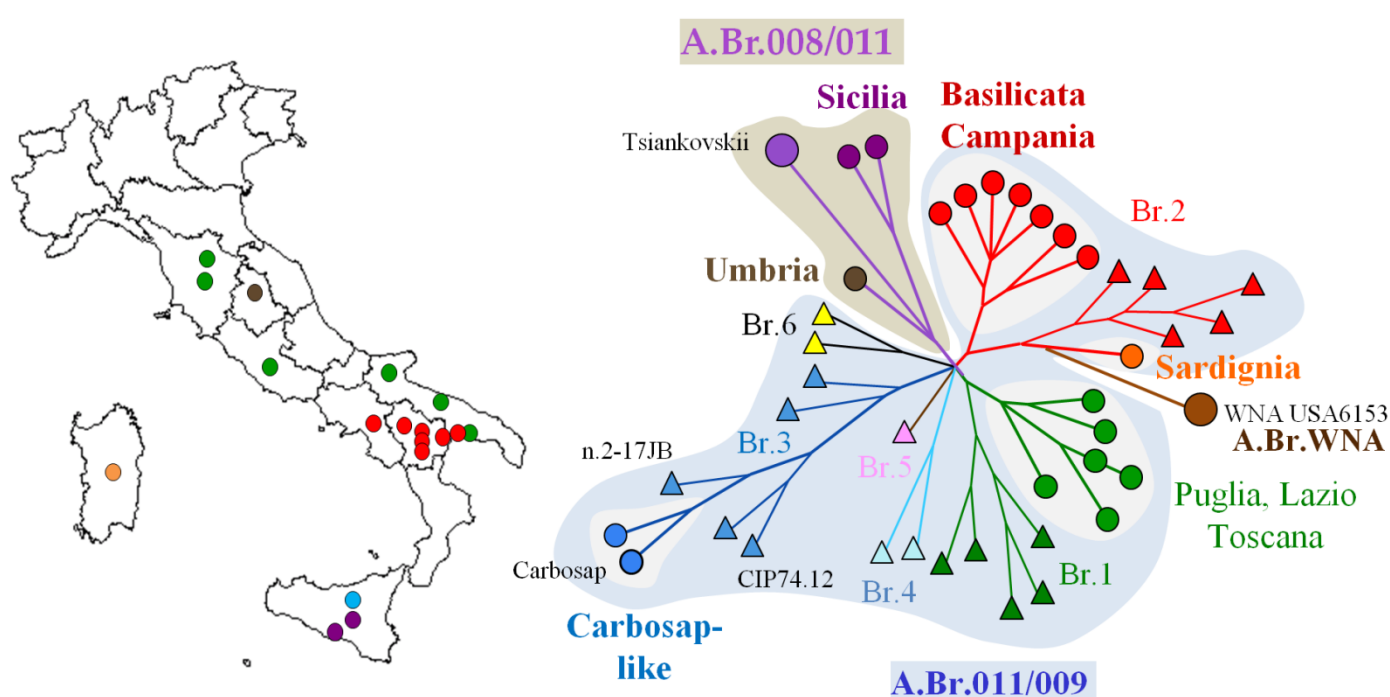


Table 1.

List of the genotypes of *Bacillus anthracis* circulating in Italy up to 2014 (MLVA with 15 VNTRs)

Origin/Region	N. strains	pXO1	vrro2	vrro1	vrro	vrro1	pXO2	vrro2	cg3	Vntr 19	vntr16 (pxo2)	vntr 32	vntr 12	vntr17 (pxo1)	vntr 35	vntr 23	GENOTYPE
Basilicata (62) - Puglia (7) - Campania (1)	70	121	154	223	306	620	133	594	148	93	262	501	110	381	109	193	GT 1
Basilicata (25) - Puglia (11) - Calabria (1)	37	121	154	223	306	620	133	594	148	90	262	501	110	381	109	193	GT 2
Toscana (15) - Sicilia (6) - Puglia (3) - Campania (2) - Lazio (1) CARBOSAP	28	124	154	223	306	620	133	594	148	93	262	501	110	381	109	193	GT 3
Toscana (16) - Puglia (2) - Sardegna (1)	19	124	154	223	306	620	133	594	148	93	271	501	110	390	116	193	GT 4
Sicilia	9	119	154	223	283	620	133	594	148	93	262	501	110	381	109	193	GT 5
Campania	8	121	154	223	306	620	133	522	148	93	262	501	110	381	109	193	GT 6
Puglia	7	124	154	223	306	620	133	594	148	93	271	501	110	390	109	193	GT 7
Basilicata (3) - Puglia (2) - Campania (1)	6	121	154	223	306	620	133	594	148	90	262	501	110	390	109	193	GT 8
Basilicata	5	121	164	223	306	620	135	594	148	90	262	501	110	381	109	193	GT 9
Sardegna (4)	4	124	154	223	306	620	133	594	148	93	262	501	113	381	109	193	GT 10
Sicilia	4	121	154	223	283	620	133	594	148	93	262	501	110	381	116	193	GT 11
Sicilia	2	121	154	223	283	620	133	594	148	93	271	501	110	381	109	193	GT 12
Puglia	2	121	154	223	306	620	135	594	148	93	262	501	110	381	109	193	GT 13
Trentino (1) - Veneto (1)	2	124	154	179	295	584	131	522	153	93	280	521	108	374	116	182	GT 14
Puglia	1	121	154	223	306	620	139	594	148	90	262	501	110	381	109	193	GT 15
Puglia	1	121	146	223	306	620	133	594	148	93	262	501	110	381	109	193	GT 16
SICILIA	1	121	154	223	283	620	133	594	148	93	262	501	110	381	109	193	GT 17
Basilicata	1	121	154	223	306	620	133	594	148	90	255	501	110	381	109	193	GT 18
Puglia	1	121	154	223	306	620	133	594	148	90	143	501	110	381	109	193	GT 19
Puglia	1	121	154	223	306	620	133	594	148	90	262	501	110	398	109	193	GT 20
Puglia	1	124	145	223	306	620	133	594	148	93	262	501	110	381	109	193	GT 21
Puglia	1	121	154	223	306	620	133	594	148	93	255	501	110	381	109	193	GT 22
Sardegna	1	124	154	223	306	620	133	594	148	90	262	501	110	381	109	193	GT 23
Veneto	1	127	154	214	295	584	131	522	153	93	295	521	108	374	116	182	GT 24
Veneto	1	121	164	223	318	620	133	594	153	93	271	461	110	381	109	182	GT 25
Lazio	1	124	154	223	306	620	133	594	148	93	262	501	110	390	109	193	GT 26
Toscana	1	121	154	223	306	620	131	594	148	93	262	501	110	381	109	193	GT 27
Puglia	1	121	154	223	306	620	133	594	148	93	271	501	110	390	109	193	GT 28
Campania	1	121	154	223	306	620	133	594	148	93	262	501	110	381	109	182	GT 29
Umbria	1	130	154	223	306	620	133	594	148	93	262	501	110	381	109	193	GT 30
Veneto	1	121	154	223	306	620	133	594	148	93	262	501	110	390	109	193	GT 31

Table 2. Whole-sequenced strains and NGS data

Strain	canSNP group	Reads (pb)	Unambiguous bases	Genome coverage*	Sequencing depth	Region	years
Ba050	A.Br.005/006	100	5225980	99,972	116	Veneto	1995
Ba187	B.Br.CNEVA	100	5214171	99,747	96	Trentino	2006
Ba188		100	5214813	99,759	82	Veneto	2007
Ba007	A.Br.008/011	100	5224728	99,949	125	Sicilia	1996/1997
Ba152		100	5226864	99,989	116	Umbria	2009
Ba180		100	5226515	99,983	104	Sicilia	2009
Ba005	A.Br.011/009	100	5226958	99,991	125	Puglia	< 1993
Ba010		250	5169504	98,892	143	Sicilia	1996/1997
Ba019		100	5226895	99,990	127	Basilicata	1999
Ba031		250	5225465	99,963	194	Puglia	< 11/1984
Ba039		250	5226425	99,981	357	Puglia	1972
Ba040		100	5226707	99,986	119	Basilicata	< 11/1984
Ba048		100	5226980	99,992	131	Sardegna	2000
Ba051		100	5226956	99,991	114	Lazio	< 1999
Ba053		100	5226836	99,989	111	Toscana	< 1999
Ba079		100	5226820	99,989	132	Basilicata	2004
Ba137		100	5226980	99,992	133	Puglia	2006
Ba157		100	5226965	99,991	141	Toscana	2009
Ba297		100	5226673	99,986	78	Campania	2011
Ba315		100	5226759	99,987	107	Basilicata	2011
Ba430		100	5226815	99,988	115	Basilicata	2012

* Genome coverage are based on the Ames Ancestor genome size (5227419 pb).

Table 3. Italian specific canSNPs and primer sequences used for HRM analysis

canSNP	position*	Target	SNP	AS Forward primers (5'-3')	Common Reverse primer (5'-3')
BA30 B2_IT	1074648	Northern Italy (B.CNEVA)	G to A	gcggCGGTTAGAAATTAAGAAAGATGCacA cgggcaggagCGGTTAGAAATTAAGAAAGATGCatG	GAGTTGCCAGTGGTTTTGT
BA31 A08/1_IT	240337	Umbria (A.008/011)	C to A	gcggCCTTGAAAACGTAGCGTTGCcT gcgggcagggcgCCTTGAAAACGTAGCGTTGCgG	TGTTGAAAGAATTCGGAGTG
A08/1KEIM _IT	3242118		A to G	gcgggcagggcgGAGCGGTTTTAAGTACAGTACTCATTTTtG GAGCGGTTTTAAGTACAGTACTCATTTTcA	TAATCCGCTGCTGTTGCTGA
BA32 A08/2_IT	1037495	Sicilia (A.008/011)	G to A	gcggCGAAATGATTTTCAAGTTAGTTAGTGCTaT GCgggctgggcgCGAAATGATTTTCAAGTTAGTTAGTGCTgC	GCTAGCCTCCTTATTGTTATATATTGC T
BA33 A11/1_IT	2752660	Toscana to Puglia (A.011/009)	G to A	gcggCTTTCACTTTATCGCTAGAAACACGATcA gcgggcagggcgCTTTCACTTTATCGCTAGAAACACGATtG	AAGGGAATAGCTATATCTGAATTTGA A
BA34 A11/2B_IT	3213630	Basilicata, Campania (A.011/009)	A to G	gcggCTGAATCGTTTTTGAATAAACCTTTACTATaA gcgggcagggcgCTGAATCGTTTTTGAATAAACCTTTACTATaC	TGAAATCAGTTGACATTACCA
BA35 A11/2C_IT	910017	Sardegna (A.011/009)	A to G	gcggGATGCTTTCAAGTAATGATGAAAGtA gcgggcagggcgGATGCTTTCAAGTAATGATGAAAGaG	GCTAATAACGAGCTTTGGCATC
BA36 A11/3A1/IT	3821849	Sicilia (A.011/009)	T to C	In revision	In revision

* Localisation on the Ames Ancestor chromosome (GenBank accession no. AE017334.2).

(Table 4. Melting temperatures determined for Italian-specific canSNP by *T_m*-shift AS-PCR assays)

canSNP	Region	Allele	T _m values (°C)
BA30	North-East (Italian Alps)	A	In progress
	Others	G	In progress
BA31	Umbria	A	In progress
	Others	C	In progress
BA32	Sicilia (A.008/011)	A	In progress
	Others	G	In progress
BA33	Puglia, Lazio, Toscana	A	In progress
	Others	G	In progress
BA34	Basilicata, Campania	G	In progress
	Others	A	In progress
BA35	Sardegna	G	In progress
	Others	A	In progress
BA36	Sicilia (A.008/011)	C	In progress
	Others	T	In progress

2. Cas importés dans la filière du textile

Article 4:

Genomic analysis of *Bacillus anthracis* from the wool industry highlights the polymorphism of lineage A.Br.Vollum

Article soumis à *BMC Genomics* le 29/10/2014

Genomic analysis of *Bacillus anthracis* from the wool industry highlights the polymorphism of lineage A.Br.Vollum

Guillaume GIRAULT¹, Joachim FREY², Pierre WATTIAU³ and Sylviane DERZELLE^{1*}

¹ University Paris-Est, Anses, Animal Health Laboratory, Av du Général De Gaulle 23, 94706 Maisons-Alfort, France

² Institute of Veterinary Bacteriology, Vetsuisse, University of Bern, Laenggassstrasse 122, 3001 Bern, Switzerland.

³ Veterinary & Agrochemical Research Centre, Dpt. of Bacteriology, CODA-CERVA, 99 Groeselenberg, B-1180 Brussels, Belgium

*corresponding author

Animal Health Laboratory of Maisons-Alfort; ANSES

23 Avenue du Général de Gaulle, 94706 Maisons Alfort cedex, France

Phone: +33 1 49 77 26 25 ; E-mail : sylviane.derzelle@anses.fr

guillaume.girault@anses.fr

joachim.frey@vetsuisse.unibe.ch

Pierre.Wattiau@coda-cerva.be

sylviane.derzelle@anses.fr

Abstract

Background. The advent of massively parallel whole genome sequencing provides an opportunity to get detailed insight into the genetic diversity and population structure of monomorphic pathogens such as *Bacillus anthracis*. Here we report the use of next-generation sequencing (NGS) technology to characterize fourteen *B. anthracis* strains isolated in two European wool factories processing goat and sheep raw wool from various regions of the Middle East and northern Asia.

Results. Comparative analyses were carried out using a total of eighteen genomes affiliated to the A.Br.Vollum lineage. The 745 chromosomal single nucleotide polymorphisms (SNPs) defining the phylogenetic relationships existing between these strains split the A.Br.Vollum lineage into five different branches. The fourteen strains from wool processing facilities clustered into three of these branches that are distinct from the branch of the terminal reference strain A0488 of lineage A.Br.Vollum. Six SNPs, representative of these phylogenetic sub-groups, were selected and developed into high-resolution melting SNP discriminative assays. Whole-genome SNP analysis places the recently published strain H9401 in a distinct evolutionarily lineage that is basal to A.Br.Vollum.

Conclusions. A NGS dataset of 14 strains that originate from Middle East and North Asia affiliated to the A.Br.Vollum lineage provides new insight into the substructure of this lineage. Novel analytic methods specific to the A.Br.Vollum lineage were developed for the purpose of rapid molecular epidemiology.

Keywords :

Bacillus anthracis, Whole genome sequencing, Single nucleotide polymorphism, Molecular typing, Comparative genomics, HRM

Introduction

Anthrax is an acute disease caused by the bacterium *Bacillus anthracis*. *B. anthracis* form spores that can remain dormant for years in many types of soil all over the world. Anthrax commonly infects wild and domesticated herbivorous mammals that ingest or inhale spores while grazing [1]. Once they entered the host, the endospores germinate and then spread by the circulation to the lymphatics. When they gain access to the circulating blood, the bacteria multiply and typically kill the animal within a few days or weeks. Production of the powerful tripartite protein toxins by *B. anthracis*, the lethal toxin and the edema toxin with a common adhesion subunit, causes cell death and tissue swelling. Anthrax is often fatal in cattle, sheep and goats.

Occupational exposure to infected animals or their contaminated products, such as hides, wool, hair, or bones, is the usual pathway of exposure for humans besides consumption of meat from infected animals. In the nineteenth and early twentieth centuries, many workers who dealt with wool and animal hides were routinely exposed to anthrax spores. Spores were brought into factories with organic matter (blood clots and skin fragments) that was contaminating the animal fibers. This industrial anthrax, including a dangerous form of inhalational anthrax, became known as "wool sorter's" or "rag picker's" disease. It was a serious threat in the 1830s when various Eastern wools such as alpaca, mohair or Cashmere goat hair were imported for use as textile fibers [2,3].

Today, this form of infection is extremely rare in Europe, largely due to widespread animal vaccination during the 20th century, better decontamination procedures and the fact that raw wool and goat hair processing have almost disappeared from Western industrialized countries. In Europe, only a very few factories still exist in regions with historical textile industries where sheep wool and goat hair originating from areas where anthrax is endemic, remains a high biological risk. Currently West Africa is the most affected area of the world followed by the traditional anthrax belt that stretches from the Middle East into Central Asia [1].

In Belgium, an industrial plant performing middle-scale cleaning and disinfection of wool and goat hair from all over the world has been operating since 1890 [4]. Microbiological and molecular epidemiological study of contaminating bacteria in this factory confirmed the presence of anthrax spores in untreated wastewater, air filter dust and raw goat hair fibers originating from Middle Eastern countries and Northern Asia [5]. However, no clinical case of anthrax had ever been recorded among the employees of this factory. In Switzerland, iterative cases of anthrax reported between 1978 to 1981 involved workers in a textile plant processing cashmere wool. Within 3 years,

25 workers had contracted anthrax. Twenty four cases had the cutaneous form and one had inhalation anthrax of which all the patients recovered [6,7].

B. anthracis has the particularity to solely replicate during short periods of time while infecting the host, which is estimated to last 20 – 40 bacterial generations [8], followed generally by long dormant periods where the bacteria survive as spores without replicating. Hence *B. anthracis* evolves very slowly and represents a genetically extremely homogenous, monomorphic species which makes it an ideal organism to study vertical evolution over extended periods of time.

Analysis of the global genetic population structure of over 1,000 strains of *B. anthracis* collected from all over the world using canonical single polymorphisms (canSNPs) defined the tree previously recognised major genetic clades A, B, and C that were further subdivided into 12 distinct SNP lineages [9,10]. SNP lineage A.Br.Vollum corresponding to MLVA cluster A4 reveals to contain a particularly large number of strains, mostly isolated from humans from various, most distantly related countries world-wide. This lineage could be traced back to strains originating from the Middle-East – North Asian area by carefully defined strains that infected workers from wool factories in Switzerland and Belgium or strains directly isolated from Cashmere wool or wool imported from Middle-east - North Asia that was processed in these factories, since all these were found to belong to phylogenetic cluster A4, lineage A.Br.Vollum [5,11].

To gain further resolution into the phylogenetic relationships of strains belonging to lineage A.Br.Vollum which contains the type strain of the species *Bacillus anthracis*, strain Vollum (NCTC10340^T; ATCC14578^T) and generally strains from wool processing industry or infections of humans related to this activity, fifteen selected strains from this cluster have been sequenced in this study, using parallel sequencing technology. Comparative genomics, with a focus on SNPs discovery, were conducted using three additional strains belonging to lineage A.Br.Vollum and 37 strains belonging to other lineages to obtain a deep insight into the worldwide population structure of this cluster. Identification of thousands of SNPs retrieved from compiled NGS sequences provided the highest possible resolution. Six SNPs that specifically discriminate sub-groups of strains within the five sub branches of lineage A.Br.Vollum were identified and developed into PCR-based genotyping assays.

Results

canSNP typing and whole genome sequencing

Fourteen strains isolated in 1981 from a wool-processing factory in northern Switzerland and from infected workers thereof (n = 6) and in 2006 from a factory located in Belgium (n = 8) that both imported bales from the Middle East and Northern Asia were selected for whole genome sequencing, as well as one additional collection strain dating from 1948 and isolated in Bradford, an area of traditional British wool industry (Table 1). All strains were first subjected to canSNP typing [12] which confirmed their placement into the A.Br.Vollum canSNP lineage [9]. The fifteen strains were then characterized by paired-end whole genome sequencing and parallel assembly. The Illumina sequencers produced 0.65 to 8.85 million reads per strain after applying the quality filter of the Illumina base-calling pipeline. The filter-passed reads were aligned to the Ames Ancestor reference genome, resulting in more than a 112-fold sequencing depth on average and a genome coverage ranging from 99.67 to 99.98 % (Table 1).

Whole genome SNP discovery and phylogenetic clustering among *B. anthracis* strains

Comparative analysis of the genomic sequences of these strains was carried out using genomic data of 40 additional *B. anthracis* strains, 32 belonging to various lineages of clade A, seven of clade B and one of clade C (Table 2). A total of 6158 SNPs were identified among this total dataset and used to draw the phylogeny of the strains from wool processing industry in comparison with the major known clades and lineages. Figure 1 illustrates the minimum spanning tree (MST) generated by the chromosomal whole-genome SNPs analysis. In general, all strains from the wool processing industry and human infections therefrom and the additional NCTC07752 strain (Bradford strain) cluster in lineage A.Br.Vollum together with the other sequenced strains Vollum, A0488 and CDC684 (Figure 1).

Interestingly, this analysis revealed that the Korean H9401 strain is clustering in close proximity with the A.Br.Vollum lineage (characterized by a derived nucleotide for the A.Br.007 SNP position). H9401 splits out from the branch leading to the A.Br.Vollum lineage at a distance of 7 SNPs from the central and most basal node within the A clade. The distance to the tip of this unique new lineage representative (tentatively termed here A.Br.H9401) is approximately 260 SNPs.

A.Br.Vollum phylogenetic analysis

Omitting the genomic data of strain H9401, a total of 745 unique chromosomal SNPs resolved the eighteen A.Br.Vollum-affiliated isolates (Figure 2). Additional SNPs (31 in pXO1 and 23 in pXO2) were also identified in both plasmids resulting in five sub-branches (Figure 1 and 2). All strains sequenced in this study belong to four sub-branches, distinct from the branch of the A.Br.Vollum terminal reference genome A0488 (sub-branch a) (Figure 2). The wool-associated isolates are clustered into three of them. Sub-branch b is represented by the six strains that originate from the Cashmere area and that were isolated in Switzerland in 1981 during a supposed human epidemic of anthrax (Pfisterer, 1991). All six strains are individually characterized and differ from each other by several SNPs. Isolates collected from the Belgian factory are grouped into sub-branch e with the exception of one. This unique strain VAR06/1 strain (recovered from goat hair mix imported from Afghanistan) composed sub-branch d. The remaining NCTC07752 strain from Bradford is part of sub-branch c.

Specific SNP discrimination genotyping assays

Canonical SNPs that define the A.Br.Vollum lineage or that lie on its various branches were selected and developed into real-time PCR assays (Table 2). Six SNP discrimination assays based on High-Resolution Melting (HRM) technology were designed. Assays targeted one SNP specific to each of the five sub-groups identified (Figure 2) and one SNP residing along the branch leading to sub-branches a to d (Figure 1). The novel canSNPs were validated *in silico* against 40 *B. anthracis* genomes available in the public NCBI database and the Sequence Read Archive (SRA). These diagnostic assays were then successfully used to screen 200 *B. anthracis* strains from our collection. The two expected alternate alleles exhibited distinct melting curves and melting temperatures (Tms). On average, differences in Tm ranged from 0.5 to 1 ° C between both allelic states (data not shown). The assays allow unambiguous grouping of each allele and accurate discrimination of the A.Br.Vollum strains from other globally diverse canSNP genotypes (data not shown).

Discussion

In 2007, seven whole genomes of *B. anthracis* were initially used to define 13 “canonical SNPs” (canSNPs), representing key phylogenetic positions along the *B. anthracis* phylogenetic SNP tree. This set of canSNP markers divides the *B. anthracis* worldwide population into three major clades (A, B and C) and 12 clonal groups or lineages [9]. Today, clustering of strains is refined and improved by whole genome sequencing as the cost and time required for sequencing is constantly lowering and the number of published genomes rapidly increasing.

Currently there is little genomic data on strains affiliated to the lineage A.Br.Vollum, in contrast to other lineages of clade A (e.g. A.Br.Ames, A.Br.WNA, A.Br.008/009) [13-16]. It is assumed that the A.Br.Vollum lineage appeared soon after the divergence of the most basal lineage, A.Br.005/006 found in Africa. Lineage A.Br.Vollum became ecologically established in the Middle East, notably in Iran, Pakistan and Afghanistan [9,11,16]. However, no strains from this area have been sequenced yet. In this study, whole genome sequencing and comparative analysis were performed on 14 strains that are associated to wool factories treating imported raw hair. The 14 isolates studied here were imported to Europe from the Middle East and Northern Asia through trade of goat hair bales and were isolated either directly at the processing plant or from patients that got infected [5,11]. All these strains revealed to belong to lineage A.Br.Vollum. Sequence data and SNP analysis allowed determining their phylogenetic relationship to the lineage’s reference strains A0488 and Vollum.

Our investigation sub-divided these strains into three new sub-branches that correlate with their geographic origin. Strains from sub-branch b were isolated in 1981 in Switzerland during an outbreak of human anthrax in a wool factory treating exclusively goat hair from the Cashmere region that was supposed to be decontaminated [11]. The fact that all 6 strains are significantly different from each other demonstrates that the putative epidemic between 1979 and 1981 was due to iterative infections occurring after various shipments of insufficiently decontaminated raw material rather than an epidemic caused by a unique strain.

Sub-branches d and e contain *B. anthracis* strains collected in 2006 at a Belgian factory processing wool mix imported from Afghanistan, Russia, Kazakhstan or Iran [5]. All wool factory associated strains are clearly distinct from strains A0488 and Vollum, both isolated from bovine cases of anthrax in the UK. These latter strains define the terminal branch of the A.Br.Vollum lineage (here

termed sub-branch a) (Figures 1 and 2). As previously described, the CDC 684 isolate is located in a node along this latter sub-branch [17].

The last strain sequenced in this study NCTC07752 representing the sole strain in sub-branch c, was isolated in Bradford in 1940, a renowned city of the British wool industry in the 19th century from where the term ‘Bradford disease’, also known in the 1830s as ‘woolsorters’ disease originated. The presence of NCTC07752 and the bovine strains A0488 and Vollum (isolated in 1963 from a cow in Edinburgh) together with the above named wool factory-associated strains isolated in Belgium and Switzerland indicate spill over from wool contaminants into the endemic bovine population in the UK.

The occurrence of strains of lineage A.Br.Vollum is thought to be due to the extensive trade of unprocessed animal fibers from the Indian subcontinent, such as mohair and cashmere [5,9,11,18]. Cashmere goat hair was processed in many wool factories during the last century. Trading of commercial quantities of raw cashmere between Asia (Nepal and Kashmir) and continental Europe began in the 19th century with Valerie Audresset SA, Louviers, France claiming to be the first European company to commercially spin cashmere [19]. The goat down was imported from Tibet and Tartary through Kasan, the capital of the Russian province Volga [20]. Today, China and Mongolia have become the largest producers of raw cashmere, while Afghanistan, Iran, Turkey, and other Central Asian Republics produce significant but lesser amounts. Concerning Mohair, the Turkish province of Ankara was the sole producer of Angora goats until 1849. Angora goat is thought to originate from the mountains of Tibet, reaching Turkey in the 16th century. Raw mohair was first exported around 1820 from Turkey to England, which then became the leading manufacturer of mohair products. The growing demand for mohair resulted in attempts on a commercial scale to introduce the goat into South Africa in 1838, the United States in 1849, Australia from 1856–1875, and later still New Zealand [21]. Today, South Africa is the world's largest mohair producer.

Molecular epidemiologic studies of *B. anthracis* have demonstrated that particular genotypes usually match particular geographic regions [13-15,22]. A.Br.Vollum appears to be dominant in Pakistan and Afghanistan [9,16,23,this study]. In the central Asian republic of Kazakhstan, most isolates belong to the A.Br.008/009 group and only a few specimens belonged to the A.Br.Vollum cluster [23]. In Georgia and the South Caucasian-Turkish region, a majority of strains are affiliated to the A.Br.Australia 94 cluster [22]. Collections from India and western China are also dominated by genotypes belonging to the A.Br.Aust94 sub-lineage [13]. The diversity unrevealed in this study

within the A.Br.Vollum lineage likely reflects regional strain patterns. Unfortunately, the tentative assignment of a particular genotype to countries of origin was here very complicated, due to the scarcity of metadata available for the sequenced strains. Further comparisons with other strains associated with detailed origin information are definitively needed for drawing sound conclusions on the evolution and natural history of the A.Br.Vollum lineage throughout the world.

Another interesting observation of this study was the placement of the Korean H9401 strain, an isolate obtained from a Korean patient with gastrointestinal anthrax in 1994, in a new lineage phylogenetically related to A.Br.Vollum. *In silico* canSNP typing using the original set of 13 canSNPs [9] wrongly assigned this strain to the African A.Br.005/006 group, as H9401 does not share the derived genotype for the A.Br.007 canSNP that defines the A.Br.Vollum lineage. H9401 neither shares the derived genotype for the A.Br.005 [24] or A05 [25] SNP position that characterizes the A.Br.005/006 canSNP group. *In silico* analysis of ten recently described SNPs that phylogenetically reside along the Vollum branch [17] indicated that H9401 is located in the most basal node defined by the analysis of 100 Vollum-like strains, sharing the derived allele at the 4752196 SNP position [17] exclusively. The bifurcation of the A.Br.H9401 lineage appears to occur along the Vollum branch at a distance of 7 SNPs from the central and most basal node within the A clade (Figure 1). The distance to the tip of this unique new lineage representative is approximately 260 SNPs, suggesting that the new lineage might have been introduced into Korea centuries ago from an ancestor that originated in the Middle East.

Conclusion

Comparative whole-genome sequencing offers a powerful way for in-depth characterization of any bacterial pathogen. It also provides an unbiased approach for informative SNPs discovery. In this study, a new picture of the genetic diversity existing within the population of *B. anthracis* strains of the lineage A.Br.Vollum has been established. The knowledge derived from this work adds unprecedented resolving power to current molecular typing tools and allows tracing of the origin and spread of anthrax during the 19th century.

Methods

Bacterial strains and biosafety procedures

A total of 15 *B. anthracis* strains belonging to A.Br.Vollum were sequenced in this study (Table 1). It included one strain from the NCTC collection (e.g. NCTC7752) isolated in Bradford in 1940, six strains collected during anthrax outbreaks that have occurred in Zurich over three years (1978–1981) [11] and eight Middle East strains isolated in a textile factory in Belgium in 2006 [5]. All *B. anthracis* manipulations were performed in a biosafety level 3 laboratory using class II type A2 biosafety cabinet.

The sequenced A0488 Vollum strain has been isolated from cattle in the UK in 1935 [GenBank: ABJC000000000]. It is presumed to be a close relative of the British isolate that was tested as a biological weapon on Gruinard Island, Scotland, in the 1940s. The bovine Vollum strain (ATCC14578, NCTC10340) has been isolated in Edinburgh in 1963 [GenBank: NZ_AAEP000000000]. CDC 684 is an attenuated *B. anthracis* strain [GenBank: CP001215.1]. H9401 is a Korean isolate obtained from a patient with gastrointestinal anthrax in 1994 [GenBank: CP002091.1].

DNA extraction

Genomic DNAs were obtained from vegetative cells grown at 37 °C on 5 % horse blood agar plates. DNA was purified using the QIAGEN® Genomic-tip 100/G columns and QIAGEN® Genomic DNA Buffer Set as previously described [15]. Viability testing was systematically performed before DNA was taken out of the BSL-3 facility. An aliquot of each DNA preparation (a quarter) was spread on Petri dishes and grown at 37 °C for 24 h.

Draft whole genome sequencing (WGS) and data analysis

In this study, 15 strains were sequenced at the IMAGIF sequencing platform (Imagif, Gif sur Yvette, France) or at the DGA CBRN sequencing platform (Vert le Petit, France). Isolates were subjected to paired-end whole genome sequencing on the Illumina HiSeq2000 platform (paired-end data of 2x100pb) or MiSeq platform (paired-end data of 2x250 pb) (Illumina Inc., San Diego, CA, USA). Genome coverage of at least 34× was obtained (Table 1, average depth ranging from 34× to 186×). The number of reads that passed Illumina quality filters varied from 0.6 to 8.8 million.

Ames Ancestor [GenBank:AE017334.2] was used as the reference genome for assembly. Ames Ancestor plasmid pXO1 [GenBank:AE017336.2] and pXO2 [GenBank:AE017335.3] were used as references for plasmids assembly. Short reads data sets were exported in the FastQ format and

mapped to the Ames Ancestor genome and both pXO1 and pXO2 plasmidic sequences using BioNumerics version 6.6 (Applied Maths, Belgium) and Power assembler module asking for a similarity of at least 90%. A set of SNPs was deduced for each genome sequence data using BioNumerics Chromosome Comparisons module. Individual lists were compiled and data filtered to remove SNP positions at which one or more isolate displayed an ambiguous residue call or missing data. Ribosomal operons and VNTR loci were also excluded from the analysis, as well as contiguous SNPs (using a window-frame of 10 pb). The list of canSNP positions specific to the five A.Br.Vollum branches is provided in Table 3.

The whole genome sequences for 40 available *B. anthracis* strains (Table 2) can be found in the NCBI microbial genome website at <http://www.ncbi.nlm.nih.gov>. The draft genome sequences of three representative strains of the new A.Br.Vollum “wool” sub-branches (c to e) have been deposited in the NCBI database.

Whole genome phylogenetic analysis

A minimum spanning tree was drawn in BioNumerics by using the filtered whole genome sequencing SNP data as input. Nodes were numbered by BioNumerics. CanSNPs along the five branches revealed within the A.Br.Vollum lineage were identified from whole genome sequencing data by searching for SNPs with allelic states shared only by these different subgroups.

SNP discrimination assays by HRM

We designed High Resolution Melting (HRM) assays for 6 specific SNPs using Primer 3⁺ software [26]. The positions of these SNPs in the Ames Ancestor genome [GenBank: AE017334.2] and the primers sequences used are listed in Table 3.

Amplification was performed on the ViiA7™ Real-Time PCR System (Life Technologies) using the LightCycler® 480 High Resolution Melting Master Mix (Roche Diagnostics). The reaction mixture consisted of 0.2 µM of each primer, 1 × LightCycler® 480 HRM master mix and 2.5 mM MgCl₂ in a 10-µl final volume. The following parameters were used: 10 min at 95 °C were followed by 40 cycles consisting of 10 s at 95 °C, 10 s at 58 °C and 20 s at 72 °C. Samples were next heated to 95 °C for 30 s, cooled down to 65 °C for 1 min and heated from 65 °C to 88 °C at a rate of 1 °C/s with 25 acquisitions/°C. HRM data were analyzed by the ViiA7™ Software (version 1.2.1).

Availability of supporting data

Sequence of genomes VAR 06/8 (Branch b), VAR 06/1 (Branch d) and JF3783 (Branch e) will be soon deposited at DDBJ/EMBL/GenBank. **(under progress)**

List of abbreviations used

DNA: deoxyribonucleic acid

SNP: single nucleotide polymorphism

canSNP: canonical single nucleotide polymorphism

VNTR: variable-number tandem repeat

MLVA: multiple-locus variable-number tandem repeat analysis

WGS: whole-genome sequencing

NGS: next generation sequencing

PCR: polymerase chain reaction

HRM: high resolution melting

T_m: melting temperature

NCBI: National Center for Biotechnology Information

SRA: Sequence Read Archive

Competing interests

The authors declare that they have no competing interests

Author's contributions

GG performed the whole-genome SNP analysis and sequences analysis. SD, JF and PW conceived the experiment, collected the biological materials and participate in analysis and interpretation of data. SD wrote the manuscript. JF and PW revised it critically. All authors read and approved the final manuscript.

Acknowledgements

This work was supported by the French National Research Agency (ANR) and Direction Générale de l'Armement (DGA) (project number ANR11ASTR0007). GG is a PhD student co-supported by Anses and DGA grants.

This work has benefited from the facilities and expertise of two high throughput sequencing platforms: IMAGIF (Centre de Recherche de Gif [<http://www.imagif.cnrs.fr>]) and DGA CBRN (Vert le Petit).

References

1. Hugh-Jones M, Blackburn J: **The ecology of *Bacillus anthracis***. *Mol Aspects Med* 2009, **30**:356-367.
2. Brachman PS, Kaufman AF, Dalldorf FG: **Industrial inhalation Anthrax**. *Bacteriol Rev* 1966, **30**:646–659.
3. Carter T: **The dissemination of anthrax from imported wool: Kidderminster 1900-14**. *Occup Environ Med* 2004, **61**:103-107.
4. Kissling E, Wattiau P, China B, Poncin M, Fretin D, Pirenne Y, Hanquet G: ***B. anthracis* in a wool-processing factory: seroprevalence and occupational risk**. *Epidemiol Infect* 2012, **140**:879-86.
5. Wattiau P, Klee SR, Fretin D, Van Hesse M, Ménart M, Franz T, Chasseur C, Butaye P, Imberechts H: **Occurrence and genetic diversity of *Bacillus anthracis* strains isolated in an active wool-cleaning factory**. *Appl Environ Microbiol* 2008, **74**: 4005-4011.
6. Winter H, Pfisterer RM: **Inhalation anthrax in a textile worker: non-fatal course**. *Schweiz Med Wochenschr* 1991, **121**:832-835.
7. Pfisterer RM: **An anthrax epidemic in Switzerland. Clinical, diagnostic and epidemiological aspects of a mostly forgotten disease**. *Schweiz Med Wochenschr* 1991, **121**:813-825.
8. Keim P, Van Ert MN, Pearson T, Vogler AJ, Huynh LY, Wagner DM: **Anthrax molecular epidemiology and forensics: using the appropriate marker for different evolutionary scales**. *Infect Genet Evol* 2004, **4**:205-213.
9. Van Ert MN, Easterday WR, Huynh LY, Okinaka RT, Hugh-Jones ME, Ravel J, Zanecki SR, Pearson T, Simonson TS, U'Ren JM, Kachur SM, Leadem-Dougherty RR, Rhoton SD, Zinser G, Farlow J, Coker

- PR, Smith KL, Wang B, Kenefic LJ, Fraser-Liggett CM, Wagner DM, Keim P: **Global genetic population structure of *Bacillus anthracis***. *PLoS ONE* 2007, **2**:e461.
10. Pilo P., Frey J: ***Bacillus anthracis*: molecular taxonomy, population genetics, phylogeny and patho-evolution**. *Infect Genet Evol* 2011, **11**:1218-1224.
 11. Pilo P, Perreten V, Frey J: **Molecular epidemiology of *Bacillus anthracis*: determining the correct origin**. *Appl Environ Microbiol* 2008, **74**:2928-2931.
 12. Derzelle S: **Single nucleotide polymorphisms discrimination using high-resolution melting analysis for the genotyping of *Bacillus anthracis***. In *Veterinary Infection Biology: Molecular Diagnostics and high-throughput strategies. Volume 1247*. Edited by Cunha MV and Inacio J. New York: Springer Science+Business Media; 2015: in press.
 13. Simonson TS, Okinaka RT, Wang B, Easterday WR, Huynh L, U'Ren JM, Dukerich M, Zanecki SR, Kenefic LJ, Beaudry J, Schupp JM, Pearson T, Wagner DM, Hoffmaster A, Ravel J, Keim P: ***Bacillus anthracis* in China and its relationship to worldwide lineages**. *BMC Microbiol* 2009, **9**:71.
 14. Kenefic LJ, Pearson T, Okinaka RT, Schupp JM, Wagner DM, Hoffmaster AR, Trim CB, Chung WK, Beaudry JA, Jiang L, Gajer P, Foster JT, Mead JI, Ravel J, Keim P: **Pre-Columbian origins for North American anthrax**. *PLoS One* 2009, **4**(3):e4813.
 15. Girault G, Blouin Y, Vergnaud G, Derzelle S: **High-throughput sequencing of *Bacillus anthracis* in France: investigating genome diversity and population structure using whole-genome SNP discovery**. *BMC genomics* 2014, **15**:288.
 16. Price EP, Seymour ML, Sarovich DS, Latham J, Wolken SR, Mason J, Vincent G, Drees KP, Beckstrom-Sternberg SM, Phillippy AM, Koren S, Okinaka RT, Chung WK, Schupp JM, Wagner DM, Vipond R, Foster JT, Bergman NH, Burans J, Pearson T, Brooks T, Keim P: **Molecular epidemiologic investigation of an anthrax outbreak among heroin users, Europe**. *Emerg Infect Dis* 2012, **18**(8):1307-13.
 17. Okinaka RT, Price EP, Wolken SR, Gruendike JM, Chung WK, Pearson T, Xie G, Munk C, Hill KK, Challacombe J, Ivins BE, Schupp JM, Beckstrom-Sternberg SM, Friedlander A, Keim P: **An attenuated strain of *Bacillus anthracis* (CDC 684) has a large chromosomal inversion and altered growth kinetics**. *BMC Genomics* 2011, **12**:477.

18. Keim P, Price LB, Klevytska AM, Smith KL, Schupp JM, Okinaka R, Jackson PJ, Hugh-Jones ME:
Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis*. *J Bacteriol* 2000,**182**(10):2928-36.
19. McGregor BA: **Australian Cashmere: Attributes and Processing.** *Report for the Rural Industries Research and Development Corporation.* Kingston, RIRDC Publication N° 02/112; 2002:p1.
20. Ripley G, Dana CA (Eds): **Cashmere.** In *The New American Cyclopaedia IV.* New York: D. Appleton & Company; 1861:p514.
21. **Mohair.** In *The Encyclopædia Britannica.* 11th edition. Edited by Chisholm H. Cambridge University Press & Hooper HE; 1911.
22. Khmaladze E, Birdsell DN, Naumann AA, Hochhalter CB, Seymour ML, Nottingham R, Beckstrom-Sternberg SM, Beckstrom-Sternberg J, Nikolich MP, Chanturia G, Zhgenti E, Zakalashvili M, Malania L, Babuadze G, Tsertsvadze N, Abazashvili N, Kekelidze M, Tsanava S, Imnadze P, Ganz HH, Getz WM, Pearson O, Gajer P, Eppinger M, Ravel J, Wagner DM, Okinaka RT, Schupp JM, Keim P, Pearson T: **Phylogeography of *Bacillus anthracis* in the country of Georgia shows evidence of population structuring and is dissimilar to other regional genotypes.** *PLoS One* 2014, **9**(7):e102651.
23. Aikembayev AM, Lukhnova L, Temiraliyeva G, Meka-Mechenko T, Pazylov Y, Zakaryan S, Denissov G, Easterday WR, Van Ert MN, Keim P, Francesconi SC, Blackburn JK, Hugh-Jones M, Hadfield T: **Historical distribution and molecular diversity of *Bacillus anthracis*, Kazakhstan.** *Emerg Infect Dis* 2010, **16**:789-96.
24. Birdsell DN, Pearson T, Price EP, Hornstra HM, Nera RD, Stone N, Gruendike J, Kaufman EL, Pettus AH, Hurbon AN, Buchhagen JL, Harms NJ, Chanturia G, Gyuranecz M, Wagner DM, Keim PS: **Melt analysis of mismatch amplification mutation assays (Melt-MAMA): a functional study of a cost-effective SNP genotyping assay in bacterial models.** *PLoS One* 2012,**7**(3):e32866.
25. Girault G, Thierry S, Cherchame E, Derzelle S: **Application of High-throughput sequencing : discovery of informative SNPs to subtype *Bacillus anthracis*.** *Adv Biosci Biotechnol* 2014, **5**:669-677.
26. Primer 3+ software. <http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/>.

Figure legends

Figure 1. Position of sequenced A.Br.Vollum-affiliated strains on the *B. anthracis* phylogenetic tree based on SNP analysis. Minimum spanning tree based on 6,158 chromosomal SNPs. The 13 different canSNP groups are color-coded: C.Br.A1055 in white, B.Br.CNEVA in yellow, B.Br.001/002 and B.Br.Kruger in orange, A.Br.011/009 in light blue, A.Br.008/011 in dark blue, A.Br.WNA in violet, A.Br.005/006 in pink, A.Br.003/004 in grey, A.Br.001/002 in light green, A.Br.Ames in dark green, A.Br.Aust94 in brown and A.Br.Vollum in purple. The new proposed A.Br.H9401 lineage is in black. The position of the 15 newly sequenced A.Br.Vollum isolates (underlined) and 40 available whole genome-sequenced strains is marked. The length of each branch is proportional (logarithmic scale) to the number of SNPs identified between strains. Indicated in red are the position and name of some published canSNPs specific to various canSNP groups: B.Br.003 (B clade), A.Br.006 (A clade), A05 (A.Br.005/006 group), A.Br.007 (A.Br.Vollum), A.Br.008 (A.Br.008/009 group), A.Br.009 (A.Br.WNA), A.Br.004 (A.Br.003/004, A.Br.Australia94, A.Br.001/002 and A.Br.Ames), A03 (A.Br.003/004), A94 (A.Br.Aust94), A.Br.002 (A.Br.001/002 and A.Br.Ames). The total tree size is 6,316 SNPs, i.e. it contains approximately 2.4 % of homoplasia.

Figure 2. Minimum spanning tree of eighteen *B. anthracis* strains belonging to the A.Br.Vollum lineage based on whole-genome SNP analysis. Data are based on 745 chromosomal SNPs (A), 31 pXO1 SNPs (B) and 23 pXO2 SNPs (C). Phylogeographic clustering of the strains is color-coded: Branch b from Pakistan in blue, Branches d and e from the Middle East in green, Branch c from England in grey and Branch a in red. Each circle represents a unique SNP genotype. The diameter of each circle varies according to the number of isolates having the same genotype. The length of each branch is proportional (logarithmic scale) to the number of SNPs identified between strains. Indicated in red are the position and name of the new canSNPs specific to subgroups of A.Br.Vollum. The tree size is 747 SNPs, i.e. it contains approximately 0.25 % of homoplasia. The tree sizes for pXO1 and pXO2 are 31 and 23 SNPs respectively. It contains no homoplasia.

Table 1. Whole-sequenced strains and NGS data

Strain	Read length (pb)	Unambiguous bases	Genome coverage (%) [*]	Sequencing depth	Region of origin	year of isolation	Source of isolation
VAR 06/5 (7570-CAP)	100	5,226,524	99,983	133	Russia - Kazakhstan	2006	Goat hairs (air filter)
VAR 06/4 (7570.4d2#4)	250	5,220,858	99,874	96	Russia - Kazakhstan	2006	Goat hairs (air filter)
VAR 06/3 (7570.4d1#3)	250	5,211,764	99,701	50	Russia - Kazakhstan	2006	Goat hairs (air filter)
VAR 06/2 (5348.1#2)	100	5,226,549	99,983	110	Afghanistan	2006	Goat hair mix (wastewater)
VAR 06/1 (1106.3#2)	250	5,210,334	99,673	34	Afghanistan	2006	Goat hair mix (wastewater)
VAR 06/8 (2093.3#1)	250	5,225,164	99,957	186	Afghanistan	2006	Goat hair mix (wastewater)
VAR 06/7 (5348.3#1)	250	5,222,657	99,909	84	Afghanistan	2006	Goat hair mix (wastewater)
VAR 06/6 (F6 BP#8)	100	5,226,565	99,984	127	Russia - Afghanistan - Iran	2006	Goat hairs (air dust)
NCTC 07752	100	5,226,506	99,983	142	Bradford, UK	1948	unknown
JF3783	250	5,224,298	99,940	152	Cashmere	1981	human infected
JF3784	250	5,224,268	99,940	136	Cashmere	1981	air filter
JF3785	100	5,226,622	99,985	140	Cashmere	1981	raw cashmere goat hair
JF3786	250	5,220,080	99,860	86	Cashmere	1981	raw cashmere goat hair
JF3787	250	5,211,345	99,693	58	Cashmere	1981	cashmere goat hair
JF3788	250	5,222,664	99,909	105	Cashmere	1981	human infected

* Genome coverage is based on the Ames Ancestor genome size (5,227,419 pb).

Table 2 Whole-genome sequences from public database used in this study

Strain	Country	canSNP	Accession number
Ames Ancestor	USA	A.Br.Ames	NC_007530.2
A2012	USA	A.Br.Ames	AAAC00000000.1
Ames	USA	A.Br.Ames	NC_003997.3
A0248	USA	A.Br.Ames	NC_012659.1
Sterne	South Africa	A.Br.001/002	AE017225.1
A0389	Indonesia	A.Br.001/002	ABL000000000.1
BA103	Japan	A.Br.001/002	DRR000183 (SRA)
A16	China	A.Br.001/002	CP001970.1
ATCC14185	Israel	A.Br.003/004	AZQ000000000.1
CZC5	Zambia	A.Br.005/006	BAVT00000000.1
Tsiankovskii-1	Soviet Union	A.Br.008/011	ABDN00000000.2
Ba3154	Bulgaria	A.Br.008/011	ANFF00000000.1
Ba3166	Bulgaria	A.Br.008/011	ANFG00000000.1
Ba4599	Scotland	A.Br.008/011	AGQP00000000.1
UR-1	Germany	A.Br.008/011	ALNY00000000.1
Carbosap	Italy	A.Br.011/009	ANAO00000000.1
Sen2Col2	Africa	A.Br.011/009	CAVC00000000.1
Sen3	Africa	A.Br.011/009	CAVD00000000.1
Gmb1	Africa	A.Br.011/009	CAVE00000000.1
Australia 94	Australia	A.Br.Australia94	AAES00000000.1
9080 G	Georgia	A.Br.Australia94	AZUE00000000.1
52 G	Georgia	A.Br.Australia94	AZUF00000000.1
8903 G	Georgia	A.Br.Australia94	AZUD00000000.1
CDC684	USA	A.Br.Vollum	NC_012581.1
H9401	Korea	A.Br.H9401	NC_017729.1
A0488	UK	A.Br.Vollum	ABJC00000000.1
Vollum	UK	A.Br.Vollum	AAEP00000000.1
USA6153	USA	A.Br.WNA	AAER00000000.1
A0174	Canada	A.Br.WNA	ABLT00000000.1
A0193	USA	A.Br.WNA	ABKF00000000.1
99-100	France	A.Br.011/009	JHDR00000000
08-8/20	France	A.Br.001/002	JHCB00000000
00-82	France	B.Br.CNEVA	JHDS00000000
CNEVA-9066	France	B.Br.CNEVA	NZ_AAEN00000000.1
A0465	France	B.Br.CNEVA	NZ_ABLH00000000.1
BF1	Germany	B.Br.CNEVA	AMDT00000000.1
Kruger B	South Africa	B.Br.Kruger	AAEQ00000000.1
A0442	South Africa	B.Br.001/002	ABKG00000000.1
SVA 11	Sweden	B.Br.001/002	CP006742.1
A1055	USA	C.Br.A1055	AAEO00000000.1

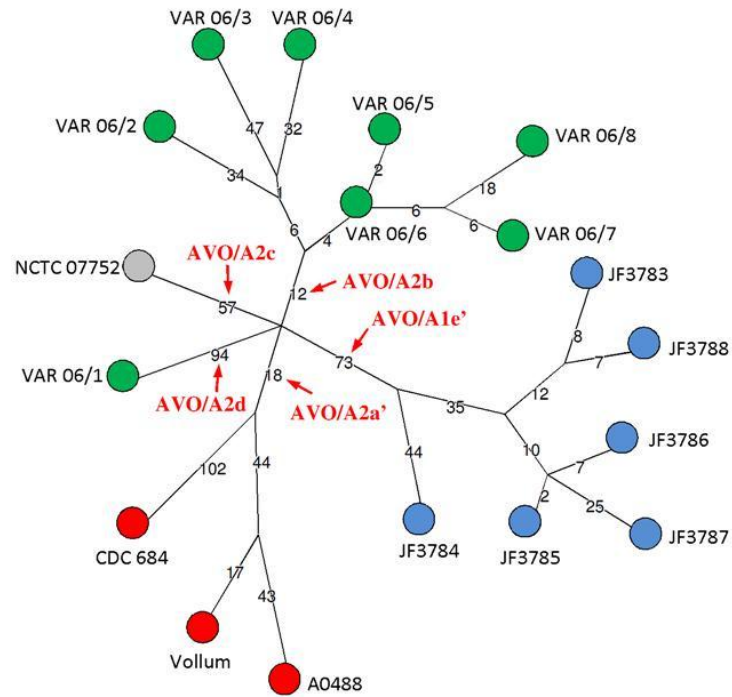
Table 3. Specific canSNPs and primer sequences used for HRM analysis

canSNP	position*	SNP	Forward primer (5' -3')	Reverse primer (5' -3')	Product size
A.Br.007	266439	A to G	CCACAAGGTGGAATTATTACTAAAGA	GGTTCACCTGTTTCGGATCT	80 bp
AVO/A2a'	41872	A to G	AAAGAACTTTGGGAAAAGATGGA	GACGATCTCCGTTGTACGAATA	64 bp
AVO/A2b	4325045	T to C	GTATAACGGCGTCTCACAACCTC	AAAACAGTTCAAGGACCAGATGA	62 bp
AVO/A2c	2041995	T to C	GCTATGTTAGAGCAAAGGAAACG	AGGAAGTCCTGTCCACTCTCATT	55 bp
AVO/A2d	5217173	T to C	AATTTGAATTGGCGAACGACT	CAAGAGCGGTTGTACGAAATG	56 bp
AVO/A1e'	1443043	C to T	GTCAAGCGACAACCTGTTACGTCT	TGTACTTCCTCAGCCGCTTC	58 bp
AVO/A2	55501	G to A	ACAGCATCTCATCTTGAATCAGAA	TTGAATAATTGTTCCCTCACCAA	70 bp

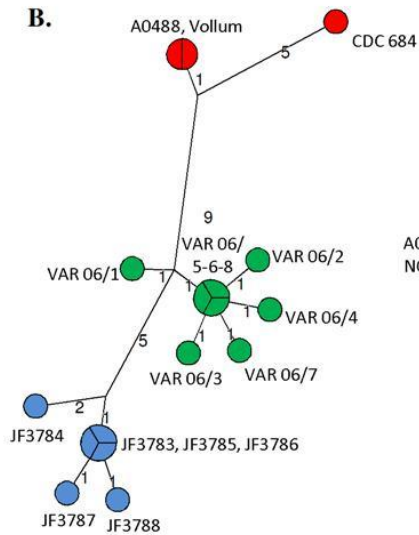
* localisation on the Ames Ancestor chromosome (GenBank accession no. AE017334.2).



A.



B.



C.

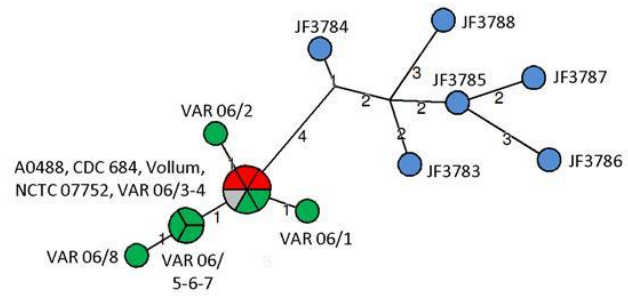


Figure 2

3. Génomotypage des souches européennes de *B. anthracis*

Article 5:

Evolution and dissemination of *B. anthracis* in continental Europe

Article en préparation

Evolution and dissemination of *B. anthracis* in continental Europe

Guillaume Girault¹, Pierre Wattiau³, Joachim Frey⁴, Øystein Angen⁵, Miriam Koene⁶, László Makrai⁹, László Fodor⁹, Joakim Ågren⁸, Aleksandra Zasada¹⁰, Gregor Grass⁷, Antonio Fasanella² and Sylviane Derzelle^{1*}

¹ University Paris-Est, Anses, Animal Health Laboratory, Av du Général De Gaulle 23, 94706 Maisons-Alfort, France

² Istituto Zooprofilattico Sperimentale of Puglia and Basilicata, Anthrax Reference Center of Italy, Via Manfredonia 20, 71100 Foggia, Italy a.fasanella@izsfg.it

³ Foodborne and Highly pathogenic zoonoses Unit, CODA-CERVA, 99 Groeselenberg, B-1180 Brussels, Belgium Pierre.Wattiau@coda-cerva.be

⁴ Institute of Veterinary Bacteriology, University of Bern, Laenggass-Str. 122 Postfach, 3001 Bern, Switzerland joachim.frey@vetsuisse.unibe.ch

⁵ National Veterinary Institute, Technical University of Denmark, Bülowsvej 27, 1870 Frederiksberg C, Denmark. oang@vet.dtu.dk

⁶ Central Veterinary Institute of Wageningen University and Research Centre, Edelhertweg 15, 8219PH Lelystad, Netherlands

⁷ Institut für Mikrobiologie der Bundeswehr, Hochsicherheitslabor/Spezialdiagnostik, Neuherbergstr.11, 80937 München GregorGrass@bundeswehr.org

⁸ National Veterinary Institute, Department of Bacteriology, Travvägen 20, SE-751 89 Uppsala, Sweden ou Department of Biomedical Sciences and Veterinary Public Health, Swedish University of Agricultural Sciences (SLU), Uppsala, Sweden

⁹ SZIU-FVS, Department of Microbiology and Infectious Diseases, Hungária krt. 23-25, 1143 Budapest, Hungary Makrai.Laszlo@aotk.szie.hu

¹⁰ National Institute of Public Health - National Institute of Hygiene, Chocimska 24, 00-791 Warsaw, POLAND rgierczynski@pzh.gov.pl ou azasada@pzh.gov.pl

*corresponding author

Bacterial Zoonosis Unit; Animal Health Laboratory; ANSES

23 Avenue du Général de Gaulle, 94706 Maisons Alfort cedex, France

Phone: +33 1 49 77 38 84 ; E-mail : sylviane.derzelle@anses.fr

Abstract:

Background *Bacillus anthracis*, the etiological agent of anthrax, is an evolutionarily young species that presents an extremely low genetic diversity. We used next-generation sequencing technology to investigate the phylogenetic relationships existing between strains of *Bacillus anthracis* in Europe. A hundred of geographically diverse strains isolated in 10 countries (France, Italy, Poland, Switzerland, Hungary, Germany, Denmark, Belgium, the Netherlands, Sweden and United Kingdom) were sequenced and comparative analyses were carried out with a focus on single nucleotide polymorphisms (SNPs) discovery.

Results European population was found to be quite diverse, with three main canSNP groups found in most countries, i.e. Trans-Eurasian (TEA) A.Br.008/009 group, B.Br.CNEVA and A.Br.001/002, and 6 other minor sub-lineages of anthrax. Whole-genome SNPs screening provide further resolution to both TEA and A.Br.001/002 canSNP groups. Ten TEA groups were resolved, six of them corresponding to the A.Br.011/009 sub-groups previously described. Two distinct phylogenetic clusters containing, respectively, the Ames strain (termed here A.Br.01) or Sterne strain (termed here A.Br.02) were also identified among strains belonging to the A.Br.001/002 canSNP groups. Positioning European strains with respect to globally diverse strains, revealed an interesting link between the A.Br.011/009 Branch 2, Western North America and Western Africa strains. Isolates from the B.Br.CNEVA lineage were found in different countries and present local settlement and differentiation profiles.

Conclusions The large Whole-genome SNPs resolved European subpopulations structure of *B. anthracis* at unprecedented level. The results improve our knowledge on the evolution and dissemination of *B. anthracis* in continental Europe and provides clues for the phylogenetic history of this monomorphic pathogen.

Introduction

Bacillus anthracis, a spore-forming pathogenic bacterium, is the causative agent of anthrax. All mammals are known to be susceptible to anthrax, but the bacterium primarily affects wild and domesticated herbivores, causing acute, often fatal disease. Humans can be infected through contact with diseased animals or by exposure to spore-contaminated animal products (such as hides, wool, or meat). Anthrax is distributed throughout the world and still endemic in many countries, particularly in the developing world.

Bacillus anthracis constitute a very monomorphic species with relatively little genetic variation within the species (Rasko *et al.*, 2005). Due to its episodic reproductive cycle, opportunities for accumulating DNA mutations are limited. *B. anthracis* evolved as a highly fit clonal pathogen that has spread globally and become ecologically established and genetically distinct in local populations (Keim *et al.*, 2013).

SNPs are important phylogenetic markers for diagnostic because of their stability, with mutation rates of approximately 10^{-10} changes per nucleotide per generation (Keim *et al.*, 2004). SNPs are rare in *B. anthracis*, but extensive whole genome sequencing has led to the discovery of SNPs that are particularly useful in defining and establishing membership in phylogenetic clades. Genome comparison of multiple strains from various origins has led to the construction of an extremely accurate phylogenetic backbone.

The substructure of the population is divided into three major clades (A, B and C), with further subdivisions into 12 major sub-lineages or groups. The A-radiation has known the most dramatic dispersal and clonal expansion around the world (Keim *et al.*, 1997; Keim *et al.*, 2000; Keim *et al.*, 2009). Its eight sublineages or groups exhibit distinct geographical patterns across the five continents. Most basal of the A clade is the A.Br.005/006 group which is common in Africa, consistent with an “out of Africa” hypothesis for the A clade. Soon after this divergence, the Vollum clade appeared, and became ecologically established in Iran, Pakistan and Afghanistan. The highly successful Transeurasian clade is slightly more recent and spread across Europe and Asia, and one lineage was introduced into North America. Another recent bifurcation led to the Australia 94 and Ames/Sterne group that contains isolates found in Southeast Europe and Asia (Keim *et al.*, 2013). Collections from western and south-central Asia (Turkey, India, Georgia) and western China are dominated by genotypes belonging to A.Br.Aust94 sub-lineage. Further into central and eastern China the genotypes are dominated by isolates belonging to the A.Br.001/002 sub-group and A.Br.Ames sub-lineage (Simonson *et al.*, 2009). North and South America contain different genetic groups, the A.Br.WNA lineage nested within TEA and the A.Br.003/004 group, respectively (Van Ert *et al.*, 2007a).

The B lineage is divided into two genetically distinct sub-clades, geographically restricted to specific regions. The B.Br.001/002 sub-clade (including the B.Br.Kruger lineage) is ecologically established in Southern Africa, in particular in the Kruger National Park, where it co-exists with strains from the main A.Br.Vollum sub-lineage. The B.Br.CNEVA lineage is primarily found in continental Europe (Derzelle *et al.*, 2011a; Garofolo *et al.*, 2011; Gierczyński *et al.*, 2004; Pilo *et al.*, 2008; Van Ert *et al.*, 2007a). Finally, the C lineage (C.Br.A1055) is an uncommon cluster consisting of only three isolates of unknown origin (Sue *et al.*, 2007; Van Ert *et al.*, 2007a).

Anthrax has been known in Europe since Antiquity. Nowadays, although an increasing number of European countries are probably free of the disease (including Cyprus, Austria, the Czech Republic, Denmark, Finland, Latvia, Luxembourg and Malta), Anthrax is not yet a disease of the past. It remains relatively common in countries adjoining the Mediterranean Sea and Eastern Europe. The population structure of *B. anthracis* is largely diversified in Europe, with the co-occurrence of up to 6 canSNP groups in a single country. Strains collections from Europe are dominated by A.Br.008/009 sub-group. The Transeurasian groups appear to be more largely distributed (Derzelle *et al.*, 2013).

The minor groups observed are likely a consequence of human activities, international trade exchanges and industrial processing of contaminated animal products, often wool in textile mills. These cases contribute to the overall genetic diversity of European *B. anthracis* isolates, but generally represent small case clusters that do not become ecologically established on a sub-continental scale.

Advances in the area of high throughput sequencing technology (and the result of genome sequencing data) have provided opportunities to get information on whole genome variability within species and subspecies groups. Today, whole genome sequences are made more available than ever before and can be obtained within a matter of days.

Sequencing efforts to increase the number of genome sequences from *B. anthracis* provides an unprecedented opportunity to create comprehensive species phylogenies. Informative polymorphisms can be identified with multiple sequence alignments and comparison to discover genome-wide SNPs that define a basic tree.

Massively parallel sequencing technology is a powerful tool for the unbiased typing of any pathogen. By interrogating nearly every base of the genome, it is now possible to discover rare polymorphisms and gain high resolution in genotyping that will facilitate comparisons between different bacterial isolates.

The aim of this study was to improve the knowledge of the existing diversity in *B. anthracis* isolates in Europe. High-resolution whole-genome sequencing and comparative genomics have been applied to 123 strains from different geographic areas to identify their key genetic features. We sought to phylogenetically place the representative European isolates by using previously published and novel WGS sequences available in databases or Sequence Read Archives.

Extensive genomic sampling through large strain collections has established one of the most accurate phylogenetic reconstructions of any bacterial species.

The knowledge acquired from the present sequencing effort provides new insight on the natural history (dissemination, local settlement and differentiation) and global population structures of lineages of this highly monomorphic pathogen.

Materials and Methods

Bacterial strains and biosafety procedures

DNA from a total of 123 *B. anthracis* strains isolated in ten countries (France, Italy, Belgium, The Netherlands, Denmark, Switzerland, Germany, Poland, Hungary, Sweden, United Kingdom) were used in this study (Table 1). Strains came from the collections of *Bacilli* of the different partners: Anses (n = 22), IZS (n= 20), CODA-CERVA (n=14), SVA (n = 2), CVI (n = 11), DTU-Vet (n=11), NIPH - NIH (n = 7), SZIU-FVS (n = 5), IVB (n = 5) and InstMikroBioBw (n = 6). It also included six reference strains from the NCTC's Health Protection Agency culture collection. All *B. anthracis* manipulations were performed in a biosafety level 3 laboratory using class II type A2 biosafety cabinet.

SNP Genotyping

CanSNP HRM assays for *B. anthracis* were used to determine the approximate phylogenetic placement of all strains on *B. anthracis* substructure population, as described in Derzelle *et al.*, 2011.

Whole Genome Sequencing

Besides 35 French (Girault *et al.*, 2014a) , 21 Italian (Fasanella, in progress) and 2 sweden isolates (Agren *et al.*, 2014; Lewerin *et al.*, 2010) previously published, We sequenced 65 new genomes to accurately establish the evolutionary relationships among these strains. With the exception of previously sequenced strains, isolates were subjected to paired-end whole genome sequencing on either the Illumina HiSeq2000 platform (paired-end data of 2 x 100 pb) (Illumina Inc., San Diego, CA,

USA) or the Illumina MiSeq platform (paired-end data of 2 x 250 pb) (Illumina Inc., San Diego, CA, USA). Genome coverage of at least 65X was obtained (average depth ranging from 65X to 357X). Ames Ancestor [GenBank: AE017334.2] was used as the reference genome for assembly. Short reads data sets were exported on the FastQ format and mapped to the Ames Ancestor using BioNumerics version 6.6 (Applied Maths, Belgium) and Power assembler module asking for a similarity of at least 90%. A set of SNPs was deduced for each genome sequence data using BioNumerics Chromosome Comparisons module. Individual lists were compiled and data filtered to remove SNP positions at which one or more isolate displayed an ambiguous residue call or missing data. Ribosomal operons and VNTR loci were also excluded from the analysis, as well as contiguous SNPs (using a window-frame of 10 pb).

The whole genome sequences of databases strains can be found in the NCBI microbial genome website at <http://www.ncbi.nlm.nih.gov> (Table 2).

Whole Genome Phylogenetic Analysis

A minimum spanning tree was drawn in BioNumerics version 7.0 by using the filtered whole genome sequencing SNP data as input. Nodes were numbered by BioNumerics. CanSNPs along the seven branches leading to the Italian strains were identified from whole genome sequencing data by searching for SNPs with allelic states shared only by these different subgroups.

Results

canSNP typing and Phylogeographic clustering

A total of 123 strains, representative of anthrax activity in ten European countries since the 1920's, were selected for this study (Table 1, Table 3). All strains were first subjected to canSNP typing by HRM to determine their approximate phylogenetic placement on the *B. anthracis* substructure population. Nine canSNP sub-lineages or subgroups were identified among European strains, i.e. B.Br CNEVA (n=20), B.Br.001/002 (n = 4), A.Br 008/011 (n=20), A.Br.011/009 (n=37), A.Br 001/002 (n=25), A.Br.Ames (n=2), A.Br.Australia94 (n=7), A.Br.Vollum (n=2), A.Br.003/004 (n=2) and A.Br 005/006 (n=4) (Table 1, Figure 1).

A fifth of the European strains examined in the present collection was part of the relatively rare B clade, most of them (20 on 24) being members of the B2 sub-clade (B.Br.CNEVA) which is exclusively found in continental Europe. A few strains isolated in Belgium and Sweden were affiliated to the Southern African B1 sub-clade (B.Br.001/002) (Smith *et al.*, 2000; Van Ert *et al.*, 2007a). None specimen from the rare C clade was found. Two canSNP groups were responsible for the vast majority of European strains: the Transeurasian group (including A.Br.008/011 and A.Br.011/009-affiliated strains) and the A.Br.001/002 group (including the A.Br.Ames lineage). TEA-affiliated strains spread across all Europe and were found in nearly all European countries (except Sweden). It was well represented in particular in Bulgaria, Hungary, Italy and France. The A.Br.001.002 also crossed most of Europe but at slightly higher latitude. It was dominant in Denmark, The Netherlands and Germany. Distribution of A.Br.Australia94 affiliated strains throughout Europe portrayed the same pattern (presence in Denmark, The Netherlands and Germany). A few strains belonging to other canSNP groups which are common in other parts of the world were also found in some disparate European countries. Strains affiliated to A.Br.005/006 have been found in Denmark, Italy and Belgium; A.Br.Vollum lineage is found in The Netherlands and UK collection, and the A.Br.003/004 group is found in Sweden and Belgium.

Whole genome sequencing and SNP discovery

All strains were subsequently characterized by paired-end whole genome sequencing using Illumina sequencers. After applying the quality filter of the Illumina base-calling pipeline, filter-passed reads

were aligned to the Ames Ancestor reference genome, resulting in more than a 65-fold sequencing depth on average and a genome coverage ranging from 97.2 to 99.9% (Table 1).

Comparative analysis of the 123 genomic sequences was next carried out for SNP identification using *in silico* analysis. Available data for 34 additional *B. anthracis* genomes from diverse geographic origins were also included (Table 2). The *B. cereus* AH820 genome was used as outgroup. By mapping each dataset against the Ames ancestor genome, a total of 9167 chromosomal SNPs were identified among the 157 *Bacilli* strains examined (with a homoplasia level of 2%). Figure 2 illustrates the minimum spanning tree (MST) deduced from the whole-genome SNPs data. European strains were positioned with respect to the 34 other diverse, fully sequenced genomes.

Phylogenomic analysis

Consistent with previous publication, three branches were observed to radiate from the A clade leading to the major TEA (A.Br.008/011 and A.Br.011/009) and A.Br.WNA groups (Figure 2). The more basal corresponds to the “Ancient A” group (A.Br.005/006) which is common in Africa. Soon after this divergence (17 SNPs), the second one is related to the A.Br.Vollum lineage which is found in the Middle East, Asia (Afghanistan, Iran, Pakistan) and Africa. Slightly more recent, the third branch leads to the A.Br.001/002 group that includes the Ames and Sterne strains. The A.Br.003/004 group split at a distance of 48 SNPs from this branching point. The secondary split between lineages A.Br.Australia94 and A.Br.001/002 occurs very shortly afterwards, 11 SNPs later. The B clade is composed of two genetically distinct sub-clades: B.Br.001/002 wherein the B.Br.Kruger lineage is rooted and B.Br.CNEVA.

A.Br.001/002 phylogenetic analysis

Whole genome SNP analysis of the A.Br.001/002 subpopulation resolved the 25 European strains into two clearly separated sub-groups (termed A01 and A02 herein, Figure 3) that share a common evolutionary times (69 SNPs) (Figure 2). The first one (A01) includes the A.Br.Ames lineage and radiates very shortly after the A01-A02 divergence (1 SNP in position 515111) into at least six sub-branches, including A.Br.Ames strains. The second (A02) sub-group includes the Sterne vaccine strain and was distantly linked (37 SNPs) to the bifurcation point with A01. Surprising, compared to publicly available genomes from Asia, the A.Br.001/002 population structure was found to be complex and a little bit confusing. Nearly all branches were constituted by a mixed population isolated from both Europe and Asia.

The A01 sub-branch leads to the A.Br.Ames lineage (sub-branch A). In addition to the Texas and Chinese Ames-related strains previously reported, two European strains isolated in Denmark were also included in this sub-branch A. The five remaining A01 sub-branches were, respectively, composed of an Indonesian strain (sub-branch B); Danish, Japanese, and French strains (sub-branch C); English, Chinese and German isolates among which the recently released A16 genome (sub-branch D); Danish, English (isolated from Chinese hide) and French strains (sub-branch E) and a single Belgian isolate (sub-branch F). The A02 sub-group could be further resolved into several clusters composed of strains with no apparent geographic clustering at any regional- or national scale-level. Beside the Sterne strain, isolates from Belgium, The Netherlands, Denmark, Germany, France and Poland are members of this A02 sub-group.

A.Br.008/009 phylogenetic analysis

As illustrated in Figure 4, the A.Br.008/011 sub-group consists in several distinct sub-branches that split from the TEA group. Bifurcation points leading to the different A.Br.008/011 sub-branches lies at a few SNPs forward the A.Br.011/009 radiation. Two SNPs, i.e. A.Br.011 canSNP [2553486] (Marston *et al.*, 2011) and SNP [1612053] were found to clearly separate the two sub-groups, A.Br.008/011 and A.Br.011/009.

In addition to the two Italian A.Br.008/011 sub-branches previously described, involving strains from Umbria (Branch D) or Sicilia (Branch B), two others phylogenetic groups (termed A and C sub-

branches, herein) were identified in this study. The two published Bulgarian genomes were grouped together into the Branch B. The C branch is composed of two Danish and one Dutch strains clustering with two strains linked, respectively, to the 2009-2010 and 2012 anthrax outbreaks that affected heroin users across Scotland, Germany and England (Grunow *et al.*, 2013; Hanczaruk *et al.*, 2014; Price *et al.*, 2012). The last branch (branch A) is more diverse and contains isolates from Belgium, Poland and UK, as well as all Hungarian strains and the Tsiankovskii strain. Most genomes from the latter branch shows the derived state for several previously published SNPs identified in the TEA branch leading to either the A0628, A0362 or Ba1168 strains (Birdsell *et al.*, 2012).

Consistent with previous whole genome SNP analyses (Girault *et al.*, 2014a), at least six sub-branches radiate simultaneously from the A.Br.011/009 branching point and compose this sub-group (Figure 4A). Beside French and Italian strains, only two European isolates from the present collection fit into this sub-group (Figure 4B). Strain CVI-260187 (The Netherlands) is linked to the A.Br.011/009 Branch 3 that clusters strains collected in the center of France, Carbosap-like Italian strains and Pasteur II vaccine strains. The Fildes strain (NCTC 01328) is related to Branch 4 and two French strains from remote geographical regions within. Branch 1 includes strains isolated in the Center-North-East of France (CNE clade) and in Italy (Puglia-Lazio-Toscana). Branches 5 and 6 are composed of French strains isolated close to the Pyrenees (South-West) or in South-East of France, respectively. Branch 2 includes two Italian clusters, Basilicata-Campania and Sardegna, in addition to French north-eastern strains (NE clade). The A.Br.WNA lineage is rooted in this latter sub-branch, as well as three Western African genomes (Gmb1, Sen3, Sen2col2) recently released in public database. African strains appear to be separated by a longer phylogenetic branch (>350 SNPs) from European near relatives compared to the Western North American lineage (about 100 SNPs).

B.Br.CNEVA phylogenetic analysis

The twenty-one B.Br.CNEVA strains isolated in France, Switzerland, Italy, Poland and Germany clustered according to their geographic origin. Except the Swiss samples which were composed of two clearly separated phylogenetic groups (JF3852 and JF3854 versus JF3887 and JF3888), each national strain form a distinct branch that split from the B.Br.CNEVA lineage leading to the CNEVA-9066 and other French strains. The bifurcation between strains ecologically established in Poland and East-Western Switzerland and the others strains was more basal, at a distance of 43 SNPs from the basis of the lineage. The French radiation into three geographic clusters previously described occurs at the tip.

A.Br.Aust94 phylogenetic analysis

Seven collection strains from different countries (Switzerland, Netherlands, Denmark and Germany) have been compared to previously published genomes (Australia94 and Georgian strains). Strain JF-3853 was only loosely related to other A.Br.Aust94-affiliated strains. One of the two Dutch strain is clustered with Georgian strains (9080-G, 52-G and 8903-G), indicating a common ancestor. The three German strains are grouped with the Australia94 strain. It has been hypothesized that anthrax was first introduced to Australia in 1847 via contaminated bone meal-based fertilizer shipped from India (Van Ert *et al.*, 2007a).

Discussion

Available evidence points to sub-Saharan Africa with its rich faunal diversity and density, as the cradle of anthrax, from whence it spread to the rest of Africa and subsequently into Eurasia via humans and their domestic animals (Hugh-Jones et De Vos, 2002).

In Europe, the population structure of *B. anthracis* is largely diversified, with the co-occurrence of both A- and B- clades and at least 9 canSNP groups. Three groups or lineages – that may reflect ancient introductions of *B. anthracis* into Europe – dominate the region on a sub-continental scale

and appear to have experienced the biggest spread across Europe, e.g. A.Br.008/009, B.Br.CNEVA, and A.Br.001/002. The present whole-genome SNP analysis has gained further resolution among those clonal lineages and provides new insights into the population substructure of *B. anthracis* in Europe, with the discovery of a much more complex phylogeny than previously thought (Keim *et al.*, 2013).

The genetic diversity in Europe appears to change with latitudes and longitudes. Countries from Central and Eastern Europe are dominated by isolates belonging to the A.Br.008/011 sub-group, while Western and South-Central Europe (France and Italy in particular) are dominated by the A.Br.011/009 sub-group. A.Br.001/002 appears to be distributed at higher latitudes in Western Europe. It accounts for a consequent part of the *B. anthracis* population in Denmark, Belgium, Netherlands, the United Kingdom (this study) and the North-East of France (Calavas *et al.*, 2008). Their dispersal in Europe and Asia has been suggested to be mediated by ancient east-west human trade routes, such as the “silk Road” (Simonson *et al.*, 2009; Van Ert *et al.*, 2007a). At last, The B.Br.CNEVA lineage is ecologically established across an axis in South-Central Europe from the Pyrenees (South-West France) to the North-East of Poland, along Northern Italy, Switzerland, Southern Germany (Bavaria), Croatia, and Slovakia (Antwerpen *et al.*, 2012; Derzelle *et al.*, 2011a; Garofolo *et al.*, 2011; Pilo *et al.*, 2008; Steiner *et al.*, 2013; Van Ert *et al.*, 2007a). The Anthrax dissemination path somehow follows the trans-Alpine route. The Alps extend for almost 700 miles from the coastline of southern France into Switzerland, northern Italy and Austria, then southeast through Slovenia, Croatia, and Bosnia-Herzegovina.

The Trans-Eurasian canSNP group (A.Br.008/009) has experienced the most global dramatic dispersal. It predominates throughout Europe, but also Russia, Kazakhstan, the Caucasus and the Western most Chinese province of Xinjiang. It has become as one of the Europe’s most successful subpopulation of *B. anthracis* both in distribution and frequency. The A.Br.008/011 sub-group appears to be basal to the TEA group leading to the A.Br.011/009 sub-group. Only a short evolutionary time period, two SNPs (2552486 (A.Br.011) and SNP 1612053), separates the branching point of the six A.Br.011/009 sub-branches from the A.Br.008/011 four groups. The phylogeography changes along a longitudinal axis. A.Br.011/009-affiliated strains were primarily identified in France and Italy, but also in England, while the A.Br.008/011 sub-group dominates in Eastern Europe (Bulgaria, Hungary, Albania and Russia, Poland (Antwerpen *et al.*, 2011; Van Ert *et al.*, 2007a)) and is part of the population in Denmark, the Netherlands, Belgium, UK (this study). One of the most diverse collection of *B. anthracis* strains existing in the world has been reported to contain 120 strains affiliated to A.Br.008/011 comprising isolates from Albania (n = 5), Argentina (1), Austria (1), China (47), Ethiopia (1), Hungary (3), Iran (1), Italy (21), Norway (1), Pakistan (2), Poland (1), Republic of Georgia (5), Russia (2), Scotland (12), Slovakia (2), Turkey (6) and Ukraine (1). Our data suggest an earlier introduction of the A.Br.008/011 sub-group into Eastern Europe than in Western Europe, with an A.Br.011/009 expansion and radiation into at least six branches occurring primarily in the West (Italy and France). To gain further resolution in this hypothesis, it will be interesting to have informations about strains of countries like Portugal and Spain. Nevertheless, the precise origin of the A.Br.008/011 ancestral populations (Europe or Western Asia) is complicated by the wide phylogeographic distribution of this sub-group and the lack of NGS data from strains isolated in Western Asia.

The present data supports a possible introduction of the derived Western North American lineage from Europe, possibly via French colonization 500 years ago and the rooting of A.Br.WNA within the A.Br.011/009 Branch 2 (defined according to French *B. anthracis* population (Girault *et al.*, 2014a)). These observations are in opposition with the pre-Columbian hypothesis dating back the introduction and establishment of the WNA lineage of anthrax in North America during the late Pleistocene epoch (13,000 ybp) from the Far East through the Bering bridge (Kenefic *et al.*, 2009). A new finding of this analysis was the phylogenetic linkage between A.Br.WNA, several A.Br.011/009-

affiliated strains from Western Africa (Senegal and Gambia) and isolates from Sardinia and North-eastern France, suggesting dispersal of anthrax in portion of Western Africa (in countries such as Senegal and Gambia) after founding by a European ancestor (French or Italian).

The link between the canSNP A.Br.011/009 lineage and the West African origin of the strains is unexpected. Up to now, only a few additional isolates from Africa has been discovered to belong in this canSNP. A few additional strains linked to contaminated animal products are believed to originate from West Africa and had been implicated in several allochthonous cases of anthrax in the United States (2006 New York inhalation case, 2007 Connecticut cutaneous cases (goat hides from Guinea) and 2009 New Hampshire gastrointestinal case (drums) (Marston *et al.*, 2011)).

However, the relatively long branch evolution (about 400 SNPs) of Western African strains compared to European strains (< 40 SNPs) or American strains (about 100 SNPs) raises the question of the divergence times between these branches and the origin of the nearest Old World relatives (Africa versus Europe). Only a few informations related to these strains are available, and we can not exclude the possibility of an intensive culture of these strains. If it is the case, the accumulation of mutations can be explained. Very little information is known regarding the molecular subtypes of isolates in the western region of Africa, but much more diversity can be expected. In this respect, CT, NY, and NH isolates have been reported to share the derived allele for one canSNP located near the base of the A.Br.011/009 lineage terminating in A0343 (Hungary collection), suggesting that additional branch likely exists.

The precise origin of the TEA subgroup (e.g. Asia versus Europe) is still difficult to predict and would need to acquire additional genome sequences from Asian or African isolates. More complete genomic data are required to substantiate both hypotheses. Phylogenetic reconstruction by incomplete taxon sampling inherently biases any analysis (Pearson *et al.*, 2004). African data expand the understanding of the geographic range of this Eurasian lineage but also provide insights into its historical incidence in the country.

Strains from the B.Br.CNEVA lineage are found in a relatively limited number of European countries, along an axis that somehow follows the trans-Alpine route in south-central Europe. The Alps extend for almost 700 miles from the coastline of southern France, through Switzerland, northern Italy and Austria, then southeast through Slovenia, Croatia and Bosnia-Herzegovina. In the mentioned countries, B.Br.CNEVA usually co-dominates with other A lineages (like TEA), although these phylogenetic groups usually exhibit distinct geographic localization (Girault *et al.*, 2014a). One possible explanation for the differences in global distribution and abundance observed among the lineages is the adaptive genetic differences that affect survival and propagation in either the environment or hosts. A comparison of A versus B isolates from Kruger National Park indicated that A strains were adapted to more diverse environments than B strains, which were restricted to more narrow environmental conditions. This trend is also reflected on a global scale, where the B type geographic distribution may arise from fitness costs associated with niche specialization (Van Ert *et al.*, 2007a). The relatively limited dispersion and expansion observed is consistent with a single introduction of this lineage, followed by ecological establishment and progressive derived local populations of strains and *in situ* differentiation (linked to host or environment). Strains collected regionally form their own unique cluster, clearly different from strains isolated elsewhere in Europe. This observation reinforced the idea of the dispersal of an ancestral population towards West and South. Nevertheless, it is still difficult to determine the origin of the lineage. These issues can eventually be resolved using additional whole genome sequencing and phylogenetic inference.

Another finding of this study was the intermixed structure of the A.Br.001/002 group. These patterns indicate a long history for the A.Br.001/002 sub-group in Europe and are consistent with a combination of possibly repeated introduction and infections. The central and eastern regions of China are dominated by isolates belonging to the A 001/002 sub-group (Simonson *et al.*, 2009). Seventy percent of all A.Br.001/002 isolates are in a world-wide collection of strains (USA) and most

of the diversity for this sub-group is from China, suggesting that the A.Br.001/002 cluster may have originated in China. The remaining was found in UK, Europe, North America and other parts of Asia. The Ames lineage descends from the A.Br.001/002 group. The lineage was introduced into the Gulf Coast, probably by early European settlers or traders during the early to mid 1800s. It became established in Texas, and perhaps Louisiana. Isolates from China and Texas share common ancestors that can originate from China (Simonson *et al.*, 2009). Up to now, none European relatives were known, leaving behind a missing historical gap within the phylogeography of the Ames lineage. Analysis of 29 (Simonson *et al.*, 2009) up to 32 (Van Ert *et al.*, 2007b) chromosomal SNPs located along the *B. anthracis* Ames lineage, indicated that the two Danish A.Br.Ames European strains have the derived state for only four SNPs of 29 original SNPs located along the Ames strain lineage (200731, 182106, 243867, 337241: Br1.4, 1.2, 1.5, 1.6). The bifurcation point leading to these Danish isolates identified in this study lies near the base the A.Br.Ames lineage.

Minor sub-groups found in Europe include A.Br.005/006, A.Br.Aust94, A.Br.Vollum, B.Br001/002, A.Br.Ames and A.Br.003/004 (Derzelle et Thierry, 2013). This cosmopolitan assortment of genetic types dominant in portions of Asia, Africa or South America is likely a consequence of human activities, international trade exchanges and industrial processing of contaminated animal products (hair and wool, hides and skins, bone meal, shaving brushes). A.Br.Vollum, in particular, in European countries as The Netherlands and UK, is thought to be due to the extensive trade of cashmere wool from the Indian subcontinent. The presence of *B. anthracis* in particular countries of Europe often reflects colonial-era importation of specific genotypes from Asia and Africa, as well as the repeated industrial importation of diverse genotypes into developed countries via spore-contaminated animal products. The presence of an A.Br.005/006 strain in relatively small area throughout Europe (Italy, Belgium,) has been suggested to be related to trade exchanges dating back to the middle Ages period. This sub-group is well represented in other part of the world (Africa, Middle East, Asia) but rare in Europe and could have been introduced through animal products (e.g. wool, hides, and bone meal) from Africa to Northern Italy in the past. These cases generally represent small case clusters that do not become ecologically founded. The victims from industries processing animal products were woolworkers but also included people and cattle that had been infected through the environmental release of contaminated waste (Carter, 2004).

Active control measures, improved industrial and animal husbandry practices and large animal vaccination programs (following the development of live vaccines for animals) succeeded in reducing the incidence of the disease. Over time, vaccination efforts prevented the replenishment of spores in soil in many European countries. In the middle and higher latitudes of Europe, Anthrax in animals is now either absent or found only in very sporadic cases, while it remains relatively common in Southern and Eastern Europe.

Molecular clock estimations are problematic in this species because of the difficulty in estimating the number of generations per year. Molecular clock models suggest that the divergence of the two major B branches from a common ancestor appears 4661-9322 ybp. The two major B lineages are spatially separated, one is found in southern Africa and the other is found in portions of Europe, suggesting that after diverging from a common ancestor, these two groups experiences independent evolutionary histories.

Anthrax transmission frequencies can vary dramatically. Because the frequency of animal transmission greatly influences the evolutionary rates, higher mutation accumulation should be associated with higher transmission frequencies. Host availability or density may be the driving ecological force behind transmission frequency and, thus, evolutionary rates. West Africa is the most affected area of the world (Schmid et Kaufmann, 2002). Anthrax, for example, is common in West Africa where there are large herds of susceptible ungulate hosts. In contrast, episodic cases are primarily found in France where the hosts may be mostly domesticated animals with many fewer opportunities for infection. Anthrax is a sporadic disease that remains relatively common in countries

adjoining the Mediterranean Sea and Eastern Europe. It almost exclusively affects animals at pasture, with a majority of cases reported during the summer months. In the rest of Europe, foci of animal disease are either absent or found only in very sporadic cases. The occurrence of new animal cases takes place in area where outbreaks have occurred in the past.

Conclusions

This work contributes to a better understanding of the *B. anthracis* population present in Europe and provides some clues about the evolution and global spread of *B. anthracis* in Europe. In this study, we have defined the evolutionary history of European anthrax using 9167 single nucleotide polymorphisms (SNPs) and 157 geographically diverse *B. anthracis* isolates, generating hypotheses about patterns and modes of disease dispersal.

The discovery of thousands informative SNPs allow isolates and outbreaks to be placed in the wider context of pathogen population structure.

Competing interests

The authors declare that they have no competing interests

Authors' contributions

SD conceived the experiments with all others partners. GG prepare genomic DNA and performed the experiments/SNPs discrimination assays experiments. GG and SD analyzed NGS data and wrote the paper. All authors read and approved the final manuscript.

Acknowledgements

This work was supported by the French National Research Agency (ANR) and Direction Générale de l'Armement (DGA) (project number ANR11ASTR0007). GG is a PhD student co-supported by Anses and DGA grants.

This work has benefited from the facilities and expertise of the high throughput sequencing platform of IMAGIF (Centre de Recherche de Gif [<http://www.imagif.cnrs.fr>]).

Figure legends

Figure 1: Distribution of canSNP lineage distribution throughout the 10 countries.

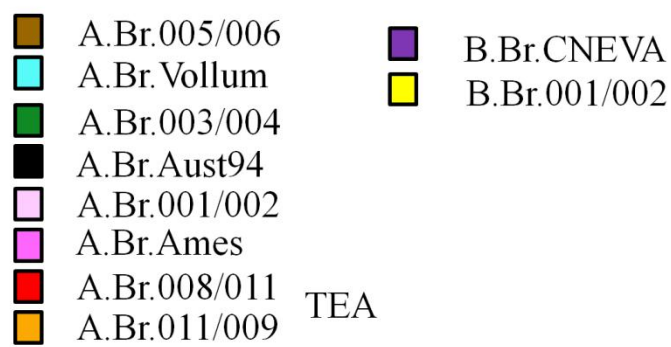
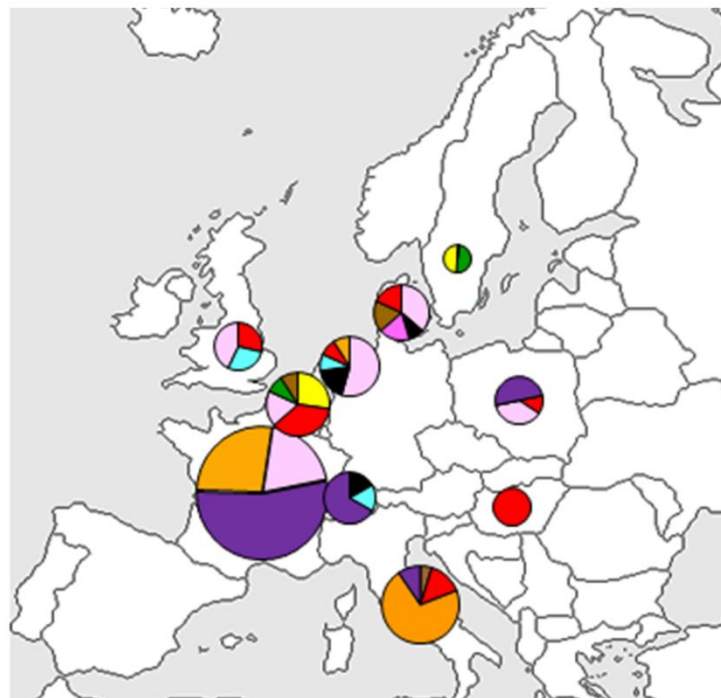


Figure 2: Minimum spanning tree of European *B. anthracis* strains based on whole-genome SNP analysis. Data are based on 9167 chromosomal SNPs. The different canSNP groups are color-coded: C.Br.A1055 in black, B.Br.001/002 in dark blue, B.Br.Kruger in dark red, B.Br.CNEVA in pink, A.Br.003/004 in red, A.Br.011/009 in light blue, A.Br.008/011 in yellow, A.Br.005/006 in light purple, A.Br.001/002 in light green, A.Br.WNA in green, A.Br.Vollum in purple, A.Br.Australia94 in brown, and A.Br.Ames in dark green. The length of each branch is proportional (logarithmic scale) to the number of SNPs identified between strains. The star marks the approximate branching point of the *B. anthracis* lineage within the *B. cereus* group.

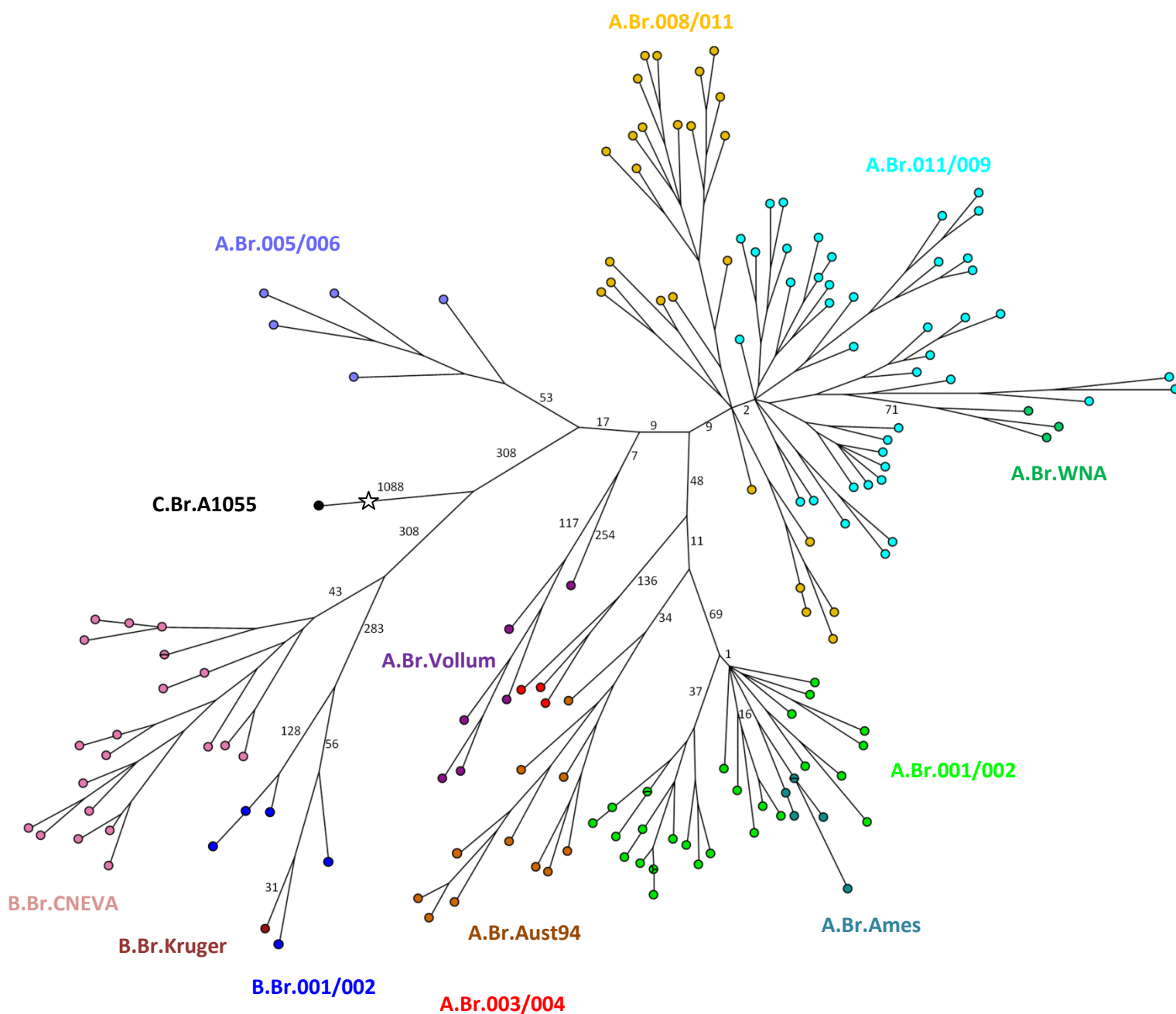
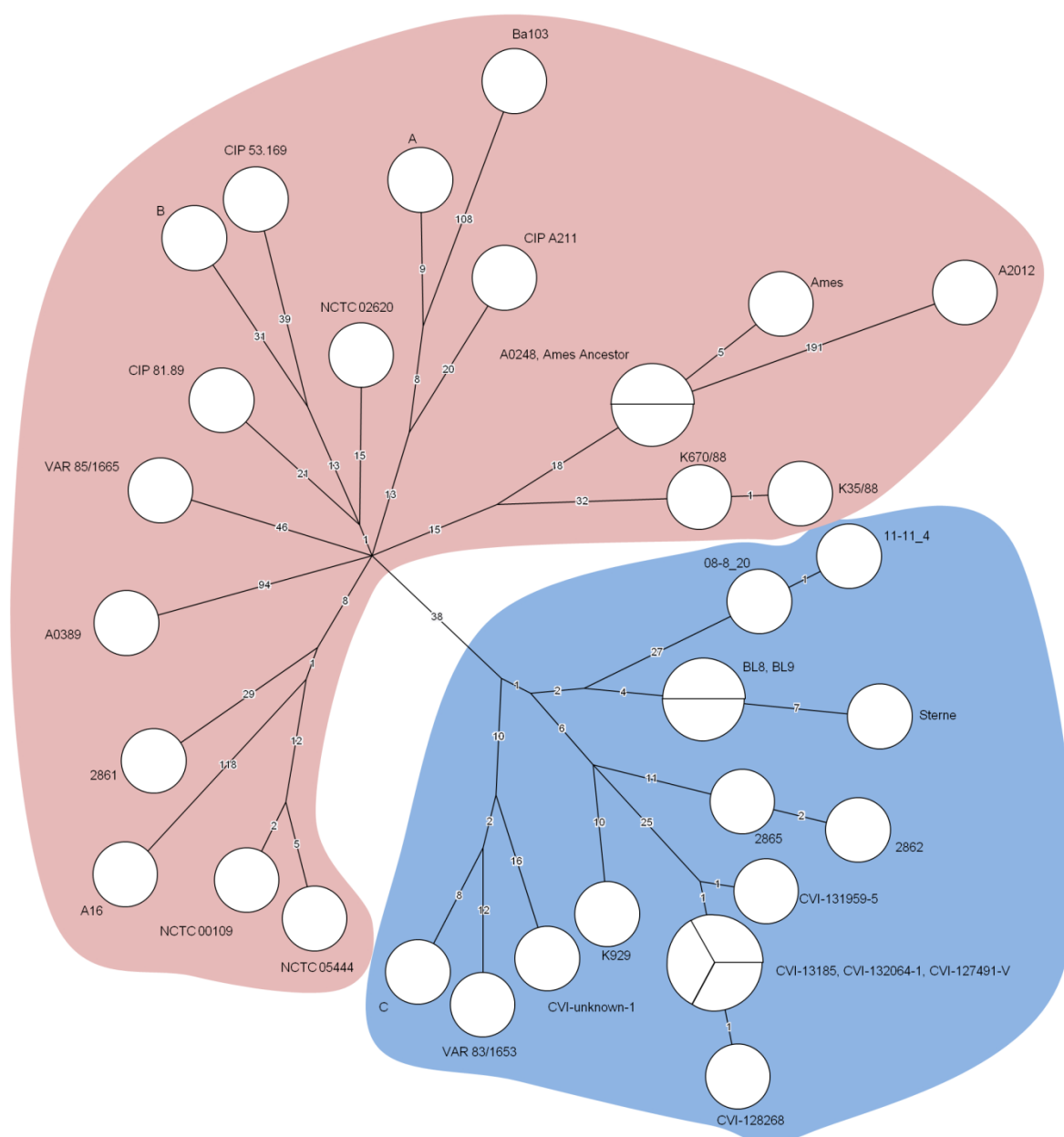


Figure 3: Minimum spanning tree based on whole-genome SNP analysis of strains affiliated to the A.Br.001/002 and A.Br.Ames canSNP group. Data are based on 1090 chromosomal SNPs. The A01 group, including Ames strains, is circled in red. The A02 group, including the Sterne vaccine strain, is circled in blue. Each circle represents a unique SNP genotype. The length of each branch is proportional (logarithmic scale) to the number of SNPs identified between strains. Homoplasmy index is 0,2%.



B.

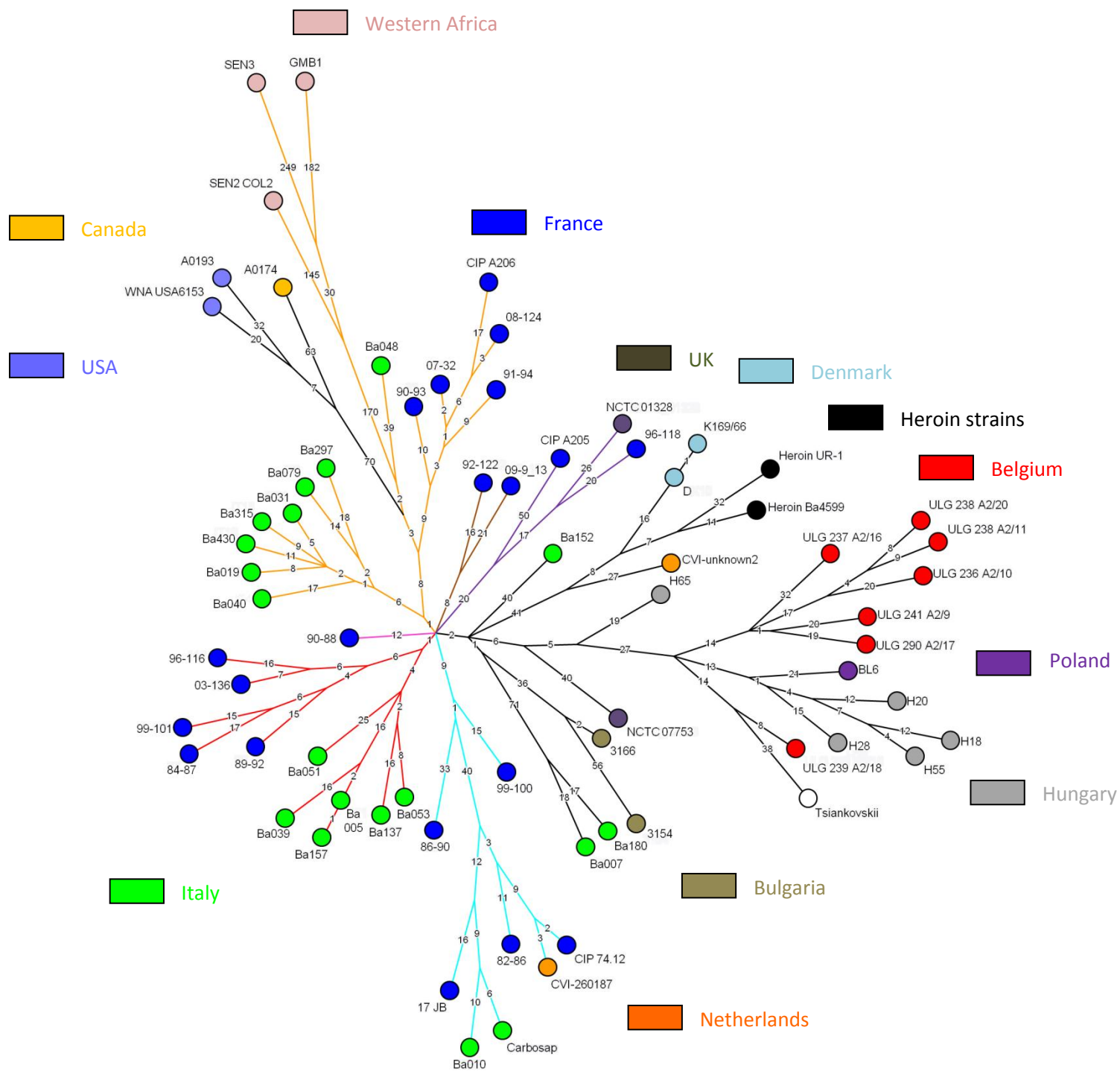


Figure 5: Minimum spanning tree based on whole-genome SNP analysis of strains affiliated to the B.Br.CNEVA canSNP group. Data are based on 586 chromosomal SNPs. Each circle represents a unique SNP genotype. Color-coded circles represent the country of origin of the strains. The length of each branch is proportional (logarithmic scale) to the number of SNPs identified between strains. Homoplasmy index is 0,34%.

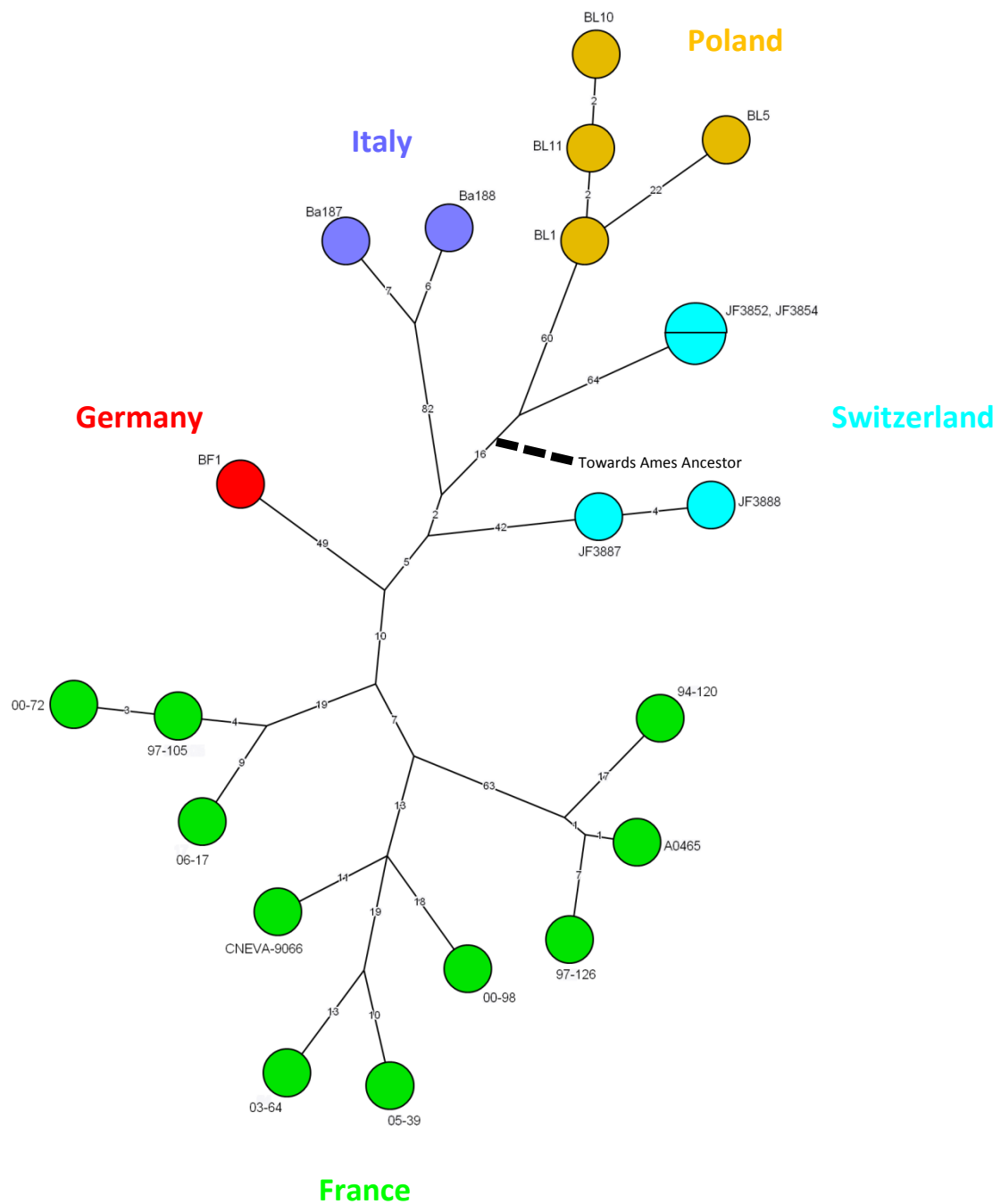


Table 1: Whole-sequenced strains and NGS data

Strain	Country	years	canSNP group	sub-group	Reads (pb)	Unambiguous bases	Genome coverage*	Sequencing depth	reference
06-17	France	2006	B.Br.CNEVA	Alps	100	5214857	99.760	231	(Girault <i>et al.</i> , 2014a)
05-39	France	2005	B.Br.CNEVA	SL	100	5214911	99.761	312	(Girault <i>et al.</i> , 2014a)
97-109#	France	1997	B.Br.CNEVA	Pyrh	100	5214843	99.759	231	(Girault <i>et al.</i> , 2014a)
03-64	France	200	B.Br.CNEVA	SL	100	5215050	99.763	288	(Girault <i>et al.</i> , 2014a)
00-72	France	2000	B.Br.CNEVA	Alps	100	5214602	99.755	297	(Girault <i>et al.</i> , 2014a)
00-98	France	2000	B.Br.CNEVA	MC	100	5215085	99.764	291	(Girault <i>et al.</i> , 2014a)
97-105	France	1997	B.Br.CNEVA	Alps	100	5214832	99.759	304	(Girault <i>et al.</i> , 2014a)
94-120	France	1994	B.Br.CNEVA	Pyrh	100	5214989	99.762	254	(Girault <i>et al.</i> , 2014a)
92-9066#	France	1992	B.Br.CNEVA	MC	100	5214839	99.759	282	(Girault <i>et al.</i> , 2014a)
97-126	France	1997	B.Br.CNEVA	Pyrh	100	5173019	98.959	227	(Girault <i>et al.</i> , 2014a)
Ba187	Italy	2006	B.Br.CNEVA	Italy	100	5214171	99,747	96	(Fasanella 2014)
Ba188	Italy	2007	B.Br.CNEVA	Italy	100	5214813	99,759	82	(Fasanella 2014)
BL1	Poland	1993	B.Br.CNEVA	Poland	100	5215578	99,773	143	(Gierczyński <i>et al.</i> , 2004)
BL5	Poland	1988	B.Br.CNEVA	Poland	250	5212522	99,715	165	(Gierczyński <i>et al.</i> , 2004)
BL10	Poland	1993	B.Br.CNEVA	Poland	100	5215493	99,772	156	(Gierczyński <i>et al.</i> , 2004)
BL11	Poland	1993	B.Br.CNEVA	Poland	100	5215647	99,775	122	(Gierczyński <i>et al.</i> , 2004)
JF3852	Switzerland	1953	B.Br.CNEVA	SW1	100	5215053	99,763	124	(Pilo <i>et al.</i> , 2008)
JF3854	Switzerland	1957	B.Br.CNEVA	SW1	100	5215771	99,777	127	(Pilo <i>et al.</i> , 2008)
JF3887	Switzerland	1960	B.Br.CNEVA	SW2	100	5214807	99,759	147	(Pilo <i>et al.</i> , 2008)
JF3888	Switzerland	1962	B.Br.CNEVA	SW2	100	5214716	99,757	147	(Pilo <i>et al.</i> , 2008)
VAR 1982/78	Belgium	1978	B.Br. 001/002		100	5213945	99,742	148	(Wattiau <i>et al.</i> , 2008)
VAR 88/28-8/1	Belgium	1988	B.Br. 001/002		100	5213452	99,733	148	(Wattiau <i>et al.</i> , 2008)
CODA	Belgium	-	B.Br. 001/002		100	5213265	99,729	140	this study
SVA11	Sweden	2011	B.Br. 001/002		xx	-	-	-	-
Ba050	Italy	1995	A.Br. 005/006	A1	100	5225980	99.972	116	(Fasanella 2014)
K836	Denmark	1968	A.Br. 005/006		100	5226729	99,987	143	this study
E	Denmark	1937	A.Br. 005/006	A111	100	5224229	99,939	124	this study
ULG 197 A2/19	Belgium	-	A.Br. 005/006	A11	100	5226246	99,978	136	this study
CVI- 23932-MS	Netherlands	1989	A.Br.Vollum	A2A	100	5226573	99,984	148	this study
NCTC 07752	UK	1948	A.Br.Vollum	A2C	100	5226506	99,983	142	this study
2054 HB2	Belgium	-	A.Br.003/004		100	5226543	99,983	134	this study
SVA08	Sweden	2008	A.Br.003/004		xx	-	-	-	(Lewerin <i>et al.</i> , 2010)
1409 dk5	Denmark	1974	A.Br.Australia94	A1	100	5224872	99,951	126	this study
CVI-188678-1	Netherlands	1993	A.Br.Australia94	A	100	5226676	99,986	146	this study
CVI-56430	Netherlands	1987	A.Br.Australia94	A	100	5204923	99,57	137	this study
JF3853	Switzerland	1952	A.Br.Australia94	B	100	5226635	99,985	135	(Pilo <i>et al.</i> , 2008)
2863	Germany	-	A.Br.Australia94	A11	250	5219921	99,857	100	this study
2864	Germany	-	A.Br.Australia94	A11	250	5225140	99,956	90	this study
2866	Germany	-	A.Br.Australia94	A11	250	5225098	99,956	104	this study
2861	Germany		A.Br. 001/002	A01/D	250	5226707	99,986	110	this study
2862	Germany		A.Br. 001/002	A02/B1A/B	250	5226202	99,977	112	this study
2865	Germany		A.Br. 001/002	A02/B1A/B	250	5226229	99,977	93	this study
A dk7	Denmark	1935	A.Br. 001/002	A01/C	100	5227288	99,997	144	this study
B	Denmark	1935	A.Br. 001/002	A01/E	100	5226754	99,987	124	this study
CIP 53.169	France	1953	A.Br. 001/002	A01/E	100	5226886	99,990	209	(Girault <i>et al.</i> , 2014a)
CIP 81.89	France	1981	A.Br. 001/002	A01/E	100	5226220	99,977	216	(Girault <i>et al.</i> , 2014a)
CIP A211	France	1954	A.Br. 001/002	A01/C	75	5227176	99,995	65	(Girault <i>et al.</i> , 2014a)
11-11_4	France	2011	A.Br. 001/002	A02/B2B	100	5226067	99,974	182	(Girault <i>et al.</i> , 2014a)
08-8_20	France	2008	A.Br. 001/002	A02/B2B	100	5226034	99,974	149	(Girault <i>et al.</i> , 2014a)
BL8	Poland	-	A.Br. 001/002	A02/B2A	100	5227236	99,996	150	(Gierczyński <i>et al.</i> , 2004)
BL9	Poland	-	A.Br. 001/002	A02/B2A	100	5227250	99,997	135	this study
K929 Dk4	Denmark	1966	A.Br. 001/002	A02/B1A/C	100	5226938	99,991	132	this study
C	Denmark	1960	A.Br. 001/002	A02/A	100	5222142	99,899	130	this study
VAR 85/1665	Belgium	1978	A.Br. 001/002	A01/F	100	5227263	99,997	150	this study
VAR 83/1653	Belgium	1983	A.Br. 001/002	A02/A	100	5222598	99,908	151	this study
CVI-unknown-1	Netherlands	1968	A.Br. 001/002	A02/A	100	5227126	99,994	151	this study
CVI-132064-1	Netherlands	1991	A.Br. 001/002	A02/B1A/A1	100	5226921	99,99	140	this study
CVI-127491 -V	Netherlands	-	A.Br. 001/002	A02/B1A/A1	100	5226821	99,989	140	this study
CVI-128268	Netherlands	1989	A.Br. 001/002	A02/B1A/A1	100	5226799	99,988	132	this study
CVI-131959-5	Netherlands	1991	A.Br. 001/002	A02/B1A/A1	100	5227270	99,997	154	this study
CVI-13185	Netherlands	1991	A.Br. 001/002	A02/B1A/A1	100	5226834	99,989	163	this study
NCTC 00109	UK	1920	A.Br. 001/002	A01/D	100	5227100	99,994	145	this study
NCTC 05444	UK	1938	A.Br. 001/002	A01/D	100	5226959	99,991	151	this study
NCTC 02620	UK	1928	A.Br. 001/002	A01/E	100	5226728	99,987	141	this study
K670/88	Denmark	1988	A.Br.Ames	A01/A	100	5227065	99,993	134	this study
K35/88	Denmark	1988	A.Br.Ames	A01/A	100	5227033	99,993	136	this study

Ba007	Italy	1996-7	A.Br.008/011	B	100	5224728	99,949	125	(Fasanella 2014)
Ba152	Italy	2009	A.Br.008/011	A	100	5226864	99,989	116	(Fasanella 2014)
Ba180	Italy	2009	A.Br.008/011	B	100	5226515	99,983	104	(Fasanella 2014)
BL6	Poland	1996	A.Br.008/011	E11/A	100	5226827	99,989	117	(Gierczyński <i>et al.</i> , 2004)
K169/66 dk1	Denmark	1966	A.Br.008/011	D11	100	5226952	99,991	137	this study
D	Denmark	1973	A.Br.008/011	D11	100	5226977	99,992	143	this study
ULG 241 A2/9	Belgium	-	A.Br.008/011	E11/B	100	5226681	99,986	133	this study
ULG 236 A2/10	Belgium	-	A.Br.008/011	E11/B	100	5103784	97,635	169	this study
ULG 238 A2/11	Belgium	-	A.Br.008/011	E11/B	100	5091586	97,402	143	this study
ULG 237 A2/16	Belgium	-	A.Br.008/011	E11/B	100	5225678	99,967	145	this study
ULG 290 A2/17	Belgium	-	A.Br.008/011	E11/B	100	5226740	99,987	131	this study
ULG 239 A2/18	Belgium	-	A.Br.008/011	E11/C	100	5226529	99,983	105	this study
ULG 238 A2/20	Belgium	-	A.Br.008/011	E11/B	250	5083634	97,249	119	this study
CVI- unknown2	Netherlands	1976	A.Br.008/011	D	100	5226515	99,983	132	this study
NCTC 07753	UK	1948	A.Br.008/011	E	100	5226919	99,99	142	this study
H18	Hungary	-	A.Br.008/011	E11/A	100	5226733	99,987	107	this study
H20	Hungary	-	A.Br.008/011	E11/A	100	5226433	99,981	129	this study
H28	Hungary	-	A.Br.008/011	E11/A	100	5213307	99,73	110	this study
H55	Hungary	-	A.Br.008/011	E11/A	100	5226778	99,988	113	this study
H65	Hungary	-	A.Br.008/011	E1	100	5226768	99,988	97	this study
99-101	France	1999	A.Br.011/009	1A CNE	100	5226789	99,988	228	(Girault <i>et al.</i> , 2014a)
84-87	France	1984	A.Br.011/009	1A CNE	100	5227090	99,994	258	(Girault <i>et al.</i> , 2014a)
89-92	France	1989	A.Br.011/009	1A CNE	100	5226833	99,989	150	(Girault <i>et al.</i> , 2014a)
96-116	France	1996	A.Br.011/009	1A CNE	100	5227103	99,994	223	(Girault <i>et al.</i> , 2014a)
03-136	France	2003	A.Br.011/009	1A CNE	100	5227069	99,993	255	(Girault <i>et al.</i> , 2014a)
08-124	France	2008	A.Br.011/009	2A NE	75	5226917	99,990	70	(Girault <i>et al.</i> , 2014a)
CIP A 206	France	1954	A.Br.011/009	2A NE	100	5219092	99,841	286	(Girault <i>et al.</i> , 2014a)
07-32	France	2007	A.Br.011/009	2A NE	100	5227041	99,993	284	(Girault <i>et al.</i> , 2014a)
90-93	France	1990	A.Br.011/009	2A NE	100	5227037	99,993	229	(Girault <i>et al.</i> , 2014a)
91-94	France	1991	A.Br.011/009	2A NE	100	5227132	99,995	302	(Girault <i>et al.</i> , 2014a)
99-100	France	1999	A.Br.011/009	3 vacc	100	5227019	99,992	144	(Girault <i>et al.</i> , 2014a)
86-90	France	1986	A.Br.011/009	3 vacc	75	5226826	99,989	77	(Girault <i>et al.</i> , 2014a)
82-86	France	1982	A.Br.011/009	3 vacc A/2	100	5202979	99,532	285	(Girault <i>et al.</i> , 2014a)
17JB	France	-	A.Br.011/009	3 vacc A/1	250	5197403	99,426	98	This study
CIP 74.12	France	1974	A.Br.011/009	3 vacc A/2	100	5200997	99,495	260	(Girault <i>et al.</i> , 2014a)
CIP A 205	France	-	A.Br.011/009	4	100	5222353	99,903	276	(Girault <i>et al.</i> , 2014a)
96-118	France	1996	A.Br.011/009	4	100	5227088	99,994	240	(Girault <i>et al.</i> , 2014a)
90-88	France	1990	A.Br.011/009	5	100	5227085	99,994	296	(Girault <i>et al.</i> , 2014a)
92-122	France	1992	A.Br.011/009	6	100	5226757	99,987	286	(Girault <i>et al.</i> , 2014a)
09-9_13	France	2009	A.Br.011/009	6	100	5226945	99,991	289	(Girault <i>et al.</i> , 2014a)
Ba005	Italy	-	A.Br.011/009	1B PLT	100	5226958	99,991	125	(Fasanella 2014)
BA010	Italy	1996-7	A.Br.011/009	3 vacc	250	5169504	98,892	143	(Fasanella 2014)
BA019	Italy	1999	A.Br.011/009	2B BC	100	5226895	99,99	127	(Fasanella 2014)
BA031	Italy	1984	A.Br.011/009	2B BC	250	5225465	99,963	194	(Fasanella 2014)
Ba039	Italy	1972	A.Br.011/009	1B PLT	250	5226425	99,981	357	(Fasanella 2014)
Ba040	Italy	1984	A.Br.011/009	2B BC	100	5226707	99,986	119	(Fasanella 2014)
BA048	Italy	2000	A.Br.011/009	2A/2C	100	5226980	99,992	131	(Fasanella 2014)
Ba051	Italy	-	A.Br.011/009	1B PLT	100	5226956	99,991	114	(Fasanella 2014)
BA053	Italy	-	A.Br.011/009	1B PLT	100	5226836	99,989	111	(Fasanella 2014)
Ba079	Italy	2004	A.Br.011/009	2B BC	100	5226820	99,989	132	(Fasanella 2014)
BA137	Italy	2006	A.Br.011/009	1B PLT	100	5226980	99,992	133	(Fasanella 2014)
BA157	Italy	2009	A.Br.011/009	1B PLT	100	5226965	99,991	141	(Fasanella 2014)
Ba297	Italy	2011	A.Br.011/009	2B BC	100	5226673	99,986	78	(Fasanella 2014)
Ba315	Italy	2011	A.Br.011/009	2B BC	100	5226759	99,987	107	(Fasanella 2014)
Ba430	Italy	2012	A.Br.011/009	2B BC	100	5226815	99,988	115	(Fasanella 2014)
CVI-260187	Netherlands	1987	A.Br.011/009	3 vacc A/2	100	5197249	99,423	143	this study
NCTC 01328	UK	1922	A.Br.011/009	4	100	5227007	99,992	149	this study

* Genome coverages are based on the Ames Ancestor genome size (5227419 pb); # also termed A0465 or CNEVA 9066

Table 2: Whole genome sequences available in public database used in this study

Strain	Country	canSNP	Accession number
Ames Ancestor	USA	A.Br.Ames	NC_007530.2
A2012	USA	A.Br.Ames	AAAC00000000.1
Ames	USA	A.Br.Ames	NC_003997.3
A0248	USA	A.Br.Ames	NC_012659.1
Sterne	South Africa	A.Br.001/002	AE017225.1
A0389	Indonesia	A.Br.001/002	ABLB00000000.1
BA103	Japan	A.Br.001/002	DRR000183 (SRA)
A16	China	A.Br.001/002	CP001970.1
V770-Np1-R	Israel	A.Br.003/004	AZQO00000000.1
CZC5	Zambia	A.Br.005/006	BAVT00000000.1
Tsiankovskii-1	Soviet Union	A.Br.008/011	ABDN00000000.2
Ba3154	Bulgaria	A.Br.008/011	ANFF00000000.1
Ba3166	Bulgaria	A.Br.008/011	ANFG00000000.1
Ba4599	Scotland	A.Br.008/011	AGQP00000000.1
UR-1	Germany	A.Br.008/011	ALNY00000000.1
Carbosap	Italy	A.Br.011/009	ANAO00000000.1
Sen2Col2	Africa	A.Br.011/009	CAVC00000000.1
Sen3	Africa	A.Br.011/009	CAVD00000000.1
Gmb1	Africa	A.Br.011/009	CAVE00000000.1
Australia 94	Australia	A.Br.Australia94	AAES00000000.1
9080 G	Georgia	A.Br.Australia94	AZUE00000000.1
52 G	Georgia	A.Br.Australia94	AZUF00000000.1
8903 G	Georgia	A.Br.Australia94	AZUD00000000.1
CDC684	USA	A.Br.Vollum	NC_012581.1
H9401	Korea	-	NC_017729.1
A0488	UK	A.Br.Vollum	ABJC00000000.1
Vollum	UK	A.Br.Vollum	AAEP00000000.1
USA6153	USA	A.Br.WNA	AAER00000000.1
A0174	Canada	A.Br.WNA	ABLT00000000.1
A0193	USA	A.Br.WNA	ABKF00000000.1
A0442	South Africa	B.Br.001/002	ABKG00000000.1
BF1	Germany	B.Br.CNEVA	AMDT00000000.1
Kruger B	South Africa	B.Br.Kruger	AAEQ00000000.1
A1055	USA	C.Br.A1055	AAEO00000000.1

Table 3: Geographical composition of *B. anthracis* isolates used in this study

Country	No. of isolates	canSNP
France	35	B.Br.CNEVA, A.Br.011/009, A.Br.001/002
Italy	21	A.Br.011/009,A.Br.008/011, B.Br.CNEVA, A.Br.005/006
The Netherlands	11	A.Br.001/002, A.Br.Australia94, A.Br.008/011
Denmark	11	A.Br.001/002,A.Br.Ames, A.Br.008/011, A.Br.005/006, A.Br.Australia94
Poland	7	B.Br.CNEVA, A.Br.008/011, A.Br.001/002
Hungary	5	A.Br.008/011
Switzerland	5	B.Br.CNEVA, A.Br.Australia94
Belgium	14	A.Br.008/011, B.Br.001/002, A.Br.001/002, A.Br.005/006, A.Br.003/004
UK	6	A.Br.001/002, A.Br.Vollum, A.Br.011/009, A.Br.008/011
Germany	6-1*	A.Br.001/002, A.Br.Australia94 - B.Br.CNEVA*
Sweden	2	B.Br.001/002, A.Br.003/004
Bulgaria	2	A.Br.008/011

ETUDE EXPERIMENTALE

Chapitre 3:

Développement de méthodes de typage de *Bacillus anthracis*

Une fois la diversité de *B. anthracis* étudiée au niveau européen, la seconde partie de mon projet de thèse consistait à identifier des marqueurs polymorphes au sein des isolats, permettant de discriminer les souches entre elles. La découverte de milliers de SNP au sein du génome du bacille du charbon a permis de sélectionner certains marqueurs afin de les développer en outil de typage.

Deux méthodes de biologie moléculaire capables de discriminer des mutations ponctuelles de type SNP ont été développées durant ce travail : la PCR-HRM (décrite précédemment dans le Chapitre 1) et le Luminex. La première permet de suivre la fusion de l'ADN en temps réel et de discriminer deux populations d'amplicons ne différant que par un seul nucléotide. Cette méthode a l'avantage d'être simple à mettre en œuvre, peu onéreuse car elle nécessite seulement un couple d'amorces, et rapide (entre 1h et 1h30 pour obtenir le résultat).

La seconde méthode développée permet le multiplexage des réactions à haut niveau (jusqu'à 100 SNP en une réaction), alors que la PCR HRM ne le permet pas.

1. Application du séquençage à haut débit: identification de marqueurs SNP informatifs permettant le génotypage de *B. anthracis*

Article 6:

Application of High-Throughput Sequencing: Discovery of Informative SNPs to Subtype *Bacillus anthracis*

Publié le 17 juin 2014 dans *Advances in Bioscience and Biotechnology*

Application of High-Throughput Sequencing: Discovery of Informative SNPs to Subtype *Bacillus anthracis*

Guillaume Girault*, Simon Thierry, Emeline Cherchame, Sylviane Derzelle

University Paris-Est, ANSES, Animal Health Laboratory, Bacterial Zoonoses Unit, Maisons-Alfort, France

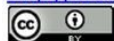
Email: guillaume.girault@anses.fr, simon.thierry@anses.fr, emeline.cherchame@anses.fr,
sylviane.derzelle@anses.fr

Received 17 April 2014; revised 29 May 2014; accepted 17 June 2014

Copyright © 2014 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Single Nucleotide Polymorphisms (SNPs) are the most common and abundant genetic variation found in the genome of any living species, from bacteria to humans. In bacterial genotyping, these evolutionarily stable point mutations represent important DNA markers that can be used to elucidate deep phylogenetic relationships among worldwide strains, but also to discriminate closely related strains. With the advent of next generation sequencing (NGS) technologies, affordable solutions are now available to get access to the complete genome sequence of an organism. Sequencing efforts of an increasing number of strains provide an unprecedented opportunity to create comprehensive species phylogenies. In this study, a comparative analysis of 161 genomes of *Bacillus anthracis* has been conducted to discover new informative SNPs that further resolves *B. anthracis* SNP-based phylogenetic tree. Nine previously unpublished SNPs that define major groups or sub-groups within the *B. anthracis* species were selected and developed into SNP discriminative assays. We report here a cost-effective high-resolution melting-based genotyping method for the screening of 27 canonical SNPs that includes these new diagnostic markers. The present assays are useful to rapidly assign an isolate to one sub-lineages or sub-groups and determine its phylogenetic placement on the *B. anthracis* substructure population.

Keywords

SNPs, *Bacillus anthracis*, Genotyping, HRM, Phylogeny

1. Introduction

Bacillus anthracis is a zoonotic pathogen that primarily affects grazing herbivores, but can be used as an agent of

*Corresponding author.

How to cite this paper: Girault, G., Thierry, S., Cherchame, E. and Derzelle, S. (2014) Application of High-Throughput Sequencing: Discovery of Informative SNPs to Subtype *Bacillus anthracis*. *Advances in Bioscience and Biotechnology*, 5, 669-677. <http://dx.doi.org/10.4236/abb.2014.57079>

bioterrorism. This Gram-positive endospore-forming bacterium is the causative agent of anthrax, an infectious and life threatening diseases in all mammals, including humans. *B. anthracis* evolved as a highly fit clonal pathogen that has spread throughout the world and became ecologically established as local, genetically distinct populations [1].

In both microbiological forensics and epidemiological investigations, obtaining reliable information regarding the identification and source of a suspicious strain is often essential to set up effective responses, as highlighted by the Amerithrax investigations following the so-called 2001 “anthrax letter attacks” in the United States [2]. The principal genomic markers used for genotyping a highly monomorphic bacterium such as *B. anthracis* are variable number tandem repeats (VNTRs) [3]–[5], single nucleotide repeats (SNRs) [6]–[8] and SNPs [9]. Since 2007, the genotyping of a set of 13 canonical SNPs has emerged as the gold standard method to be used for determining phylogenetic relationships within this bacterium [10]–[13]. SNPs are generated by nucleotide substitutions, probably associated to errors during DNA replication that are not repaired. These rare events occur at an estimated rate of approximately 10^{-10} changes per nucleotide per generation in the *B. anthracis* species [14]. Because of their low mutation rates, SNPs represent ideal targets for phylogenetic analyses. SNPs are evolutionarily stable markers that are unlikely to mutate back to their ancestral state [9]. This stability can be used for defining canonical markers representative of groups of related strains, such as phylogenetic sub-groups [9].

The substructure of the *B. anthracis* population is divided into three major clades (A, B and C), with further subdivisions into 13 major lineages or groups [10] [12] [13]. The A-radiation has known the most dramatic dispersal and clonal expansion. It represents nowadays the majority of anthrax cases reported around the world [3] [15] [16]. The eight A-lineages or groups exhibit distinct geographical patterns across the five continents. Most basal of the A-clade is the African A.Br.005/006 group, consistent with an “out of Africa” hypothesis for *B. anthracis*. Soon after this divergence, the Vollum-clade appeared, with isolates currently ecologically established in Iran, Pakistan and Afghanistan, for instance. The highly successful TransEurasian (TEA) group (A.Br.008/009) is slightly more recent and spread across Europe and Asia from which a derived lineage (A.Br. WNA) was introduced into North America. Another recent bifurcation led to the A.Br.003/004, A.Br. Australia-94 and Ames/Sterne (A.Br.001/002) groups which contain isolates found, respectively, in South America (A.Br.003/004), Asia and part of Europe (A.Br.Aust94 and A.Br.001/002) [1] [10]. Collections from western and south-central Asia (Turkey, India) and western China are dominated by genotypes belonging to the A.Br.Aust94 lineage. Further into central and eastern China the genotypes are dominated by isolates belonging to the A.Br.001/002 group and its derived A.Br.Ames lineage. The B clade is divided into two genetically distinct lineages with geographical restricted repartition. The B.Br.001/002 group (including the derived B.Br.Kruger lineage) is ecologically established in Southern Africa, in particular in the Kruger National Park (South Africa), where it co-exists with other strains from the A-clade [10] [17]. The B.Br.CNEVA lineage is exclusively found in continental Europe [10] [18]–[21]. The C-clade is composed of an uncommon lineage, C.Br.A1055, clustering only three collection’s isolates of unknown origin [10] [22].

Whole genome sequencing is the ultimate genotyping tool, giving the highest possible resolution, with comparison information spanning from distant phylogenetic relationships to the highest level of subtyping. Identification of thousands of SNPs retrieved from compiled NGS sequences facilitates high-resolution strain tracking and provides the level of discrimination required for microbial forensics. Canonical SNP typing and whole genome comparisons of multiple strains will soon become the method of choice in *B. anthracis* genotyping [23]. We have recently reported the high throughput sequencing and phylogeographical analysis of 122 strains of *B. anthracis* isolated in France [24]. The resulting whole-genome SNPs analysis determined relationships existing between French strains at a level of discrimination never reached before and reconstructed the French population substructure in details. Eight new canonical SNPs have been identified that allow classifying French *B. anthracis* isolates at a higher resolution [24]. In this study, we compared the chromosomal sequences of 161 globally diverse strains, focusing on identifying additional informative SNPs, to provide further resolution into current SNP-based typing of *B. anthracis*. Twenty-seven SNP discrimination assays, including nine newly described markers, are reported. The PCR-HRM-based method presented here is a cost-effective tool to genotype the highly clonal *B. anthracis* species.

2. Materials and Methods

2.1. Extraction of Whole-Genome SNPs among *B. anthracis* Strains

In this study, the genome of one African A.Br.005/006-affiliated strain (IEMVT 89) and a hundreds of French

strains ($n = 123$) belonging to 3 groups or lineages (A.Br.001/002, A.Br.011/009 (A.Br.008/009 sub-group), and B.Br.CNEVA) [24] were whole-genome compared to thirty-seven globally diverse genomes publically available in database [25] (Table 1). Ames Ancestor [GenBank: AE017334.2] was used as reference genome. Sequences alignment and whole-genome SNPs discovery were carried out using the BioNumerics version 7.0 software (Applied Maths, Belgium) and its Chromosome Comparisons module. SNPs were next filtered to avoid erroneous and biased data. Missing sequences, ribosomal operons, VNTR loci and contiguous SNPs in a 10 pb window were eliminated from the analysis.

Table 1. Whole genome sequences available in public database used in this study.

Strain	Country	canSNP	Accession number
Ames Ancestor	USA	A.Br.Ames	NC_007530.2
A2012	USA	A.Br.Ames	AAAC00000000.1
Ames	USA	A.Br.Ames	NC_003997.3
A0248	USA	A.Br.Ames	NC_012659.1
Sterne	South Africa	A.Br.001/002	AE017225.1
A0389	Indonesia	A.Br.001/002	ABLB00000000.1
Ba103	Japan	A.Br.001/002	DRR000183 (SRA)
A16	China	A.Br.001/002	CP001970.1
V770-Np1-R (ATCC 14185)	Israel	A.Br.003/004	AZQO00000000.1
CZC5	Zambia	A.Br.005/006	BAVT00000000.1
Tsiankovskii	Soviet Union	A.Br.008/011	ABDN00000000.2
Ba 3154	Bulgaria	A.Br.008/011	ANFF00000000.1
Ba 3166	Bulgaria	A.Br.008/011	ANFG00000000.1
Heroin Ba4599	Scotland	A.Br.008/011	AGQP00000000.1
UR-1	Germany	A.Br.008/011	ALNY00000000.1
Carbosap	Italy	A.Br.011/009	ANAO00000000.1
Sen2Col2	Africa	A.Br.011/009	CAVC00000000.1
Sen3	Africa	A.Br.011/009	CAVD00000000.1
Gmb1	Africa	A.Br.011/009	CAVE00000000.1
Australia 94	Australia	A.Br.Australia94	AAES00000000.1
9080 G	Georgia	A.Br.Australia94	AZUE00000000.1
52 G	Georgia	A.Br.Australia94	AZUF00000000.1
8903 G	Georgia	A.Br.Australia94	AZUD00000000.1
CDC684	USA	A.Br.Vollum	NC_012581.1
H9401	Korea	A.Br.Vollum	NC_017729.1
A0488	UK	A.Br.Vollum	ABJC00000000.1
Vollum	UK	A.Br.Vollum	AAEP00000000.1
USA6153	USA	A.Br.WNA	AAER00000000.1
A0174	Canada	A.Br.WNA	ABLT00000000.1
A0193	USA	A.Br.WNA	ABKF00000000.1
A0442	South Africa	B.Br.001/002	ABKG00000000.1
SVA11	Sweden	B.Br.001/002	CP006742.1
BF1	Germany	B.Br.CNEVA	AMDT00000000.1
CNEVA-9066	France	B.Br.CNEVA	AAEN00000000.1
A0465	France	B.Br.CNEVA	ABLH00000000.1
Kruger B	South Africa	B.Br.KrugerB	AAEQ00000000.1
A1055	USA	C.Br.A1055	AAEO00000000.1

2.2. Phylogenetic Analysis

A minimum spanning tree was drawn in BioNumerics by using the filtered whole genome sequencing SNP data as input. Nodes were automatically numbered by the software. Canonical SNPs that defining major group, sub-groups or clusters of strains along the SNP tree were identified by searching for key signatures with allelic states shared only by these selected subgroups.

2.3. SNP Discrimination Assays by HRM

After identification, primer sequences suitable for HRM analysis were designed for each SNP marker using the Primer 3+ software (<http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/>) (Table 2). Short amplicons (<100 pb) were selected. High-resolution melting (HRM) is a post-PCR technique that determines with high precision the melt profile of PCR products. A new generation dye is incorporated into a double-stranded DNA. Using a slow constant increase in temperature, fluorescence acquisition allows the distinction between two different populations of amplicons.

Amplification was performed on the ViiA7™ Real-Time PCR System (Life Technologies) using the LightCycler® 480 High Resolution Melting Master Mix (Roche Diagnostics). The reaction mixture consisted of 0.2 µM of each primer, 1 × LightCycler® 480 HRM master mix and 2.5 mM MgCl₂ in a 10 µl final volume. The following parameters were used: 10 min at 95°C were followed by 40 cycles consisting of 10 s at 95°C, 10 s at 58°C and 20 s at 72°C. Samples were next heated to 95°C for 30 s, cooled down to 65°C for 1 min and heated from 65°C to 88°C at a rate of 1°C/s with 25 acquisitions/°C. HRM data were analyzed by the ViiA7™ Software (version 1.2.1).

3. Results

3.1. In Silico SNP Typing

Thirty-seven *B. anthracis* genomes available in public NCBI database or Sequence Read Archive (SRA) were used for this study [25] (Table 1). These globally diverse genomes were first subjected to *in silico* canSNP typing using the 14 canSNPs references set previously published [10] [12] to determine their phylogenetic placement on the *B. anthracis* canSNP tree (Table 1). French strains were previously shown to belong to A.Br.011/009 (a sub-group of the TEA population) (n = 32), A.Br.001/002 (n = 24) and B.Br.CNEVA (n = 67) lineages or groups. The African IEMVT 89 strain is affiliated to A.Br.005/006 [24].

3.2. Phylogenetic Analysis

A total of 161 genomes of *B. anthracis* were aligned to the Ames ancestor reference sequence and whole-genome compared with a focus on SNPs discovery. SNPs that define the major A.Br.Aust94 lineage (four genomes), A.Br.003/004 group (one genome) and A.Br.005/006 group (two genomes) were identified and one canonical marker selected for each subpopulation (Table 2, Figure 1). Interestingly, the whole-genome SNP analysis further resolved the A.Br.001/002 subpopulation (32 genomes) into two distinct phylogenetic sub-groups which were named A01 and A02 (Figure 1, inset A). The A01 sub-group (also termed “Ames sub-group”) radiates very shortly after the A01-A02 divergence (1 SNP at position 515111) into at least 5 sub-branches. The first one contains the A.Br.Ames lineage (strains Ames Ancestor, Ames, A2012 and A0248). The second one is composed by the A0389 strain. The third one includes the Japanese strain Ba103 and one old French strain (CIP A211). The two last branches are composed of, respectively, two old French isolates (CIP 81.89 and CIP 53.169), and the A16 Chinese strain. The A02 sub-group (also termed “Sterne sub-group”) clusters the Sterne vaccine strain and all recent strains isolated in France from the Doubs department, including the 08-8_20 genome [24]. Noteworthy to mention, the Carbosap vaccine was found to be phylogenetically related to five French genomes positioned within the third branch of the six sub-branches previously discovered within the A.Br.011/009 sub-groups. SNPs specific to these A.Br.011/009 branches were also selected (Table 2, Figure 1, inset B).

3.3. canSNP Discrimination Assays Using HRM

Twenty-seven canSNP discrimination assays, including nine unpublished markers, were designed using the

Table 2. canSNPs and primer sequences used for HRM analysis.

canSNP name	Target lineage	Position*	Forward Primer (5' - 3')	Reverse Primer (5' - 3')	SNP	Reference
A.Br.001	A.Br.Ames	182106	CAAGCGGAACCAAAT TTAATCTTT	TTCACCGTACGTCATTG TATAATACG	T/C	[10]
A.Br.002	A.Br.Ames, A.Br.001/002	947760	AACGATACCTAAAAT CGATAAAG	GGCAGAAGGAGCAAGT AATGTT	G/A	[10]
A01	A.Br.Ames, A.Br.01	515111	CGTGGCGTAAAAATA AAGATCC	TCGTTTGCTGCAACATA ATTTC	A/G	[11]; this study
A02	A.Br.02	240050	AGCTGCTGAAAATGA TGTAGCA	GAACCATCCCAAAGTTT GATTG	T/C	[11]; this study
Ba20	A.Br.02 (Fr-Doubs)	3434997	AGCGAGCCAATTTTG GAACCGA	AGGCGGGATTGTTGGTG GATGT	A/C	[24]
A.Br.003	A.Br.Ames, A.Br.001/002, A.Br.Aust94	1493157	GCTACTGTCATTGTAT AAAAACCTCCTTT	CGCTTGCCAAGCTTTTTT TC	A/G	[10]
A94	A.Br.Aust94	569295	CCCTGCACCTAACATA ACAACA	CCGCAATGTACCCACTGT ATAA	C/T	[11]; this study
A.Br.004	A.Br.Ames, A.Br.001/002, A.Br.Aust94, A.Br.003/004	3600659	CCGATACCAGTAAAC GACGACAT	CTGGAATTGGTGGAGCTA TGGA	T/C	[10]
A03	A.Br.003/004	3956945	CGAAAGAACATATTG CACTCGTAA	GCGTACAAGTACAGGCTC ACCT	A/G	this study
A05	A.Br.005/006	405303	CTATATTTTGGACTTG TGCGACAG	GCAATGTTTTGGGAGCTT AGTG	C/T	this study
A.Br.006	A-Clade	162509	CCGAAATTTGCTATTA GAACGAA	TCCCAATCTAGCGTTTTTA AGTTCA	C/A	[10]
A.Br.007	A.Br.Vollum	266439	TTGGTAACGAGACGAT AAACTGAATAA	GCCTTGGATTGGCGATTG	T/C	[10]
A.Br.008	A.Br.008/009	3947248	TTGCGAACTACGCTATA CGTTTTAGAT	CAAAACGGTGAAAAAGTT ACAAATATACG	T/G	[10]
A.Br.009	A.Br.WNA	2589823	GGCAATCGGCCACTGT TT	GGGTTTCTACTGTGTATG TGTGTAATAAAAAAG	A/G	[10]
A.Br.011	A.Br.011/009	2552486	CTAAAAAGAAACGAAT TCCCGCTGA	CGATAAAAAATCGGAATT GAAGCAGGAG	G/A	[12]; [13]
A11/1 (Ba21)	A.Br.011/009 (Branch 1 Fr)	3562427	AGCAAAAAGTCGGCAA AGAA	ACAGAGCTTCCTCCGAA CTG	A/C	[24]
A11/2 (Ba22)	A.Br.011/009 (Branch 2 Fr)	820195	AGTGGTGCAATCCCAA TTTC	CGCAGCAATATTCGCTA TCA	T/C	[24]
A11/3	A.Br.011/009 (Branch 3)	1975689	CACATTGTCAATATTA AATCGCTGA	TGCTTCTCCACCTAAATC AAATG	G/A	this study
A11/4	A.Br.011/009 (Branch 4)	2571863	TCCTACACCTGTTTCAC CACAA	AAGCTTGTATCACCAACT GATGC	C/T	this study
A11/5	A.Br.011/009 (Branch 5)	5021456	TATTTGTTGAAAGAGC GTCATCC	TGTGTCAGTCTCCGCATA CAATA	T/C	this study
A11/6	A.Br.011/009 (Branch 6)	1391599	ACCAAGTAATTAAGTT CCAAGCTATTG	ATTGCGCTTACAAAATGG TTCTC	C/T	this study
B.Br.001	B.Br.KrugerB	1455279	TGCATGCTTCTTCTTAC AGAGTAGTTAAT	CGGTCATAAAAGAAATCG GTACAA	T/C	[10]
B.Br.002	B.Br.KrugerB, B.Br.001/002	1056740	TGTTGCACCTTCTGTGT TCGTT	GTAGTGGCTTACCGAATG GA	G/T	[10]
B.Br.003	B-Clade	1494269	CATTATTCGCATAGAA GCAGATGA	TGTGCCATCAAATAACTCT TTCTCAA	G/A	[10]
B.Br.004	B.Br.CNEVA	69952	GAAGTTAAGTATCAACC AGCAGAAGAAA	CCGCCGCCTTGAGCTT	T/C	[10]
Ba 19	B.Br.CNEVA (France)	2573536	CATATATTTACCTCTT TTATGAACA	GATAAAAGGCTGTCGGATGG	A/G	[24]
A/B.Br.001	C.Br.A1055	3697886	GAAGGTCCTCAATTTGG ATTTAAAAAT	CGTGTGAACCTTTCGGTAA ATAGTC	A/G	[10]

*Position based on the Ames Ancestor genome (NC_007530.2).

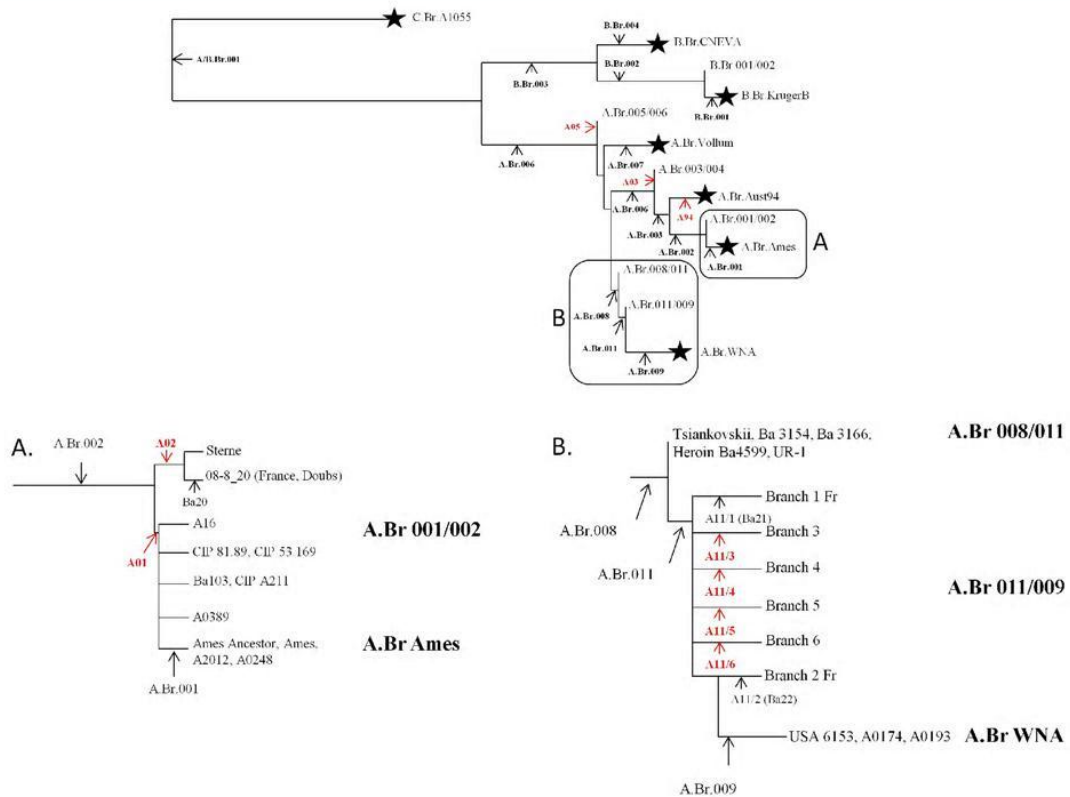


Figure 1. Phylogeny of the major canSNP groups of *B. anthracis*. Dendrogram of 14 canSNP analysis of *Bacillus anthracis* isolates after Van Ert *et al.* (2007) and Marston *et al.* (2011). Canonical SNP positions and names that define lineages, groups or sub-groups are indicated in black arrows (previously published canSNP) and red arrows (newly discovered canSNPs). The A.Br.001/002 group is now subdivided into two sub-groups: A.Br.01 and A.Br.02 (inset A). The A.Br.011/009 sub-group is further resolved into six branches as previously described for the French TEA subpopulation (inset B). These new groups are named after the canSNP marker and assay that define their position.

HRM-PCR technology (Table 2). Novel assays targeted four canSNPs specific to the A.Br.011/009 sub-branches 3 to 6 (A11/3, A11/4, A11/5, A11/6), two canSNPs specific to the new A01 and A02 sub-groups within A.Br.001/002 (A01, A02) and three canSNPs specific to the main A.Br.Aust94, A.Br.003/004 and A.Br.005/006 subdivisions (A94, A03 and A05, respectively). These diagnostic assays were successfully validated against all DNA samples of our *B. anthracis* collection (about 250), including a hundred of non-French *B. anthracis* DNAs of diverse origins (described by 10 canSNP typing). Average melting temperatures calculated for each bi-allelic assay are reported in Table 3. All non-sequenced genomes fell on the correct group (data not shown), confirming HRM-based assays' specificity.

4. Discussion

Cost is an important issue for all diagnostic assays. In this respect, real-time PCR platforms and high-resolution DNA melting analysis (HRM) are attractive technologies for SNP-interrogation. HRM techniques can determine with high precision the melt profile of PCR products using accurate fluorescence data acquisition over small temperature increments [26]. Two standard primers are used to amplify short segments flanking each SNP. The melting profile of the resulting amplicon is characteristic of its GC content, *i.e.*, a substitution of a G or C to an A or T reduces the melting temperature (T_m) while a substitution of an A or T to a G or C increases the T_m . Based on its simplicity, low cost, non-destructive nature, high sensitivity and specificity, the popularity of HRM analysis has grown considerably in the last few years. HRM analysis is an effective alternative to other

Table 3. Melting temperature of new canonical SNPs.

canSNP name	Target	Allele	T _m values (°C)
A01	A.Br.Ames, A.Br.001/002 subgroup A01	G	77.2 ± 0.1
	Others	A	76.6 ± 0.1
A02	A.Br.001/002 subgroup A02	C	75.9 ± 0.1
	Others	T	75.3 ± 0.1
Ba20	A.Br.001/002 subgroup A02 (France, Doubs)	C	80.4 ± 0.1
	Others	A	79.8 ± 0.1
A94	A.Br.Aust94	T	77.3 ± 0.1
	Others	C	78.0 ± 0.1
A03	A.Br.003/004	G	76.6 ± 0.1
	Others	A	75.8 ± 0.1
A05	A.Br.005/006	T	76.4 ± 0.1
	Others	C	77.6 ± 0.1
A11/1 (Ba21)	French A.Br.011/009 Branch 1	C	77.9 ± 0.1
	Others	A	77.1 ± 0.1
A11/2 (Ba22)	French A.Br.011/009 Branch 2	C	77.5 ± 0.1
	Others	T	76.9 ± 0.1
A11/3	A.Br.011/009 Branch 3	A	72.1 ± 0.1
	Others	G	72.9 ± 0.1
A11/4	A.Br.011/009 Branch 4	A	74.6 ± 0.1
	Others	G	75.8 ± 0.1
A11/5	A.Br.011/009 Branch 5	C	76.2 ± 0.1
	Others	T	75.3 ± 0.1
A11/6	A.Br.011/009 Branch 6	T	75.8 ± 0.1
	Others	C	76.6 ± 0.1
Ba19	French B.Br.CNEVA group	G	75.8 ± 0.1
	Others	A	75.2 ± 0.1

PCR-based methods such as Dual Probe TaqMan PCR assays [27] and Mismatch Amplification Mutation Assays (MAMA) [28] [29]. HRM is an ideal format for scoring a small number of SNPs. PCR-HRM assays can be rapidly designed around SNPs to determine the extent to which each marker varies in the *B. anthracis* population.

Over the last decade, significant research efforts have been undertaken to develop genotyping methods with increased resolving power for *B. anthracis* strain differentiation. The absence of horizontal transfer of genetic material, a slow rate of accumulation of mutations and the paucity of large genome rearrangements within the genome of *B. anthracis* [9] make SNPs markers of choice for genotyping a species. The advent of next-generation sequencing technologies and comparative genome analyses offers a powerful approach for identifying rare polymorphisms for the unbiased typing of highly clonal pathogens. Since 2007, and the identification by Van Ert *et al.* of 13 “canonical” SNPs representing key phylogenetic positions along the *B. anthracis* phylogenetic SNP tree, a growing number of diagnostic SNPs has been discovered and used to survey *B. anthracis* diversity

in nature [11] [24] [30]-[32]. In this study, a same but large-scale approach, comparing 161 *B. anthracis* genomes of diverse origins, were conducted to discover novel informative SNPs that can further discriminate within lineages or groups of strains. Nine novel DNA signatures were selected and developed into SNP discrimination assays. Combined with previously established canSNPs [10]-[12] [24], the set of 27 markers described here allows a more precise and reliable phylogenetic assignment of any *B. anthracis* strain into 20 sub-lineages or sub-groups. The present method will contribute to improve our ability to characterize *B. anthracis* strains and to react rapidly when the identity and origin of a suspicious strain need to be established.

Acknowledgements

This work was supported by the French National Research Agency (ANR) and Direction Générale de l'Armement (DGA) (project number ANR11ASTR0007). GG is a PhD student co-supported by ANSES and DGA grants.

This work has benefited from the facilities and expertise of the high throughput sequencing platform of IMAGIF (Centre de Recherche de Gif, <http://www.imagif.cnrs.fr>).

References

- [1] Keim, P., Okinaka, R. and Pearson, T. (2013) The Evolution and Global Dissemination of *Bacillus anthracis*. *The International Conference on Bacillus anthracis, B. cereus, and B. thuringiensis*, Victoria, 1-5 September 2013.
- [2] Rasko, D.A., Worsham, P.L., Abshire, T.G., Stanley, S.T., Bannan, J.D., Wilson, M.R., Langham, R.J., Decker, R.S., Jiang, L., Read, T.D., *et al.* (2011) *Bacillus anthracis* Comparative Genome Analysis in Support of the Amerithrax Investigation. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 5027-5032. <http://dx.doi.org/10.1073/pnas.1016657108>
- [3] Keim, P., Price, L.B., Klevytska, A.M., Smith, K.L., Shupp, J.M., Okinaka, R., Jackson, P.J. and Hugh-Jones, M.E. (2000) Multiple-Locus Variable-Number Tandem Repeat Analysis Reveals Genetic Relationships within *Bacillus anthracis*. *Journal of Bacteriology*, **182**, 2928-2936. <http://dx.doi.org/10.1128/JB.182.10.2928-2936.2000>
- [4] Lista, F., Faggioni, G., Valjevac, S., Ciammaruconi, A., Vaissaire, J., le Doujet, C., Gorge, O., De Santis, R., Carattoli, A., Ciervo, A., *et al.* (2006) Genotyping of *Bacillus anthracis* Strains Based on Automated Capillary 25-Loci Multiple Locus Variable-Number Tandem Repeats Analysis. *BMC Microbiology*, **6**, 1-15.
- [5] Le Flèche, P., Hauck, Y., Onteniente, L., Prieur, A., Denoeud, F., Ramisse, V., Sylvestre, P., Benson, G., Ramisse, F. and Vergnaud, G. (2001) A Tandem Repeats Database for Bacterial Genomes: Application to the Genotyping of *Yersinia pestis* and *Bacillus anthracis*. *BMC Microbiology*, **1**, 2.
- [6] Stratilo, C.W., Lewis, C.T., Bryden, L., Mulvey, M.R. and Bader, D. (2006) Single-Nucleotide Repeat Analysis for Subtyping *Bacillus anthracis* Isolates. *Journal of Clinical Microbiology*, **44**, 777-782. <http://dx.doi.org/10.1128/JCM.44.3.777-782.2006>
- [7] Kenefic, L.J., Beaudry, J., Trim, C., Daly, R., Parmar, R., Zanecki, S., Huynh, L., Van Ert, M.N., Wagner, D.M., Graham, T., *et al.* (2008) High Resolution Genotyping of *Bacillus anthracis* Outbreak Strains Using Four Highly Mutable Single Nucleotide Repeat Markers. *Letters in Applied Microbiology*, **46**, 600-603. <http://dx.doi.org/10.1111/j.1472-765X.2008.02353.x>
- [8] Garofolo, G., Ciammaruconi, A., Fasanella, A., Scasciamacchia, S., Adone, R., Pittiglio, V. and Lista, F. (2010) SNR Analysis: Molecular Investigation of an Anthrax Epidemic. *BMC Veterinary Research*, **6**, 11. <http://dx.doi.org/10.1186/1746-6148-6-11>
- [9] Keim, P., Van Ert, M.N., Pearson, T., Vogler, A.J., Huynh, L.Y. and Wagner, D.M. (2004) Anthrax Molecular Epidemiology and Forensics: Using the Appropriate Marker for Different Evolutionary Scales. *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, **4**, 205-213.
- [10] Van Ert, M.N., Easterday, W.R., Huynh, L.Y., Okinaka, R.T., Hugh-Jones, M.E., Ravel, J., Zanecki, S.R., Pearson, T., Simonson, T.S., U'Ren, J.M., *et al.* (2007) Global Genetic Population Structure of *Bacillus anthracis*. *PLoS One*, **2**, Article ID: e461.
- [11] Kuroda, M., Serizawa, M., Okutani, A., Sekizuka, T., Banno, S. and Inoue, S. (2010) Genome-Wide Single Nucleotide Polymorphism Typing Method for Identification of *Bacillus anthracis* Species and Strains among *B. cereus* Group Species. *Journal of Clinical Microbiology*, **48**, 2821-2829. <http://dx.doi.org/10.1128/JCM.00137-10>
- [12] Marston, C.K., Allen, C.A., Beaudry, J., Price, E.P., Wolken, S.R., Pearson, T., Keim, P. and Hoffmaster, A.R. (2011) Molecular Epidemiology of Anthrax Cases Associated with Recreational Use of Animal Hides and Yarn in the United States. *PLoS One*, **6**, Article ID: e28274. <http://dx.doi.org/10.1371/journal.pone.0028274>

- [13] Price, E.P., Seymour, M.L., Sarovich, D.S., Latham, J., Wolken, S.R., Mason, J., Vincent, G., Drees, K.P., Beckstrom-Sternberg, S.M., Phillippy, A.M., *et al.* (2012) Molecular Epidemiologic Investigation of an Anthrax Outbreak among Heroin Users, Europe. *Emerging Infectious Diseases*, **18**, 1307-1313. <http://dx.doi.org/10.3201/eid1808.111343>
- [14] Vogler, A.J., Busch, J.D., Percy-Fine, S., Tipton-Huntton, C., Smith, K.L. and Keim, P. (2012) Molecular Analysis of Rifampicin Resistance in *Bacillus anthracis* and *Bacillus cereus*. *Antimicrobial Agents and Chemotherapy*, **46**, 511-513. <http://dx.doi.org/10.1128/AAC.46.2.511-513.2002>
- [15] Keim, P., Kalif, A., Shupp, J., Hill, K., Travis, S.E., Richmond, K., Adair, D.M., Hugh-Jones, M., Kuske, C.R. and Jackson, P. (1997) Molecular Evolution and Diversity in *Bacillus anthracis* as Detected by Amplified Fragment Length Polymorphism Markers. *Journal of Bacteriology*, **179**, 818-824.
- [16] Keim, P., Gruendike, J.M., Klevytska, A.M., Schupp, J.M., Challacombe, J. and Okinaka, R. (2009) The Genome and Variation of *Bacillus anthracis*. *Molecular Aspects of Medicine*, **30**, 397-405. <http://dx.doi.org/10.1016/j.mam.2009.08.005>
- [17] Smith, K.L., DeVos, V., Bryden, H., Price, L.B., Hugh-Jones, M.E. and Keim, P. (2000) *Bacillus anthracis* Diversity in Kruger National Park. *Journal of Clinical Microbiology*, **38**, 3780-3784.
- [18] Derzelle, S., Laroche, S., Le Fleche, P., Hauck, Y., Thierry, S., Vergnaud, G. and Madani, N. (2011) Characterization of Genetic Diversity of *Bacillus anthracis* in France by Using High-Resolution Melting Assays and Multilocus Variable-Number Tandem-Repeat Analysis. *Journal of Clinical Microbiology*, **49**, 4286-4292. <http://dx.doi.org/10.1128/JCM.05439-11>
- [19] Garofolo, G., Serrecchia, L., Corro, M. and Fasanella, A. (2011) Anthrax Phylogenetic Structure in Northern Italy. *BMC Research Notes*, **4**, 273. <http://dx.doi.org/10.1186/1756-0500-4-273>
- [20] Pilo, P., Perreten, V. and Frey, J. (2008) Molecular Epidemiology of *Bacillus anthracis*: Determining the Correct Origin. *Applied and Environmental Microbiology*, **74**, 2928-2931. <http://dx.doi.org/10.1128/AEM.02574-07>
- [21] Gierczyński, R., Kałużewski, S., Rakin, A., Jagielski, M., Zasada, A., Jakubczak, A., Borkowska-Opacka, B. and Rastawicki, W. (2004) Intriguing Diversity of *Bacillus anthracis* in Eastern Poland—The Molecular Echoes of the Past Outbreaks. *FEMS Microbiology Letters*, **239**, 235-240.
- [22] Sue, D., Marston, C.K., Hoffmaster, A.R. and Wilkins, P.P. (2007) Genetic Diversity in a *Bacillus anthracis* Historical Collection (1954 to 1988). *Journal of Clinical Microbiology*, **45**, 1777-1782.
- [23] Derzelle, S. and Thierry, S. (2013) Genetic Diversity of *Bacillus anthracis* in Europe: Genotyping Methods in Forensic and Epidemiologic Investigations. *Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science*, **11**, S166-S176.
- [24] Girault, G., Blouin, Y., Vergnaud, G. and Derzelle, S. (2014) High-Throughput Sequencing of *Bacillus anthracis* in France: Investigating Genome Diversity and Population Structure Using Whole-Genome SNP Discovery. *BMC Genomics*, **15**, 288.
- [25] NCBI. <http://www.ncbi.nlm.nih.gov/>
- [26] Erali, M., Voelkerding, K.V. and Wittwer, C.T. (2008) High Resolution Melting Applications for Clinical Laboratory Medicine. *Experimental and Molecular Pathology*, **85**, 50-58.
- [27] Hurtle, W., Bode, E., Kulesh, D.A., Kaplan, R.S., Garrison, J., Bridge, D., House, M., Frye, M.S., Loveless, B. and Nordwood, D. (2004) Detection of the *Bacillus anthracis* *gyrA* Gene by Using a Minor Groove Binder Probe. *Journal of Clinical Microbiology*, **42**, 179-185.
- [28] Easterday, W.R., Van Ert, M.N., Zanecki, S. and Keim, P. (2005) Specific Detection of *Bacillus anthracis* Using a TaqMan Mismatch Amplification Mutation Assay. *Biotechniques*, **38**, 731-735.
- [29] Birdsell, D.N., Pearson, T., Price, E.P., Hornstra, H.M., Nera, R.D., Stone, N., Gruendike, J., Kaufman, E.L., Pettus, A.H., Hurbon, A.N., *et al.* (2012) Melt Analysis of Mismatch Amplification Mutation Assays (Melt-MAMA): A Functional Study of a Cost-Effective SNP Genotyping Assay in Bacterial Models. *PLoS One*, **7**, Article ID: e32866.
- [30] Kenefic, L.J., Pearson, T., Okinaka, R.T., Schupp, J.M., Wagner, D.M., Hoffmaster, A.R., Trim, C.B., Chung, W.K., Beaudry, J.A., Jiang, L., *et al.* (2009) Pre-Columbian Origins for North American Anthrax. *PLoS One*, **4**, Article ID: e4813.
- [31] Simonson, T.S., Okinaka, R.T., Wang, B., Easterday, W.R., Huynh, L., U'Ren, J.M., Dukerich, M., Zanecki, S.R., Kenefic, L.J., Beaudry, J., *et al.* (2009) *Bacillus anthracis* in China and Its Relationship to Worldwide Lineages. *BMC Microbiology*, **9**, 71.
- [32] Okinaka, R.T., Henrie, M., Hill, K.K., Lowery, K.S., Van Ert, M., Pearson, T., Shupp, J., Kenefic, L., Beaudry, J., Hofstadler, S.A., *et al.* (2008) Single Nucleotide Polymorphism Typing of *Bacillus anthracis* from Sverdlovsk Tissue. *Emerging Infectious Diseases*, **14**, 653-656.

Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

Other selected journals from SCIRP are listed as below. Submit your manuscript to us via either submit@scirp.org or [Online Submission Portal](#).



2. Génomage de *B. anthracis*: utilisation des SNP via une approche par multiplexage

Article 7:

A multiplex bead-based suspension array assay for interrogation of phylogenetically informative single nucleotide polymorphisms for *Bacillus anthracis*

Publié le 16 octobre 2013 dans *Journal of Microbiology Methods*



A multiplex bead-based suspension array assay for interrogation of phylogenetically informative single nucleotide polymorphisms for *Bacillus anthracis*



Simon Thierry^a, Raditijo A. Hamidjaja^b, Guillaume Girault^a, Charlotta Löfström^c, Robin Ruuls^d, Derzelle Sylviane^{a,*}

^a University Paris-Est, ANSES, Animal Health Laboratory, Av du Général De Gaulle 23, Fr-94706 Maisons-Alfort, France

^b National Institute for Public Health and the Environment, Centre for Infectious Disease Control, Laboratory for Zoonoses and Environmental Microbiology, Antonie van Leeuwenhoeklaan 9, 3730 MA, Bilthoven, The Netherlands

^c National Food Institute, Technical University of Denmark, Mørkhøj Bygade 19, 2860 Søborg, Denmark

^d Central Veterinary Institute of Wageningen University and Research Centre, Edelhertweg 15, 8219 PH Lelystad, The Netherlands

ARTICLE INFO

Article history:

Received 5 August 2013

Received in revised form 3 October 2013

Accepted 6 October 2013

Available online 16 October 2013

Keywords:

Bacillus anthracis

Genotyping

SNP

MOL-PCR

Suspension microarray

Luminex

ABSTRACT

Single nucleotide polymorphisms (SNPs) are abundant in genomes of all species and represent informative DNA markers extensively used to analyze phylogenetic relationships between strains. Medium to high throughput, open methodologies able to test many SNPs in a minimum time are therefore in great need. By using the versatile Luminex® xTAG technology, we developed an efficient multiplexed SNP genotyping assay to score 13 phylogenetically informative SNPs within the genome of *Bacillus anthracis*. The Multiplex Oligonucleotide Ligation-PCR procedure (MOL-PCR) described by Deshpande et al., 2010 has been modified and adapted for simultaneous interrogation of 13 biallelic canonical SNPs in a 13-plex assay. Changes made to the originally published method include the design of allele-specific dual-priming-oligonucleotides (DPOs) as competing detection probes (MOLigo probes) and use of asymmetric PCR reaction for signal amplification and labeling of ligation products carrying SNP targets. These innovations significantly reduce cross-reactivity observed when initial MOLigo probes were used and enhance hybridization efficiency onto the microsphere array, respectively. When evaluated on 73 representative samples, the 13-plex assay yielded unambiguous SNP calls and lineage affiliation. Assay limit of detection was determined to be 2 ng of genomic DNA. The reproducibility, robustness and easy-of-use of the present method were validated by a small-scale proficiency testing performed between four European laboratories. While cost-effective compared to other singleplex methods, the present MOL-PCR method offers a high degree of flexibility and scalability. It can easily accommodate newly identified SNPs to increase resolving power to the canSNP typing of *B. anthracis*.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Bacillus anthracis, the causal agent of anthrax, is an evolutionarily young species that presents very low molecular diversity among strains (Keim et al., 2000, 2009; Pearson et al., 2004). Single nucleotide polymorphisms (SNPs) represent a major source of genetic variation in *B. anthracis*. Nucleotide substitutions, with mutation rates of approximately 10^{-10} changes per nucleotide per generation, are important biologically informative markers (Vogler et al., 2002). Due to their evolutionary stability, SNPs have been extensively used to determine phylogenetic groups in clonal pathogen species such as *B. anthracis* and elucidate relationships among worldwide

strains (Pearson et al., 2004; Kuroda et al., 2010). By querying a large number of them against collections of diverse strains, a set of 13 representative SNPs that define major clades within the *B. anthracis* species has been selected and used for assigning an isolate to one sub-lineage or sub-group (Pearson et al., 2004). These canonical SNPs (canSNPs) subdivided all of the *B. anthracis* isolates into three major lineages (A, B and C), with further subdivisions into 7 distinct sub-lineages (C.Br.A1055, B.Br.KrugerB, B.Br.CNEVA, A.Br.Ames, A.Br.Australia94, A.Br.Vollum, A.Br.Western North America) and 5 sub-groups (B.Br.001/002, A.Br.001/002, A.Br.003/004, A.Br.005/006, A.Br.008/009). An additional 14th canSNP that further divides the A.Br.008/009 group into two sub-groups (A.Br.008/011 and A.Br.011/009) was also recently reported (Marston et al., 2011). CanSNP analysis is now considered the reference method in *B. anthracis* genotyping. Current *B. anthracis* canSNP typing methods are all based on singleplex real-time technologies, i.e. Dual Probe TaqMan PCR (Van Ert et al., 2007), Mismatch Amplification Mutation

* Corresponding author at: Bacterial Zoonosis Unit, Animal Health Laboratory, ANSES, 23 Avenue du Général de Gaulle, 94706 Maisons Alfort Cedex, France. Tel.: +33 1 49 77 38 84.

E-mail address: sylviane.derzelle@anses.fr (D. Sylviane).

Assays (MAMA) (Birdsell et al., 2012) and High Resolution Melting (HRM) (Derzelle et al., 2011).

SNPs are the most common genetic variation found in genomes of all species. It is therefore not surprising that the development of technologies for SNP-based genotyping has been the subject of intense activity (Kwok, 2001; Shi, 2001; Kim and Misra, 2007; Ding and Jin, 2009; Martino et al., 2010). Several platforms and methods now exist (Kwok, 2001; Kim and Misra, 2007), including ultra-high throughput array-based genotyping technologies, such as those offered by Illumina (Gunderson et al., 2006) and Affymetrix (Kennedy et al., 2003), and many other strategies (Germer and Higuchi, 1999; Livak, 1999; Ahmadian et al., 2000; Kutayvin et al., 2000; Mishima et al., 2005; Tobler et al., 2005; Bruse et al., 2008; Edwards et al., 2009). These methods vary largely in their throughput, cost, technical difficulty or laboriousness, subjectivity in allele interpretation and flexibility. But only few of them are at the same time flexible, rapid (<1 day), cost-effective, and capable of detecting multiplexed signals simultaneously with medium to high throughput (Dunbar, 2006; Bruse et al., 2008; Price et al., 2010). The Luminex bead-based technology is one of these platforms with broad applications on many assay formats, including nucleic acid, receptor–ligand and immune-assays. This suspension array format implies the use of microsphere sets coupled to probes that recognize and bind the target DNAs. Once bound, the target DNA molecules are fluorescently tagged with Streptavidin–R-phycoerythrin, and the beads are individually analyzed by flow cytometry on the Luminex® platform. Each microsphere set is characterized by a distinct spectral address given by the combination of red and infrared fluorophores within the spheres. A red laser recognizes the microsphere set and a green laser provides a quantitative readout of the bound target (Dunbar et al., 2003). Application of Luminex® xTAG technology for multiple SNP genotyping offers several advantages. Based on a suspension of magnetic beads conjugated with probes, the Luminex xTAG microarray format exhibits rapid hybridization kinetics, flexibility in assay design and is cost-effective (Dunbar, 2006). It can also be compiled as desired by adding or replacing beads and probes without having to reformat and print new arrays (Janse et al., 2012). This system has been increasingly used for the design of several diagnostic assays based on various approaches, namely, Direct Hybridization, Allele-Specific Primer Extension, Single-Base-Extension and Oligonucleotide Ligation (Lee et al., 2004; Dunbar, 2006; Ducey et al., 2007; Ward et al., 2008; Stucki et al., 2012).

In 2010, a smart and more flexible assay method than any other approaches adapted to the Luminex® platform has been introduced by Deshpande et al. for successful pathogen detection and SNP-typing of *B. anthracis*, *Yersinia pestis*, and *Francisella tularensis* (Deshpande et al., 2010). Conceptually related to Multiplex Ligation-dependent Probe Amplification (MLPA) technique (Schouten et al., 2002), this method, called Multiplex Oligonucleotide Ligation-PCR (MOL-PCR), enables direct detection of multiple nucleic acid signatures in a single tube reaction (Deshpande et al., 2010). In MOL-PCR, detection probes consist of modular components that enable target detection, probe amplification, and subsequent capture onto microsphere arrays. MOL-PCR uses allele-specific ligation for allele discrimination, singleplex PCR for signal amplification and hybridization to fluorescent microspheres (beads) for signal detection on a flow cytometer. The ability to discriminate base pair mismatches flanking the ligation site is conferred by the ligase used in the reaction. An enhanced MOL-PCR procedure that makes the method easier to perform than similar published methods to carry out SNP multilocus genotyping has recently been described (Stucki et al., 2012).

In the present work, we adapted and modified the MOL-PCR method for the simultaneous typing of a set of 13 *B. anthracis* lineage-specific canSNPs previously identified. The developed multiplex assay was primarily built on the MOL-PCR concept as described by Stucki et al. (2012) but with a key modification for SNP discrimination and oligonucleotide design improvement. A dual priming oligonucleotide

(DPO) system (Chun et al., 2007) was coupled to the MOL-PCR method to increase assay specificity and allow multiplexing reactions to be more easily designed and produced. The setup and validation of a 13-plex SNP-typing assay for the identification of the main phylogenetic lineages of *B. anthracis* are presented.

2. Material and methods

2.1. Bacterial strains

Three different strain panels were used in this study. A training panel of 5 strains was first chosen to setup single and multiplex SNP-typing assays based on MOL-PCR concept. Strains AFSSA#31, AFSSA#99, AFSSA#08-27, CIP 66.17 and IEMVT 89-1620 (Derzelle et al., 2011) are affiliated to 5 distinct phylogenetic sub-lineages of *B. anthracis*: B.Br.CNEVA, A.Br.011/009, A.Br.001/002, A.Br.Vollum and A.Br.005/006, respectively.

A test panel was selected for validation of the 13-plex assay under blind conditions. This panel contained 60 isolates representing the 10 canSNP branches found in Eurasia (B.Br.CNEVA, B.Br.001/002, A.Br.Ames, A.Br.Aust94, A.Br.Vollum, A.Br.001/002, A.Br.003/004, A.Br.005/006, A.Br.008/011 and A.Br.011/009), including the 5 strains of the first panel.

A ring-trial test panel consisted of 13 strains representing 8 canSNP branches (B.Br.CNEVA, B.Br.001/002, A.Br.Aust94, A.Br.Vollum, A.Br.001/002, A.Br.003/004, A.Br.008/011 and A.Br.011/009) was also used. All these strains were previously characterized by PCR–HRM (Derzelle et al., 2011) in our laboratory and were available as purified DNA.

2.2. MOL-PCR probe design

For each individual SNP assay, the developed MOL-PCR procedure required a pair of competing allele-specific probes (MOLigoP1) and one common probe (MOLigoP2) that anneal adjacent to each other on specific target DNA. When hybridized to the specific target DNA, ligated MOLigo probes are amplified using a pair of universal PCR primers, labeled with biotin and incubated with Streptavidin–Phycoerythrin (SA–PE), followed by detection and quantification of amplicons on color coded Luminex microspheres carrying anti-tag sequences.

Allele-specific MOLigoP1 probes are modular 68–80 nucleotide-long, single-stranded DNA oligonucleotides that contain three functional components that allow (i) detection of target sequences, (ii) universal amplification of successfully ligated probes, and (iii) capture of amplified products onto a microsphere array. Each pair contains a 5'-end universal PCR reverse primer sequence common for all the different MOLigoP1 for amplification, an internal 24-bp xTAG sequence unique for each MOLigoP1 probe for capture to a reverse complement anti-tag conjugated xTAG microsphere, and a 3'-end sequence complementary to the specific target DNA including the allele-specific nucleotide (SNP) at its 3' end. To improve allele-specificity, the 3' specific target sequence was separated into two distinct priming regions by a polydeoxyinosine (Poly(I)) linker, resulting in a 5'-segment 11–23 nucleotides in length and a 3'-segment 8-nucleotides in length. This unequal distribution of nucleotides leads to different annealing preferences for each segment, according to the DPO principle (Chun et al., 2007). The longer 5' segment initiates stable priming. The shorter 3' segment determines target-specific extension and SNP discrimination (see upper part of Fig. 1). The common MOLigoP2 probe sequence consists of two sequences: the 5'-end reverse complement of a universal-forward primer and a 3'-end locus-specific portion complementary to the specific target sequence located just after the SNP. The 5'-end of each MOLigoP2 probe is phosphorylated to enable covalent linkage to MOLigoP1 in the presence of DNA ligase.

MOLigo probes for each biallelic SNP marker were primarily designed using MOLigo Designer online software (<http://moligodesigner.lanl.gov>) using default parameters (Song et al., 2010). Universal primers utilized in downstream PCR amplification and labeling reaction were previously

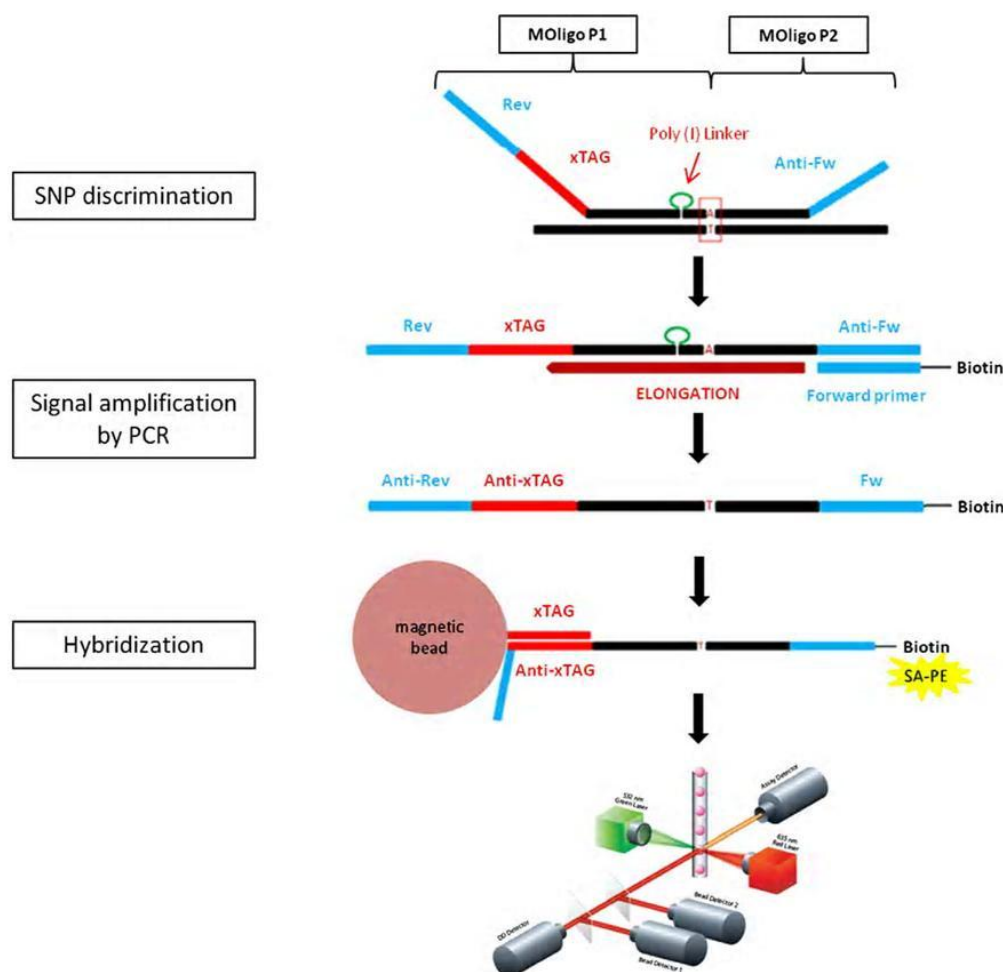


Fig. 1. MOL-PCR general scheme. As shown, MOLigo probes contain functional components that allow detection of target sequences (black sequences), universal amplification of successfully ligated probes (Rev and Anti-fw sequences in blue) and capture of amplified products onto a specific microsphere (xTAG sequence in red). MOL-PCR assay consists of three main steps. (i) SNP discrimination by annealing of one of two competitive allele-specific MOLigoP1 adjacent to the common MOLigoP2 on the target SNP locus, and ligation of both probes. (ii) Amplification of the ligated oligonucleotides by PCR using a pair of universal primers including a biotinylated forward primer. (iii) Hybridization of the generated labeled amplicons on color-coded magnetic beads specific to each SNP allele based on the xTAG technology and readout of beads carrying Streptavidin-Phycoerythrin reporter with a flow cytometric device on the Luminex® platform. Three MOLigo probes and two beads are used for interrogation of one SNP. For the 13-plex assay, a total of 39 MOLigo probes and 26 beads are used in the same reaction.

described (Song et al., 2010). All MOLigo probe sets were designed so that the locus-specific sequence of both the common and allele-specific probes has an optimal melting temperature ranging from 55 °C to 60 °C. A Poly(I) linker composed of 5 deoxyinosines was next inserted into each MOLigoP1 sequence 8 bp upstream of the 3' terminal SNP nucleotide. Possible cross-interaction between MOLigo probes was checked by the MOLigo Designer software and cross-reactivity monitored in each experiment using a no-template PCR control. The final set of MOLigo nucleotides and both universal primers (Table 1) were ordered from MWG Operon. The universal-forward primer was triply biotinylated by the manufacturer.

2.3. MOL-PCR array description

A modified MOL-PCR procedure was carried out as described by Stucki et al. (2012). First, the tag sequences complementary to anti-tag sequences on microspheres were included on the MOLigoP1, i.e. preceding the allele-discriminating sequence. With this modification in oligonucleotide design, two competing probes were used, differing

only by the 3'-end (the SNP allele) and the tag sequence. This change allowed the calculation of an allelic ratio rather than a signal-to-noise ratio, and improved the sensitivity by discriminating background signal from uncalled allele signal. Second, the hybridization/ligation step was separated from the PCR step (Deshpande et al., 2010) to avoid high background signal levels when combining these two steps.

The assay therefore consisted of three main steps (Fig. 1):

1. Annealing of the MOLigo pairs P1 and P2 adjacent to each other on the target DNA sequence (SNP) and ligation of the MOLigo by a thermostable DNA ligase that is functional at the annealing temperature. SNP discrimination is achieved by competitive hybridization of both allele-specific MOLigoP1 and ligation to the common, 3'-adjacent MOLigoP2 if MOLigo P1 3' terminal base is complementary to the target SNP. This step creates multiple single-stranded DNA molecules of approximately 110–130 nucleotides long that will serve as PCR templates for signal amplification.
2. Amplification of the ligation products by single-plex PCR using a universal primer pair for all SNPs. The high level of multiplexing is

Table 1
Oligonucleotide sequences of all probes and primers used.

xTAG	SNP	Allele	MOLigo	Sequence (5'–3')
B007	1	C	P1	ACTCGTAGGGAATAAACCGTAAATTTGTAAGAGATTGTTCTGTACAAATTAATCTTTAAAGGAAIIIIACCGAAAC
A014		T	P1	ACTCGTAGGGAATAAACCGTATTGTGAAGAGAGAGAAATTCAAATTAATCTTTAAAGGAAIIIIACCGAAAT
–		–	P2	P-TTGAAGTCGATGATAAAGGGAACCGTATTATATCTCACTTCTTACTACCGCG
A021	2	C	P1	ACTCGTAGGGAATAAACCGTATTAGTAAGAATTGAGAGTTTGAAGAGGAGCAAGTAATGTTATAGIIIIITTTAGGC
A022		T	P1	ACTCGTAGGGAATAAACCGTGATTGATATTGAATTTGTTTGAGAAGGAGCAAGTAATGTTATAGIIIIITTTAGGT
–		–	P2	P-TGGCGGGCAGTCGCTTATCTCTCACTTCTTACTACCGCG
A025	3	G	P1	ACTCGTAGGGAATAAACCGTGATGTGTAATGTATTAAGAAAGTGATAAAACCTCTTTTCTIIIIACCTCAAG
A026		A	P1	ACTCGTAGGGAATAAACCGTTTGATTAGAGTCTTGAATGTATATAAAACCTCTTTTCTIIIIACCTCAAA
–		–	P2	P-TTGAAGTCGATAAAGGAGGTTTATACAATGACATCTCACTTCTTACTACCGCG
A027	4	T	P1	ACTCGTAGGGAATAAACCGTAAGATGATAGTTAAGTGAAGTTAGACATCGCGCTCATACCTIIIIITGGAATGT
A028		C	P1	ACTCGTAGGGAATAAACCGTGATAGATTAGAATGAATTAAGTGACATCGCGCTCATACCTIIIIITGGAATGC
–		–	P2	P-CCCTAATCTTCCATAGCTCCACCATCTCACTTCTTACTACCGCG
B008	5	T	P1	ACTCGTAGGGAATAAACCGTTGTAAGTGAATAGTGAGTTATTTCTGTTTAAAGTTCATCATACCTIIIIATGCACCT
A015		G	P1	ACTCGTAGGGAATAAACCGTTGTGTAATGTAGTAAGAAAGTACGTTTAAAGTTCATCATACCTIIIIATGCACCG
–		–	P2	P-AGGCGATGGAATGATCAACAACATATTGATCTCACTTCTTACTACCGCG
A029	6	T	P1	ACTCGTAGGGAATAAACCGTTTAAAGTGAAGTTAGAGTAGTAAGACGATAAAGTAATAATACCTIIIIATCCTTAT
A030		C	P1	ACTCGTAGGGAATAAACCGTGTTGATAGAGTTAAATGTTAAAGAGACGATAAAGTAATAATACCTIIIIATCCTTAC
–		–	P2	P-ATTCAGCTCGAATACCTACCTTGTAAATCTCTCACTTCTTACTACCGCG
B009	7	T	P1	ACTCGTAGGGAATAAACCGTGAATTTGATAAAGTATTAGATGTGTATACGTTTATAGTGGAGATAIIIIITCTTCT
A018		G	P1	ACTCGTAGGGAATAAACCGTGAATTTGATAAAGTATTAGATGTGTATACGTTTATAGTGGAGATAIIIIITCTTCTG
–		–	P2	P-CCGCTTGTAAACGTATATTGTAACTTTTCACTCTCACTTCTTACTACCGCG
A033	8	C	P1	ACTCGTAGGGAATAAACCGTTATTAGAGTTTGAGAAATAAGTAGTAGGTATATTAACTGCGGATGIIIIATGCAAGC
A034		T	P1	ACTCGTAGGGAATAAACCGTTGATATAGTAGTGAAGAAATAAGTAGGTATATTAACTGCGGATGIIIIATGCAAGT
–		–	P2	P-AAAGCGTTCAAAACAGTGGCCTCTCACTTCTTACTACCGCG
A035	9	G	P1	ACTCGTAGGGAATAAACCGTAATAAGAGAAATGATATGAAGATGCGGATATGATACCGATACCTIIIIITCTTATCC
A036		T	P1	ACTCGTAGGGAATAAACCGTTGTGATTAAGAGTTGTTTAAATGCGGATATGATACCGATACCTIIIIITCTTATCT
–		–	P2	P-TCTTCTATTGTACCGATTCTTTTATGACCGTCTCACTTCTTACTACCGCG
A037	10	T	P1	ACTCGTAGGGAATAAACCGTTGATATGTAATGAGATGTTGTAACCTTCTGTGTCGTGTTAAIIIIICGTACTT
A038		G	P1	ACTCGTAGGGAATAAACCGTAGTAAGTGTAGATAGTATGAATACCTTCTGTGTCGTGTTAAIIIIICGTACTG
–		–	P2	P-CTGTCTTTTGTCACTTCTCTCCATCTCACTTCTTACTACCGCG
A012	11	G	P1	ACTCGTAGGGAATAAACCGTAGTAGAAAGTTGAAATGATTATGGCATAGAGCAGATGAGCTTAAIIIIICATATCCG
A019		A	P1	ACTCGTAGGGAATAAACCGTGCTGTTATTGTTGTAAGATATGCATAGAGCAGATGAGCTTAAIIIIICATATCCA
–		–	P2	P-CTTCAGCTTATGGTTCGTTGAAGTCTGAGTCTCACTTCTTACTACCGCG
A013	12	A	P1	ACTCGTAGGGAATAAACCGTAGTGAATGTAAGATTATGATTGTAATGAAGATATGACAAIIIIICGGGATGA
A020		G	P1	ACTCGTAGGGAATAAACCGTAATTAAGTGAAGATGAGAAAGTAAATGAAGATATGACAAIIIIICGGGATGA
–		–	P2	P-TAGAAGTAAAGAGGTTACCAAGCACTTGTCTCACTTCTTACTACCGCG
A039	14	T	P1	ACTCGTAGGGAATAAACCGTTGTGATAGTAGTATGATTTTGAAGCAGGAIIIIIICGCCCTT
A042		C	P1	ACTCGTAGGGAATAAACCGTATTGTTATGATAAATGTAGTGTGAAGCAGGAIIIIIICGCCCTT
–		–	P2	P-ATTATTTTACGGGGAATTCGTTCTTTTATGTCCTCACTTCTTACTACCGCG
–		Forward		BIOT-CGCGGTAGTAAGAAGTGAGA
–		Reverse		ACTCGTAGGGAATAAACCGT

Primer, xTAG and DNA target sequences are indicated, respectively, by underlined, italic and bold, or light grey sequences. The forward primer is biotinylated at its 5' end and at two internal T nucleotides (positions are underlined and indicated in bold).

made possible by PCR-amplification of all ligated oligonucleotides rather than amplification of template DNA. The universal forward primer is tagged with three biotin moieties which guarantee the sensitivity at the readout step.

- Hybridization of the amplicons to microspheres using xTAG sequences complementary to anti-tag probes and signal detection on a flow cytometer. Bead-(PCR product) complexes are incubated with Streptavidin-R-phycoerythrin (SA-PE) which associated to Biotin will produce fluorescence when analyzed by the Luminex system. The BioPlex 200 (Bio-Rad, Hercules, USA) identifies each bead type and measures fluorescence intensity emitted by all bead-(PCR product) complexes. A universal set of 100 capture probes (anti-tag) covalently linked to magnetic beads allows a wide variety of assays to utilize the same array elements. Each SNP allele is attributed to one specific xTAG sequence and unique bead color. MicroPlex® microspheres were obtained from Luminex Corporation (Austin, USA).

2.4. Multiplex assay protocol

Each experiment included two controls for the calculation of signal-to-noise ratio: a bead-only control that reports background fluorescence obtained from the microspheres alone and a no template PCR control that reports cross reactivity between MOLigo pairs in the absence of any DNA template (H₂O control).

For the 13-plex assay, a total of 39 MOLigo probes and 26 beads are used in the same reaction. Annealing and ligation reactions were performed in a final volume of 10 µl containing 5 units of Ampligase® DNA ligase (Epicentre, Madison, USA), 0.1 nmol of each MOLigo probes (MWG Operon), 1 × Taq DNA Ligase Buffer (Epicentre, Madison, USA) and 2 µl of template DNA (approximately 5 ng/µl). The thermal cycling profile included DNA denaturation at 95 °C for 5 min, followed by 10 ligation cycles of 60 °C for 5 min and 95 °C for 2 min, and a final ligation-denaturation step at 98 °C for 5 min.

The PCR reaction was performed in a final volume of 20 µl. Eight microliters of ligation products was mixed with 2.5 units of Hot Start Taq polymerase (Qiagen, Courtaboeuf, France), 2.5 µM of each dNTPs, 1 × Taq Buffer (Qiagen, Courtaboeuf, France), 100 pmol of biotin-labeled universal-forward primer and 2 pmol of universal-reverse primer (MWG Operon, Ebersberg, Germany). The following thermal cycling profile was run: 95 °C for 15 min, 45 cycles of 95 °C for 30 s, 55 °C for 30 s and 72 °C for 30 s. Reactions were then cooled to 4 °C, and either used immediately in the bead-hybridization step or stored at –20 °C before proceeding with the hybridization step.

A microsphere mix consisting of anti-tag-coated Luminex xTAG beads of each set specific to each MOLigo pair used in the assay was prepared. Hybridization and measurements were carried out guided by the manufacturer's recommendations. Briefly, the bead-mix contained 26 MicroPlex® microspheres were diluted in Tm Hybridization Buffer (Luminex Corporation, 2010) to a concentration of 1250 beads of each

2.5. SNP calling

$$AR = \left(MFI_{\text{allele1}} / MFI_{H_2O_{\text{allele1}}} \right) / \left(\left(MFI_{\text{allele1}} / MFI_{H_2O_{\text{allele1}}} \right) + \left(MFI_{\text{allele2}} / MFI_{H_2O_{\text{allele2}}} \right) \right). \quad (1)$$

2.6. Ring-trial

3. Result

In the original MOL-PCR procedure (Deshpande et al., 2010), ligation and PCR were conducted in a single reaction using an Amplitaq Gold DNA polymerase activated through a slow release mechanism that perform the amplification of ligated MOLigo pairs following the ligation step (Deshpande et al., 2010). Although we had tested several enzymes (DNA ligases and polymerases), thermocycling profiles, primer concentrations, and MOLigo pairs (designed using the online software described by the same authors (Song et al., 2010), we were unsuccessful in our attempts to obtain any signal, measured as MFI, using the one step (ligation/PCR) procedure. Therefore, we separated the ligation step and were able to get weak signals compared to the background noise. The optimal concentration of MOLigo probes was established at a very low level (0.1 nmol of each probe) as higher concentration

3.2. MOLigo probe and array design

Despite the enhanced allele specificity endowed by using two alternate single assays for each SNP, most of assays displayed cross-reactivity and resulted in poor to inaccurate allele discrimination. In addition, designing MOLigoP1 binding at the 5' end of the SNP position was sometimes problematic since the probe binding site could be of poor quality. A novel design strategy than simplified design from the opposite strand was required to ensure true allele-specific ligation. The dual priming oligonucleotide (DPO) principle (Chun et al., 2007) was therefore used to design a full set of new competing MOLigoP1 pairs and rescue the poor performing MOL-PCR assays. Cross-reactivity which previously made it difficult to discriminate SNP in some duplex assays was almost eliminated using the new DPO-MOLigoP1 pairs. A signal-to-noise ratio sufficient for accurate genotype calling was obtained for each duplex assays.

When combined together to implement the 13-plex panel, a 15 to 25% decrease in individual duplex assays signals was observed according to the SNP locus. A few adjustments to the developed protocol were necessary to optimize the multiplex assay. The universal forward primer was labeled with three biotin moieties for post-PCR capture of Streptavidin-Phycoerythrin (SA-PE) fluorophores. SA-PE concentration was optimized to provide additional fluorescence (signal increase by a factor 2 to 3) with only a slight increase of the background noise when analyzed by the Luminex system. SA-PE conjugates from two different manufacturers (InterChim and Bio-Rad) were also evaluated. The highest signal-to-noise ratio was obtained using SA-PE from Bio-Rad (data not shown). Appropriate conditions included a higher amount of SA-PE (3-fold) and a lower concentration of beads than recommended by the manufacturer (2-fold) (Luminex Corporation, 2010).

3.4. 13-Plex assay validation

The 13-plex assay was applied to a panel of 60 samples (including the five isolates used to develop the assays) representative of ten

different canSNP lineages. All samples and controls were analyzed in duplicates. MOL-PCR generated complete allelic information for all 13 SNPs of 60 samples. Allelic ratios (AR) calculated for each canSNP lineage are indicated in Table 2. Molecular typing results obtained with the MOL-PCR array were fully consistent with previous genotyping data determined by HRM (Derzelle et al., 2011). Both methods agreed 100% in SNP calling and lineage assignment. It has to be mentioned, however, that as expected, three A.Br.005/006 samples displayed an AR value of 0.5 for both assays targeting the B.Br.004 canSNP, scoring as negative for both alleles. These strains isolated in Africa carry a second mutation contiguous to the B.Br.004 canSNP position (Derzelle et al., 2011; Pilo et al., 2011). This additional base difference just upstream of the target canSNP prevents specific hybridization and ligation of both MOLigoP1 probes at this locus, resulting in a third distinct SNP calling that was correctly interpreted as the T allele.

A specificity control was also performed by genotyping *Bacillus cereus* (strain ATCC-14579). All experiments using this template resulted in a negative result (data not shown). No typing result higher than the signal level of H₂O control reactions was observed using this sample for template for any SNP.

3.5. Typing sensitivity

We investigated the accuracy and sensitivity of the 13-plex MOL-PCR array against ten-fold serial dilutions of DNA template as input material, using two different strains (*B. anthracis* lineages A.Br.005/006 and A.Br.Vollum). As illustrated in Fig. 2, decreasing amount of template DNA resulted in decreasing signal intensity. A minimum quantity of 2 ng of genomic DNA (approximately 10⁵ genome copies per reaction) was necessary in the ligation reactions to obtain reliable results and calculate confident allelic ratios. A direct correlation between input quantity and Luminex MFI values was observed in a relatively small dynamic range.

3.6. Reproducibility

To further challenge the assay, 13 additional DNA samples affiliated to eight sub-lineages of *B. anthracis* found in Europe were blind-tested in duplicates in a small-scale European interlaboratory trial. Four Luminex-inexperienced users were asked to run the 13-plex assay in their respective laboratories (ANSES, RIVM, CVI and DTU) to assess analytical performances such as robustness and ease of use of the MOL-PCR typing assay.

Inter-operator and inter-machine reproducibility was found to be very poor. Absolute MFI data showed considerable variations, with signal levels for control reactions relatively high for two laboratories (data not shown). But, the use of MFI ratios instead of net MFI values for SNP calling was shown to minimize the variation among different tests and yielded consistent SNP calling results, even if manual inspection of raw data was necessary to determine a few alleles. The rate of automatic SNP calling ranged from 78.7% to 99.4%, according to the laboratories (Table 3). It increased to more than 94.1% after manual inspection of the results. In contrast, the successful assignment of a sample to a lineage was much more dependent of manual inspection, with results ranging from 0% to 100% using automatic assignment to 92.5% to 100% after manual checking (Table 3). While one of the 13 SNPs failed the automatic lineage assignment, complete allelic information for all SNPs is not always necessary for lineage calling. Redundancy exists between the canSNPs set defining different sub-lineages and they were taken into account in the manual assignment. This one-shot trial evaluation demonstrated that, by making some small parameter adjustments to optimize Luminex-based readout on each individual flow cytometer, the present MOL-PCR Luminex array can be easily implemented in any laboratory equipped with the appropriate device.

4. Discussion

As the cost and time required for whole-genome sequencing and bioinformatic analyses are continuously being reduced, thousands of SNPs retrieved from compiled whole genome sequences become available for *B. anthracis*, allowing the identification of additional SNP markers for genotyping. There is an urgent need for medium to high throughput multiplexed methods to test many targets in a minimum time.

MOL-PCR coupled to Luminex xTAG technology-based detection provides an open and attractive approach for multiplexed SNP locus analyses. Composed of a set of universal anti-tag conjugated microspheres, this suspension array format offers several technical advantages in terms of assay development such as speed, flexibility, moderate to high multiplexing capability and ease-of-use. The color-encoded magnetic beads exhibit rapid hybridization kinetics and are readily manipulated. The microarray format is amendable to hundreds of SNPs and a variable number of strains. It is also able to accommodate new markers while preserving all other system parameters. Indeed, addition of new MOLigo pairs to existing assays does not require redesign of an entire array. All it takes is to incorporate two beads,

Table 2
SNP allelic ratio for the 60 strains tested.

	A.Br.001	A.Br.002	A.Br.003	A.Br.004	A.Br.006	A.Br.007	A.Br.008	A.Br.009	B.Br.001	B.Br.002	B.Br.003	B.Br.004	A.Br.011
A.Br.005/006 (n = 5)	0.14–0.15	0.14–0.17	0.61–0.63	0.88–0.92	0.96–0.98	0.04–0.07	0.90–0.97	0.60–0.63	0.09–0.17	0.10–0.13	0.61–0.68	0.50 ^a –0.71	0.06–0.10
A.Br.001/002 (n = 11)	0.08–0.14	0.86–0.97	0.08–0.16	0.088–0.26	0.96–1.02	0.01–0.07	0.92–1.23	0.61–0.76	0.02–0.11	0.02–0.11	0.65–0.82	0.71–0.77	0.01–0.10
A.Br.011/009 (n = 8)	0.10–0.17	0.12–0.17	0.59–0.64	0.89–0.93	0.96–0.98	0.04–0.07	0.07–0.14	0.61–0.65	0.07–0.15	0.07–0.14	0.53–0.73	0.68–0.76	0.76–0.87
B.Br.CNEVA (n = 9)	0.08–0.18	0.12–0.19	0.57–0.63	0.87–0.93	0.07–0.12	0.04–0.09	0.85–0.98	0.62–0.65	0.07–0.22	0.08–0.14	0.10–0.18	0.03–0.10	0.06–0.13
A.Br.008/011 (n = 6)	0.10–0.13	0.09–0.17	0.57–0.66	0.89–0.95	0.96–0.99	0.03–0.08	0.07–0.12	0.58–0.73	0.06–0.13	0.07–0.13	0.66–0.76	0.69–0.76	0.02–0.09
A.Br.Ames (n = 2)	0.36–0.38	0.93–0.92	0.91–0.93	0.13–0.14	0.99–1.00	0.01–0.02	1.04–1.05	0.61–0.63	0.06–0.07	0.05	0.75–0.77	0.68–0.74	0.01–0.03
A.Br.Aust94 (n = 7)	0.04–0.11	0.01–0.06	0.02–0.44	0.06–0.39	1.01–1.07	0.01–0.02	1.12–1.39	0.64–0.88	0.01–0.40	0.03–0.34	0.62–0.93	0.61–0.92	0.03–0.25
B.Br.001/002 (n = 3)	0.03–0.34	0.02–0.19	0.61–0.78	0.86–1.02	0.01–0.16	0.01–0.18	0.65–1.67	0.62–0.87	0.02–0.11	0.82–1.07	0.01–0.29	0.63–0.96	0.01–0.26
A.Br.003/004 (n = 2)	0.06–0.25	0.03–0.16	0.64–0.70	0.09–0.24	0.95–1.02	0.01–0.10	0.71–1.24	0.68–0.79	0.02–0.06	0.02–0.11	0.74–0.85	0.64–0.79	0.01–0.08
A.Br.Vollum (n = 7)	0.04–0.08	0.02–0.09	0.62–0.73	0.95–1.02	1.00–1.05	0.65–0.76	1.01–1.48	0.67–0.93	0.01–0.04	0.01–0.04	0.76–0.95	0.67–0.87	0.01–0.26

^a An allelic ratio of 0.5 (corresponding to negative signal for both alleles) was observed for 3 strains.

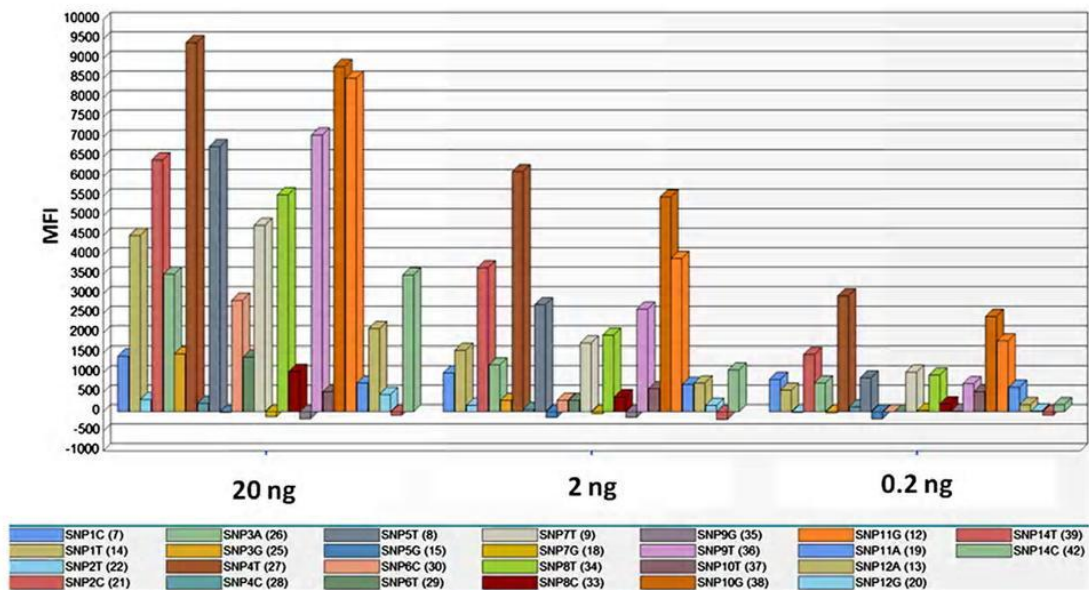


Fig. 2. MOL-PCR analysis of an A.Br.Vollum strain (CIP 66.17) using the 13-plex assays and ten-fold series dilution of genomic DNA. Read-out plots showing the median fluorescence intensities (MFI) measured for the 26 beads (color-coded according to the biallelic SNP marker tested) and 3 DNA concentrations. Sensitivity testing data indicated that more than 2 ng of input DNA are required for discrimination of all 13 canSNPs and confident allelic ratio calculation. Data and plots were produced by the BioPlex 200 System using the BioPlex Manager software.

each one coupled to unique anti-tag sequences, making inclusion of new targets straight-forward, without affecting the performance of any other pairs within the assay.

We report several changes to the original MOL-PCR method (Deshpande et al., 2010) to improve efficient SNP discrimination and multiplexing. First, two MOLigo oligonucleotides competing for annealing to the 5'-end of the template DNA were used per SNP, allowing an allelic call based on allelic ratio rather than signal-to-noise ratio. Second, the DPO design of MOLigo primers, in addition to reducing cross-reactivity, effectively increased the assay design success rate. Standard MOLigo probes have been reported to show unfavorable allelic or signal-to-noise ratios occasionally. MOLigo probes have to anneal directly adjacent to the SNP of interest, requiring the need for a suitable sequence up- and downstream of the SNP. Moreover, the use of long oligonucleotides makes hairpin and homodimer formations likely, and multiplexing probes add a level of complexity due to a high combinatorial number of possible heterodimer formations (Song et al., 2010). DPOs that contain two separate priming regions joined by a polydeoxyinosine linker for stable annealing and target-specific extension are in many aspects easier to design than standard primers and more effective to prevent cross-reactivity. Third, altering the ratios of universal primer concentrations for signal amplification by asymmetric PCR significantly enhanced bead-hybridization efficiency between microsphere-bound anti-tags

and labeled-PCR product target strands. It also improved assay's performance and sensitivity.

To date, except the recent CUMA methodology (i.e. Capillary electrophoresis Universal tail Mismatch amplification mutation Assay) (Price et al., 2010), the available techniques to interrogate *B. anthracis*-specific canSNP loci (Dual Probe TaqMan technology, Melt-MAMA or HRM) are relatively difficult to multiplex. CUMA, that combines a primer extension method (i.e. MAMA) with a universal tail labeling system and amplified detection signal discrimination from multiplex reactions using fragment sizing on capillary electrophoresis, is capable of accommodating up to 40 assays (Price et al., 2010). Compared to CUMA, the present MOL-PCR procedure achieves greater flexibility and higher multiplexing capability. It allows for simultaneous interrogation of 50 or 250 biallelic SNPs (using the Luminex® xTAG or xMAP technology, respectively) and uses an open microarray format to accommodate the addition of new markers. While CUMA uses multiplex-PCR for amplification of template DNAs, multiplexing was made possible by segregating assays into mixes based on dye set and amplicon size, MOL-PCR performs detection in a pre-PCR ligation step, followed by single amplification and labeling of the multiple ligation products with a pair of universal primers. As MOL-PCR probes are designed to be similar in length, the assay is not susceptible to the amplification bias commonly seen in multiplex PCR (Vora et al., 2004).

Nevertheless, both approaches are somehow quite similar and display common caveats. (i) Both CUMA and MOL-PCR strategies use universal labeling system that reduces analysis costs by eliminating the dependency on specifically labeled allele-specific probes and primers. Five oligonucleotides per SNP are needed. In CUMA, depending on which of the two competitive UT-MAMA primer anneals to the DNA target, the corresponding UT sequence is integrated into the PCR products during amplification and acts as a hybridization site for the corresponding complementary fluorescently labeled UT primer. (ii) In each case, the method requires an expensive piece of equipment for signal readout (in the form of a Luminex or capillary electrophoresis instrument) that is, however, within the budget of many academic labs. Although both technologies can be used for a wide variety of applications in addition to SNP-typing, a significant limitation in the development of both assays

Table 3
SNP calling and lineage assignment ring trial results.

	Laboratory 1	Laboratory 2	Laboratory 3	Laboratory 4
SNP				
Automatic assignment	99.4%	89.3%	87.6%	74.5%
	(168/169)	(151/169)	(148/169)	(126/169)
Manual inspection	100%	100%	97%	94.1%
			(164/169)	(159/169)
Lineage				
Automatic assignment	100%	46.1%	15.4%	0%
		(6/13)	(2/13)	
Manual inspection	100%	100%	92.3%	92.3%
			(12/13)	(12/13)

is perhaps the relatively high initial setup costs due to the purchase of long oligonucleotides (and beads with the xTAG technology). In MOL-PCR, only one universal biotin-labeled primer is employed for signal quantification in all assays, thus saving the cost of synthesizing different fluorescently-labeled UT oligonucleotide assays. Since reagent and running costs per SNP drop with increasing number of multiplexed SNPs, the cost per assay should be significantly less than other similar assays designed to examine a single locus. (iii) While MAMA system relies on the inefficiency of Taq polymerase to extend primers containing nucleotide mismatches at their 3' end (Huang et al., 1992; Rhodes et al., 2001), MOL-PCR array is based on the inefficiency of DNA ligase to covalently link such primers. (iv) Original MOLigo or MAMA primers traditionally suffer from high rates of assay design failures and knowledge gaps on assay robustness and sensitivity (Birdsell et al., 2012). In a small number of MAMA assays, one allele can exhibit amplification for both UT-MAMA primers, resulting in a dual fluorescent signal. In some instances, primer dimers and nonspecific amplification can also result in spurious or erroneous electrophoretic peaks. To increase the allele-specific stringency in CUMA, it was necessary to alter the stoichiometry of each UT-MAMA primer. DPO design was used in MOL-PCR.

The described MOL-PCR assay was compared with real-time PCR–HRM which was used as standard SNP-typing method in our laboratory. We found an agreement of 100% in allele-calling and classifying strains into the main phylogenetic lineages, when typing a diverse panel of DNAs. Even though we could not always generate complete allelic information for all SNPs of the 13-plex assay, redundancy existing between the canSNP set, allows neglecting the lack of an allelic call for one SNP if the alleles at other loci are called successfully. With a detection limit of MOL-PCR as low as 2 ng of genomic DNA, the modified MOL-PCR array is an effective SNP genotyping tool that provides high confidence results under 6 h to complete. In conclusion, the present 13-plex MOL-PCR assay is accurate, robust, less expensive than previously published singleplex methods, and easy to perform.

Acknowledgments

The authors thank Joakim Ågren (SVA, Sweden) and Miriam Koene (CVI, Netherlands) for providing DNA samples, as well as Pia Engelsmann (DTU, Denmark) for excellent technical assistance.

This research was supported by/executed in the framework of the EU-project AniBioThreat (Grant Agreement: Home/2009/ISEC/AG/191) with the financial support from the Prevention of and Fight against Crime Programme of the European Union, European Commission – Directorate General Home Affairs. This publication reflects the views only of the author, and the European Commission cannot be held responsible for any use which may be made of the information contained therein.

References

Ahmadian, A., Gharizadeh, B., et al., 2000. Single-nucleotide polymorphism analysis by pyrosequencing. *Anal. Biochem.* 280 (1), 103–110. <http://dx.doi.org/10.1006/abio.2000.4493>.

Birdsell, D.N., Pearson, T., et al., 2012. Melt analysis of mismatch amplification mutation assays (Melt-MAMA): a functional study of a cost-effective SNP genotyping assay in bacterial models. *PLoS One* 7 (3), e32866. <http://dx.doi.org/10.1371/journal.pone.0032866>.

Bruse, S., Moreau, M., et al., 2008. Improvements to bead-based oligonucleotide ligation SNP genotyping assays. *Biotechniques* 45 (5), 559–571. <http://dx.doi.org/10.2144/000112960>.

Chun, J.Y., Kim, K.J., et al., 2007. Dual priming oligonucleotide system for the multiplex detection of respiratory viruses and SNP genotyping of CYP2C19 gene. *Nucleic Acids Res.* 35 (6), e40. <http://dx.doi.org/10.1093/nar/gkm051>.

Derzelle, S., Laroche, S., et al., 2011. Characterization of genetic diversity of *Bacillus anthracis* in France by using high-resolution melting assays and multilocus variable-number tandem-repeat analysis. *J. Clin. Microbiol.* 49 (12), 4286–4292. <http://dx.doi.org/10.1128/JCM.05439-11>.

Deshpande, A., Gans, J., et al., 2010. A rapid multiplex assay for nucleic acid-based diagnostics. *J. Microbiol. Methods* 80 (2), 155–163. <http://dx.doi.org/10.1016/j.mimet.2009.12.001>.

Ding, C., Jin, S., 2009. High-throughput methods for SNP genotyping. *Methods Mol. Biol.* 578, 245–254. http://dx.doi.org/10.1007/978-1-60327-411-1_16.

Ducey, T.F., Page, B., et al., 2007. A single-nucleotide-polymorphism-based multilocus genotyping assay for subtyping lineage I isolates of *Listeria monocytogenes*. *Appl. Environ. Microbiol.* 73 (1), 133–147. <http://dx.doi.org/10.1128/AEM.01453-06>.

Dunbar, S.A., 2006. Applications of Luminex xMAP technology for rapid, high-throughput multiplexed nucleic acid detection. *Clin. Chim. Acta* 363 (1–2), 71–82. <http://dx.doi.org/10.1016/j.cccn.2005.06.023>.

Dunbar, S.A., Vander Zee, C.A., et al., 2003. Quantitative, multiplexed detection of bacterial pathogens: DNA and protein applications of the Luminex LabMAP system. *J. Microbiol. Methods* 53 (2), 245–252 (doi: S0167701203000289).

Edwards, K.J., Reid, A.L., et al., 2009. Multiplex single nucleotide polymorphism (SNP)-based genotyping in allohexaploid wheat using padlock probes. *Plant Biotechnol. J.* 7 (4), 375–390. <http://dx.doi.org/10.1111/j.1467-7652.2009.00413.x>.

Gerner, S., Higuchi, R., 1999. Single-tube genotyping without oligonucleotide probes. *Genome Res.* 9 (1), 72–78. <http://dx.doi.org/10.1101/gr.9.1.72>.

Gunderson, K.L., Steemers, F.J., et al., 2006. Whole-genome genotyping. *Methods Enzymol.* 410, 359–376. [http://dx.doi.org/10.1016/S0076-6879\(06\)10017-8](http://dx.doi.org/10.1016/S0076-6879(06)10017-8).

Huang, M.M., Amheim, N., et al., 1992. Extension of base mispairs by Taq DNA polymerase: implications for single nucleotide discrimination in PCR. *Nucleic Acids Res.* 20 (17), 4567–4573. <http://dx.doi.org/10.1093/nar/20.17.4567>.

Janse, I., Bok, J.M., et al., 2012. Development and comparison of two assay formats for parallel detection of four biothreat pathogens by using suspension microarrays. *PLoS One* 7 (2), e31958. <http://dx.doi.org/10.1371/journal.pone.0031958>.

Keim, P., Price, L.B., et al., 2000. Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis*. *J. Bacteriol.* 182 (10), 2928–2936. <http://dx.doi.org/10.1128/JB.182.10.2928-2936.2000>.

Keim, P., Gruendike, J.M., et al., 2009. The genome and variation of *Bacillus anthracis*. *Mol. Aspects Med.* 30 (6), 397–405. <http://dx.doi.org/10.1016/j.mam.2009.08.005>.

Kennedy, G.C., Matsuzaki, H., et al., 2003. Large-scale genotyping of complex DNA. *Nat. Biotechnol.* 21 (10), 1233–1237. <http://dx.doi.org/10.1038/nbt869>.

Kim, S., Misra, A., 2007. SNP genotyping: technologies and biomedical applications. *Annu. Rev. Biomed. Eng.* 9, 289–320. <http://dx.doi.org/10.1146/annurev.bioeng.9.060906.152037>.

Kuroda, M., Serizawa, M., et al., 2010. Genome-wide single nucleotide polymorphism typing method for identification of *Bacillus anthracis* species and strains among *B. cereus* group species. *J. Clin. Microbiol.* 48 (8), 2821–2829. <http://dx.doi.org/10.1128/JCM.00137-10>.

Kutyavin, I.V., Afonina, I.A., et al., 2000. 3'-Minor groove binder-DNA probes increase sequence specificity at PCR extension temperatures. *Nucleic Acids Res.* 28 (2), 655–661. <http://dx.doi.org/10.1093/nar/28.2.655>.

Kwok, P.Y., 2001. Methods for genotyping single nucleotide polymorphisms. *Annu. Rev. Genomics Hum. Genet.* 2, 235–258. <http://dx.doi.org/10.1146/annurev.genom.2.1.235>.

Lee, S.H., Walker, D.R., et al., 2004. Comparison of four flow cytometric SNP detection assays and their use in plant improvement. *Theor. Appl. Genet.* 110 (1), 167–174. <http://dx.doi.org/10.1007/s00122-004-1827-1>.

Livak, K.J., 1999. Allelic discrimination using fluorogenic probes and the 5' nuclease assay. *Genet. Anal.* 14 (5–6), 143–149.

Luminex Corporation, 2010. Sample protocols for nucleic acid detection using MagPlex®-TAG™ microspheres. <http://www.luminexcorp.com/prod/groups/public/documents/lmncorp/custom-nucleic-acid-detection.pdf>.

Marston, C.K., Allen, C.A., et al., 2011. Molecular epidemiology of anthrax cases associated with recreational use of animal hides and yarn in the United States. *PLoS One* 6 (12), e28274. <http://dx.doi.org/10.1371/journal.pone.0028274>.

Martino, A., Mancuso, T., et al., 2010. Application of high-resolution melting to large-scale, high-throughput SNP genotyping: a comparison with the TaqMan method. *J. Biomol. Screen.* 15 (6), 623–629. <http://dx.doi.org/10.1177/1087057110365900>.

Mishima, K., Takarada, T., et al., 2005. Capillary electrophoretic discrimination of single nucleotide polymorphisms using an oligodeoxyribonucleotide-polyacrylamide conjugate as a pseudo-immobilized affinity ligand: optimum ligand length predicted by the melting temperature values. *Anal. Sci.* 21 (1), 25–29. <http://dx.doi.org/10.2116/analsci.21.25>.

Pearson, T., Busch, J.D., et al., 2004. Phylogenetic discovery bias in *Bacillus anthracis* using single-nucleotide polymorphisms from whole-genome sequencing. *Proc. Natl. Acad. Sci. U. S. A.* 101 (37), 13536–13541. <http://dx.doi.org/10.1073/pnas.0403844101>.

Pilo, P., Rossano, A., et al., 2011. Bovine *Bacillus anthracis* in Cameroon. *Appl. Environ. Microbiol.* 77 (16), 5818–5821. <http://dx.doi.org/10.1128/AEM.00074-11>.

Price, E.P., Matthews, M.A., et al., 2010. Cost-effective interrogation of single nucleotide polymorphisms using the mismatch amplification mutation assay and capillary electrophoresis. *Electrophoresis* 31 (23–24), 3881–3888. <http://dx.doi.org/10.1002/elps.201000379>.

Rhodes, R.B., Lewis, K., et al., 2001. Analysis of the factor V Leiden mutation using the READIT assay. *Mol. Diagn.* 6 (1), 55–61. <http://dx.doi.org/10.1054/modi.2001.22327>.

Schouten, J.P., McElgunn, C.J., et al., 2002. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res.* 30 (12), e57. <http://dx.doi.org/10.1093/nar/gnf056>.

Shi, M.M., 2001. Enabling large-scale pharmacogenetic studies by high-throughput mutation detection and genotyping technologies. *Clin. Chem.* 47 (2), 164–172.

Song, J., Li, P.E., et al., 2010. Simultaneous pathogen detection and antibiotic resistance characterization using SNP-based multiplexed oligonucleotide ligation-PCR (MOL-PCR). *Adv. Exp. Med. Biol.* 680, 455–464. http://dx.doi.org/10.1007/978-1-4419-5913-3_51.

Stucki, D., Malla, B., et al., 2012. Two new rapid SNP-typing methods for classifying *Mycobacterium tuberculosis* complex into the main phylogenetic lineages. *PLoS One* 7 (7), e41253. <http://dx.doi.org/10.1371/journal.pone.0041253>.

- Tobler, A.R., Short, S., et al., 2005. The SNPlex genotyping system: a flexible and scalable platform for SNP genotyping. *J. Biomol. Tech.* 16 (4), 398–406 (doi: PMC2291745).
- Van Ert, M.N., Easterday, W.R., et al., 2007. Global genetic population structure of *Bacillus anthracis*. *PLoS One* 2 (5), e461. <http://dx.doi.org/10.1371/journal.pone.0000461>.
- Vogler, A.J., Busch, J.D., et al., 2002. Molecular analysis of rifampin resistance in *Bacillus anthracis* and *Bacillus cereus*. *Antimicrob. Agents Chemother.* 46 (2), 511–513. <http://dx.doi.org/10.1128/AAC.46.2.511-513.2002>.
- Vora, G.J., Meador, C.E., et al., 2004. Nucleic acid amplification strategies for DNA microarray-based pathogen detection. *Appl. Environ. Microbiol.* 70 (5), 3047–3054. <http://dx.doi.org/10.1128/AEM.70.5.3047-3054.2004>.
- Ward, T.J., Ducey, T.F., et al., 2008. Multilocus genotyping assays for single nucleotide polymorphism-based subtyping of *Listeria monocytogenes* isolates. *Appl. Environ. Microbiol.* 74 (24), 7629–7642. <http://dx.doi.org/10.1128/AEM.01127-08>.

Discussion Générale

1. La génomique, un outil puissant pour caractériser les relations phylogénétiques de *B. anthracis*

B. anthracis est considérée comme une espèce jeune sur le plan évolutif, ayant récemment émergé du groupe *B. cereus* (Keim *et al.*, 2009). En tant que bactérie sporulante, *B. anthracis* passe la majeure partie de son cycle sous la forme de spore dormante, ce qui explique la très grande homogénéité observée au sein de cette espèce, très peu d'accumulation de mutations ou de transferts étant possibles sous cette forme. De ce fait, seulement quelques méthodes de typage permettent une discrimination entre isolats (**Tableau 5**). Parmi celles-ci, la MLVA permet d'atteindre un niveau de résolution croissant selon le panel et le nombre de VNTR sélectionnés (Keim *et al.*, 2000; Le Flèche *et al.*, 2001; Lista *et al.*, 2006; Thierry *et al.*, 2014). Les résultats obtenus sont généralement satisfaisants, puisque cette méthode rend possible le génotypage de toute souche d'origine inconnue. Cependant, elle ne permet pas la réalisation d'études phylogéniques (homoplasie et pression de sélection de certains VNTR). Les SNR permettent une bonne différenciation des isolats considérés comme identiques en MLVA, mais ils possèdent un taux de mutation très élevé et induisent des erreurs de type homoplasie lors d'études phylogéniques. Les canSNP donnent également de bons résultats, mais renvoient une image « écrasée » de la population étudiée : toute souche isolée dans le monde sera intégrée à l'une des 13 sous-lignées, mais sans être différenciée des autres isolats appartenant au même groupe. Lorsque l'on souhaite étudier l'histoire évolutive d'une espèce bactérienne tout en réalisant le génotypage de cette espèce, une méthode s'impose petit à petit comme le « gold standard » : il s'agit du séquençage à haut débit du génome complet, qui permet l'identification de marqueurs en tant que signatures diagnostiques, en étudiant les SNP répartis sur l'ensemble du génome.

Tableau 5 : Synthèse des pouvoirs résolutifs des principales méthodes de typage chez *B. anthracis*

Niveau d'identification	14 canSNP	MLVA	SNR	Whole-genome SNP (ce travail)
Lignée	Oui	Oui	Non	Oui
Sous-lignée	Oui	Oui	Non	Oui
Branches	Non	Oui	Oui	Oui
Souche	Non	Non	Oui	Oui

Les méthodes de typage ont évolué avec le temps, les technologies employées ainsi que les coûts associés permettant l'emploi de méthodes à haut pouvoir de résolution, telles que le séquençage complet du génome. L'AFLP avait permis d'identifier 2 grandes lignées (A et B) au sein de *B. anthracis* (Keim *et al.*, 1997). La MLVA a permis d'ajouter la lignée C (Keim *et al.*, 2000), observation confirmée par l'approche basée sur les canSNP (Van Ert *et al.*, 2007a). Dorénavant, le séquençage du génome complet de plusieurs souches ou collections de souches à travers le monde ouvre de nouvelles perspectives.

Au cours de ce projet, j'ai obtenu une cartographie à très haute résolution de *B. anthracis*. Le niveau de résolution obtenu est sans précédent en comparaison des méthodes utilisées auparavant. Basé sur le séquençage du génome complet de près de 250 isolats, ce travail a analysé les polymorphismes de type SNP sur la totalité du génome afin de mettre en évidence les relations phylogénétiques existant entre souches au niveau local, régional ou national. Dans un premier temps, l'étude a été réalisée en France. Afin d'élargir la connaissance de la diversité de *B. anthracis* en Europe, plusieurs collaborations avec des pays Européens ont été menées. Les bases de données sont en constante évolution et une cinquantaine de génomes sont actuellement disponibles (<http://www.ncbi.nlm.nih.gov/genome/genomes/181>). De trois sous-lignées présentes en France, l'étude a été élargie aux 13 sous-lignées identifiées au sein de l'espèce *B. anthracis* par les canSNP. Dans cette discussion, une image de la population de *B. anthracis* au niveau mondial est présentée

(292 génomes). L'arbre phylogénétique obtenu à partir des SNP chromosomiques est présenté en **Figure 31**. La liste des génomes utilisés dans cette étude est présentée en **Annexe 3**. Les **Annexes 4 et 5** présentent l'arbre phylogénétique de manière détaillée.

Lors de mon projet de thèse, j'ai appliqué une méthode d'analyse phylogénétique particulière. Les SNP identifiés sur génome complet ont été concaténés pour générer une matrice de SNP. Une approche par parcimonie a ensuite été utilisée. L'application du principe de parcimonie à la phylogénie est à l'origine de l'utilisation du Minimum Spanning Tree (MST) ou arbre de recouvrement minimal (concept introduit par Edwards et Cavalli-Sforza en 1963) (Sobral, 2012). Selon ce concept, « *l'arbre évolutif à privilégier est celui induisant la quantité minimum d'évolution* », c'est-à-dire celui impliquant le moins d'évènements évolutifs. Techniquement, l'arbre de recouvrement est celui dont la somme des longueurs des branches est minimale. Ce type d'arbre permet la définition de génotypes identifiés comme « ancestraux » à partir desquels des variants auraient divergé (Sobral, 2012). Cette méthode de clustering est beaucoup plus robuste et stable que celles fondées sur l'UPGMA ou le Neighbor Joining : selon la taille de la population, certains génotypes sont assignés à un agrégat plutôt qu'à un autre avec ces approches, alors que l'arbre de recouvrement minimal (MST) n'est pas influencé par la taille de la population analysée ou par l'ajout de nouveaux génotypes (Sobral, 2012). Le principe de parcimonie sous-jacent permet la conservation des points de branchements des différents clusters. Les MST sont graphiquement très informatifs et il est possible de prédire le génome ancestral à partir duquel un complexe clonal aurait divergé. L'algorithme utilisé pour la génération des arbres présentés durant ma thèse permet la création de nœuds hypothétiques, toujours basés sur le principe de parcimonie.

J'ai choisi de baser mon étude sur les SNP chromosomiques plutôt que sur les SNP plasmidiques. Bien que les principaux facteurs de virulence soient retrouvés sur les 2 plasmides, il arrive que certaines souches perdent ces dits-plasmides, ce qui biaise alors les relations phylogénétiques observées. Même si le séquençage du génome a permis la récupération des génomes plasmidiques, les travaux présentés concernent uniquement les mutations identifiées sur le chromosome.

1.1. La lignée C : une lignée atypique

La lignée C.Br.A1055 est anecdotique à travers le monde (seuls quelques isolats ont été clairement identifiés à ce jour) (Pearson *et al.*, 2004; Van Ert *et al.*, 2007a). Elle comprend le point d'insertion menant vers *B. cereus*, qui est situé à environ 550 SNP sur la branche C (**Figure 31**). *B. cereus* est situé à près de 20000 SNP de ce point. La lignée C est distante d'environ 1100 SNP du point d'insertion menant aux branches A et B (**Figure 31**). Cela représente près d'un dixième du nombre total de SNP identifiés dans l'arbre global (10583 au total). Très peu d'informations sont disponibles sur ces souches, mais elles sont clairement identifiées comme étant assez différentes au niveau génétique des autres isolats de *B. anthracis*.

Deux branches majeures sont issues de la lignée C et partent dans des directions opposées avec près de 300 SNP chacune : les branches A et B. D'après nos observations, l'ancêtre commun le plus récent à ces 2 lignées est situé à une distance quasi identique du point de séparation de ces lignées (297 SNP vers la B et 292 SNP vers la A) (**Figure 31**). Pearson et collaborateurs avaient également rapporté des distances quasi identiques (312 SNP vers la B et 308 SNP vers la A), indiquant des procédés de filtre des SNP similaires (Pearson *et al.*, 2004). Cette observation évoque une divergence de ces 2 lignées à peu près à la même période (il y a environ 8700 ans d'après Van Ert *et al.*, 2007).

1.2. La lignée B : une lignée régionale

Représentant environ 15 % des isolats de *B. anthracis* à travers le monde (Pearson *et al.*, 2004; Van Ert *et al.*, 2007a), la lignée B comporte trois sous-lignées identifiées par canSNP: B.Br.001/002, B.Br.Kruger et B.Br.CNEVA.

La sous-lignée B.Br.CNEVA est retrouvée exclusivement en Europe (Antwerpen *et al.*, 2012; Derzelle *et al.*, 2011a; Garofolo *et al.*, 2011; Gierczyński *et al.*, 2004; Pilo *et al.*, 2008; Steiner *et al.*, 2013; Van Ert *et al.*, 2007a), à l'exception d'un cas observé en Californie (Van Ert *et al.*, 2007a). Sous-lignée majoritaire en France (Derzelle *et al.*, 2011a), mon travail a montré qu'une corrélation étroite existe entre sa répartition géographique et le génotype associé. Trois grands groupes géographiques ont été définis (Massif-Central, Alpes, Pyrénées) ainsi qu'un sous-groupe au sein du Massif Central (Saône-et-Loire). Contrairement à la MLVA 30 qui sépare notamment certaines souches de Saône et Loire (données non publiées), le groupement géographique observé via l'étude des SNP sur génome complet permet une parfaite corrélation entre le génotype et la région géographique.

Présente en Europe dans différents pays (Pologne, Suisse, Italie, Allemagne) (Antwerpen *et al.*, 2012; Garofolo *et al.*, 2011; Gierczyński *et al.*, 2004; Pilo *et al.*, 2008), la sous-lignée B.Br.CNEVA forme cinq branches phylogénétiques distinctes. La filiation entre la région géographique d'origine et le génotype observée en France est retrouvée au niveau Européen, chaque branche étant caractérisée par un pays ou une région au sein d'un pays. Il est intéressant de noter que les souches isolées dans les Alpes Italiennes diffèrent significativement des souches retrouvées dans les Alpes Françaises (plus de 130 SNP de différence). Elles sont définies comme étant plus anciennes que les souches Françaises par cette étude.

Les souches autochtones Françaises sont identifiées ici comme étant les plus récentes de la sous-lignée B.Br.CNEVA (plus précisément les souches du Massif Central), puisqu'elles sont retrouvées aux extrémités de la branche phylogénétique.

Ces observations montrent une régionalisation de la sous-lignée B.Br.CNEVA. Chaque Pays, ou région au sein d'un pays, semble posséder une sous-population propre et clairement identifiable. Ceci concorde avec la dispersion d'un ancêtre commun vers l'Ouest et le Sud, avec l'apparition de populations dérivées en fonction du lieu d'implantation. En France, l'hypothèse serait celle d'une introduction unique de cette sous-lignée, suivie d'une dispersion de la population vers l'Ouest (Massif Central-Saône et Loire) et le Sud (Pyrénées). Ceci est montré dans cette étude par l'émergence des souches Pyrénéennes et du Massif Central à partir de celles des Alpes. De nouvelles populations locales sont ainsi apparues, conduisant à la structure de population observée en France.

Les deux autres sous-lignées de la branche B sont B.Br.001/002 et B.Br.Kruger (Van Ert *et al.*, 2007a). La première est généralement retrouvée en Afrique (Zimbabwe, Afrique du Sud) (Van Ert *et al.*, 2007a), mais certains isolats ont été identifiés en Europe (Suède, Belgique) et en Asie (Corée, Bhoutan) (Thapa *et al.*, 2014). Les génomes de cette sous-lignée sont assez distants les uns des autres, à l'exception de trois souches Belges formant un cluster (7 SNP de différence au maximum) et indiquant vraisemblablement une même origine d'introduction en Belgique, qui pourrait être due à des importations provenant d'anciennes colonies Africaines. Les souches de ces sous-lignées sont majoritairement originaires d'Afrique. Leur présence dans d'autres pays est vraisemblablement due aux échanges commerciaux. Une hypothèse probable serait l'importation de farines animales originaires d'Afrique du Sud, conduisant aux cas sporadiques observés dans les différents pays. Une fois provoqués par l'utilisation de ces farines, les foyers déclarés dans les différents pays resteraient ainsi locaux, ne tendant pas à s'étendre et conduisant ainsi à la régionalisation observée ici.

La branche B, seconde branche la plus importante en termes d'isolats mondiaux de *B. anthracis*, est assez régionalisée en termes de lieux d'isolements. Elle semble implantée dans certaines régions particulières du monde, telles que les régions montagneuses en France, ou l'Afrique du Sud et ses parcs (comme le Parc National Kruger). Il semble y avoir une certaine notion d'adaptation aux sols ou

aux conditions de persistance des spores dans l'environnement pour cette lignée B. Seulement retrouvée dans quelques régions du globe, elle semble être moins adaptée à des environnements variés. Il serait ainsi intéressant d'étudier avec précision la composition des sols et les écosystèmes associés dans les régions où les spores sont retrouvées dans ces pays. Néanmoins, ces régions sont propices à l'apparition récurrente de la maladie : au sein des parcs naturels d'Afrique, de nombreux troupeaux d'animaux sauvages sont susceptibles d'entretenir le cycle de *B. anthracis* (Smith *et al.*, 2000). De même, en France, les régions concernées sont montagneuses et caractérisées par l'élevage extensif de bétail et la pratique de la transhumance. Des sols adaptés à la persistance des spores ainsi qu'une faune sauvage ou domestique permettant à la bactérie de réaliser pleinement son cycle sont les deux hypothèses principales, et certainement les plus probables, à l'explication de la régionalisation observée pour cette lignée B.

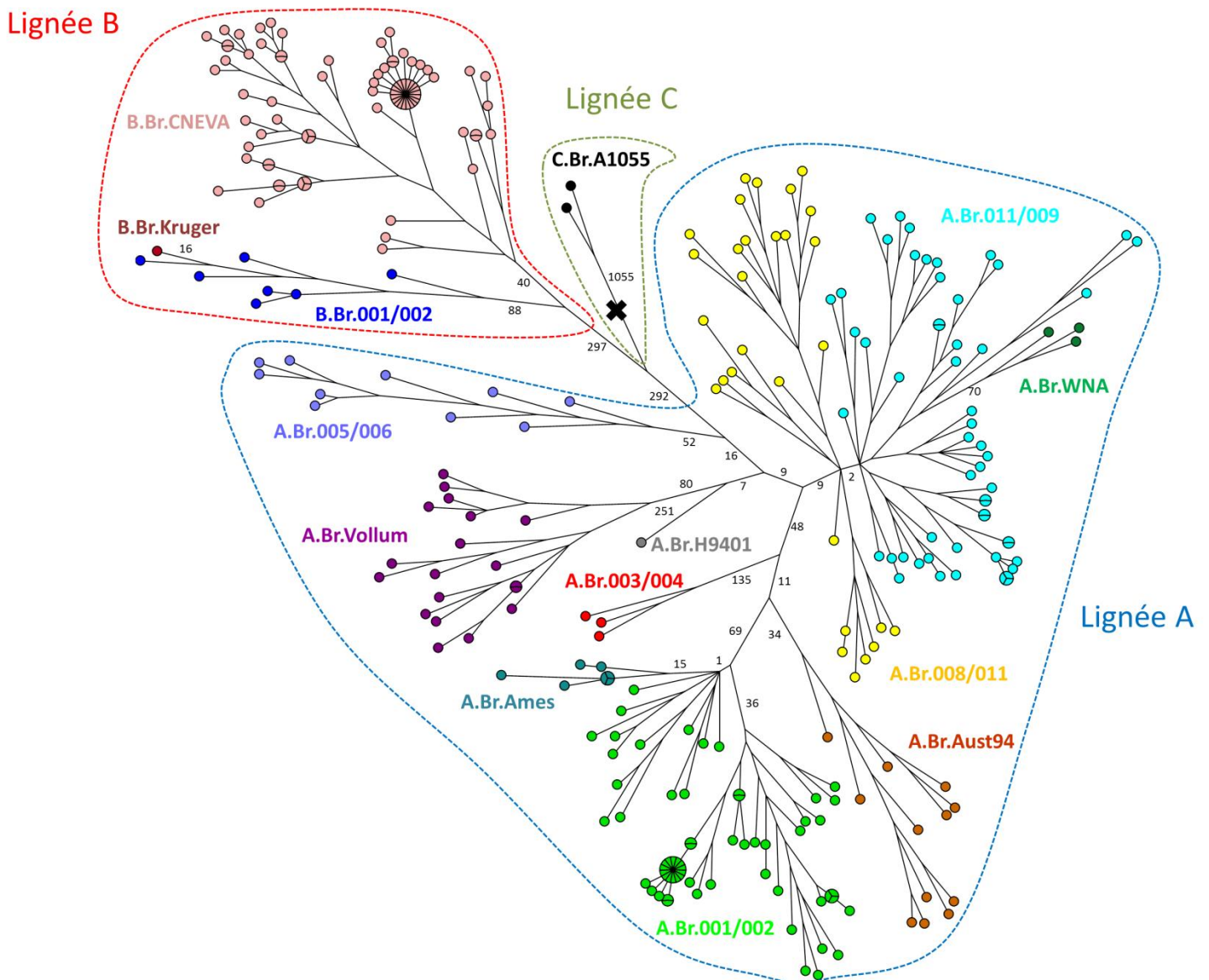


Figure 31 : Arbre de recouvrement minimal (Minimum Spanning Tree) avec nœuds hypothétiques généré à partir de 292 génomes de *B. anthracis* (Matrice de 10583 SNP, Approche par parcimonie). Les sous-lignées sont représentées par couleur. Le nombre de SNP est indiqué pour les branches majeures et le point de branchement vers *B. cereus* est indiqué par la croix noire.

1.3. La lignée A : une présence au niveau mondial

Représentant près de 85 % des isolats de *B. anthracis* retrouvés sur Terre (Pearson *et al.*, 2004; Van Ert *et al.*, 2007a), la lignée A a connu la plus grande expansion de toutes les lignées. Neuf sous-lignées retrouvées sur tous les continents la constituent.

La plus ancienne sous-lignée est A.Br.005/006. Retrouvée principalement en Afrique (Van Ert *et al.*, 2007a), elle a été recensée dans certains pays Européens (Danemark, Italie, Belgique, France et Suisse). Mon étude a mis en évidence une grande hétérogénéité avec plusieurs branches uniques et assez longues en termes de SNP (≈ 100 SNP). La présence de cette sous-lignée en Italie est vraisemblablement liée aux échanges commerciaux maritimes datant du Moyen-âge. Une introduction via des produits animaux (laine, farine animale,...) est ainsi suspectée (Fasanella *et al.*, en préparation). L'observation la plus intéressante au sein de cette sous-lignée provient de cinq souches formant un cluster quasi clonal: trois souches Suisses originaires du Tchad (centre-ouest de l'Afrique), une souche originaire du Cameroun et une souche d'origine indéterminée (IEMVT89, collection ANSES). Ces cinq souches proviennent d'Afrique de l'Est et partagent un ancêtre commun récent. Les trois souches du Tchad sont regroupées, alors que les deux autres souches sont quasi identiques (3 SNP). Cette dernière observation permet de définir avec quasi-certitude l'origine de la souche IEMVT89 comme provenant du Cameroun. Ceci illustre un intérêt applicatif de ce genre d'étude, à savoir, la possibilité de remonter à l'origine d'une souche.

La sous-lignée A.Br.Vollum est connue pour être retrouvée principalement dans le sud de l'Afrique ainsi qu'en Asie (Van Ert *et al.*, 2007a). Elle a été impliquée dans des épisodes de FC chez les travailleurs de la laine en Europe (Suisse et Belgique) (Pilo *et al.*, 2008; Wattiau *et al.*, 2008). Comme pour la sous-lignée B.Br.CNEVA, mon étude montre une relation entre le génotype et le lieu d'isolement pour certaines souches. Les souches Suisses originaires du Pakistan sont regroupées au sein d'une même branche, de même que les souches isolées dans la région Iran-Russie-Kazakhstan-Afghanistan. Ces souches proviennent de cas observés dans les usines de travailleurs de la laine: lors de cas importés, les souches sont souvent très proches voire identiques, ce qui cause souvent un ou plusieurs foyers localisés (pas ou peu d'expansion), expliquant cette forte proximité génétique.

Peu de données sont disponibles concernant la sous-lignée A.Br.003/004, retrouvée majoritairement en Amérique du Sud, bien que quelques laboratoires l'aient identifiée en Europe (Suisse, Belgique) et en Israël (Cohen-Gihon *et al.*, 2014). Seuls trois génomes ont été étudiés au cours de cette étude, montrant des branches distinctes entre isolats. Il serait extrêmement intéressant de pouvoir investiguer plus en détail cette sous-lignée qui semble principalement cantonnée en Amérique du Sud. Si elle a été exportée, elle n'a pas réussi à s'implanter durablement dans les pays touchés. Ceci pourrait refléter un certain déficit dans la capacité d'adaptation à de nouveaux environnements, voire d'hôtes pour cette sous-lignée.

La sous-lignée A.Br.Aust94 est retrouvée à différents endroits dans le monde (Australie, Suisse, Géorgie, Allemagne, Danemark, Pays-Bas, Turquie, Inde, Namibie). Les génomes étudiés n'ont pas montré de relation marquante entre isolats, en dehors de deux souches allemandes quasi identiques. Cette sous-lignée est assez diverse, mais un cas observé de régionalisation est décrit : les souches originaires de Géorgie sont apparentées, indiquant une implantation locale de la sous-lignée (Khmaladze *et al.*, 2014).

Les deux sous-lignées suivantes sont étroitement liées : A.Br.Ames et A.Br.001/002, la première étant issue de la seconde. La dénomination commune « sous-lignée Ames-Sterne » est de plus en plus utilisée. Nos travaux montrent que la sous-lignée A.Br.001/002 est clairement scindée en deux groupes. Le premier comprend les souches Ames et certaines souches A.Br.001/002. Il est nommé

groupe A01 et est défini à la base par un seul SNP (position 515111). Le second groupe comprend la souche Sterne et d'autres souches A.Br.001/002. Il est nommé groupe A02.

La sous-lignée Ames-Sterne est très hétérogène en termes d'origines. Le groupe A01 contient des souches Asiatiques (Indonésie, Chine, Japon), les souches Ames (USA, Géorgie-Union Soviétique, Danemark) mais également quelques souches Européennes (France, Royaume-Uni, Danemark, Allemagne, Belgique). Le groupe A02 est également très divers avec des souches Européennes (France, Pays-Bas, Allemagne, Danemark, Pologne, Belgique), mais également plusieurs souches du Bangladesh et la souche vaccinale animale Sterne.

Le groupe A01 semble regrouper des souches Asiatiques (Chine, Japon, Indonésie). Il a été montré que la sous-lignée A.Br.001/002 est retrouvée majoritairement en Chine (Simonson *et al.*, 2009). Dans cette étude (Simonson *et al.*, 2009), il est également observé que la sous-lignée A.Br.Ames est issue d'A.Br.001/002 et que les souches retrouvées dans le Texas partagent des ancêtres communs avec les souches de Chine. Mon travail montre que le groupe A01 est basal à la sous-lignée Ames-Sterne (il intervient précocement au sein de l'arbre) et contient donc des souches plus anciennes que celles retrouvées dans le groupe A02.

Ce dernier apparaît 36 SNP plus loin. Il est également très divers et majoritairement constitué de souches Européennes et de souches retrouvées au sud de la Chine (Bangladesh). Bien que le groupe A01 contienne également quelques isolats Européens, le groupe A02 en contient un nombre plus important dans cette étude. L'hypothèse d'une dispersion de la sous-lignée A.Br.001/002 venant des Pays Asiatiques vers l'Ouest de l'Asie (Bangladesh) et l'Europe, avec établissement de populations locales pour le groupe A02, peut être posée. Les quelques souches du groupe A01 retrouvées en Europe pourraient être dues à des cas importés d'Asie via le commerce. N'ayant pas pu s'établir écologiquement (à l'inverse de ce qui est observé pour le groupe A02), ces cas seraient restés sporadiques. Cette hypothèse pourrait être vérifiée en incluant un plus grand nombre de souches provenant de collections à travers le monde.

Situées dans le groupe A02, les souches isolées en France dans le Doubs (en 2008 et 2011) forment un complexe clonal, illustrant la persistance d'un même clone. La très forte proximité entre isolats observée ici indique une introduction unique dans la région, conduisant à cette homogénéité. Les souches Françaises isolées dans le Doubs partagent leur ancêtre commun le plus récent avec trois souches originaires du Bangladesh, ce qui pourrait indiquer une possible introduction en France via cette région du Globe. Les trois souches Françaises de la CIP retrouvées dans le groupe A01 sont éloignées d'une centaine de SNP du complexe clonal observé dans la région du Doubs et ne font pas partie du même groupe (A01 contre A02, respectivement). Ces observations suggèrent, au moins, une double introduction de cette sous-lignée en France, celle conduisant au groupe A01 étant plus ancienne.

Un isolat présente des propriétés hémolytiques sur gélose au sang (BL9 hem) dans le groupe A02. Isolée en Pologne, cette souche possède le plasmide pXO1 mais pas pXO2. Elle fait partie des isolats de *B. anthracis* de type hémolytiques. Ce critère phénotypique est habituellement caractéristique des souches de type *B. cereus*, permettant de distinguer les deux espèces. Cette observation illustre les limites des méthodes d'identification classiques qui peuvent parfois induire quelques erreurs et la nécessité de développer des approches génomiques plus résolutes.

La sous-lignée A.Br.Ames comprend plusieurs isolats américains, notamment la souche référence Ames Ancestor, deux isolats danois quasi identiques (1 SNP de différence), indiquant probablement une même origine d'introduction au Danemark, mais également la souche isolée en Floride en 2001 lors de l'épisode Amerithrax (A2012). Cette dernière possède une très grande longueur de branche (190 SNP) : cette accumulation de mutations pourrait être due à une culture intensive de la bactérie. En effet, il a été montré que la souche utilisée lors des attentats de 2001 était une souche de laboratoire, ce qui pourrait expliquer cette longueur de branche conséquente. Cette observation

montre une des limites à l'étude des SNP sur génome complet : lorsque des données sont publiées et déposées dans les bases de données publiques (NCBI, Genbank,...), elles sont représentatives de la souche à un instant donné. Or, lorsque l'on réalise des études phylogénétiques, il est important de partir de données les plus homogènes possibles. Chez une espèce clonale comme l'est *B. anthracis*, le taux de mutation approche de 10^{-10} par génération (espèce très stable) (Keim *et al.*, 2004). Lorsque l'on observe une longueur de branche très importante (ici 190 SNP), il est possible d'émettre des hypothèses erronées (souches très éloignées,...). De plus, il a été montré que très peu de SNP sont observés au sein des souches Ames à travers le monde (Simonson *et al.*, 2009; Van Ert *et al.*, 2007a). Dans ce type d'étude incluant les génomes des bases de données publiques, il convient de considérer l'historique lorsqu'il est disponible et de rester prudent lors de la formulation des hypothèses. Ici, l'hypothèse la plus probable semble donc être la culture intensive de la souche plutôt qu'une grande distance évolutive.

Un autre exemple de ces limites concerne la souche STI-1 (souche vaccinale en Géorgie/Union Soviétique) déposée récemment dans les bases de données en ligne (Okinaka *et al.*, 2014). Il est indiqué qu'elle appartient à la sous-lignée TEA. Or, le génome disponible est identique à celui de la souche référence Ames Ancestor, indiquant vraisemblablement une erreur d'import des données. Pour pallier ce type d'erreurs, la meilleure solution reste l'accès direct aux données brutes de séquençage, ce qui est le cas avec les Sequence Read Archive (SRA).

La lignée A a connu la plus grande expansion de toutes les lignées de *B. anthracis*. A quoi est due cette expansion remarquable ? Possède-t-elle de meilleures capacités d'adaptation aux différents sols rencontrés ? A-t-elle un plus large spectre d'hôtes que la lignée B ou la lignée C ? Toutes ces hypothèses sont vraisemblables mais demandent à être vérifiées.

Les dernières sous-lignées de la lignée A sont décrites en détail dans le paragraphe suivant (A.Br.008/011, A.Br.011/009 et A.Br.WNA).

1.3.1. La sous-lignée Trans-Eurasienne : une polytomie remarquable

La sous-lignée TEA est la plus représentée à travers le monde (Derzelle et Thierry, 2013; Garofolo *et al.*, 2011; Price *et al.*, 2012; Van Ert *et al.*, 2007a). Cette sous-lignée est séparée en deux sous-groupes (A.Br.008/011 et A.Br.011/009) (Marston *et al.*, 2011). Deux SNP séparent ces deux sous-branches (**Figure 31**).

En France, la sous-lignée TEA (A.Br.011/009) est beaucoup plus dispersée sur le territoire que les sous-lignées B.Br.CNEVA et A.Br.001/002, démontrant une certaine capacité d'adaptation à des sols divers. Elle présente une structure polytomique à six branches issues d'un même point de départ (Girault *et al.*, 2014a), suggérant une introduction unique à l'origine de ces radiations. Ces branches sont représentées en nombre d'isolats de manière variable (entre 1 et 14 souches) et leur longueur varie de 19 à 82 SNP (branches 5 et 4 respectivement), pouvant suggérer une vitesse d'évolution différente selon les branches durant un même laps de temps.

L'incorporation de souches d'horizons variés (Italie, Afrique, Royaume-Uni) a permis de valider l'existence de ces six branches au niveau mondial, toutes les souches étudiées s'incorporant dans les six branches identifiées en France.

Notre hypothèse de départ était qu'une souche de la sous-lignée TEA originaire d'Italie aurait été introduite en France, conduisant à cette extrême diversité. Pour vérifier cette hypothèse, des données ont été collectées à partir de souches Italiennes, mais également d'autres pays en Europe. Une partie de la collection de l'IZS de Foggia (Italie) a permis d'étayer cette hypothèse : les souches italiennes étudiées sont retrouvées dans les branches identifiées en France. Au sein de la branche 1 (souches françaises et souches italiennes Pouilles-Lazio-Toscane), une distinction nette entre souches

françaises et italiennes est cependant observée. La séparation entre les deux intervient très précocement après le nœud de la branche (1 SNP) indiquant un ancêtre commun. Les longueurs de branche sont très similaires, avec un minimum et un maximum de 14 et 35 SNP pour l'Italie, et de 18 et 33 SNP en France. Ces valeurs semblent indiquer que les souches italiennes ont divergé peu de temps avant les souches françaises. Le même type d'observation est fait pour la branche 2 (souches françaises et italiennes Basilicate-Campanie (BC)). La séparation entre les souches françaises et italiennes intervient très près du nœud de la branche (1 SNP) indiquant un ancêtre commun. Les longueurs de branche sont cependant différentes : avec un minimum et un maximum de 13 et 27 SNP, les souches italiennes BC sont plus anciennes que les françaises (minimum de 23 et maximum de 45 SNP). Ces valeurs semblent indiquer que les souches italiennes ont divergé antérieurement par rapport aux souches françaises.

Ces observations semblent donc suggérer la présence d'un ancêtre commun aux souches italiennes et françaises pour chaque branche, vraisemblablement originaire d'Italie puisque les souches italiennes semblent diverger plus précocement que les souches françaises.

Dans la branche 2, une souche isolée en Sardaigne est retrouvée seule, partageant un ancêtre commun avec les souches françaises. Trois souches d'Afrique de l'Ouest sont issues de cette souche de Sardaigne. Cependant, une longueur de branche très importante est observée et il est de nouveau difficile de formuler une interprétation (grande longueur de branche due à une culture intensive ? réelle différence ?) et ce, d'autant plus que ces données de séquences proviennent des bases de données en ligne (contigs). L'étude présentée montre clairement que la sous-lignée A.Br.WNA est issue de la sous-lignée TEA (branche A.Br.011/009). Elle provient d'une branche phylogénétique où sont retrouvées des souches Françaises et Italiennes. Elle pourrait être issue d'une souche Française se trouvant à la base de la branche menant d'une part à la souche Italienne isolée en Sardaigne et d'autre part aux souches Nord-Américaines. Cette hypothèse est en opposition avec celle émise par Kenefic en 2009 qui suggérerait une introduction de la TEA conduisant à la sous-lignée A.Br.WNA via le détroit de Behring et datant du Pléistocène il y a 12000 ans (Kenefic *et al.*, 2009). Une différenciation du Nord vers le Sud du continent Nord-Américain est ainsi observée, ayant pour origine le Canada. L'hypothèse formulée au cours de mon étude conduit à une introduction beaucoup plus récente, pouvant venir de la colonisation du Canada par la France il y a 500 ans de cela. Toutefois, ces observations demandent à être confirmées par une étude plus variée en termes de pays d'origine de souches de la sous-lignée TEA afin d'être validées. Il serait ainsi très intéressant d'étudier la diversité de souches provenant d'Espagne et du Portugal. En effet, les Espagnols ont fait partie des pays colonisateurs au 16^{ème} siècle alors que le Portugal était une grande puissance coloniale et maritime. Il se pourrait que les souches espagnoles et portugaises puissent apporter de précieuses informations concernant l'histoire évolutive de cette sous-lignée, et valider ou infirmer l'hypothèse proposée ici.

Une observation intéressante est faite au sein de la branche 3 qui comprend des souches Européennes (France, Italie, Pays-Bas) et une souche du Congo. Un cluster est clairement distingué au sein de ce groupe, regroupant des souches vaccinales animales possédant les 2 plasmides (pXO1 et pXO2) : la souche Carbosap est la souche vaccinale utilisée en Italie ; la souche 95014 a été identifiée comme très proche de la Carbosap ; la souche CIP 74.12 est une Pasteur II, alors que la 17 JB a été utilisée pour des tests de vaccinations animales. Il semble donc que ce cluster regroupe des souches dites vaccinales. Comme on pouvait s'y attendre, les plus grandes longueurs de branche observées pour la branche 3 sont observées pour les souches dites vaccinales. Ceci illustre le fait que les souches cultivées de manière intensive acquièrent plus de mutations.

Trois délétions chromosomiques de 29, 24 and 3.5 kb ont été décrites chez la Carbosap (Harrington *et al.*, 2013). De manière intéressante, les délétions de 3.5 kb (position 1235796-1239220, incluant les loci tags GBAA_1286 à GBAA_1290 dans le génome Ames Ancestor) et de 24 kb (position 1640915-1665378, GBAA_1751-GBAA_1781 incluant le gène associé à la virulence GBAA_1760 (Papazisi *et al.*, 2011)) sont retrouvées chez les souches de type Pasteur II et celles identifiées comme proches de la Carbosap. Cependant, la troisième région (position 267162-294257) est intacte chez les

Pasteur II. Ces souches vaccinales animales possèdent les plasmides de virulence identifiés chez *B. anthracis*. De ce fait, les délétions observées au sein du génome de ces souches pourraient être associées à une perte de facteurs de virulence autres que ceux portés sur les plasmides.

La sous-lignée A.Br.011/009 vient d'être détaillée, il reste désormais son pendant, la sous-lignée A.Br.008/011. Une polytomie est également observée ici avec quatre branches identifiées.

Les origines des souches de cette sous-lignée A.Br.008/011 sont très diverses et variées : l'Europe (Belgique, Hongrie, Pologne, Royaume-Uni, ex-Union Soviétique, Italie, Bulgarie, Pays-Bas, Danemark, Turquie, Ecosse, Allemagne), la Chine et l'Afrique (Ethiopie). Les cas de charbon d'injection ont été causés par des souches appartenant à cette sous-lignée. Les souches isolées à partir d'héroïne sont apparentées, mais présentent tout de même quelques différences permettant de les distinguer. Les investigations épidémiologiques menées pendant les cas de charbon d'injection en Europe avaient conduit à supposer que la contamination avait eu lieu en Turquie (Berger *et al.*, 2014; Hanczaruk *et al.*, 2014; Price *et al.*, 2012). La souche la plus proche des souches incriminées provenait alors de Turquie. L'étude menée ici montre que l'ancêtre commun le plus récent des souches provenant des cas de charbon d'injection est partagé avec les souches danoises plutôt qu'avec la souche turque. L'origine précise des souches danoises est inconnue. Elles pourraient avoir été importées de Turquie dans les années 1970 (années de leur isolement sur le sol danois). Une nouvelle limite de ce type d'étude est atteinte ici, liée à la qualité des méta-données des isolats. Les informations peuvent être soit manquantes, soit incomplètes. Par moment, l'hôte est manquant, ce qui empêche tout type d'hypothèse concernant un hôte privilégié plutôt qu'un autre. La plupart des souches de cette étude sont issues de bovins. Les méta-données sont très importantes lorsque l'on souhaite reconstruire une phylogénie, et elles doivent, dans la mesure du possible, être le plus complètes possibles.

La sous-lignée TEA est particulière au sein de l'espèce *B. anthracis*. Elle est la plus représentée dans le monde en termes d'isolats et de pays où elle est recensée. L'étude des polymorphismes sur génome complet a montré une structure polytomique pour les deux groupes A.Br.008/011 et A.Br.011/009. Quatre branches sont identifiées pour la première et six pour la seconde. L'hypothèse de deux souches distinctes à l'origine de ces polytomies est intéressante. Les structures de population observées montrent une radiation à partir d'un point unique. Toutes les souches issues de ce point possèderaient le même ancêtre commun (un pour A.Br.008/011 et un pour A.Br.011/009). L'ancêtre de ces deux souches serait alors un isolat de la sous-lignée A.Br.008/011 comme le suggère l'arbre phylogénétique généré, A.Br.008/011 étant antérieure à A.Br.011/009 (**Figure 31**). Les souches de cette sous-lignée ont des origines très diverses. Cependant, plusieurs isolats proviennent d'Asie ou d'Europe de l'Est (Chine, Russie, Bulgarie) alors que les souches de la sous-lignée A.Br.011/009 semblent venir majoritairement d'Europe de l'Ouest (France, Italie, Royaume-Uni). L'hypothèse d'une introduction A.Br.008/011 venant de l'Est du continent Eurasien conduisant à une différenciation locale, puis en A.Br.011/009 semble être envisageable. La sous-lignée A.Br.WNA serait ensuite apparue à partir d'A.Br.011/009, possiblement via une souche française exportée au Canada il y a de cela 500 ans.

1.3.2. Proposition d'une nouvelle sous-lignée : A.Br.H9401

La souche Coréenne H9401 (Chun *et al.*, 2012) est un cas à part au sein de l'arbre phylogénétique de *B. anthracis*. En typage canSNP *in silico*, elle appartient à la sous-lignée A.Br.005/006. Or, cette dernière n'est pas définie clairement par un canSNP unique, mais par un marqueur regroupant plusieurs sous-lignées (A.Br.006) (Van Ert *et al.*, 2007a). Si l'on se base sur la **Figure 31**, la souche H9401 est apparentée à la sous-lignée A.Br.Vollum (et non A.Br.005/006), mais elle reste tout de même bien éloignée avec plus de 300 SNP de différence. Elle n'est pas définie par le canSNP discriminant les souches A.Br.Vollum (A.Br.007) (Van Ert *et al.*, 2007a). Sa position dans l'arbre *B.*

anthracis porte ainsi à discussion. Doit-on redéfinir un nouveau canSNP pour les souches A.Br.Vollum ? Forme-t-elle une nouvelle sous-lignée ?

Ce génome provient des bases de données et la longueur de branche pourrait être due à une culture intensive de la souche. Cependant, malgré la longueur de branche pouvant biaiser l'interprétation, elle reste assez mal définie par les canSNP décrits précédemment (Marston *et al.*, 2011; Van Ert *et al.*, 2007a). Nos résultats conduisent donc à considérer cette souche comme faisant partie d'une nouvelle sous-lignée, nommée A.Br.H9401. Pour cela, un nouveau canSNP a été défini (position 2591768). Cette souche partage un ancêtre commun avec les souches Vollum. Elle aurait été introduite en Corée il y a plusieurs siècles de cela, bien avant la différenciation des souches A.Br.Vollum observée au Moyen-Orient.

En conclusion, ce travail a conduit à affiner de façon conséquente les connaissances phylogénétiques de *B. anthracis*. L'analyse des SNP sur génome complet a permis de reconstruire la phylogénie de *B. anthracis* à un niveau sans précédent, grâce à l'utilisation du séquençage sur un ensemble original de souches issues de différentes parties du monde. Cette combinaison particulière (méthode de séquençage x large panel de souches) a permis d'obtenir un pouvoir résolutif encore jamais atteint pour cette espèce bactérienne : cette approche permet de discriminer de la lignée jusqu'à la souche. Des relations fines entre isolats ont été caractérisées et ont été décrites pour la première fois au cours de ce projet : l'identification des origines géographiques des souches présentes sur le sol français puis dans différents pays d'Europe, le caractère polytomique de la sous-lignée TEA, une nouvelle hypothèse sur l'origine de la sous-lignée A.Br.WNA, la caractérisation de 2 groupes au sein de la sous-lignée Ames-Sterne et la proposition d'une nouvelle sous-lignée A.Br.H9401.

Cette méthode présente donc toutes les qualités requises à l'analyse de la diversité d'une espèce. Cependant, quelques limites inhérentes aux données ont été identifiées (nombre de souches limité pour certaines sous-lignées, données publiques sous forme de contigs, méta-données manquantes,...) et il est ainsi nécessaire de rester prudent lors de la formulation d'hypothèses pour ce type d'études.

Nonobstant la contribution à l'étude des relations phylogénétiques, ce travail descriptif a également permis l'identification de marqueurs génétiques utilisables pour le développement d'outils de typage, qui a constitué le second volet de ce projet et qui est discuté ci-après.

2. Développement d'outils de typage robustes pour caractériser la diversité génétique et optimiser le diagnostic moléculaire de *B. anthracis*

La première partie de mon travail consistait à étudier la diversité de *B. anthracis*. L'étape suivante a consisté à mettre au point un outil permettant la mise en évidence de cette diversité, sans avoir à recourir au séquençage du génome. Le pouvoir de discrimination obtenu après séquençage du génome complet est optimal, mais, bien que les coûts soient en constante diminution, tous les laboratoires ne sont pas en mesure de séquencer un génome. L'étude de la diversité de *B. anthracis* a permis l'obtention d'un arbre phylogénétique montrant les relations entre les divers isolats. Choisir des marqueurs représentatifs de chaque branche, sous-branche ou cluster est une approche moins coûteuse qui permet de positionner tout isolat de *B. anthracis* isolé dans le monde.

2.1. Méthodologie pour définir les signatures moléculaires caractéristiques de lignées

Identifier un marqueur moléculaire représentatif d'un groupe ou d'une souche (marqueur canonique) nécessite de mettre en œuvre une démarche analytique complexe. Dans un premier temps, la totalité des génomes de l'étude sont alignés sur le génome de référence Ames Ancestor (via le logiciel BioNumerics et le module Chromosome Comparisons). Les SNP sur génome complet sont ensuite identifiés via un criblage du logiciel. Plusieurs milliers de SNP sont ainsi récupérés avec leur position et les nucléotides impliqués (données brutes).

A l'heure actuelle, il n'existe pas de méthode standardisée pour l'identification phylogénétique des SNP à partir des données de séquençage du génome complet. De manière fréquente, les SNP déduits de l'alignement de lectures courtes, de contigs ou de génomes complets sur une séquence référence sont filtrés afin de masquer différentes sortes d'éléments répétés (Blouin *et al.*, 2012; Comas *et al.*, 2010; Ford *et al.*, 2012; Gardy *et al.*, 2011; Weiner *et al.*, 2012). Ces étapes de filtres peuvent se justifier de plusieurs manières. Les lectures de séquençage courtes ne permettent pas de reconstituer fidèlement certaines régions du génome (régions répétées, opérons ribosomiques, transposases,...) (Blouin *et al.*, 2012). En particulier, les VNTR et les ARNr ne sont pas reconstitués en un seul fragment si leur taille dépasse celle des lectures de séquençage. De plus, il a été montré chez *Mycobacterium tuberculosis* que, dans certains cas, il existe des recombinaisons homologues intrachromosomiques entre ce type d'éléments du génome (Uplekar *et al.*, 2011). De ce fait, les SNP identifiés dans les régions précitées n'ont pas été pris en compte lors de mon analyse basée sur les comparaisons entre génomes complets et données issues de NGS (lectures courtes).

Il est également communément admis que les SNP identifiés en clusters (dans une fenêtre de 10 pb) sont éliminés lors de ce type d'étude (Blouin *et al.*, 2012; Gardy *et al.*, 2011). Ces clusters de SNP sont souvent associés à des régions répétées du génome (Gardy *et al.*, 2011). Les positions où les données sont incomplètes ont également été retirées de mon étude. Une fois ces filtres appliqués, près d'un tiers du nombre total de SNP a été masqué de mon étude. On obtient ainsi une matrice de SNP.

Un autre facteur à prendre en compte pour valider un SNP concerne la profondeur de séquençage associée. Certaines régions du génome seront moins séquencées, et, dans certains cas, seules quelques lectures de séquençages seront alignées sur le génome de référence, pouvant induire un biais d'analyse en cas d'erreur de séquençage. Afin de valider un SNP, une profondeur de séquençage de 30X minimum est conseillée par Illumina pour l'utilisation de leur outil (http://res.illumina.com/documents/products/technotes/technote_snp_caller_sequencing.pdf). Cependant, cet outil est calibré pour les génomes diploïdes. Nous conseillons ici une profondeur de séquençage d'environ 60X.

Afin d'obtenir une image précise et aussi proche que possible de la réalité biologique et phylogénique, une donnée supplémentaire est à prendre en compte pour une bactérie hautement

clonale telle que l'est *B. anthracis*. Il est important de garder à l'esprit qu'une très faible part d'homoplasie doit être identifiée. L'homoplasie est la similitude d'un état de caractère chez différentes espèces qui, contrairement à l'homologie, ne provient pas d'un ancêtre commun. La plupart des SNP retrouvés dans plusieurs branches ont été éliminés de mon étude. Une approche par parcimonie a permis de vérifier cet indice d'homoplasie lors de la génération de chaque arbre phylogénétique, qui est inférieur à 2 % dans la plupart des cas.

Après avoir filtré les données brutes puis vérifié la profondeur de séquençage, les SNP reliés aux différents groupes d'intérêts ont été identifiés. La démarche suivie lors de mon projet a été de sélectionner des groupes contenant un certain nombre d'isolats ou, le cas échéant, une branche unique si peu de génomes sont regroupées (cas de la sous-lignée TEA ou A.Br.H9401 par exemple). Plus le nombre d'isolats est important, plus le marqueur a de chances d'être canonique. A l'inverse, si un seul génome est présent dans le groupe, le marqueur risque de ne pas être canonique si de nouvelles données apparaissent. Cette situation a été rencontrée durant mon projet, avec la définition d'un marqueur pour les souches de type « Carbosap-like » retrouvées en Italie. Un génome dans les bases de données a été libéré et possédait l'allèle « Carbosap-like », bien que ne faisant pas partie du groupe ciblé à l'origine : il a donc fallu définir un nouveau marqueur « Carbosap-like » (BA36). Ceci met en lumière la limite de l'étude d'un nombre limité de marqueurs, à savoir la perte de « canonicité » de ce SNP. Il est bien entendu beaucoup plus robuste d'utiliser les SNP sur l'ensemble du génome, mais le séquençage n'est pas possible pour tous les laboratoires. C'est pourquoi il est important de définir des marqueurs canoniques sur un nombre le plus important possible d'isolats. Si ce n'est pas le cas, le risque d'erreur est accru.

D'une manière générale, plusieurs classes de SNP sont définies : la classe 1 (C/T et G/A), la classe 2 (C/A et G/T), la classe 3 (C/G) et la classe 4 (A/T) (Liew *et al.*, 2004). Ces différentes classes impactent les résultats de typage moléculaire, la classe 1 étant la plus simple à discriminer alors que la classe 4 est la plus complexe. Un biais dans le choix des SNP est ainsi introduit : les SNP de classe 1 ont été privilégiés (54/60), les 6 restants étant de classe 2, car ce sont les plus faciles à discriminer dans l'optique de développer un outil de typage performant. Une autre limite dans le choix des SNP à développer en outil de typage concerne la région entourant la mutation : dessiner des amorces d'amplification ne sera pas forcément toujours possible en fonction de la région dans le génome (amorces non efficaces, stretch de T, incompatibilité au niveau du Tm des amorces dans la région ciblée,...).

2.2. Identification de nouvelles signatures moléculaires spécifiques

La première étape a été de choisir des SNP représentatifs de chaque sous-lignée et/ou sous-groupe retrouvés dans l'arbre de *B. anthracis*. Pour cela, l'étude détaillée de chaque sous-lignée a permis de définir des groupes géographiques lorsque cela était possible, ou de distinguer des branches contenant un certain nombre d'isolats. Le **Tableau 6** présente 60 nouveaux canSNP identifiés au cours de mon étude ainsi que leurs localisations (gènes,...). Tous ces SNP ont été testés et validés *in silico* contre la totalité des 292 génomes utilisés au cours de cette étude (**Tableau 6**). L'**Annexe 6** présente les positions de ces nouveaux marqueurs avec les 14 canSNP déjà décrits sur l'arbre phylogénétique de *B. anthracis*.

Ces 60 nouveaux SNP sont retrouvés à différentes localisations sur le chromosome. Après étude, il s'avère que sur 60 SNP, 30 sont responsables d'une mutation non-sens, 25 causent une mutation synonyme et 5 sont intergéniques. Les résultats sont présentés dans le **Tableau 6**. Il semble donc que pour 50 % des SNP choisis, un changement d'acide aminé soit lié à la mutation SNP. Les gènes impliqués sont divers et variés (**Tableau 6**). Il se pourrait que les mutations identifiées confèrent un avantage aux souches impliquées : adaptation aux sols, adaptation à l'hôte, virulence accrue... Afin d'approfondir l'étude fonctionnelle des mutations, des études de la biochimie des protéines issues

des changements d'acides aminés (biosynthèse, interactions, perte de fonctionnalité ?...) pourraient être menées. Il serait également intéressant de comparer l'ensemble des mutations présentes sur une même branche et d'étudier les gènes impliqués ainsi que le type de mutations associées. Il pourrait être observé que certaines branches soient liées à un type de mutation intervenant à un endroit particulier du génome ou affectant un type de gène spécifique (système de réparation de l'ADN, ...). Concernant les mutations intergéniques, étudier d'éventuelles modifications de sites de transcription est une piste à envisager. Pour 3 SNP, 3 protéines identifiées sont concernées (TetR family transcriptional regulator, acétyltransférase et hydroxylamine reductase), le reste concernant des protéines hypothétiques ou identifiées par le logiciel BioNumerics. Pour la famille des TetR, il semble que ces protéines aient un rôle dans le contrôle de gènes dont les produits sont impliqués dans les résistances multiples aux antibiotiques, dans différentes voies cataboliques, les stress osmotiques ou encore la pathogénicité (Krushkal *et al.*, 2011; Ramos *et al.*, 2005). Certaines acétyltransférases sont susceptibles d'induire une résistance aux antibiotiques, tels que la chloramphénicol acétyltransférase. Elles pourraient donc conférer un avantage sélectif aux souches possédant cet allèle. Cette piste est donc à investiguer.

2.3. Limites rencontrées lors de l'étude de la diversité génomique et perspectives d'étude sur les SNP

Dans un cadre d'analyse plus global, les SNP pourraient avoir un impact sur la virulence d'une souche. Dans le cadre de mon projet, toutes les souches ont causé la mort d'un animal ou ont provoqué la maladie (cas des héroïnomanes). Elles sont donc toutes virulentes, les souches vaccinales mises à part (cas de la Sterne). On ne peut donc pas réellement juger de la présence d'un SNP pouvant augmenter la virulence d'une souche par rapport à une autre. Cependant, comme décrit dans l'analyse bibliographique, plusieurs formes de FC sont identifiées chez les bovins, et notamment la forme suraigüe. Il serait intéressant de pouvoir isoler une souche ayant causé cette forme clinique particulière et l'étudier afin de voir si certaines mutations sont directement liées à une augmentation de la virulence. L'état général de l'animal peut également avoir un impact sur la maladie, le rendant plus sensible à la bactérie par exemple.

Une limite de mon projet est également liée au fait d'avoir limité l'étude aux SNP chromosomiques. Les principaux facteurs de virulence sont localisés sur les plasmides pXO1 et pXO2. De ce fait, rechercher des SNP pouvant accroître (ou diminuer) la virulence nécessiterait une étude complémentaire approfondie au niveau plasmidique, ce qui est tout à fait envisageable puisque les génomes plasmidiques ont été identifiés.

Des cas de souches résistantes aux antibiotiques ont été observés chez *B. anthracis* (Agren *et al.*, 2014; Athamna *et al.*, 2004; Pilo *et al.*, 2011; Serizawa *et al.*, 2010; Vogler *et al.*, 2012). L'implication de réarrangements chromosomiques dans ces résistances est une piste à investiguer dans un avenir proche (Opal, 2009). Pour cela, une approche par assemblage *de novo* sera à privilégier, permettant ainsi de visualiser d'éventuels événements d'insertion-délétion. De même pourrait-il y avoir une mutation de type SNP susceptible de causer une multirésistance ? Ces questions ouvrent des perspectives intéressantes pour l'avenir.

Tableau 6 : Liste des 60 nouveaux SNP canoniques définis dans ce travail au sein de l'espèce *B. anthracis* ainsi que leurs caractéristiques

canSNP	Position *	Type	ADN	AA	Label	Gènes encadrant le SNP si mutation intergénique	Locus Tag	Gène	Discrimination	Validation	
										<i>in silico</i>	<i>in vitro</i>
AVO/A2a'	41872	Non-sens	A->G	Y->C	methionyl-tRNA synthetase		GBAA_0036	<i>metG</i>	A.Br.Vollum, Branche A	OK	OK
AVO/A2	55501	Silencieuse	G->A		transcription-repair coupling factor		GBAA_0052	<i>mfd</i>	A.Br.Vollum, Branches ABCD	OK	OK
A08/2	72398	Non-sens	G->T	A->S	pantothenate kinase		GBAA_0065		A.Br.008/011, Branche 2	OK	En cours
BA15	126639	Silencieuse	A->G		50S ribosomal protein L24		GBAA_0121	<i>rplX</i>	B.Br.CNEVA, France, Alpes	OK	OK
A02	240050	Silencieuse	T->C		alanine racemase		GBAA_0252	<i>dal-1</i>	A.Br.02	OK	OK
BA31 (A08/4)	240337	Non-sens	C->A	A->D	alanine racemase		GBAA_0252	<i>dal-1</i>	A.Br.008/011, Branche 4 (Italie, Ombrie)	OK	OK
A05	405303	Non-sens	C->T	G->D	chitinase		GBAA_0385		A.Br.005/006	OK	OK
A02/1	513052	Silencieuse	C->T		Protein encoding sequence predicted by BioNumerics				A.Br.02, Branche 1	OK	En cours
A01	515111	Silencieuse	G->A		A/G-specific adenine glycosylase		GBAA_0522	<i>mutY</i>	A.Br.Ames, A.Br.01	OK	OK
A94	569295	Non-sens	C->T	R->H	citrate transporter		GBAA_0561		A.Br.Aust94	OK	OK
A08/3	602705	Silencieuse	T->C		glycerophosphoryl diester phosphodiesterase		GBAA_0591		A.Br.008/011, Branche 3	OK	En cours
A11/2	658847	Silencieuse	G->T		Protein encoding sequence predicted by BioNumerics				A.Br.011/009, Branche 2	OK	En cours
A08/1_part	692623	Silencieuse	T->C		multidrug resistance protein		GBAA_0674		A.Br.008/011, Branche 1 sans H65 et NCTC	OK	En cours
A08/1_2	812271	Non-sens	T->C	T->A	hypothetical protein		GBAA_0796		A.Br.008/011, Branche 1, Hongrie et Pologne	OK	En cours
BA22	820195	Non-sens	T->C	I->V	branched chain amino acid ABC transporter carrier protein		GBAA_0802		A.Br.011/009, NE France	OK	OK
BA35	910017	Non-sens	A->G	K->E	enoyl-CoA hydratase		GBAA_0894		A.Br.011/009, Italie, Sardaigne	OK	OK
A08/3_1	945588	Non-sens	T->C	V->A	replicative DNA helicase		GBAA_0934		A.Br.008/011, Branche 3, Ethiopie et NL	OK	En cours
A08/1_3	970810	Intergénique	C->T			2 protéines hypothétiques			A.Br.008/011, Tsiankovskii, Cvac02 et Belgique	OK	En cours
BA32 (A08/2_2)	1037495	Intergénique	G->A			protéine hypothétique et TetR family transcriptional regulator			A.Br.008/011, Italie, Sicile	OK	OK
BA30	1074648	Silencieuse	G->A		S-layer protein		GBAA_1093		B.Br.CNEVA, Italie	OK	OK

A08/1_2_Hun	1145595	Non-sens	C->T	T->M	hypothetical protein	GBAA_1180		A.Br.008/011, Branche 1, Hongrie	OK	En cours
A08/3_2	1314546	Silencieuse	G->A		AsnC family transcriptional regulator	GBAA_1380		A.Br.008/011, DK, Turquie et Héroïne	OK	En cours
B2/Pol_Swi1	1357067	Silencieuse	A->G		glycerate dehydrogenase	GBAA_1434		B.Br.CNEVA, Pologne et Suisse Branche 2	OK	En cours
A11/6	1391599	Silencieuse	C->T		phosphate transporter family protein	GBAA_1471		A.Br.011/009, Branche 6	OK	OK
A01/5	1412473	Non-sens	A->G	I->V	spore maturation protein	GBAA_1492		A.Br.01, Branche 5 (A0389)	OK	En cours
AVO/A1e'	1443043	Non-sens	C->T	R->C	NAD(P)H-dependent glycerol-3-phosphate dehydrogenase	GBAA_1526	<i>gpsA</i>	A.Br.Vollum, Branche E	OK	OK
A08/1	1669436	Silencieuse	C->T		hypothetical protein	GBAA_1785		A.Br.008/011, Branche 1	OK	En cours
A01/3	1828921	Non-sens	T->C	IS->TP	cytochrome d ubiquinol oxidase subunit I	GBAA_1943	<i>cydA-1</i>	A.Br.01, Branche 3	OK	En cours
B2/Swi1	1942496	Non-sens	C->T	A->T	hypothetical protein	GBAA_2073		B.Br.CNEVA, Suisse Branche 1	OK	En cours
A11/3	1975689	Non-sens	G->A	H->Y	membrane-bound protease	GBAA_2120		A.Br.011/009, Branche 3	OK	OK
AVO/A2c	2041995	Non-sens	T->C	L->S	hypothetical protein	GBAA_2188		A.Br.Vollum, Branche C	OK	OK
A01/4	2261440	Silencieuse	C->T		hypothetical protein	GBAA_2427		A.Br.01, Branche 4 (Belgique)	OK	En cours
A08/2_Bul	2268287	Non-sens	A->G	H->R	hypothetical protein	GBAA_2437		A.Br.008/011, Bulgarie	OK	En cours
BA18	2390832	Silencieuse	T->C		neutral protease	GBAA_2568		B.Br.CNEVA, France, SL	OK	OK
B1/Bel	2490460	Silencieuse	C->T		GntR family transcriptional regulator	GBAA_2671		B.Br.001/002, Belgique	OK	En cours
A11/4	2571863	Non-sens	C->T	V->I	acetoin operon transcriptional activator	GBAA_2761		A.Br.011/009, Branche 4	OK	OK
BA19	2573536	Intergénique	A->G					protéine hypothétique et acétyltransférase	OK	OK
A.Br.H9401	2591768	Non-sens	T->C	Y->H	glycine betaine/L-proline ABC transporter substrate-binding protein	GBAA_2787		A.Br.H9401	OK	En cours
A08/2_1	2668654	Non-sens	T->C	I->M	solute-binding family 5 protein	GBAA_2881		A.Br.008/011, Bulgarie et Smith 1013	OK	En cours
BA33	2752660	Intergénique	G->A					2 protéines hypothétiques prédites par BioNumerics	OK	OK
A01/Ames_DK	2973500	Non-sens	T->C	H->R	lipoprotein	GBAA_3228		A.Br.Ames, Danemark	OK	En cours

BA34	3213630	Intergénique	C->A		protéine hypothétique et hydroxylamine reductase		A.Br.011/009, Italie, BC	OK	OK
B2/Swi2	3311458	Non-sens	A->G	N->S	short chain dehydrogenase	GBAA_3602	B.Br.CNEVA, Suisse Branche 2	OK	En cours
A05/AfEst	3408237	Non-sens	G->A	L->F	Oye family NADH-dependent flavin oxidoreductase	GBAA_3707	A.Br.005/006 (Afrique de l'Est)	OK	En cours
BA20	3434997	Silencieuse	A->C		Protein encoding sequence predicted by BioNumerics	GBAA_3735	A.Br.02, France, Doubs	OK	OK
BA21	3562427	Non-sens	A->C	D->A	hypothetical protein	GBAA_3894	A.Br.011/009, Centre NE France	OK	OK
BA16	3765357	Non-sens	A->G	A->AR	hypothetical protein	GBAA_4087	B.Br.CNEVA, France Pyrénées	OK	OK
B2/BF1	3791952	Silencieuse	T->C		pdz domain-containing protein	GBAA_4136	B.Br.CNEVA, BF1	OK	En cours
BA36	3821849	Silencieuse	T->C		hypothetical protein	GBAA_4170	A.Br.011/009, Italie, Sicile	OK	En cours
A11/1	3952578	Non-sens	G->A	M->I	hypothetical protein	GBAA_4328	A.Br.011/009, Branche 1	OK	En cours
A03	3956945	Non-sens	A->G	T->A	bifunctional 3,4-dihydroxy-2-butanone 4-phosphate synthase/GTP cyclohydrolase II protein	GBAA_4333	<i>ribBA</i> A.Br.003/004	OK	OK
B2/Pol	4196256	Silencieuse	T->C		hypothetical protein	GBAA_4618	B.Br.CNEVA, Pologne	OK	En cours
A02/2	4264917	Non-sens	C->T	S->N	valyl-tRNA synthetase	GBAA_4690	<i>valS</i> A.Br.02, Branche 2	OK	En cours
AVO/A2b	4325045	Silencieuse	T->C		excinuclease ABC subunit C	GBAA_4757	<i>uvrC</i> A.Br.Vollum, Branche B	OK	OK
A01/1	4407325	Non-sens	G->A	P->L	pyruvate kinase	GBAA_4843	<i>pykA-2</i> A.Br.01, Branche 1	OK	En cours
A01/2	4481211	Non-sens	G->A	R->C	ribosomal-protein-serine acetyltransferase	GBAA_4937	A.Br.01, Branche 2	OK	En cours
BA17	4719494	Silencieuse	G->A		Protein encoding sequence predicted by BioNumerics	GBAA_5197	B.Br.CNEVA, France, MC + SL	OK	OK
A11/5	5021456	Silencieuse	T->C		hypothetical protein	GBAA_5530	A.Br.011/009, Branche 5	OK	OK
A08/1_1	5064151	Silencieuse	G->A		fructose-bisphosphate aldolase	GBAA_5580	<i>fba-2</i> A.Br.008/011, Branche 1, Belgique	OK	En cours
AVO/A2d	5217173	Silencieuse	T->C		mechanosensitive ion channel MscS	GBAA_5726	A.Br.Vollum, Branche D	OK	OK

* Position sur le génome Ames Ancestor ; NL : Pays-Bas ; DK : Danemark ; NE : Nord-est ; PLT : Pouilles-Lazio-Toscane ; NE : Nord-est ; BC : Basilicate-Campanie ; MC : Massif Central ; SL : Saône-et-Loire

2.4. Développement d'outils de typage moléculaire

Au cours de mon projet de thèse, deux méthodes de biologie moléculaire ont été développées. La première est la PCR-HRM. Basée sur des amplicons courts (inférieurs à 100 pb), cette méthode présente plusieurs avantages : elle nécessite seulement un couple d'amorces (peu coûteuse de ce fait), elle est rapide (résultat obtenu en moins de 2h) et précise (discrimination inférieure à 1°C). Elle peut ainsi être utilisée en routine au sein d'un laboratoire. Pour les laboratoires ne disposant pas d'appareil HRM, une alternative est possible par amorces MAMA : un thermocycleur en temps réel est alors suffisant, seul le design d'amorces est plus complexe (Birdsell *et al.*, 2012). A l'heure actuelle, seule une partie des marqueurs a été testée avec cet outil (cf. **Tableau 6**). Les résultats observés pour les marqueurs testés sont en parfaite corrélation avec ceux observés *in silico* (Girault *et al.*, 2014a; Girault *et al.*, 2014c).

Bien que la PCR-HRM présente plusieurs avantages, elle ne permet pas de multiplexer les tests. Au mieux, il est possible de tester deux SNP en même temps, selon leurs températures de fusion respectives. Même si l'intégralité des 60 SNP identifiés n'est pas nécessaire pour un typage moléculaire « satisfaisant » à fine échelle, il serait intéressant de pouvoir multiplexer les résultats. La technologie Luminex xTAG permet ce type de multiplexage. Des billes magnétiques sont utilisées pour fixer les fragments amplifiés par PCR. Le complexe bille-produit PCR est incubé avec de la Streptavidine-R-phycoerythrine (SA-PE), qui couplé avec de la biotine (portée par le fragment PCR) va produire de la fluorescence qui sera analysée par le système Luminex (la détection a lieu en cytomètre de flux). Ce système permet de tester 100 marqueurs simultanément. Cependant, il est nécessaire d'incorporer le test pour les deux allèles différents (allèle sauvage/allèle muté) de chaque SNP : 50 marqueurs peuvent donc être testés simultanément (théorie). Cette méthode a été validée pour 13 des 14 canSNP originaux décrits par Van Ert en 2007 dans l'article 7 de ma thèse (Thierry *et al.*, 2013) et pourrait être envisagée pour tester le nouveau panel de marqueurs identifiés.

Cependant, le coût associé à la méthode Luminex est encore assez élevé lorsque le nombre de billes est conséquent. Mon étude préconise donc l'utilisation de la PCR-HRM en première intention, ce qui explique que son développement ait été privilégié durant mon projet de thèse.

Les nouveaux marqueurs identifiés permettent de positionner tout isolat de *B. anthracis* isolé dans le monde dans un contexte global et peuvent être vérifiés soit *in silico* si un génome est disponible, soit par PCR-HRM, soit via la technologie Luminex.

2.5. Intégration de ces nouveaux outils dans les activités de recherche et de référence

Comme indiqué précédemment, l'apport de méthodes de typage à haut pouvoir résolutif de *B. anthracis* a une finalité duale : d'un côté pour la santé animale, et d'un autre pour la défense du pays. Au sein du LNR Charbon, les méthodes développées au cours de ma thèse apportent des outils capables de tracer avec précision les souches impliquées dans les différents foyers identifiés en France. Le meilleur exemple de cette application directe a eu lieu en août 2013. Un cas de fièvre charbonneuse a été déclaré en Haute-Marne (exploitation mixte laitier et allaitant de plus de 400 bovins), rappelant que cette maladie est toujours d'actualité en France (**Annexe 2**) (Girault *et al.*, 2014b). Dans un premier temps, la maladie du Charbon n'a pas été suspectée, conduisant à l'ouverture de cadavres à proximité des hangars de la ferme. Devant les signes cliniques et l'historique local (cas de charbon observés une dizaine d'années auparavant dans un champ voisin), l'orientation du diagnostic a été rapide. Durant ce laps de temps, 13 autres bovins sont morts. Une fois la maladie confirmée, une partie de l'équipe du LNR Charbon s'est déplacée pour effectuer des

prélèvements de sols afin de tenter d'isoler des spores. Différents sites (autopsie, champ, zone d'affleurement) ont été testés et des spores de *B. anthracis* ont pu être isolées en utilisant la méthode GABRI développée en Italie (Fasanella *et al.*, 2013). Une fois l'ADN extrait, l'analyse des canSNP de référence (14 marqueurs, (Marston *et al.*, 2011; Van Ert *et al.*, 2007a)) a permis de placer les souches incriminées au sein de la sous-lignée TEA (A.Br.011/009). Puis, le panel de 12 SNP A.Br.011/009 a permis d'identifier la présence de l'allèle positif pour le SNP BA22 : la souche incriminée présentait le génotype Nord-est de la France, ce qui corrèle avec sa localisation géographique. Il a ainsi été confirmé que le cas de FC observé était dû à une souche « locale ». Ce test « grandeur nature » a permis de valider l'utilisation de ce panel lors d'un cas réel. La démarche appliquée dans le cadre de ce diagnostic réalisé par le LNR est présentée en **Figure 32** avec l'illustration du cas pré-cité.

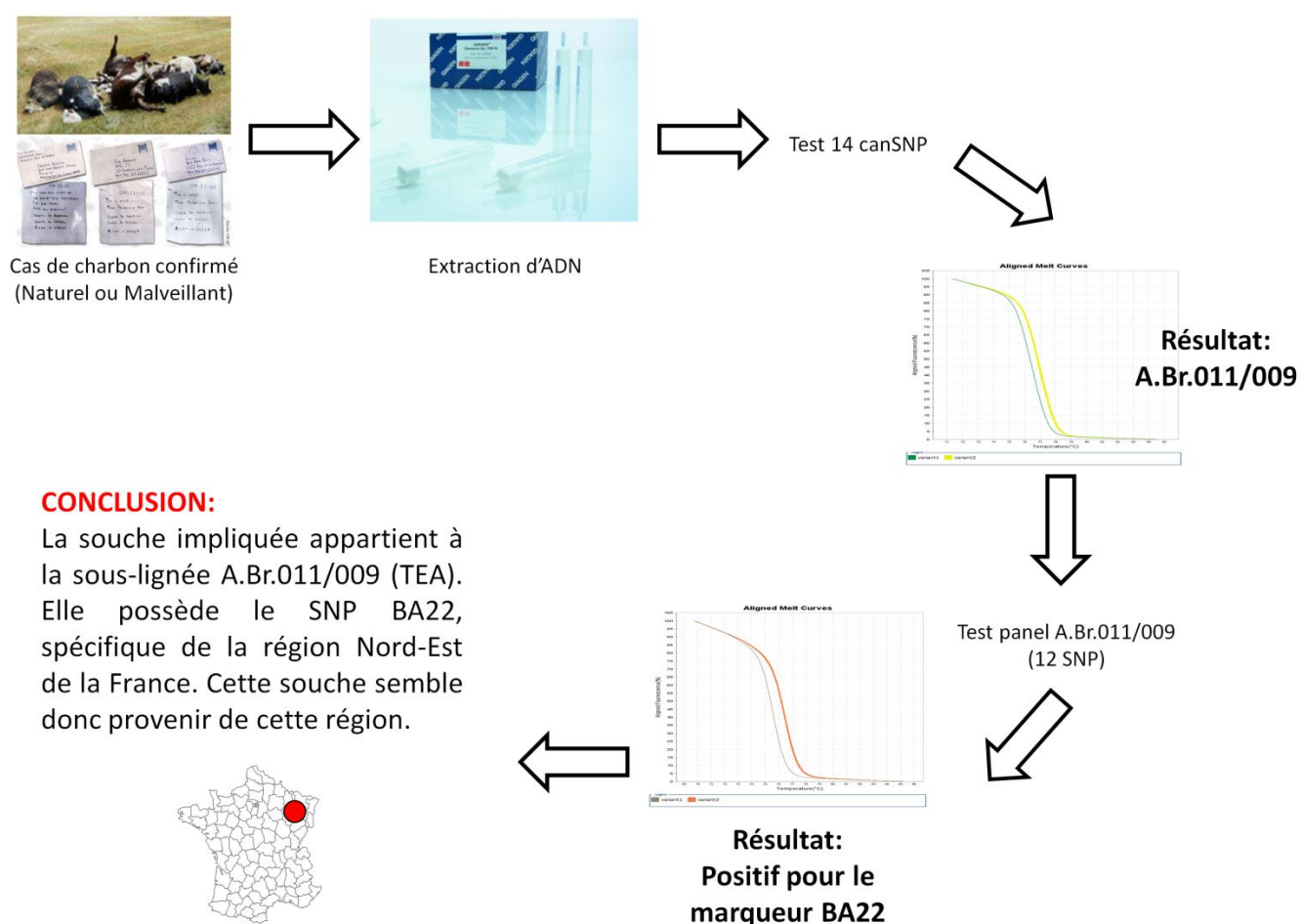


Figure 32 : Démarche suivie lors d'un cas de fièvre charbonneuse en Haute-Marne en 2013 pour identifier la souche (Girault *et al.*, 2014b).

Le panel des 60 nouveaux marqueurs SNP identifiés permet de relier une souche d'origine inconnue de *B. anthracis* à une région, un pays ou une branche de l'arbre phylogénétique obtenu au cours de cette étude (**Tableau 7**). Les panels les plus développés sont ceux comprenant le plus d'isolats étudiés au cours de mon étude, et notamment les 3 sous-lignées identifiées en France (Derzelle *et al.*, 2011a), ce qui permet d'envisager une application directe dans le cadre des missions de traçabilité et de surveillance réalisées par le laboratoire national de référence (**Figure 33**).

Tableau 7 : Synthèse des différents panels proposés et discrimination associée à chaque sous-lignée de *B. anthracis* (cf. **Annexe 6**)

Sous-lignée <i>B. anthracis</i>	Nombre de marqueurs	Panel de marqueurs proposé	Discrimination obtenue
A.Br.Ames	2	A01 et A01/Ames_DK	Groupe A01 + appartenance Danemark
A.Br.001/002	10	A01, A01/1, A01/2, A01/3, A01/4, A01/5 A02, A02/1, BA20, A02/2	Groupe A01 + appartenance à l'une des 5 branches Groupe A02 + appartenance à l'une des 2 branches + cluster du Doubs
A.Br.003/004	1	A03	Spécifique de la sous-lignée
A.Br.Aust94	1	A94	Spécifique de la sous-lignée
A.Br.005/006	2	A05 et A05/AfEst	Spécifique de la sous-lignée + cluster Afrique de l'Est
A.Br.H9401	1	A.Br.H9401	Spécifique de la sous-lignée
A.Br.Vollum	6	AVO/A2a', AVO/A2b, AVO/A2c, AVO/A2d, AVO/A1e', AVO/A2	Appartenance à l'une des 5 branches + groupe de branches
A.Br.008/011	14	A08/1, A08/1_part, A08/1_1, A08/1_2, A08/1_2_Hun, A08/1_3, A08/2, A08/2_1, A08/2_Bul, BA32, A08/3, A08/3_1, A08/3_2, BA31	Appartenance à l'une des 4 branches + sous-groupes spécifiques
A.Br.011/009	12	A11/1, BA21, BA33, A11/2, BA22, BA34, BA35, A11/3, BA36, A11/4, A11/5, A11/6	Appartenance à l'une des 6 branches + sous-groupes spécifiques
B.Br.001/002	1	B1/Bel	Cluster Belgique
B.Br.CNEVA	11	B2/Pol_Swi1, B2/Pol, B2/Sw1, BA30, B2/Swi2, B2/BF1, BA19, BA15, BA16, BA17, BA18	Appartenance à l'une des 5 branches + sous-groupes régionaux spécifiques

Cependant, cet outil n'est pas exhaustif. Afin d'enrichir les panels de SNP associés aux sous-lignées les moins représentées, il serait très intéressant d'enrichir l'arbre phylogénétique généré lors de cette étude par l'ajout de nouveaux génomes. Pour ce faire, le logigramme proposé en **Figure 33** illustre bien la complémentarité des activités de recherche et de référence : ces dernières permettent d'identifier de nouvelles souches, qui, valorisées dans le cadre de projets de recherche, aboutissent à l'identification de nouveaux marqueurs ou de nouvelles ramifications. L'outil développé pourrait alors atteindre un niveau encore plus élevé de discrimination et permettrait de rendre l'identification encore plus précise.

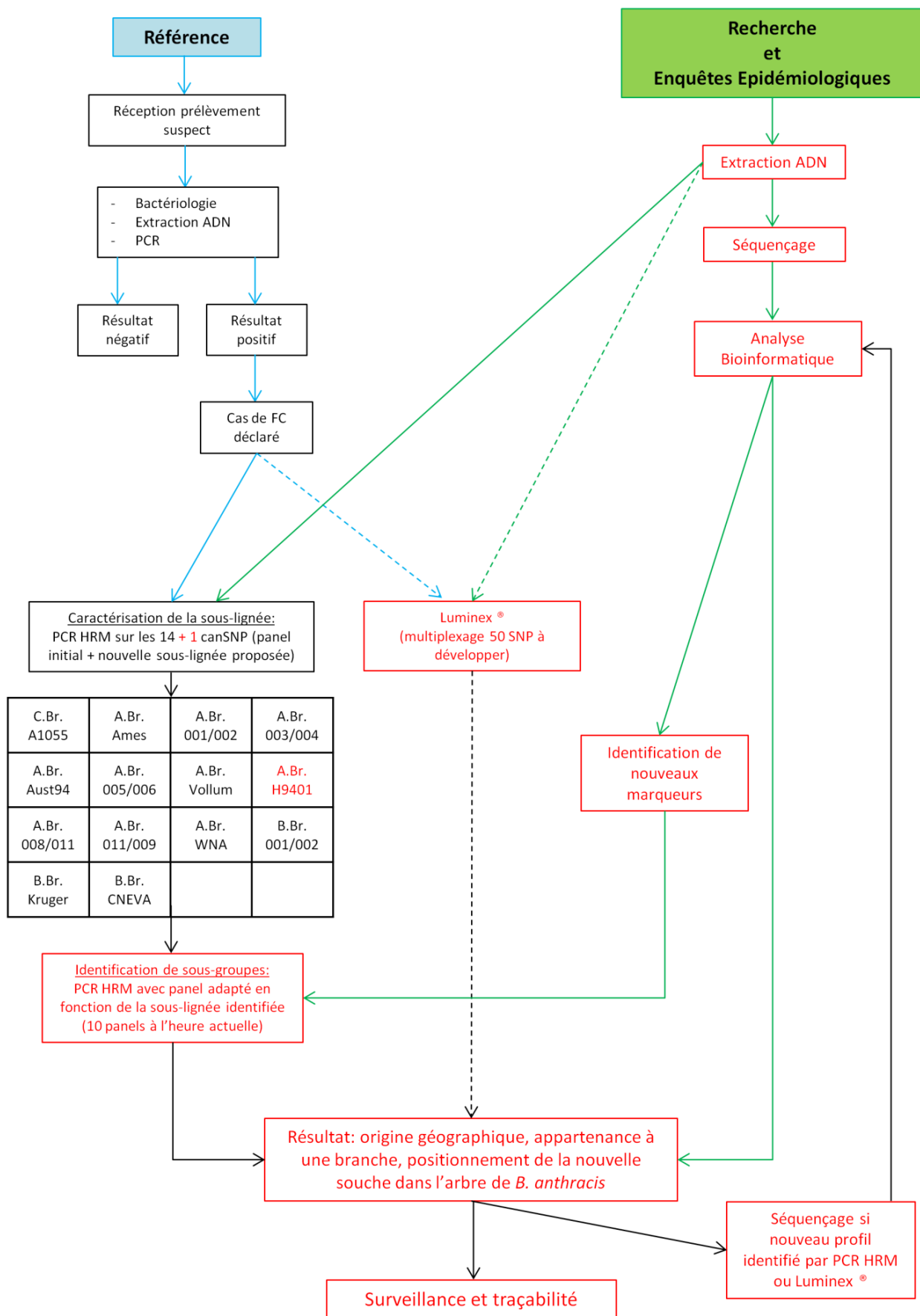


Figure 33 : Apport des outils développés durant le doctorat pour les activités de référence et de recherche (les nouveaux outils développés sont indiqués en rouge)

Conclusions et Perspectives

Le premier objectif de ce travail a consisté à étudier la diversité génétique de souches de *B. anthracis* isolées en France. Le séquençage du génome complet a été réalisé et l'identification de marqueurs polymorphes de type SNP a permis de cartographier le bacille du charbon à un niveau de résolution sans précédent en France. Afin d'étendre la vision obtenue au niveau Français, l'étude a été portée au niveau Européen. Au total, le séquençage du génome complet d'environ 250 souches (France et Europe) a permis d'étudier les relations phylogénétiques entre isolats à travers le monde. Les génomes disponibles dans les bases de données mais également dans les Sequence Read Archive (SRA) ont été incorporés à l'étude, permettant de comparer entre eux 292 génomes et de couvrir les différentes sous-lignées identifiées au sein de l'espèce *B. anthracis*.

L'étude des SNP sur génome complet a permis de positionner les souches Françaises dans un contexte global et d'étudier la diversité observée à travers le monde (première étude de cette envergure avec les SNP sur génome complet pour *B. anthracis*). Chaque sous-lignée a été caractérisée finement, permettant d'émettre plusieurs hypothèses sur les origines des souches ou des sous-lignées. Une nouvelle sous-lignée a également été proposée: A.Br.H9401.

Une étude détaillée a été réalisée à l'échelle de plusieurs pays. La population de *B. anthracis* en France a été caractérisée. Trois sous-lignées présentes à l'état naturel ont été clairement définies et 8 marqueurs SNP permettent d'identifier avec précision différents sous-groupes Français parmi la diversité mondiale de génomes.

Le même type d'étude a été effectué en Italie (7 marqueurs) et pour des cas de fièvre charbonneuse observés dans l'industrie de la laine (6 marqueurs). Ceci a permis de mieux définir les relations existant entre souches au niveau national et international. Un panel de 60 marqueurs SNP représentatifs de cette diversité a été défini, permettant de situer tout isolat d'origine inconnue dans un contexte global.

Le second objectif a été de développer une méthode de typage permettant de mettre en évidence les relations observées entre isolats au niveau génétique. Afin de tester les différents SNP identifiés préalablement, une approche rapide par PCR HRM a été mise au point, basée sur un panel de 60 SNP. Un multiplexage de ces tests a été développé sur la plateforme Luminex pour les 13 canSNP. Ainsi, en une seule réaction, 50 SNP différents (les 2 allèles, soit 100 tags) pourront désormais être testés en parallèle. Cependant, la mise au point de la méthode reste nécessaire pour l'étape avec 100 tags. La PCR HRM est ainsi privilégiée (moins chère, tout aussi précise, rapide).

Afin d'identifier l'origine d'une souche, le cheminement proposé à adopter lors d'un cas de charbon (animal ou humain) est le suivant :

- a) Cas de charbon confirmé (bactériologie, PCR)
- b) Extraction d'ADN à partir des prélèvements
- c) PCR HRM basée sur les 14 canSNP décrits par Van Ert *et al.* en 2007, et par Marston *et al.* en 2011 → **Identification de l'une des 13 sous-lignées**
- d) Selon le résultat observé en c), utilisation de l'un des sous-panels proposé par sous-lignée → **Identification d'un cluster, d'un pays voire d'une région**
- e) Traçabilité de la souche

Le séquençage du génome complet était il y a encore quelques années une approche réservée à certains laboratoires, du fait de son coût important. Avec le temps, les coûts ont diminué de manière drastique, et de plus en plus d'études sont basées sur cette technique. Méthode de typage -omique, elle permet l'accès à un volume conséquent d'informations, ouvrant des perspectives d'analyses colossales. Cependant, une difficulté supplémentaire est ajoutée pour l'étude d'un pathogène tel que *B. anthracis* faisant partie des MOT (Micro-organismes et Toxines), étant donnée la réglementation très contraignante liée à son étude.

Les polymorphismes de séquence de type SNP sont des marqueurs stables permettant de définir avec précision des groupes phylogénétiques au sein d'une espèce bactérienne. Pour *B. anthracis*, un set de 14 canSNP a permis de définir 13 sous-lignées. Cependant, la diversité au sein de ces sous-lignées reste importante, et l'utilisation de ces 14 marqueurs ne permet pas de remonter à l'origine d'une souche. L'étude présentée ici a montré que l'utilisation des SNP sur l'ensemble du génome permet l'obtention d'une image fidèle de la structure d'une population bactérienne à l'instant t (t=séquençage du génome). En identifiant des marqueurs positionnés à des endroits clés de cette structure, il est possible de tracer la provenance ou l'appartenance d'un génome à un cluster. Pour travailler sur un pathogène tel que *B. anthracis*, arme biologique potentielle, l'utilisation d'un tel outil est une nécessité : identifier clairement la souche ou son appartenance permet d'élucider une enquête (cas du bioterrorisme). Les marqueurs identifiés au cours de cette étude ont permis d'enrichir la connaissance de ce pathogène et serviront à la caractérisation des souches à travers le monde.

Les outils permettant de tester ces marqueurs sont utilisables dans tout laboratoire. Ils constituent des outils simples à mettre en place (la méthode HRM étant à privilégier en première intention, par comparaison à l'approche Luminex qui reste malgré tout onéreuse) pour les laboratoires ne souhaitant pas séquencer leurs souches et voulant tester leur collection ou les cas de charbon détectés. Concrètement, au sein du LNR Charbon, les méthodes développées permettent désormais de réaliser une traçabilité des souches impliquées dans les foyers de FC en France.

Les données sur la bactérie seront vraisemblablement enrichies davantage dans les années à venir, le nombre de génomes déposés dans les bases de données augmentant d'année en année, notamment par cette étude. Ainsi, certains groupes devraient s'agrandir et de nouvelles branches pourraient apparaître. Dans un futur proche, la méthode de typage sera certainement la même pour la plupart des laboratoires : le séquençage du génome. De ce fait, les marqueurs identifiés dans cette étude pourront (et peuvent déjà) être comparés. L'existence d'un génome de référence sur lequel ont été ciblées les positions des mutations permet la réalisation de comparaison inter-laboratoires.

Dans une approche plus fondamentale, l'analyse des données brutes de séquençage est également un sujet d'étude pour le futur. Etudier différentes caractéristiques telles que la résistance aux antibiotiques, la capacité hémolytique, la virulence et bien d'autres sont des pistes à privilégier. Une souche isolée dans les Alpes Françaises présente une résistance à la pénicilline, alors qu'une souche Polonaise présente une capacité hémolytique. Connaître les mécanismes impliqués dans ces phénomènes pourrait apporter de précieuses informations sur la bactérie.

La notion d'épigénome est également intéressante à étudier. Certaines espèces sont plus résistantes que d'autres à la FC. La forme suraigüe de la maladie est observée chez les bovins : il serait intéressant d'étudier une souche ayant causé cette forme de FC et d'identifier des marqueurs spécifiques (pourquoi la maladie est aussi rapide ? faiblesse de l'hôte ? virulence accrue de la souche ?). De plus, mener des études sur l'adaptation aux sols des différentes souches est également une piste à envisager. Cela pourrait permettre d'expliquer l'implantation préférentielle de sous-lignées telles que B.Br.CNEVA dans les régions montagneuses.

L'étude présentée ici avait pour but d'étudier la diversité de *B. anthracis* à différentes échelles (France, Europe, Monde) et de développer un outil de typage permettant de tester cette diversité. L'arbre phylogénétique de 292 génomes complets et le panel de 60 nouveaux marqueurs à tester *in silico*, en PCR HRM ou en Luminex sont des résultats présentant des retombées directes pour les activités de référence et de recherche. En outre, la connaissance de ce pathogène zoonotique présente des retombées à la fois pour le domaine vétérinaire et de la santé publique, mais également pour le domaine de la Bio-défense puisque l'identification et la caractérisation des souches est désormais facilitée.

REFERENCES

- Adair, D.M., Worsham, P.L., Hill, K.K., Klevytska, A.M., Jackson, P.J., Friedlander, A.M., Keim, P., 2000. Diversity in a Variable-Number Tandem Repeat from *Yersinia pestis*. *Journal of clinical microbiology* 38, 1516-1519.
- Afanasev, M., Kravets, E., Kugorjapova, Z., Balakhonov, S. 2013. The first molecular snapshot of the *Bacillus anthracis* strains circulating in Siberia and Far East. (Medical Biodefense Conference 2013, Munich, Germany).
- Agren, J., Finn, M., Bengtsson, B., Segerman, B., 2014. Microevolution during an Anthrax outbreak leading to clinical heterogeneity and penicillin resistance. *PLoS ONE* 9.
- Aikembayev, A.M., Lukhnova, L., Temiraliyeva, G., Meka-Mechenko, T., Pazylov, Y., Zakaryan, S., Denissov, G., Easterday, W.R., Van Ert, M.N., Keim, P., Francesconi, S.C., Blackburn, J.K., Hugh-Jones, M., Hadfield, T., 2010. Historical distribution and molecular diversity of *Bacillus anthracis*, Kazakhstan. *Emerg Infect Dis* 16, 789-796.
- Antwerpen, M., Ilin, D., Georgieva, E., Meyer, H., Savov, E., Frangoulidis, D., 2011. MLVA- and SNP analysis identified an unique genetic cluster in Bulgarian *Bacillus anthracis* strains. *European Journal of Clinical Microbiology & Infectious Diseases* 30, 923-930.
- Antwerpen, M., Proença, D.N., Rückert, C., Licht, K., Kalinowski, J., Hanczaruk, M., Tiemann, C., Grass, G., 2012. Draft Genome Sequence of *Bacillus anthracis* BF-1, Isolated from Bavarian Cattle. *J of Bacteriology* 194, 6360-6361.
- Antwerpen, M.H., Zimmermann, P., Bewley, K., Frangoulidis, D., Meyer, H., 2008. Real-time PCR system targeting a chromosomal marker specific for *Bacillus anthracis*. *Molecular and cellular probes* 22, 313-315.
- Athamna, A., Athamna, M., Abu-Rashed, N., Medlej, B., Bast, D.J., Rubinstein, E., 2004. Selection of *Bacillus anthracis* isolates resistant to antibiotics. *The Journal of antimicrobial chemotherapy* 54, 424-428.
- Beattie, S.H., Williams, A.G., 1999. Detection of toxigenic strains of *Bacillus cereus* and other *Bacillus* spp. with an improved cytotoxicity assay. *Letters in applied microbiology* 28, 221-225.
- Beecher, D.J., 2006. Forensic application of microbiological culture analysis to identify mail intentionally contaminated with *Bacillus anthracis* spores. *Applied and environmental microbiology* 72, 5304-5310.
- Berger, T., Kassirer, M., Aran, A.A., 2014. Injectional anthrax - new presentation of an old disease. *Euro Surveill* 19.
- Beyer, W., Bellan, S., Eberle, G., Ganz, H.H., Getz, W.M., Haumacher, R., Hilss, K.A., Kilian, W., Lazak, J., Turner, W.C., Turnbull, P.C., 2012. Distribution and molecular evolution of *bacillus anthracis* genotypes in Namibia. *PLoS neglected tropical diseases* 6, e1534.
- Birdsell, D.N., Antwerpen, M., Keim, P., Hanczaruk, M., Foster, J.T., Sahl, J.W., Wagner, D.M., Grass, G., 2013. Draft Genome Sequences of Two Bulgarian *Bacillus anthracis* Strains. *Genome announcements* 1, e0015213.
- Birdsell, D.N., Pearson, T., Price, E.P., Hornstra, H.M., Nera, R.D., Stone, N., Gruendike, J., Kaufman, E.L., Pettus, A.H., Hurbon, A.N., Buchhagen, J.L., Harms, N.J., Chanturia, G., Gyuranecz, M., Wagner, D.M., Keim, P.S., 2012. Melt analysis of mismatch amplification mutation assays (Melt-MAMA): a functional study of a cost-effective SNP genotyping assay in bacterial models. *PLoS ONE* 7, e32866.
- Blouin, Y., Hauck, Y., Soler, C., Fabre, M., Vong, R., Dehan, C., Cazajous, G., Massoure, P.L., Kraemer, P., Jenkins, A., Garnotel, E., Pourcel, C., Vergnaud, G., 2012. Significance of the identification in the Horn of Africa of an exceptionally deep branching *Mycobacterium tuberculosis* clade. *PLoS ONE* 7, e52841.
- Booth, M., Donaldson, L., Cui, X., Sun, J., Cole, S., Dailsey, S., Hart, A., Johns, N., McConnell, P., McLennan, T., Parcell, B., Robb, H., Shippey, B., Sim, M., Wallis, C., Eichacket, P.Q., 2014. Confirmed *Bacillus anthracis* infection among persons who inject drugs, Scotland, 2009-2010. *Emerging Infectious Diseases* 20.
- Bottone, E.J., 2010. *Bacillus cereus*, a volatile human pathogen. *Clinical microbiology reviews* 23, 382-398.

- Bourgogne, A., Drysdale, M., Hilsenbeck, S.G., Peterson, S.N., Koehler, T.M., 2003. Global Effects of virulence gene regulators in a *Bacillus anthracis* strain with both virulence plasmids. *Infection and Immunity* 71, 2736-2743.
- Braack, L.E., De Vos, V., 1990. Feeding habits and flight range of blow-flies (*Chrysomya* spp.) in relation to anthrax transmission in the Kruger National Park, South Africa. *Onderstepoort J Vet Res* 57, 141-142.
- Brachman, P.S., 1980. Inhalation Anthrax. *Annals of the New York Academy of Sciences* 353, 83-93.
- Brachman, P.S., Kaufman, A.F., Dalldorf, F.G., 1966. Industrial Inhalation Anthrax. *Bacteriological reviews* 30.
- Bradaric, N., Punda-Polic, V., 1992. Cutaneous anthrax due to penicillin-resistant *Bacillus anthracis* transmitted by an insect bite. *The Lancet* 340, 306-307.
- Bravo, A., Gill, S.S., Soberon, M., 2007. Mode of action of *Bacillus thuringiensis* Cry and Cyt toxins and their potential for insect control. *Toxicon* 49, 423-435.
- Brey, R.N., 2005. Molecular basis for improved anthrax vaccines. *Advanced drug delivery reviews* 57, 1266-1292.
- Brodzik, A.K., Francoeur, J., 2011. A new approach to in silico SNP detection and some new SNPs in the *Bacillus anthracis* genome. *BMC research notes* 4, 114.
- Brossier, F., Mock, M., 2001. Toxins of *Bacillus anthracis*. *Toxicon* 39, 1747-1755.
- Bush, L.M., Abrams, B.H., Beall, A., Johnson, C.C., 2001. Index case of fatal inhalational anthrax due to bioterrorism in the United States. *N Engl J Med* 345, 1607-1610.
- Calavas, D., Sala, C., Vaissaire, J., Condé, J., Thien-Aubert, H., Hessemann, M., Woronoff-Rehn, N., 2008. Retour d'expérience sur un épisode de fièvre charbonneuse chez les bovins dans le Doubs au cours de l'été 2008. *Bulletin épidémiologique* 32, 1-6.
- Candela, T., Mock, M., Fouet, A., 2005. CapE, a 47-amino-acid peptide, is necessary for *Bacillus anthracis* polyglutamate capsule synthesis. *Journal of bacteriology* 187, 7765-7772.
- Carter, K.C., 1988. The Koch-Pasteur dispute on establishing the cause of anthrax. *Bull Hist Med* 62, 42-57.
- Carter, T., 2004. The dissemination of anthrax from imported wool: Kidderminster 1900–14. *Occup Environ Med* 61, 103-107.
- Chen, C. 2011. De novo assembly of high-throughput short read sequences. In NECC Third skate Genome Annotation Workshop (Center for Bioinformatics and Computational Biology (CBCB), University of Delaware).
- Chen, P.E., Willner, K.M., Butani, A., Dorsey, S., George, M., Stewart, A., Lentz, S.M., Cook, C.E., Akmal, A., Price, L.B., Keim, P.S., Meteczun, A., Brahmbhatt, T.N., Bishop-Lilly, K.A., Zwick, M.E., Read, T.D., Sozhamannan, S., 2010. Rapid Identification of Genetic Modifications in *Bacillus anthracis* Using Whole Genome Draft Sequences Generated by 454 Pyrosequencing. *PLoS ONE* 5.
- Cherkasskiy, B.L., 1999. A national register of historic and contemporary anthrax foci. *Journal of applied microbiology* 89, 192-195.
- Chun, J.H., Hong, K.J., Cha, S.H., Cho, M.H., Lee, K.J., Jeong, D.H., Yoo, C.K., Rhie, G.E., 2012. Complete genome sequence of *Bacillus anthracis* H9401, an isolate from a Korean patient with anthrax. *Journal of bacteriology* 194, 4116-4117.
- Cohen-Gihon, I., Israeli, O., Beth-Din, A., Levy, H., Cohen, O., Shafferman, A., Zvi, A., Chitlarua, T., 2014. Whole-Genome Sequencing of the Nonproteolytic *Bacillus anthracis* V770-NP1-R Strain Reveals Multiple Mutations in Peptidase Loci. *Genome announcements* 2.
- Comas, I., Chakravarti, J., Small, P.M., Galagan, J., Niemann, S., Kremer, K., Ernst, J.D., Gagneux, S., 2010. Human T cell epitopes of *Mycobacterium tuberculosis* are evolutionarily hyperconserved. *Nature genetics* 42, 498-503.
- Cote, C.K., Welkos, S.L., Bozue, J., 2011. Key aspects of the molecular and cellular basis of inhalational anthrax. *Microbes and infection / Institut Pasteur* 13, 1146-1155.
- Cummings, C.A., Bormann Chung, C.A., Fang, R., Barker, M., Brzoska, P., Williamson, P.C., Beaudry, J., Matthews, M., Schupp, J., Wagner, D.M., Birdsell, D., Vogler, A.J., Furtado, M.R., Keim, P.,

- Budowle, B., 2010. Accurate, rapid and high-throughput detection of strain-specific polymorphisms in *Bacillus anthracis* and *Yersinia pestis* by next-generation sequencing. *Investigative genetics* 1, 5.
- Cybulski, R.J., Jr., Sanz, P., O'Brien, A.D., 2009. Anthrax vaccination strategies. *Molecular aspects of medicine* 30, 490-502.
- Derzelle, S. 2014. Single nucleotide polymorphisms discrimination using high-resolution melting analysis for the genotyping of *Bacillus anthracis*, In: Inácio, M.C.a.J. (Ed.) *Molecular Diagnosis in Veterinary Laboratory Practice* Humana Press, New York.
- Derzelle, S., Laroche, S., Le Fleche, P., Hauck, Y., Thierry, S., Vergnaud, G., Madani, N., 2011a. Characterization of genetic diversity of *Bacillus anthracis* in France by using high-resolution melting assays and multilocus variable-number tandem-repeat analysis. *Journal of clinical microbiology* 49, 4286-4292.
- Derzelle, S., Mendy, C., Laroche, S., Madani, N., 2011b. Use of high-resolution melting and melting temperature-shift assays for specific detection and identification of *Bacillus anthracis* based on single nucleotide discrimination. *Journal of microbiological methods* 87, 195-201.
- Derzelle, S., Thierry, S., 2013. Genetic diversity of *Bacillus anthracis* in Europe: genotyping methods in forensic and epidemiologic investigations. *Biosecurity and bioterrorism : biodefense strategy, practice, and science* 11 Suppl 1, S166-176.
- Dirckx, J.H., 1981. Virgil on Anthrax. *The American Journal of Dermatopathology* 3, 191-196.
- Dragon, D.C., Rennie, R.P., 1995. The ecology of anthrax spores: tough but not invincible. *Can Vet* 36.
- Drobniewski, F.A., 1993. *Bacillus cereus* and Related Species. *Clinical microbiology reviews* 6, 324-338.
- Durmaz, R., Doganay, M., Sahin, M., Percin, D., Karahocagil, M.K., Kayabas, U., Otlu, B., Karagoz, A., Buyuk, F., Celebi, O., Ozturk, Z., Ertek, M., 2012. Molecular epidemiology of the *Bacillus anthracis* isolates collected throughout Turkey from 1983 to 2011. *European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology* 31, 2783-2790.
- Easterday, W.R., Van Ert, M.N., Zanecki, S., Keim, P., 2005. Specific detection of *Bacillus anthracis* using a TaqMan mismatch amplification mutation assay. *Biotechniques* 38, 731-735.
- Ehling-Schulz, M., Fricker, M., Grallert, H., Rieck, P., Wagner, M., Scherer, S., 2006. Cereulide synthetase gene cluster from emetic *Bacillus cereus*: structure and location on a mega virulence plasmid related to *Bacillus anthracis* toxin plasmid pXO1. *BMC microbiology* 6, 20.
- Ekman, J., Tsitko, I., Weber, A., Nielsen-LeRoux, C., Lereclus, D., Salkinoja-Salonen, M., 2009. Transfer of *Bacillus cereus* spores from packaging paper into food. *Journal of Food Protection* 72, 2236-2242.
- Erali, M., Voelkerding, K.V., Wittwer, C.T., 2008. High resolution melting applications for clinical laboratory medicine. *Experimental and molecular pathology* 85, 50-58.
- Fasanella, A., Di Taranto, P., Garofolo, G., Colao, V., Marino, L., Buonavoglia, D., Pedarra, C., Adone, R., Hugh-Jones, M., 2013. Ground Anthrax *Bacillus* Refined Isolation (GABRI) method for analyzing environmental samples with low levels of *Bacillus anthracis* contamination. *BMC microbiology* 13.
- Fasanella, A., Garofolo, G., Galante, D., Quaranta, V., Palazzo, L., Lista, F., Adone, R., Hugh-Jones, M., 2010. Sever anthrax outbreaks in Italy in 2004: considerations on factors involved in the spread of infection. *New Microbiologica* 33, 83-86.
- Fasanella, A., Palazzo, L., Petrella, A., Quaranta, V., Romanelli, B., Garofolo, G., 2007. Anthrax in Red Deer (*Cervus elaphus*), Italy. *Emerging Infectious Diseases* 13, 1118-1119.
- Fasanella, A., Van Ert, M., Altamura, S.A., Garofolo, G., Buonavoglia, C., Leori, G., Huynh, L., Zanecki, S., Keim, P., 2005. Molecular diversity of *Bacillus anthracis* in Italy. *Journal of clinical microbiology* 43, 3398-3401.
- FBI. Amerithrax or Anthrax Investigation.

- Ford, C., Yusim, K., Ioerger, T., Feng, S., Chase, M., Greene, M., Korber, B., Fortune, S., 2012. Mycobacterium tuberculosis--heterogeneity revealed through whole genome sequencing. *Tuberculosis (Edinb)* 92, 194-201.
- Fortini, D., Ciammaruconi, A., De Santis, R., Fasanella, A., Battisti, A., D'Amelio, R., Lista, F., Cassone, A., Carattoli, A., 2007. Optimization of High-Resolution Melting Analysis for low-cost and rapid screening of allelic variants of *Bacillus anthracis* by Multiple-Locus Variable-Number Tandem Repeat Analysis. *Clinical Chemistry* 53, 1377-1380.
- Fouet, A., Smith, K.L., Keys, C., Vaissaire, J., Le Doujet, C., Lévy, M., Mock, M., Keim, P., 2002. Diversity among french *Bacillus anthracis* Isolates. *Journal of clinical microbiology* 40, 4732-4734.
- Friedlander, A.M., Welkos, S.L., Pitt, M.L.M., Ezzell, J.W., Worsham, P.L., Rose, K.J., Ivins, B.E., Lowe, J.R., Howe, G.B., Mikesell, P., Lawrence, W.B., 1993. Postexposure Prophylaxis against Experimental Inhalation Anthrax *J Infect Dis* 167, 1239-1243.
- Frothingham, R., Meeker-O'Connell, W.A., 1998. Genetic diversity in the Mycobacterium tuberculosis complex based on variable numbers of tandem DNA repeats. *Microbiology* 144, 1189-1196.
- Furelaud, G., Esnault, Y. Le séquençage d'un génome: comment ça marche?
- Gardy, J.L., Johnston, J.C., Ho Sui, S.J., Cook, V.J., Shah, L., Brodtkin, E., Rempel, S., Moore, R., Zhao, Y., Holt, R., Varhol, R., Birol, I., Lem, M., Sharma, M.K., Elwood, K., Jones, S.J.M., Brinkman, F.S.L., Brunham, R.C., Tang, P., 2011. Whole-Genome Sequencing and Social-Network analysis of a tuberculosis outbreak. *N Engl J Med* 364.
- Garofolo, G., Ciammaruconi, A., Fasanella, A., Scasciamacchia, S., Adone, R., Pittiglio, V., Lista, F., 2010. SNR analysis: molecular investigation of an anthrax epidemic. *BMC veterinary research* 6, 11.
- Garofolo, G., Serrecchia, L., Corro, M., Fasanella, A., 2011. Anthrax phylogenetic structure in Northern Italy. *BMC research notes* 4, 273.
- Ghelardi, E., Celandroni, F., Salvetti, S., Fiscarelli, E., Senesi, S., 2007. *Bacillus thuringiensis* pulmonary infection: critical role for bacterial membrane-damaging toxins and host neutrophils. *Microbes and infection / Institut Pasteur* 9, 591-598.
- Gibrat, J.-F. 2014. Les méthodes de séquençage. Les algorithmes d'assemblage et d'alignement. In *Module 8 NGS (INRA, Jouy-en-Josas)*.
- Gierczyński, R., Kafuzewski, S., Rakin, A., Jagielski, M., Zasada, A., Jakubczak, A., Borkowska-Opacka, B., Rastawicki, W., 2004. Intriguing diversity of *Bacillus anthracis* in eastern Poland--the molecular echoes of the past outbreaks. *FEMS Microbiol Lett* 239, 235-240.
- Gimenez, A.P., 2006. Dynamique de l'infection cutanée par *B. anthracis* chez la souris et rôle bactéricide de la phospholipase A2 type IIA. Université Paris 7,
- Girault, G., Blouin, Y., Vergnaud, G., Derzelle, S., 2014a. High-throughput sequencing of *Bacillus anthracis* in France: investigating genome diversity and population structure using whole-genome SNP discovery. *BMC genomics* 15.
- Girault, G., Kuhse, S., Vin, H., Marchal-Nguyen, R., Laroucau, K., Derzelle, S., 2014b. Fièvre charbonneuse en 2013 : un foyer en Haute-Marne rappelle que cette zoonose est toujours d'actualité en France. *Bulletin épidémiologique, Santé animale et alimentation* 63, 6-9.
- Girault, G., Thierry, S., Cherchame, E., Derzelle, S., 2014c. Application of High-Throughput Sequencing: Discovery of Informative SNPs to Subtype *Bacillus anthracis*. *Adv Biosci Biotechnol* 5, 669-677.
- Gleiser, C.A., 1967. Pathology of anthrax infection in animal hosts. *Fed Proc.* 26, 1518-1521.
- Goodman, J.W., Nitecki, D.E., 1966. Immunochemical studies on the poly-gamma-D-glutamyl capsule of *Bacillus anthracis*. I. Characterization of the polypeptide and of the specificity of its reaction with rabbit antisera. *Biochemistry* 5, 657-665.
- Grada, A., Weinbrecht, K., 2013. Next-generation sequencing: methodology and application. *The Journal of investigative dermatology* 133, e11.
- Grunow, R., Klee, S.R., Beyer, W., George, M., Grunow, D., Barduhn, A., Klar, S., Jacob, D., Elschner, M., Sandven, P., Kjerulf, A., Jensen, J.S., Cal, W., Zimmermann, R., Schaade, L., 2013. Anthrax

- among heroin users in Europe possibly caused by same *Bacillus anthracis* strain since 2000. Euro Surveill 18.
- Guinebreteire, M.H., Auger, S., Galleron, N., Contzen, M., De Sarrau, B., De Buyser, M.L., Lamberet, G., Fagerlund, A., Granum, P.E., Lereclus, D., De Vos, P., Nguyen-The, C., Sorokin, A., 2013. *Bacillus cytotoxicus* sp. nov. is a novel thermotolerant species of the *Bacillus cereus* Group occasionally associated with food poisoning. International journal of systematic and evolutionary microbiology 63, 31-40.
- Hanczaruk, M., Reischl, U., Holzmann, T., Frangoulidis, D., Wagner, D.M., Keim, P.S., Antwerpen, M.H., Meyer, H., Grass, G., 2014. Injectional anthrax in Heroin users Europe 2000-2012. Emerging Infectious Diseases 20, 322-323.
- Harrington, R., Ondov, B.D., Radune, D., Friss, M.B., Klubnik, J., Diviak, L., Hnath, J., Cendrowski, S.R., Blank, T.E., Karaolis, D., Friedlander, A.M., Burans, J.P., Rosovitz, M.J., Treangen, T., Phillippy, A.M., Bergman, N.H., 2013. Genome Sequence of the Attenuated Carbosap Vaccine Strain of *Bacillus anthracis*. Genome announcements 1, 1-2.
- Heine, H.S., Bassett, J., Miller, L., Hartings, J.M., Ivins, B.E., Pitt, M.L., Fritz, D., Norris, S.L., Byrne, W.R., 2007. Determination of antibiotic efficacy against *Bacillus anthracis* in a mouse aerosol challenge model. Antimicrobial agents and chemotherapy 51, 1373-1379.
- Helgason, E., Okstad, O.A., Caugant, D.A., Johansen, H.A., Fouet, A., Mock, M., Hegna, I., Kolsto, A.-B., 2000. *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis* — One Species on the Basis of Genetic Evidence. Applied and environmental microbiology 66, 2627-2630.
- Henderson, I., Duggleby, C.J., Turnbull, P.C.B., 1994. Differentiation of *Bacillus anthracis* from other *Bacillus cereus* group bacteria with the PCR. int J of Systematic Bacteriology 44, 99-105.
- Hendriksen, N.B., Hansen, B.M., Johansen, J.E., 2006. Occurrence and pathogenic potential of *Bacillus cereus* group bacteria in a sandy loam. Antonie van Leeuwenhoek 89, 239-249.
- Hernandez, E., Ramisse, F., Ducoureaux, J.-P., Cruel, T., Cavallo, J.-D., 1998. *Bacillus thuringiensis* subsp. *konkukian* (Serotype H34) Superinfection: Case Report and Experimental Evidence of Pathogenicity in Immunosuppressed Mice. Journal of clinical microbiology 36, 2138-2139.
- Hoffmaster, A.R., Meyer, R.F., Bowen, M.P., Marston, C.K., Weyant, R.S., Barnett, G.A., Sejvar, J.J., Jernigan, J.A., Perkins, B.A., Popovic, T., 2002. Evaluation and Validation of a Real-Time Polymerase Chain Reaction Assay for Rapid Identification of *Bacillus anthracis*. Emerging Infectious Diseases 8, 1178-1182.
- Hoton, F., Andrup, L., Swiecicka, I., Mahillon, J., 2005. The cereulide genetic determinants of emetic *Bacillus cereus* are plasmid-borne. Microbiology 151, 2121-2124.
- Hugh-Jones, M., 1999. 1996-97 global anthrax report. Journal of applied microbiology 1999, 189-191.
- Hugh-Jones, M.E., De Vos, V., 2002. Anthrax and wildlife. Rev. sci. tech. Off. int. Epiz. 21, 359-383.
- Hurtle, W., Bode, E., Kulesh, D.A., Kaplan, R.S., Garrison, J., Bridge, D., House, M., Frye, M.S., Loveless, B., Nordwood, D., 2004. Detection of the *Bacillus anthracis gyrA* Gene by using a minor groove binder probe. Journal of clinical microbiology 42, 179-185.
- Imperiale, M.J., Casadevall, A., 2011. Bioterrorism: lessons learned since the anthrax mailings. mBio 2, e00232-00211.
- INRS, 2001. Le charbon, une maladie professionnelle "oubliée" devenue menace bioterroriste. Documents pour le médecin du travail 88.
- INVS, 2005. Recommandations pour la surveillance et la lutte contre le charbon animal et humain. Institut de Veille Sanitaire Département des maladies infectieuses.
- Ivanova, N., Sorokin, A., Anderson, I., Galleron, N., Candelon, B., Kapatal, V., Bhattacharyya, A., Reznik, G., Mikhailova, N., Lapidus, A., Chu, L., Mazur, M., Goltsman, E., Larsen, N., D'Souza, M., Walunas, T., Grechkin, Y., Pusch, G., Haselkorn, R., Fonstein, M., Ehrlich, S.D., Overbeek, R., Kyrpides, N., 2003. Genome sequence of *Bacillus cereus* and comparative analysis with *Bacillus anthracis*. Nature 423, 87-91.
- Janse, I., Hamidjaja, R.A., Bok, J.M., van Rotterdam, B.J., 2010. Reliable detection of *Bacillus anthracis*, *Francisella tularensis* and *Yersinia pestis* by using multiplex qPCR including internal controls for nucleic acid extraction and amplification. BMC microbiology 10, 314.

- Jernigan, J.A., Stephens, D.S., Ashford, D.A., Omenaca, C., Topiel, M.S., Galbraith, M., Tapper, M., Fisk, T.L., Zaki, S., Popovic, T., Meyer, R.F., Quinn, C.P., Harper, S.A., Fridkin, S.K., Sejvar, J.J., Shepard, C.W., McConnell, M., Guarner, J., Shieh, W.-J., Malecki, J.M., Gerberding, J.L., Hughes, J.M., Perkins, B.A., Members of the Anthrax Bioterrorism Investigation Team, 2001. Bioterrorism-Related Inhalational Anthrax: The First 10 Cases Reported in the United States. *Emerging Infectious Diseases* 7, 933-944.
- Keim, P., Gruendike, J.M., Klevytska, A.M., Schupp, J.M., Challacombe, J., Okinaka, R., 2009. The genome and variation of *Bacillus anthracis*. *Molecular aspects of medicine* 30, 397-405.
- Keim, P., Kalif, A., Shupp, J., Hill, K., Travis, S.E., Richmond, K., Adair, D.M., Hugh-Jones, M., Kuske, C.R., Jackson, P., 1997. Molecular evolution and diversity in *Bacillus anthracis* as detected by Amplified Fragment Length Polymorphism Markers. *J of Bacteriology* 179, 818-824.
- Keim, P., Okinaka, R., Pearson, T. 2013. The Evolution and global dissemination of *Bacillus anthracis*. In *The International Conference on Bacillus anthracis, B. cereus, and B. thuringiensis* (Victoria, BC (Canada)).
- Keim, P., Price, L.B., Klevytska, A.M., Smith, K.L., Shupp, J.M., Okinaka, R., Jackson, P.J., Hugh-Jones, M.E., 2000. Multiple-Locus Variable-Number Tandem Repeat Analysis Reveals Genetic Relationships within *Bacillus anthracis*. *J of Bacteriology* 182, 2928-2936.
- Keim, P., Smith, K.L., Keys, C., Takahashi, H., Kurata, T., Kaufmann, A., 2001. Molecular investigation of the Aum Shinrikyo anthrax release in Kameido, Japan. *Journal of clinical microbiology* 39, 4566-4567.
- Keim, P., Van Ert, M.N., Pearson, T., Vogler, A.J., Huynh, L.Y., Wagner, D.M., 2004. Anthrax molecular epidemiology and forensics: using the appropriate marker for different evolutionary scales. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* 4, 205-213.
- Keim, P.S., Wagner, D.M., 2009. Humans and evolutionary and ecological forces shaped the phylogeography of recently emerged diseases. *Nature reviews. Microbiology* 7, 813-821.
- Kenefic, L.J., Beaudry, J., Trim, C., Daly, R., Parmar, R., Zanecki, S., Huynh, L., Van Ert, M.N., Wagner, D.M., Graham, T., Keim, P., 2008a. High resolution genotyping of *Bacillus anthracis* outbreak strains using four highly mutable single nucleotide repeat markers. *Letters in applied microbiology* 46, 600-603.
- Kenefic, L.J., Beaudry, J., Trim, C., Huynh, L., Zanecki, S., Matthews, M., Schupp, J., Van Ert, M., Keim, P., 2008b. A high resolution four-locus multiplex single nucleotide repeat (SNR) genotyping system in *Bacillus anthracis*. *Journal of microbiological methods* 73, 269-272.
- Kenefic, L.J., Pearson, T., Okinaka, R.T., Schupp, J.M., Wagner, D.M., Hoffmaster, A.R., Trim, C.B., Chung, W.K., Beaudry, J.A., Jiang, L., Gajer, P., Foster, J.T., Mead, J.I., Ravel, J., Keim, P., 2009. Pre-Columbian origins for North American anthrax. *PLoS ONE* 4, e4813.
- Khmaladze, E., Birdsell, D.N., Naumann, A.A., Hochhalter, C.B., Seymour, M.L., Nottingham, R., Beckstrom-Sternberg, S.M., Beckstrom-Sternberg, J., Nikolich, M.P., Chanturia, G., Zhgenti, E., Zakalashvili, M., Malania, L., Babuadze, G., Tsertsvadze, N., Abazashvili, N., Kekelidze, M., Tsanava, S., Imnadze, P., Ganz, H.H., Getz, W.M., Pearson, O., Gajer, P., Eppinger, M., Ravel, J., Wagner, D.M., Okinaka, R.T., Schupp, J.M., Keim, P., Pearson, T., 2014. Phylogeography of *Bacillus anthracis* in the Country of Georgia Shows Evidence of Population Structuring and Is Dissimilar to Other Regional Genotypes. *PLoS ONE* 9.
- Kim, S.K., Chung, W.-H., Kim, S.H., Jung, K.H., Kim, N., Chai, Y.G., 2014. Complete Genome Sequence of *Bacillus anthracis* HYU01, Isolated from Soil Samples in the Korean Peninsula. *Genome announcements* 2.
- Kissling, E., Wattiau, P., China, B., Poncin, M., Fretin, D., Pirenne, Y., Hanquet, G., 2012. *B. anthracis* in a wool-processing factory: seroprevalence and occupational risk. *Epidemiology and infection* 140, 879-886.
- Koch, R., 1876. Die Ätiologie der Milzbrand-Krankheit, begründet auf die Entwicklungsgeschichte des *Bacillus Anthracis*. *Cohns Beiträge zur Biologie der Pflanzen* 2.

- Koehler, T.M., 2009. *Bacillus anthracis* physiology and genetics. *Molecular aspects of medicine* 30, 386-396.
- Kolstø, A.-B., Tourasse, N.J., Økstad, O.A., 2009. What Sets *Bacillus anthracis* Apart from Other *Bacillus* Species? *Annual Review of Microbiology* 63, 451-476.
- Kotiranta, A., Lounatmaa, K., Haapasalo, M., 2000. Epidemiology and pathogenesis of *Bacillus cereus* infections. *Microbes and infection / Institut Pasteur* 2, 189-198.
- Krushkal, J., Sontineni, S., Leang, C., Qu, Y., Adkins, R.M., Lovley, D.R., 2011. Genome diversity of the TetR family of transcriptional regulators in a metal-reducing bacterial family Geobacteraceae and other microbial species. *Omics : a journal of integrative biology* 15, 495-506.
- Kuroda, M., Serizawa, M., Okutani, A., Sekizuka, T., Banno, S., Inoue, S., 2010. Genome-wide single nucleotide polymorphism typing method for identification of *Bacillus anthracis* species and strains among *B. cereus* group species. *Journal of clinical microbiology* 48, 2821-2829.
- Le Flèche, P., Hauck, Y., Onteniente, L., Prieur, A., Denoeud, F., Ramisse, V., Sylvestre, P., Benson, G., Ramisse, F., Vergnaud, G., 2001. A tandem repeats database for bacterial genomes: application to the genotyping of *Yersinia pestis* and *Bacillus anthracis*. *BMC microbiology* 1.
- Lechner, S., Mayr, R., Francis, K.P., Prüss, B.M., Kaplan, T., Wiessner-Gunkel, E., Stewart, G.S.A.B., Scherer, S., 1998. *Bacillus weihenstephanensis* sp. nov. is a new psychrotolerant species of the *Bacillus cereus* group. *Int J of Systematic Bacteriology* 48, 1373-1382.
- Lewerin, S.S., Elvander, M., Westermarck, T., Hartzell, L.N., Norstrom, A.K., Ehlers, S., Knutsson, R., Englund, S., Andersson, A.C., Granberg, M., Backman, S., Wikstrom, P., Sandstedt, K., 2010. Anthrax outbreak in a Swedish beef cattle herd--1st case in 27 years: Case report. *Acta veterinaria Scandinavica* 52, 7.
- Li, B., Kadura, I., Fu, D.-J., Watson, D.E., 2004. Genotyping with TaqMAMA. *Genomics* 83, 311-320.
- Liew, M., Pryor, R., Palais, R., Meadows, C., Erali, M., Lyon, E., Wittwer, C., 2004. Genotyping of single-nucleotide polymorphisms by high-resolution melting of small amplicons. *Clinical chemistry* 50, 1156-1164.
- Lindeque, P.M., Turnbull, P.C., 1994. Ecology and epidemiology of anthrax in the Etosha National Park, Namibia. *Onderstepoort J Vet Res* 61, 71-83.
- Lista, F., Faggioni, G., Valjevac, S., Ciammaruconi, A., Vaissaire, J., le Doujet, C., Gorge, O., De Santis, R., Carattoli, A., Ciervo, A., Fasanella, A., Orsini, F., D'Amelio, R., Pourcel, C., Cassone, A., Vergnaud, G., 2006. Genotyping of *Bacillus anthracis* strains based on automated capillary 25-loci multiple locus variable-number tandem repeats analysis. *BMC microbiology* 6, 33.
- Madani, N., Mendy, C., Moutou, F., Garin-Bastuji, B., 2009. La fièvre charbonneuse en France. Épisodes de l'été 2009 et foyers enregistrés sur la dernière décennie (1999-2009). *Bulletin épidémiologique* 38, 1-3.
- Maho, A., Rossano, A., Hachler, H., Holzer, A., Schelling, E., Zinsstag, J., Hassane, M.H., Toguebaye, B.S., Akakpo, A.J., Van Ert, M., Keim, P., Kenefic, L., Frey, J., Perreten, V., 2006. Antibiotic susceptibility and molecular diversity of *Bacillus anthracis* strains in Chad: detection of a new phylogenetic subgroup. *Journal of clinical microbiology* 44, 3422-3425.
- Manchee, R.J., Broster, M.G., Melling, J., Henstridge, R.M., Stagg, A.J., 1981. *Bacillus anthracis* on Gruinard Island. *Nature* 294, 254-255.
- Marston, C.K., Allen, C.A., Beaudry, J., Price, E.P., Wolken, S.R., Pearson, T., Keim, P., Hoffmaster, A.R., 2011. Molecular epidemiology of anthrax cases associated with recreational use of animal hides and yarn in the United States. *PLoS ONE* 6, e28274.
- Marston, C.K., Gee, J.E., Popovic, T., Hoffmaster, A.R., 2006. Molecular approaches to identify and differentiate *Bacillus anthracis* from phenotypically similar *Bacillus* species isolates. *BMC microbiology* 6, 22.
- Maxam, A.M., Gilbert, W., 1977. A new method for sequencing DNA. *Proceedings of the National Academy of Sciences of the United States of America* 74, 560-564.
- McIntyre, L., Bernard, K., Beniac, D., Isaac-Renton, J.L., Naseby, D.C., 2008. Identification of *Bacillus cereus* group species associated with food poisoning outbreaks in British Columbia, Canada. *Applied and environmental microbiology* 74, 7451-7453.

- Merabishvili, M., Natidze, M., Riggava, S., Brusetti, L., Raddadi, N., Borin, S., Chanishvili, N., Tediashvili, M., Sharp, R., Barbeschi, M., Visca, P., Daffonchio, D., 2006. Diversity of *Bacillus anthracis* strains in Georgia and of vaccine strains from the former Soviet Union. *Applied and environmental microbiology* 72, 5631-5636.
- Meselson, M., Guillemin, J., Hugh-Jones, M., Langmuir, A., Popova, I., Shelokov, A., Yampolskaya, O., 1994. The Sverdlovsk Anthrax Outbreak of 1979. *Science* 266, 1202-1208.
- Mikesell, P., Ivins, B.E., Ristoph, J.D., Dreier, T.M., 1983. Evidence for Plasmid-Mediated Toxin Production in *Bacillus anthracis*. *Infection and Immunity* 39, 371-376.
- Morozova, O., Marra, M.A., 2008. Applications of next-generation sequencing technologies in functional genomics. *Genomics* 92, 255-264.
- Ohnishi, N., Fumito, M., Hirohito, O., Hirokazu, K., Shunsuke, Y., Daisuke, F., Ichiro, N., Mudenda, B.H.o., Yuka, T., Aaron, S.M., Hideaki, H., 2014. Genome Sequence of a *Bacillus anthracis* Outbreak Strain from Zambia, 2011. *Genome announcements* 2.
- Okinaka, R.T., Challacombe, J., Drees, K., Birdsell, D.N., Janke, N., Naumann, A., Seymour, M., Hornstra, H., Schupp, J., Sahl, J., Foster, J.T., Pearson, T., Turnbull, P., Keim, P., 2014. Genome sequence of *Bacillus anthracis* ST1, a Sterne-like Georgian/Soviet vaccine strain. *Genome announcements* 2.
- Okinaka, R.T., Henrie, M., Hill, K.K., Lowery, K.S., Van Ert, M., Pearson, T., Shupp, J., Kenefic, L., Beaudry, J., Hofstadler, S.A., Jackson, P.J., Keim, P., 2008. Single Nucleotide Polymorphism typing of *Bacillus anthracis* from Sverdlovsk Tissue. *Emerging Infectious Diseases* 14, 653-656.
- Økstad, O.A., Kolstø, A.-B. 2011. Genomics of *Bacillus* Species, In: Springer (Ed.) *Genomics of Foodborne Bacterial Pathogens*. 29-53.
- Okutani, A., Sekizuka, T., Boldbaatar, B., Yamada, A., Kuroda, M., Inoue, S., 2010. Phylogenetic Typing of *Bacillus anthracis* Isolated in Japan by Multiple Locus Variable-Number Tandem Repeats and the Comprehensive Single Nucleotide Polymorphism. *J Vet Med Sci* 72, 93-97.
- OMS, 2008. Anthrax in humans and animals.
- ONU. <http://www.un.org/fr/disarmament/wmd/biological/science.shtml>.
- Opal, S.M. 2009. Molecular Mechanisms of Antibiotic Resistance in Bacteria, In: Livingstone, C. (Ed.) *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases*.
- Palazzo, L., De Carlo, E., Santagada, G., Serrecchia, L., Aceti, A., Guarino, A., Adone, R., Fasanella, A., 2012. Recent Epidemic-Like Anthrax Outbreaks in Italy: What Are the Probable Causes? *Open Journal of Veterinary Medicine* 2, 74-76.
- Papazisi, L., Rasko, D.A., Ratnayake, S., Bock, G.R., Remortel, B.G., Appalla, L., Liu, J., Dracheva, T., Braisted, J.C., Shallom, S., Jarrahi, B., Sniesrud, E., Ahn, S., Sun, Q., Rilstone, J., Økstad, O.A., Kolsto, A.B., Fleischmann, R.D., Peterson, S.N., 2011. Investigating the genome diversity of *B. cereus* and evolutionary aspects of *B. anthracis* emergence. *Genomics* 98, 26-39.
- Papp, A.C., Pinsonneault, J.K., Cooke, G., Sadée, W., 2003. Single nucleotide polymorphism genotyping using allele-specific PCR and fluorescence melting curves. *Biotechniques* 34, 1068-1072.
- Pasteur, L., Chamberland, R., Roux, E., 1881. *Compte-rendu sommaire de L. Pasteur des expériences faites à Pouilly-le-Fort, près Melun, sur la vaccination charbonneuse*. *Académie des Sciences (Archives)*, 1379-1383.
- Patra, G., Vaissaire, J., Weber-Levy, M., Le Doujet, C., Mock, M., 1998. Molecular characterization of *Bacillus* strains involved in outbreaks of Anthrax in France in 1997. *Journal of clinical microbiology* 36, 3412-3414.
- Pearson, T., Busch, J.D., Ravel, J., Read, T.D., Rhoton, S.D., U'Ren, J.M., Simonson, T.S., Kachur, S.M., Leadem, R.R., Cardon, M.L., Van Ert, M.N., Huynh, L.Y., Fraser, C.M., Keim, P., 2004. Phylogenetic discovery bias in *Bacillus anthracis* using single-nucleotide polymorphisms from whole-genome sequencing. *Proceedings of the National Academy of Sciences of the United States of America* 101, 13536-13541.

- Pearson, T., Okinaka, R.T., Foster, J.T., Keim, P., 2009. Phylogenetic understanding of clonal populations in an era of whole genome sequencing. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* 9, 1010-1019.
- Pilo, P., Perreten, V., Frey, J., 2008. Molecular epidemiology of *Bacillus anthracis*: determining the correct origin. *Applied and environmental microbiology* 74, 2928-2931.
- Pilo, P., Rossano, A., Bamanga, H., Abdoukadi, S., Perreten, V., Frey, J., 2011. Bovine *Bacillus anthracis* in Cameroon. *Applied and environmental microbiology* 77, 5818-5821.
- Price, E.P., Matthews, M.A., Beaudry, J.A., Allred, J.L., Schupp, J.M., Birdsell, D.N., Pearson, T., Keim, P., 2010. Cost-effective interrogation of single nucleotide polymorphisms using the mismatch amplification mutation assay and capillary electrophoresis. *Electrophoresis* 31, 3881-3888.
- Price, E.P., Seymour, M.L., Sarovich, D.S., Latham, J., Wolken, S.R., Mason, J., Vincent, G., Drees, K.P., Beckstrom-Sternberg, S.M., Phillippy, A.M., Koren, S., Okinaka, R.T., Chung, W.K., Schupp, J.M., Wagner, D.M., Vipond, R., Foster, J.T., Bergman, N.H., Burans, J., Pearson, T., Brooks, T., Keim, P., 2012. Molecular epidemiologic investigation of an anthrax outbreak among heroin users, Europe. *Emerg Infect Dis* 18, 1307-1313.
- Price, L.B., Hugh-Jones, M., Jackson, P.J., Keim, P., 1999. Genetic Diversity in the Protective Antigen Gene of *Bacillus anthracis*. *J of Bacteriology* 181, 2358-2362.
- Prüss, B.M., Dietrich, R., Nibler, B., Märtlaubauer, E., Scherer, S., 1999. The Hemolytic Enterotoxin HBL Is Broadly Distributed among Species of the *Bacillus cereus* Group. *Applied and environmental microbiology* 65, 5346-5442.
- Purcell, B.K., Worsham, P.L., Friedlander, A.M., 2006. Anthrax. *Medical Aspects of Biological Warfare*, 69-90.
- Ramos, J.L., Martinez-Bueno, M., Molina-Henares, A.J., Teran, W., Watanabe, K., Zhang, X., Gallegos, M.T., Brennan, R., Tobes, R., 2005. The TetR family of transcriptional repressors. *Microbiol Mol Biol Rev* 69, 326-356.
- Rao, S.S., Mohan, K.V., Atreya, C.D., 2010. Detection technologies for *Bacillus anthracis*: prospects and challenges. *Journal of microbiological methods* 82, 1-10.
- Rasko, D.A., Altherr, M.R., Han, C.S., Ravel, J., 2005. Genomics of the *Bacillus cereus* group of organisms. *FEMS microbiology reviews* 29, 303-329.
- Rasko, D.A., Worsham, P.L., Abshire, T.G., Stanley, S.T., Bannan, J.D., Wilson, M.R., Langham, R.J., Decker, R.S., Jiang, L., Read, T.D., Phillippy, A.M., Salzberg, S.L., Pop, M., Van Ert, M.N., Kenefic, L.J., Keim, P.S., Fraser-Liggett, C.M., Ravel, J., 2011. *Bacillus anthracis* comparative genome analysis in support of the Amerithrax investigation. *Proceedings of the National Academy of Sciences of the United States of America* 108, 5027-5032.
- Ravel, J., Jiang, L., Stanley, S.T., Wilson, M.R., Decker, R.S., Read, T.D., Worsham, P., Keim, P.S., Salzberg, S.L., Fraser-Liggett, C.M., Rasko, D.A., 2009. The complete genome sequence of *Bacillus anthracis* Ames "Ancestor". *Journal of bacteriology* 191, 445-446.
- Read, T.D., Petersons, S.N., Tourasse, N., Baillie, L.W., Paulsen, I.T., Nelson, K.E., Tettelin, H., Fouts, D.E., Eisen, J.A., Gill, S.R., Holtzapple, E.K., Økstad, O.A., Helgason, E., Rilstone, J., Wu, M., Kolonay, J.F., Beanan, M.J., Dodson, R.J., Brinkac, L.M., Gwinn, M., DeBoy, R.T., Madpu, R., Daugherty, S.C., Durkin, A.S., Haft, D.H., Nelson, W.C., Peterson, J.D., Pop, M., Khouri, H.M., Radune, D., Benton, J.L., Mahamoud, Y., Jiang, L., Hance, I.R., Weidman, J.F., Berry, K.J., Plaut, R.D., Wolf, A.M., Watkins, K.L., Nierman, W.C., Hazen, A., Cline, R., Redmond, C., Thwaite, J.E., White, O., Salzberg, S.L., Thomason, B., Friedlander, A.M., Koehler, T.M., Hanna, P.C., Kolstø, A.-B., Fraser, C.M., 2003. The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria. *Nature* 423, 77-81.
- Read, T.D., Salzberg, S.L., Pop, M., Shumway, M., Umayam, L., Jiang, L., Holtzapple, E., Busch, J.D., Smith, K.L., Schupp, J.M., Solomon, D., Keim, P., Fraser, C.M., 2002. Comparative genome sequencing for discovery of novel polymorphisms in *Bacillus anthracis*. *Science* 296, 2028-2033.

- Sabat, A.J., Budimir, A., Nashev, D., Sa-Leao, R., Van Dijl, J.M., Laurent, F., Grundmann, H., Friedrich, A.W., (ESGEM), o.b.o.t.E.S.G.o.E.M., 2013. Overview of molecular typing methods for outbreak detection and epidemiological surveillance. *euro Surveill* 18.
- Salb, A., Stephen, C., Ribble, C., Elkin, B., 2014. Descriptive Epidemiology of Detected Anthrax Outbreaks in Wild Wood Bison (*Bison Bison Athabasca*) in Northern Canada, 1962-2008. *Journal of wildlife diseases* 50.
- Sanger, F., Coulson, A.R., 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol* 94, 441-448.
- Sanger, F., Nicklen, S., Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* 74, 5463-5467.
- Schild, A.L., Sallis, E.S.V., Soares, M.P., Ladeira, S.R.L., Schramm, R., Priebe, A.P., Almeida, M.B., Riet-Correa, F., 2006. Anthrax in cattle in southern Brazil: 1978-2006. *Presq. Vet. Bras.* 26, 243-248.
- Schmid, G., Kaufmann, A., 2002. Anthrax in Europe: its epidemiology, clinical characteristics, and role in bioterrorism. *Clin Microbiol Infect* 8, 479-488.
- Schnepf, E., Crickmore, N., Van Rie, J., Lereclus, D., Baum, J., Feitelson, J., Zeigler, D.R., Dean, D.H., 1998. *Bacillus thuringiensis* and Its Pesticidal Crystal Proteins. *Microbiol Mol Biol Rev* 62, 775-806.
- Sejvar, J.J., Tenover, F.C., Stephens, D.S., 2005. Management of anthrax meningitis. *Lancet Infect Dis* 5, 287-295.
- Serizawa, M., Sekizuka, T., Okutani, A., Banno, S., Sata, T., Inoue, S., Kuroda, M., 2010. Genomewide screening for novel genetic variations associated with ciprofloxacin resistance in *Bacillus anthracis*. *Antimicrobial agents and chemotherapy* 54, 2787-2792.
- Shafazand, S., Doyle, R., Ruoss, S., Weinacker, A., Rafin, T.A., 1999. Inhalational Anthrax: Epidemiology, Diagnosis and Management. *CHEST* 116, 1369-1376.
- Simonson, T.S., Okinaka, R.T., Wang, B., Easterday, W.R., Huynh, L., U'Ren, J.M., Dukerich, M., Zanecki, S.R., Kenefic, L.J., Beaudry, J., Schupp, J.M., Pearson, T., Wagner, D.M., Hoffmaster, A., Ravel, J., Keim, P., 2009. *Bacillus anthracis* in China and its relationship to worldwide lineages. *BMC microbiology* 9, 71.
- Sirisanthana, T., Brown, A.E., 2002. Anthrax of the Gastrointestinal Tract. *Emerging Infectious Diseases* 8, 649-651.
- Smith, K.L., DeVos, V., Bryden, H., Price, L.B., Hugh-Jones, M.E., Keim, P., 2000. *Bacillus anthracis* diversity in Kruger National Park. *Journal of clinical microbiology* 38, 3780-3784.
- Soberon, M., Gill, S.S., Bravo, A., 2009. Signaling versus punching hole: How do *Bacillus thuringiensis* toxins kill insect midgut cells? *Cellular and Molecular Life Sciences* 66, 1337-1349.
- Soberon, M., Pardo-Lopez, L., Lopez, I., Gomez, I., Tabashnik, B.E., Bravo, A., 2007. Engineering modified Bt toxins to counter insect resistance. *Science* 318, 1640-1642.
- Sobral, D., 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. *Biology*. Université Paris Sud,
- Steiner, I., Racic, I., Spicic, S., Habrun, B., 2013. Genotyping of *Bacillus anthracis* isolated from Croatia and Bosnia and Herzegovina. *Zoonoses and public health* 60, 202-208.
- Stenfors Arnesen, L.P., Fagerlund, A., Granum, P.E., 2008. From soil to gut: *Bacillus cereus* and its food poisoning toxins. *FEMS microbiology reviews* 32, 579-606.
- Stenfors, L.P., Mayr, R., Scherer, S., Granum, P.E., 2002. Pathogenic potential of fifty *Bacillus weihenstephanensis* strains. *FEMS Microbiol Lett* 215, 47-51.
- Sternbach, G., 2003. The history of anthrax. *The Journal of Emergency Medicine* 24, 463-467.
- Stratilo, C.W., Lewis, C.T., Bryden, L., Mulvey, M.R., Bader, D., 2006. Single-nucleotide repeat analysis for subtyping *Bacillus anthracis* isolates. *Journal of clinical microbiology* 44, 777-782.
- Sue, D., Marston, C.K., Hoffmaster, A.R., Wilkins, P.P., 2007. Genetic diversity in a *Bacillus anthracis* historical collection (1954 to 1988). *Journal of clinical microbiology* 45, 1777-1782.

- Tasota, F.J., Henker, R.A., Hoffman, L.A., 2002. Anthrax as a biological weapon: an old disease that poses a new threat. *Critical Care Nurse* 22, 21-36.
- Thapa, N.K., Tenzin, Wangdi, K., Dorji, T., Migma, Dorjee, J., Marston, C.K., Hoffmaster, A.R., 2014. Investigation and control of anthrax outbreak at the human-animal interface, Bhutan, 2010. *Emerg Infect Dis* 20, 1524-1526.
- Thierry, S., Hamidjaja, R.A., Girault, G., Lofstrom, C., Ruuls, R., Sylviane, D., 2013. A multiplex bead-based suspension array assay for interrogation of phylogenetically informative single nucleotide polymorphisms for *Bacillus anthracis*. *Journal of microbiological methods* 95, 357-365.
- Thierry, S., Tourterel, C., Le Fleche, P., Derzelle, S., Dekhil, N., Mendy, C., Colaneri, C., Vergnaud, G., Madani, N., 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. *PLoS ONE* 9, e95131.
- Thorsen, L., Budde, B.B., Henriksen, L., Martinussen, T., Jakobsen, M., 2009. Cereulide formation by *Bacillus weihenstephanensis* and mesophilic emetic *Bacillus cereus* at temperature abuse depends on pre-incubation conditions. *International journal of food microbiology* 134, 133-139.
- Thorsen, L., Hansen, B.M., Nielsen, K.F., Hendriksen, N.B., Phipps, R.K., Budde, B.B., 2006. Characterization of emetic *Bacillus weihenstephanensis*, a new cereulide-producing bacterium. *Applied and environmental microbiology* 72, 5118-5121.
- Tourasse, N.J., Helgason, E., Klevan, A., Sylvestre, P., Moya, M., Haustant, M., Okstad, O.A., Fouet, A., Mock, M., Kolsto, A.B., 2011. Extended and global phylogenetic view of the *Bacillus cereus* group population by combination of MLST, AFLP, and MLEE genotyping data. *Food microbiology* 28, 236-244.
- Tourasse, N.J., Helgason, E., Okstad, O.A., Hegna, I.K., Kolsto, A.B., 2006. The *Bacillus cereus* group: novel aspects of population structure and genome dynamics. *Journal of applied microbiology* 101, 579-593.
- Turell, M.J., Knudson, G.B., 1987. Mechanical Transmission of *Bacillus anthracis* by Stable Flies (*Stomoxys calcitrans*) and Mosquitoes (*Aedes aegypti* and *Aedes taeniorhynchus*). *Infection and Immunity* 55, 1859-1861.
- Turnbull, P.C., 2000. Current status of immunization against anthrax: old vaccines may be here to stay for a while. *Curr Opin Infect Dis*. 13, 113-120.
- Turnbull, P.C.B., Doganay, M., Lindeque, P.M., Aygen, B., McLaughlin, J., 1992. Serology and anthrax in humans, livestock and Etosha National Park wildlife. *Epidemiology and infection* 108, 299-313.
- Turnbull, P.C.B., Hugh-Jones, M.E., Cosivi, O., 1999. World Health Organization activities on anthrax surveillance and control. *Journal of applied microbiology* 87, 318-320.
- Uchida, I., Sekizaki, T., Hashimoto, K., Terakado, N., 1985. Association of the Encapsulation of *Bacillus anthracis* with a 60 Megadalton Plasmid. *Journal of General Microbiology* 131.
- United-Nations 1972. Convention on the prohibition of the development, production and stockpiling of bacteriological (biological) and toxin weapons and on their destruction. Opened for signature at London, Moscow and Washington on 10 April 1972. In UNited Nations - Treaty Series.
- Uplekar, S., Heym, B., Friocourt, V., Rougemont, J., Cole, S.T., 2011. Comparative genomics of *Esx* genes from clinical isolates of *Mycobacterium tuberculosis* provides evidence for gene conversion and epitope variation. *Infect Immun* 79, 4042-4049.
- Vaissaire, J., Mock, M., Le Doujet, C., Levy, M., 2001. Le charbon bactérien. *Epidémiologie de la maladie en France. Médecine et maladies infectieuses* 31, 257-271.
- Valade, E., Tournier, J.N., Vidal, D., Morillon, M., 2009. Maladie du charbon. *EMC Maladies infectieuses*.

- Van Belkum, A., Melchers, W.J., Ijsseldijk, C., Nohlmans, L., Verbrugh, H., Meis, J.G.M., 1997. Outbreak of Amoxicillin-Resistant *Haemophilus influenzae* Type b: Variable Number of Tandem Repeats as Novel Molecular Markers. *Journal of clinical microbiology* 35, 1517-1520.
- Van Ert, M.N., Easterday, W.R., Huynh, L.Y., Okinaka, R.T., Hugh-Jones, M.E., Ravel, J., Zanecki, S.R., Pearson, T., Simonson, T.S., U'Ren, J.M., Kachur, S.M., Leadem-Dougherty, R.R., Rhoton, S.D., Zinser, G., Forlow, J., Coker, P.R., Smith, K.L., Wang, B., Kenefic, L.J., Fraser-Liggett, C.M., Wagner, D.M., Keim, P., 2007a. Global Genetic Population Structure of *Bacillus anthracis*. *PLoS ONE*.
- Van Ert, M.N., Easterday, W.R., Simonson, T.S., U'Ren, J.M., Pearson, T., Kenefic, L.J., Busch, J.D., Huynh, L.Y., Dukerich, M., Trim, C.B., Beaudry, J., Welty-Bernard, A., Read, T., Fraser, C.M., Ravel, J., Keim, P., 2007b. Strain-specific single-nucleotide polymorphism assays for the *Bacillus anthracis* Ames strain. *Journal of clinical microbiology* 45, 47-53.
- Vergnaud, G., Pourcel, C., 2009. Multiple locus variable number of tandem repeats analysis. *Methods Mol Biol* 551, 141-158.
- Vogler, A.J., Busch, J.D., Percy-Fine, S., Tipton-Hunton, C., Smith, K.L., Keim, P., 2012. Molecular analysis of rifampicin resistance in *Bacillus anthracis* and *Bacillus cereus*. *Antimicrobial agents and chemotherapy* 46, 511-513.
- Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Pelerman, J., Kuiper, M., Zabeau, M., 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* 23, 4407-4414.
- Wang, J., Chuang, K., Ahluwalia, M., Patel, S., Umblas, N., Mirel, D., Higuchi, R., Germe, S., 2005. High-throughput SNP genotyping by single-tube PCR with Tm-shift primers. *Biotechniques* 39, 885-893.
- Wattiau, P., Klee, S.R., Fretin, D., Van Hessche, M., Menart, M., Franz, T., Chasseur, C., Butaye, P., Imberechts, H., 2008. Occurrence and genetic diversity of *Bacillus anthracis* strains isolated in an active wool-cleaning factory. *Applied and environmental microbiology* 74, 4005-4011.
- Weiner, B., Gomez, J., Victor, T.C., Warren, R.M., Sloutsky, A., Plikaytis, B.B., Posey, J.E., van Helden, P.D., Gey van Pittius, N.C., Koehrsen, M., Sisk, P., Stolte, C., White, J., Gagneux, S., Birren, B., Hung, D., Murray, M., Galagan, J., 2012. Independent Large Scale Duplications in Multiple *M. tuberculosis* Lineages Overlapping the Same Genomic Region. *PLoS ONE* 7.
- Zhong, W., Shou, Y., Yoshida, T.M., Marrone, B.L., 2007. Differentiation of *Bacillus anthracis*, *B. cereus*, and *B. thuringiensis* by using pulsed-field gel electrophoresis. *Applied and environmental microbiology* 73, 3446-3449.

ANNEXES

Annexe 1 :

Apport de la biologie moléculaire pour la connaissance de la fièvre charbonneuse en France

Publié en Septembre 2014 dans Bulletin Epidémiologique Santé animale-alimentation n°63

Apport de la biologie moléculaire pour la connaissance de la fièvre charbonneuse en France

Guillaume Girault (1), Sylviane Derzelle (1) (sylviane.derzelle@anses.fr)

(1) Université Paris-Est, Anses, Laboratoire de santé animale, Unité Zoonoses bactériennes, Maisons-Alfort, France

Résumé

Zoonose sporadique en France, la fièvre charbonneuse (FC) n'en reste pas moins présente sur le territoire français. Au cours de la dernière décennie (2001-2013), soixante-neuf foyers animaux de FC ont été recensés. Le nombre de foyers déclarés chaque année est extrêmement variable, allant d'années exemptes (2012) à des années propices à la résurgence de cas de FC (19 foyers en 2008 ou 22 foyers en 2009). Aujourd'hui, le diagnostic des cas de FC se limite à la détection et à la confirmation de l'espèce de l'agent causal, *Bacillus anthracis*. *B. anthracis* est une espèce bactérienne très monomorphe, présentant peu de variabilité d'un point de vue génétique. Seules certaines techniques de biologie moléculaire à haut pouvoir résolutif permettent une discrimination des souches entre elles. Afin de mieux connaître la diversité des souches de *B. anthracis* présentes en France et les relations phylogénétiques qu'entretiennent ces souches entre elles, une analyse génomique ambitieuse a été menée au sein du laboratoire national de référence, à l'Anses. Le séquençage à haut débit de 125 génomes de *B. anthracis* collectés en France depuis soixante ans a été réalisé. Cette étude a permis l'obtention d'une cartographie extrêmement précise des génotypes de *B. anthracis* présents sur le territoire. Au cours de ce travail, l'identification d'un grand nombre de marqueurs diagnostiques de type SNP (polymorphisme de séquence unique) permettra à l'avenir une caractérisation fine de cet agent pathogène important, ainsi qu'une surveillance adéquate de cette zoonose grave lors des foyers de FC.

Mots-clés

Bacillus anthracis, France, polymorphismes de séquence uniques (SNP), génotypage

Abstract

*Although anthrax is a sporadic zoonosis in France, it nevertheless is still found in the country. Over the last twelve years (2001-2013), 69 anthrax outbreaks have been reported in animals. The number of outbreaks declared each year is extremely variable, certain years being entirely free of cases (2012, for example) and others showing a resurgence of the disease (19 outbreaks in 2008 and 22 outbreaks in 2009, for example). Today, diagnosis of anthrax is limited to the detection and confirmation of the causative agent of the disease, *Bacillus anthracis*, which is a monomorphic bacterial species with very low genetic variability. Only a few high-resolution typing methods are able to discriminate between strains. To increase knowledge of the diversity of *B. anthracis* strains in France and of the phylogenetic relationships that they share, an ambitious genomic analysis campaign has been conducted at the ANSES National Reference Laboratory for anthrax. This included the high-throughput DNA sequencing of 125 *B. anthracis* genomes isolated in France over a 60-year period. This study provided an extremely precise picture of the distribution of the *B. anthracis* genotypes found in France. Within this framework, the identification of a diagnostic marker pattern (Single Nucleotide Polymorphisms) will provide accurate characterization of this important pathogen, including appropriate surveillance during anthrax outbreaks.*

Keywords

Bacillus anthracis, France, Single Nucleotide Polymorphisms, Genotyping

La fièvre charbonneuse (FC) (ou charbon bactérien) est une zoonose grave de répartition mondiale causée par une bactérie sporulée « tellurique », *Bacillus anthracis* (OIE, 2008) (Encadré 1). Outre sa capacité de sporulation, *B. anthracis* possède plusieurs facteurs de virulence codés par des gènes situés sur ses deux plasmides (pXO1 et pXO2) : une capsule aux propriétés anti-phagocytaires et deux exotoxines de type A-B, le facteur Oedématogène (EF) et le Facteur Létal (LF), partageant un même facteur assurant la translocation à travers les membranes cellulaires vers le cytosol des deux autres facteurs, le facteur PA (Protective Antigen) (OIE, 2008).

L'infection d'un hôte peut avoir lieu par inhalation, ingestion ou contamination cutanée par le biais d'une plaie ou d'une piqûre d'insecte. La forme respiratoire est la plus fulgurante. Après pénétration des spores dans le système respiratoire de l'hôte, elles germent en présence de CO₂. Les bacilles vont ensuite synthétiser leur capsule, se multiplier rapidement et libérer leurs toxines. Bactéries et toxines finissent par se disséminer dans la circulation sanguine et provoquent la mort par septicémie et choc exotoxinique. Les animaux les plus exposés sont les herbivores (bovins, ovins, caprins, équins), et plus particulièrement les ruminants qui contractent généralement la maladie en ingérant des spores au cours de leur pâture (ingestion d'herbe ou de terre contaminée). En France, les cas humains sont excessivement rares. Pour les éleveurs et les personnes en contact avec les animaux, le principal risque réside dans la contraction d'un charbon cutané suite à la manipulation sans précaution d'animaux malades ou morts.

Encadré 1. Fièvre charbonneuse

Classée dans la liste de l'Organisation mondiale de la santé animale (OIE), la fièvre charbonneuse fait partie en France des dangers sanitaires de catégorie 1 en médecine vétérinaire (articles L. 223-2 et suivants, et D. 223-2 du Code rural) qui donnent lieu à l'application de mesures de police sanitaire. C'est également une maladie à déclaration obligatoire en médecine humaine depuis 2001. Elle est reconnue comme maladie professionnelle (éleveurs, vétérinaires, microbiologistes...). Il s'agit d'une maladie potentiellement mortelle qui touche tous les mammifères, y compris l'Homme. Les herbivores (bovins, ovins, caprins, équins) sont les plus sensibles. Ils contractent généralement la maladie en ingérant des spores au cours de leur pâture (ingestion d'herbe ou de terre contaminées) et développent le plus souvent des formes suraiguës ou aiguës de la maladie. Chez l'Homme, le contact d'animaux infectés, l'ingestion d'aliments contaminés ou la respiration de poussières contenant des spores peuvent être responsables de la maladie. Les formes cutanées, suites à des piqûres, des coupures ou des éraflures, sont majoritaires (95 % des cas) et le plus souvent bénignes.

Encadré 2. Définition des cas et des foyers animaux

Un cas confirmé de FC correspond à un animal dont un prélèvement a permis l'isolement de *B. anthracis*. Un cas présumé de FC correspond à un animal dont la mort a été imputée à la FC, sans confirmation par une analyse biologique. Enfin, un foyer correspond à un lot d'animaux dans lequel au moins un cas de FC a été observé.

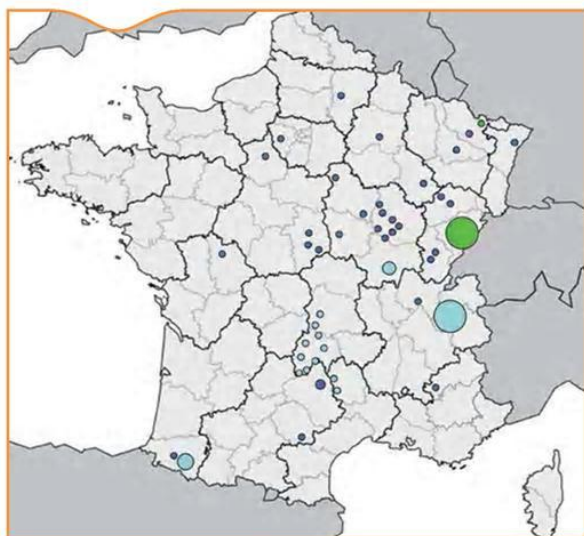


Figure 1. Répartition des foyers de FC en France et des sous-lignées associées (1953-2013)

La figure représente la répartition géographique des foyers de FC observés en France depuis soixante ans. Les trois sous-lignées présentes à l'état naturel sont représentées selon le code couleur suivant: la sous-lignée Transeurasienne A.Br.011/009 en violet, la sous-lignée majoritaire B.Br.Cneva en bleu ciel et enfin, la sous-lignée minoritaire A.Br.001/002 en vert. La taille des cercles est proportionnelle au nombre de souches qu'ils représentent.

Aujourd'hui, en France, cette zoonose est observée de manière assez sporadique chez les animaux. Au cours de la dernière décennie (2001-2013), soixante-neuf foyers animaux de FC ont été recensés (Encadré 2). Leur nombre est extrêmement variable d'une année sur l'autre, avec des années sans foyer confirmé (comme 2010 et 2012) et des années plus marquées (2008 et 2009). Les épisodes de FC sont généralement de faible importance avec un ou deux foyers et un ou deux animaux morts par foyer. Cependant, il n'est pas exceptionnel que des épisodes plus conséquents se produisent. Ainsi, en 2009, un taux record de vingt-deux foyers répartis dans cinq départements a été observé. En tout, cinquante-trois bovins, quinze caprins et deux chevaux sont morts (Fediaevsky *et al.*, 2009). En 2008, un épisode de dix-sept foyers se déclarant concomitamment sur plusieurs communes proches a été rapporté dans le Doubs. Au total, trente-neuf bovins sont morts et quatre-vingt-onze autres ont été traités par antibiotiques. Une vaste opération de vaccination de près de 12 000 bovins a été entreprise sur quinze communes (Calavas *et al.*, 2008). La répartition des cas de FC concerne quasiment tout le territoire français, à l'exception de certaines régions, comme la Bretagne (Figure 1). La grande majorité des foyers s'observe cependant dans les départements français ayant un lourd passé charbonneux, essentiellement dans le Massif Central, les Alpes, la Bourgogne et la Franche-Comté (Madani *et al.*, 2009; Vaissaire *et al.*, 2001). En effet, les spores de *B. anthracis* peuvent persister dans le sol pendant de longues périodes (demi-vie pouvant aller jusqu'à 100 ans), assurant de ce fait la pérennité de l'espèce.

Bacillus anthracis, un pathogène clonal

L'agent causal de la FC est considéré comme une espèce bactérienne relativement jeune sur le plan évolutif, ayant émergé de l'espèce *Bacillus cereus* suite à l'acquisition de deux plasmides de virulence (Keim and Wagner, 2009). D'un point de vue génétique, *B. anthracis* est l'une des bactéries les plus homogènes qui soient. C'est une espèce clonale dont il est difficile, voire impossible, de différencier les souches entre-elles au moyen de méthodes de typage classiques (biochimique, bactériologie/phénotypique, séquençage de l'ARNr 16S ou 23S). Cette homogénéité inter-souches exceptionnelle est le résultat d'un cycle de vie particulier.

B. anthracis passe la majorité de sa vie sous forme de spores dormantes dans le sol. Des décennies peuvent se passer entre deux cycles infectieux. Durant cette phase, l'organisme est peu exposé aux modifications et variations que subit en continu le matériel génétique (liées à la présence de phages, à la réplication de l'ADN). L'évolution du génome est ainsi statique ou du moins grandement réduite.

La biologie moléculaire au service du typage

Acquérir la capacité de distinguer suffisamment de différences entre deux souches est très important pour déterminer avec certitude qu'une souche particulière est à l'origine d'un foyer de FC ou d'un cas humain. Le défi majeur pour développer une méthode de typage de haute résolution pour une espèce bactérienne clonale comme *B. anthracis* réside en l'identification de caractères ou d'une combinaison de caractères uniques à chaque souche.

Depuis les années 2000, plusieurs méthodes existent pour différencier les souches de *B. anthracis* entre elles: certaines sont basées sur les séquences répétées présentes dans le génome (Multiple Locus Variable number of tandem repeats Analysis: MLVA) (Derzelle *et al.*, 2011; Keim *et al.*, 2000) et d'autres sur un petit nombre de mutations ponctuelles de type Single Nucleotide Polymorphisms (SNP) réparties le long du génome (Girault *et al.*, 2014a; Van Ert *et al.*, 2007) (Encadré 3). Les polymorphismes nucléotidiques de type SNP sont des mutations extrêmement stables d'un point de vue évolutif, ce qui en fait des marqueurs de choix pour le génotypage. Un nombre limité de 13 à 14 SNP dits « canoniques » permet de déterminer précisément la position phylogénétique de tout isolat au sein des groupes et lignées majeures de *B. anthracis* retrouvées de par le monde (Marston *et al.*, 2011; Van Ert *et al.*, 2007). Il existe trois lignées majeures (A, B et C) et douze sous-groupes phylogénétiques au sein de l'espèce *B. anthracis*. La branche A présente une dispersion globale assez étendue et est la plus diversifiée. La majorité des isolats actuels se positionne dans cette lignée. La branche B ne représenterait qu'environ 15 % des cas mondiaux de FC récents. Sa distribution géographique se limite principalement à l'Afrique du Sud et à une partie de l'Europe dont la France. La lignée C est quant à elle plus anecdotique (seulement quelques isolats de laboratoire).

Ces approches (MLVA et analyse de canSNP (un set d'une dizaine de SNP phylogénétiquement informatifs) ont cependant un pouvoir résolutif quelque peu limité. Aucune n'a le pouvoir résolutif du séquençage complet de génomes. Avec l'accessibilité des nouvelles technologies de séquençage à haut débit (plateformes, diminution des coûts), les approches basées sur l'analyse de génomes complets prennent une part nouvelle et de plus en plus importante dans le génotypage des bactéries. Afin de mieux connaître la diversité existante en France au sein de l'espèce ainsi que la distribution naturelle des souches à travers le pays, le séquençage global de 125 souches « indigènes » de *B. anthracis* issues de la collection du Laboratoire national de référence pour la FC,

Encadré 3. Qu'est ce qu'un SNP ?

Le polymorphisme nucléotidique ou polymorphisme d'un seul nucléotide (SNP, Single Nucleotide Polymorphism) est, en génétique, la variation d'une seule paire de bases du génome, entre individus d'une même espèce (exemple représenté ci-dessous, avec un SNP différenciant la souche Ames Ancestor de la souche Sterne). Ces variations sont très fréquentes chez tous les organismes vivants.

Les SNP sont générés par des erreurs de réplication de l'ADN non réparées par la suite, qui prennent place dans le génome à une très faible fréquence (10^{-10} changement par nucléotide et par génération pour *B. anthracis*). Les SNP sont des marqueurs diagnostiques importants, car ils sont stables d'un point de vue évolutif et présentent une probabilité faible de réversion.

Ames Ancestor
Sterne

GTTGTTG
GTTATTG

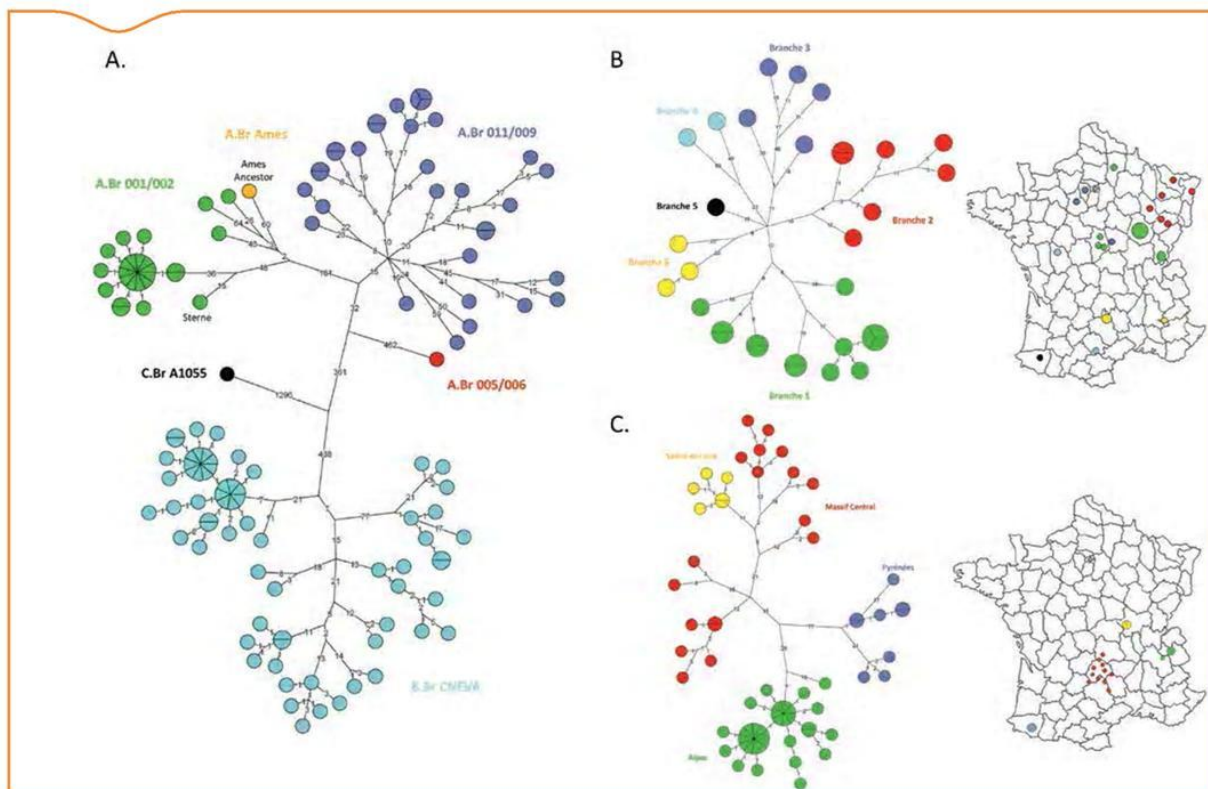


Figure 2. Cartographie à haute résolution de *B. anthracis* en France

- A. Le Minimum Spanning Tree (MST) représenté ici a été obtenu à partir de l'analyse des SNP sur génome complet des souches de la collection du LNR Charbon séquencées (125 souches) ainsi que de trois souches disponibles dans les bases de données génomiques (Ames Ancestor, Sterne et A1055). Une des souches de la collection LNR est une souche d'origine africaine (A.Br.005/006). Cet arbre phylogénétique a été obtenu à partir de 4018 SNP chromosomiques. Les sous-lignées sont représentées selon le code couleur suivant : A.Br.Ames en jaune, A.Br.001/002 en vert, A.Br.005/006 en rouge, A.Br.011/009 en violet, B.Br.Cneva en bleu ciel et C.Br.A1055 en noir.
- B. Représentation de la sous-lignée A.Br.011/009 en France. Le MST représente les six branches retrouvées en France. La carte de France représente leurs localisations géographiques.
- C. Représentation de la sous-lignée B.Br.Cneva en France. Le MST représente les trois grands groupes (Alpes, Pyrénées, Massif central) ainsi que le sous-groupe Saône-et-Loire. La carte de France représente leurs localisations géographiques.

à l'Anses, a été entrepris. La production de données génomiques a permis d'acquérir une somme considérable d'informations sur les souches de *B. anthracis* présentes dans les différentes régions de France. La connaissance exhaustive du polymorphisme existant au sein des séquences génomiques de ces souches a également permis de mettre en évidence des milliers de signatures ADN de type SNP uniques aux souches retrouvées en France. Près de 4000 SNP ont été identifiés, ce qui a permis l'obtention d'une cartographie à très haute résolution de la FC sur le territoire français (Figure 2A). La résolution obtenue est telle que l'on peut différencier presque toutes les souches sur la base de ce jeu de SNP spécifiques.

Distribution des génotypes de *B. anthracis* à l'échelle française

L'ensemble des isolats français se positionne au sein de trois des douze grandes sous-lignées phylogénétiques décrites dans le monde. Il s'agit des branches A.Br.001/002, A.Br.011/009 (lignée Transeurasienne) et B.Br.Cneva (Figure 1). La sous-lignée minoritaire A.Br.001/002 (16 % des souches) est exclusivement retrouvée dans le Nord-Est de la France. Elle est impliquée dans l'épisode de l'été 2008 du Doubs. La branche A.Br.011/009 (30 %) est à l'origine de nombreux cas sporadiques de FC dans différentes régions du territoire. La branche majoritaire B.Br.Cneva (54 %) compte près de la moitié des isolats de la collection de l'Anses. Les foyers qui lui sont attribués se localisent principalement dans la moitié Sud du pays, près des grands massifs montagneux (Alpes, Pyrénées, Massif Central). Elle est notamment impliquée dans l'épisode de l'été 2009 en Savoie.

Un regroupement géographique en trois ou quatre grands sous-groupes est obtenu pour la lignée B.Br.Cneva. Les souches isolées dans les Alpes sont toutes apparentées et se distinguent clairement des souches pyrénéennes ou de celles du Massif central. Un sous-ensemble incluant les souches isolées en Saône et Loire se détache au sein de l'ensemble Massif Central (Figure 2C). Un regroupement géographique moins apparent en six sous-branches phylogénétiques est également observé pour les souches appartenant de la lignée Transeurasienne A.Br.011/009 dispersées à travers le territoire (Figure 2B). Les branches 1 et 2 regroupent des souches isolées dans un quart Nord-Est de la France. La branche 3 comprend des souches vaccinales de type Pasteur II ou Carbosap et quelques isolats collectés dans des départements proches de Paris. Les rares souches des branches 4, 5 et 6 sont issues, respectivement, de l'Ouest ou du Sud de la France, des Pyrénées, et du quart Sud-Est.

Développement d'un outil de typage par PCR

Parmi le jeu de plus de quatre mille mutations ponctuelles identifiées par séquençage, douze SNP canoniques représentatifs des sous-groupes génétiques et géographiques majeurs mis en évidence ont été sélectionnés pour développer un outil de typage moléculaire par réaction de polymérase en chaîne (PCR) permettant la discrimination des isolats français.

Parmi les différentes méthodes existantes pour cribler des SNP, une technique post-PCR peu onéreuse, ne nécessitant pas l'achat de sondes

nucléotidiques marquées spécifiques à chacun des allèles de SNP à discriminer, a été choisie. Il s'agit de la PCR-HRM (High Resolution Melting) dont le principe est d'analyser les amplicons obtenus par PCR au moyen d'une courbe de fusion de grande précision. Après une étape de PCR en temps réel classique réalisée avec un intercalant fluorescent, une montée progressive en température permet de séparer deux populations d'amplicons courts de température de fusion différente, ne différant que par une seule base nucléotidique.

Douze marqueurs représentatifs des sous-groupes phylogénétiques présents en France ont ainsi été validés *in silico* et *in vitro* (Girault *et al.*, 2014a; Girault *et al.*, 2014b). Ils permettent d'affiner le typage des souches et confirmer l'origine (géographique) locale des souches isolées lors de foyers.

Analyse rétrospective des récents épisodes majeurs de FC

La France a récemment connu deux épisodes importants de FC: en 2008 dans le Doubs et en 2009 en Savoie (Calavas *et al.*, 2008; Madani *et al.*, 2009). Les souches impliquées dans ces épisodes appartiennent aux lignées A.Br.001/002 et B.Br.CNEVA, respectivement.

L'analyse génétique *a posteriori* des souches du Doubs démontre le caractère strictement clonal des bactéries impliquées dans les dix-sept foyers concomitants de FC de l'épisode de 2008, ainsi que dans le foyer de 2011 du Doubs. Sur les vingt-quatre souches françaises de la lignée A.Br.001/002 séquencées par l'Anses, vingt et une provenaient des foyers de 2008 et 2011 dans le Doubs. Contrairement aux trois autres souches issues de la collection de l'institut Pasteur (CIP), il n'existe que très peu, voire pas du tout, de variabilité au sein des souches du Doubs (un maximum de deux SNP de différence a été identifié parmi les 21 souches). Toutes les souches sont regroupées en un complexe clonal. À l'inverse, les trois souches de la collection de l'Institut Pasteur (CIP), plus anciennes, dont le département d'isolement est inconnu, sont bien distinctes entre elles et très éloignées phylogénétiquement des souches du Doubs, avec plus de 100 SNP de différence (Figure 2). Le génotypage détaillé confirme l'hypothèse émise à l'époque d'une résurgence multifocale de spores liée aux conditions météorologiques particulière de 2008 (Calavas *et al.*, 2008). Il renseigne en outre que l'évolution génétique des spores conservées dans cette région est particulièrement statique, la souche de 2011 étant identique au génotype majeur rencontré en 2008.

L'analyse génétique des souches collectées dans les Alpes démontre quant à elle, une bien plus grande hétérogénéité au sein de la population de spores de *B. anthracis* dans cette région, même si presque toutes les souches des Alpes (quels que soient leur lieu d'isolement et l'année) présentent une parenté forte entre elles. Cette observation est en faveur d'un ancêtre commun ayant évolué localement et de manière distincte aux souches de la lignée B.Br.CNEVA retrouvées dans d'autres massifs montagneux. L'épisode de 2009 de Savoie (17 foyers de FC confirmés sur une quinzaine de communes limitrophes) implique différents génotypes très apparentés mais présentant autant de différences (SNP) entre souches que celles qu'on peut observer avec des souches isolées dans la même vallée en 1997 ou dans d'autres localisations (foyers de 2000, 2006 et 2007), suggérant une histoire de FC plus ancienne dans les Alpes.

Conclusions

Le séquençage d'un grand nombre de génomes combiné à l'identification exhaustive des polymorphismes de type SNP qu'ils contiennent a permis de cartographier avec une grande précision les divers génotypes de *B. anthracis* présents en France. Un outil de typage résolutif, basé sur une technique de PCR a été développé. Il permet de déterminer l'origine géographique probable de n'importe quelle souche de FC présente sur le sol français et d'établir rapidement sa filiation génétique avec tout autre foyer.

Bien que les coûts de séquençage d'un génome bactérien complet soient en constante diminution, le séquençage systématique des souches isolées au cours des foyers de FC n'est pas envisageable (dans un proche avenir). Disposer d'une large banque de données génomiques représentative de la diversité rencontrée sur un territoire est, de ce fait, fort utile pour toute investigation épidémiologique. Cela permet en outre de mettre rapidement en évidence tout événement inhabituel impliquant des génotypes « inattendus » qui nécessiterait de plus amples investigations.

Le maintien de la mémoire des épisodes survenus dans une région et la cartographie des terrains sur lesquels des animaux atteints de FC ont été détenus est un autre aspect essentiel. Tant que des spores persisteront dans les sols, la maladie sera susceptible de ré-émerger périodiquement en ces endroits. Il convient donc de maintenir une sensibilisation vis-à-vis de la FC à l'échelle locale, tant les conséquences en termes de santé animale et de santé humaine peuvent être importantes, nécessitant une surveillance et une prise en charge adaptées.

Remerciements

Les auteurs remercient Nora Madani (Anses) pour l'accès aux souches du LNR fièvre charbonneuse ainsi que la plateforme de séquençage à haut débit IMAGIF (Centre de Recherche de Gif-sur-Yvette (<http://www.imagif.cnrs.fr>)). Ce travail a été financé par l'Agence nationale de recherche (ANR) et la Direction générale de l'armement (DGA) (projet ANR11ASTR0007). Guillaume Girault est un doctorant co-financé par l'Anses et la DGA.

Références bibliographiques

- Calavas, D., Sala, C., Vaissaire, J., Condé, J., Thien-Aubert, H., Hesseemann, M., Woronoff-Rehn, N., 2008. Retour d'expérience sur un épisode de fièvre charbonneuse chez les bovins dans le Doubs au cours de l'été 2008. *Bull. Epid. Santé Anim. Alim.* 32 32, 1-6.
- Derzelle, S., Laroche, S., Le Fleche, P., Hauck, Y., Thierry, S., Vergnaud, G., Madani, N., 2011. Characterization of genetic diversity of *Bacillus anthracis* in France by using high-resolution melting assays and multilocus variable-number tandem-repeat analysis. *J. Clin. Microbiol.* 49, 4286-4292.
- Fediaevsky, A., Madani, N., Garin-Bastuji, B., Moutou, F., 2009. Bilan de la surveillance de la fièvre charbonneuse en 2009: détection de quelques foyers sporadiques. *Bull. Epid. Santé Anim. Alim.* 40, 1-2.
- Girault, G., Blouin, Y., Vergnaud, G., Derzelle, S., 2014a. High-throughput sequencing of *Bacillus anthracis* in France: investigating genome diversity and population structure using whole-genome SNP discovery. *BMC genomics* 15.
- Girault, G., Thierry, S., Cherchame, E., Derzelle, S., 2014b. Application of high-throughput sequencing: discovery of informative SNPs to subtype *Bacillus anthracis*. *Adv Biosci Biotechnol* 5, 669-677.
- Keim, P., Price, L.B., Klevytska, A.M., Smith, K.L., Shupp, J.M., Okinaka, R., Jackson, P.J., Hugh-Jones, M.E., 2000. Multiple-Locus Variable-Number Tandem Repeat Analysis Reveals Genetic Relationships within *Bacillus anthracis*. *J. of Bacteriology* 182, 2928-2936.
- Keim, P.S., Wagner, D.M., 2009. Humans and evolutionary and ecological forces shaped the phylogeography of recently emerged diseases. *Nature reviews. Microbiology* 7, 813-821.
- Madani, N., Mendy, C., Moutou, F., Garin-Bastuji, B., 2009. La fièvre charbonneuse en France. Épisodes de l'été 2009 et foyers enregistrés sur la dernière décennie (1999-2009). *Bull. Epid. Santé Anim. Alim.* 38, 1-3.
- Marston, C.K., Allen, C.A., Beaudry, J., Price, E.P., Wolken, S.R., Pearson, T., Keim, P., Hoffmaster, A.R., 2011. Molecular epidemiology of anthrax cases associated with recreational use of animal hides and yarn in the United States. *PLoS ONE* 6, e28274.
- OIE, 2008. Anthrax in humans and animals.
- Vaissaire, J., Mock, M., Le Doujet, C., Levy, M., 2001. Le charbon bactérien. *Epidémiologie de la maladie en France. Médecine et maladies infectieuses* 31, 257-271.
- Van Ert, M.N., Easterday, W.R., Huynh, L.Y., Okinaka, R.T., Hugh-Jones, M.E., Ravel, J., Zanecki, S.R., Pearson, T., Simonson, T.S., U'Ren, J.M., Kachur, S.M., Leadem-Dougherty, R.R., Rhoton, S.D., Zinser, G., Forlow, J., Coker, P.R., Smith, K.L., Wang, B., Kenefic, L.J., Fraser-Liggett, C.M., Wagner, D.M., Keim, P., 2007. Global Genetic Population Structure of *Bacillus anthracis*. *PLoS ONE*.

Annexe 2 :

Fièvre charbonneuse en 2013 : un foyer en Haute-Marne rappelle que cette zoonose est toujours d'actualité en France

Publié en Septembre 2014 dans Bulletin Epidémiologique Santé animale-alimentation n°63

Fièvre charbonneuse en 2013: un foyer en Haute-Marne rappelle que cette zoonose est toujours d'actualité en France

Guillaume Girault (1), Solveig Kuhse (2), Hubert Vin (3), Régine Marchal-Nguyen (2), Karine Laroucau (1), Sylviane Derzelle (1) (sylviane.derzelle@anses.fr)

(1) Université Paris-Est, Anses, Laboratoire de santé animale, Unité Zoonoses bactériennes, Maisons-Alfort, France

(2) Service Santé et protection animales, Direction de la protection des populations de Haute-Marne, Chaumont, France

(3) Groupe vétérinaire néocastrien, Neufchâteau, France

Résumé

En 2013, la France a connu un foyer confirmé de fièvre charbonneuse (FC) dans le département de la Haute-Marne. Du 22 juillet au 13 août 2013, douze cas de morts suspectes de bovins imputables à la FC ont été signalés dans une exploitation qui avait déjà connu en 1998 un épisode de FC. Tous les animaux morts pâturaient dans un unique pré, distinct de la parcelle incriminée en 1998. Malgré le passé de l'exploitation, la suspicion de FC n'a pas été immédiate et les services vétérinaires du département ont été prévenus de manière très tardive (12 août). Des conditions climatiques exceptionnelles, une période de forte chaleur succédant à un printemps particulièrement pluvieux au cours duquel le pré suspecté est resté inondé de nombreux mois, pourraient expliquer la résurgence. Des spores de *Bacillus anthracis* ont été isolées en faible quantité d'un affleurement d'eau souterraine au centre de la pâture. Les souches isolées sur l'exploitation appartenaient à une branche phylogénétique de la lignée A.Br.011/009 spécifique du Nord-Est de la France. Ce foyer nous rappelle que la vigilance face à cette zoonose grave reste de mise. Il est important d'améliorer la réactivité des vétérinaires en cas de suspicion de FC afin d'établir rapidement un diagnostic et de mettre en place des mesures de protection.

Mots-clés

Fièvre charbonneuse, *Bacillus anthracis*, investigation épidémiologique, génotypage

Abstract

Anthrax in 2013: an outbreak in Haute-Marne is a reminder that this zoonosis is still a topical issue for France

In 2013, a confirmed outbreak of anthrax occurred in the Haute Marne department of France. From 22 July to 13 August 2013, 12 cows died suddenly of anthrax in a beef-cattle farm that had already experienced an anthrax episode in 1998. All the affected animals had been kept on pasture in a single field, but not the same field that was involved in the 1998 outbreak. Despite the farm's anthrax history, a correct diagnosis was not made immediately and the departmental veterinary services were contacted late (12 August). Unusual climatic conditions, i.e. a period of very hot weather following a particularly rainy spring during which the suspected field had been flooded for months, could explain the resurgence of the disease. Spores of *Bacillus anthracis* were isolated in small quantities from a groundwater outcrop area located in the center of the field. Strains isolated in the farm belong to a phylogenetic branch of the A.Br.011/009 lineage that is specific to the northeast of France. This outbreak reminds us that vigilance has to be maintained against this serious disease and that improving veterinarians' reactivity is important in cases of an anthrax outbreak suspicion.

Keywords

Anthrax, *Bacillus anthracis*, Epidemiological investigation, Genotyping

La fièvre charbonneuse (FC) est une zoonose aiguë qui peut atteindre tous les mammifères. Les animaux les plus exposés sont les herbivores (bovins, ovins, caprins, équins) et en particulier les ruminants qui développent des formes suraiguës ou aiguës de la maladie (Encadré). *Bacillus anthracis*, l'agent responsable de la maladie, existe dans la nature sous deux formes: une forme végétative, retrouvée dans l'organisme hôte, et une forme sporulée, présente dans le sol ou les produits d'animaux infectés (os, peau, laine...). Les herbivores contractent généralement la maladie en ingérant des spores au cours de leur pâture (ingestion d'herbe ou de terre contaminée); cette contamination peut également se faire à l'étable, par ingestion d'aliments contaminés (foin en particulier) (OIE, 2008). Chez l'Homme, le contact d'animaux infectés, l'ingestion d'aliments contaminés ou la respiration de poussières contenant des spores peuvent être responsables de la maladie. Les formes cutanées, les plus bénignes, sont néanmoins majoritaires (95 % des cas).

Sous sa forme sporulée, la bactérie peut persister durant de longues années, voire des décennies, dans un sol adéquat (de pH neutre ou légèrement alcalin, à haut degré d'azote et d'ions Ca^{2+}). Si les spores sont hautement résistantes à de nombreux stress (température, dessiccation, pH, radiation, certains désinfectants), il n'en est pas de même de la forme végétative très fragile. La bactérie, qui se multiplie de façon extrêmement rapide dans les organismes hôtes, disparaît très vite après leur mort, concurrencée par les germes de putréfaction. Le sol contaminé par les spores constitue le véritable réservoir de la maladie. Les spores assurent la transmission et les résurgences cycliques de la maladie dans les terres « charbonneuses » (dites « champs maudits »).

La FC a sévit en France pendant des siècles. Le développement d'un vaccin vivant atténué à usage vétérinaire dès la fin du XIX^e siècle et la mise en place concomitante d'un service d'équarrissage,

évitant l'enfouissement des cadavres d'animaux malades et donc la contamination des sols, ont fait progressivement et spectaculairement régresser la maladie au XX^e siècle (Vaissaire et al., 2001). Aujourd'hui, la FC n'apparaît plus que de façon inconstante d'une année sur l'autre dans certaines zones d'élevage ayant connu par le passé des épisodes charbonneux, en particulier dans le Massif Central, les Alpes, la Bourgogne et la Franche-Comté (Calavas et al., 2008; Madani et al., 2009; Vaissaire et al., 2001). Ces épisodes sont généralement de faible importance avec un ou deux foyers et un ou deux animaux morts par foyer. Ils surviennent essentiellement en été au moment de l'apparition de chaleurs (de la fin du printemps à la mi-automne), en général après des périodes pluvieuses (hiver/printemps), suivie de périodes de sécheresse relativement longues. Une pluie soudaine après un épisode de sécheresse favorise tout particulièrement la remontée des spores à la surface (Vaissaire et al., 2001).

Description du foyer de Haute-Marne

Des cas de FC ont été identifiés lors de l'été 2013 en Haute-Marne dans une exploitation mixte laitier et allaitant de plus de 400 bovins. L'enquête dans l'exploitation a permis de reconstruire la chronologie probable du foyer avec une première mortalité imputable à la FC dès le 22 juillet 2013 dans le troupeau allaitant, plus précoce que les premiers diagnostics (16 août) (Tableau 1). Douze bovins sont morts entre le 22 juillet et le 12 août. Le traitement antibiotique des animaux le 12 août, combinée à une surveillance accrue des animaux, s'est avérée efficace pour maîtriser l'épizootie. Aucun autre cas de mortalité n'a été déploré, à l'exception d'une vache laitière qui n'était pas au contact des autres bovins malades. Ce treizième animal a été retrouvé mort le 13 août dans une autre parcelle. L'analyse sanguine s'est révélée négative pour la FC.

Encadré. La biologie de *Bacillus anthracis* en quelques lignes

Agent pathogène

B. anthracis est une bactérie à Gram⁺, aéro-anaérobie sporulante. La bactérie synthétise une capsule polypeptidique aux propriétés anti-phagocytaires et deux exotoxines de type A-B: le facteur Oedémogène (EF) et le Facteur Létal (LF), associé au facteur PA (protective antigen).

La forme végétative (bactérie) se développe et se multiplie rapidement suite à la pénétration de spores dans un organisme (germination en présence de CO₂).

Elle est très fragile et sensible à la compétition des germes de putréfaction (*Proteus*, *Pseudomonas* et bactéries anaérobies). La bactérie disparaît assez rapidement après la mort de l'animal.

La sporulation est rapide en conditions favorables. Elle nécessite une température comprise entre 15 et 42°C, une atmosphère humide et la présence d'oxygène: il y a formation des spores en très grandes quantités lors de l'ouverture d'un cadavre infecté ou d'écoulement de liquides biologiques.

Écologie et Réservoir

Un sol contaminé par des spores constitue un réservoir permanent de la maladie.

Les spores peuvent persister très longtemps dans certains types de sol (géologie de la zone avec sol à pH neutre ou basique et sous-sol calcaire, matières organiques abondantes).

Les spores étant sensibles aux UV, la charge de surface diminue avec le temps. Une remise en surface des spores est nécessaire, mais non suffisante à l'apparition de la maladie.

Les hivers doux ou les épisodes pluvieux importants favorisent la prolifération de parasites (paramphistomes: altération des muqueuses) et d'insectes (taons: transmission passive) en été qui peuvent jouer un rôle dans la propagation de la maladie.

Certains facteurs météo-hydro-géologiques contribuent à la ré-émergence de la maladie en favorisant la remontée de spores: alternance d'épisodes secs et humides avec forte pluviométrie, hydro-géologie de la zone avec présence d'eaux de surface ou de remontées d'eaux souterraines.

La maladie ne se déclare pas chez tous les animaux d'une même pâture. Nécessité d'une porte d'entrée dans l'organisme pour la spore (effraction digestive ou de la sphère respiratoire principalement, éventuellement cutanée par le biais d'une plaie ou de piqûres d'insecte).

Maladie animale et symptômes

La mort survient par septicémie (hyperthermie, formation d'œdèmes, troubles de l'hémostase) et/ou choc exotoxinique.

Par ordre de sensibilité décroissante à l'infection, on trouve les petits ruminants, les grands ruminants et les équidés; viennent ensuite les rongeurs, les lagomorphes et les suidés ainsi que certains oiseaux (autruches, canards de barbarie).

L'évolution de la maladie (formes subaiguë à suraiguë) est fonction de l'espèce animale et de la porte d'entrée du germe (digestive, respiratoire ou cutanée).

Période d'incubation de un à cinq jours.

Lésions pathognomoniques de la maladie mais souvent absentes: sang incoagulable, hémorragies, rate hypertrophiée à pulpe « boueuse », absence de rigidité cadavérique. Lésions anatomo-pathologiques fréquentes. Tableau nécropsique possiblement très incomplet dans les formes suraiguës (pas de lésions macroscopiques ou présence d'hémorragies seulement).

Chez les ruminants, la forme suraiguë septicémique est prédominante. Mort brutale sans prodrome fréquente. La FC peut être aisément confondue avec une septicémie à *Clostridium* sp.

Vaccination

Le seul moyen de prévention vis-à-vis de la FC est la vaccination. Pour être efficace, les animaux à protéger doivent être vaccinés au cours de l'hiver et au moins quinze jours avant la mise à l'herbe. Les rappels se font tous les six mois dans les cheptels « à risque ».

Aucun vaccin n'est produit en France. L'importation du vaccin doit se faire avec une autorisation spéciale de l'Agence nationale du médicament vétérinaire (ANMV). Il s'agit d'un vaccin vivant atténué (souche Sterne).

Tableau 1. Chronologie de l'épisode 2013

Dates	Événement
22/07	1 ^{re} mortalité imputable à la FC dans l'élevage allaitant (en pâturage sur îlot 16)
02/08	4 ^e mortalité brutale dans le même pré (trois vaches et un veau). Autopsie d'une vache Charolaise sur une plateforme de l'exploitation: suspicion d'infection à Clostridies
09/08	Le nombre de vaches Charolaises mortes dans le même pré s'élève à 10. Autopsie d'une vache qui vient de vêler et suspicion de FC, salmonellose ou infection à Clostridies; prélèvement de la rate (FC +) et de contenu de l'intestin grêle pour analyse
12/08	12 ^e mortalité brutale. Signalement de suspicion de FC (DDecPP, LVD52 et LNR contacté). Arrêté préfectoral de mise sous surveillance (APMS). Mise sous antibiotiques du cheptel
13/08	Mortalité d'une vache laitière ayant vêlé la veille dans un autre pré, prélèvement de sang (négatif FC). Intervention des services vétérinaires (inspection terrain, mise en place des mesures de nettoyage et de désinfection pour les locaux, le matériel et les véhicules, réponse aux questions de l'éleveur)
16/08	Confirmation par PCR de FC pour la vache autopsiée (sur rate prélevée le 9 août), résultat négatif pour la vache laitière (sur sang prélevé le 13 août)
20/08	Arrêté préfectoral portant déclaration d'infection d'une exploitation pour FC (APDI)
21/08	Prélèvements de sol et d'eau sur l'îlot 16. Prise de sang sur le veau né de la vache morte le 9 août: négatif en FC. L'animal est toujours vivant et en bonne santé.
Du 23 au 29/08	Vaccination de 153 animaux de l'exploitation

La fabrication du vaccin contre la FC n'est plus assurée en France. Le stock de vaccins disponibles sur le territoire au moment du foyer était très faible, inférieur à 500 doses. Cent cinquante doses ont été commandées le 17 août à la centrale d'achat. Le vaccin (composé d'une souche vaccinale atténuée vivante) a été acheminé par le transporteur sans respecter la chaîne du froid et a, par conséquence, été refusé à réception (21 août). Les dernières doses disponibles (n=25) ont été reçues le 22 août et les animaux du pré suspect ont été vaccinés en priorité. Une demande d'importation de vaccins a été faite en urgence pour assurer la vaccination du reste du cheptel.

Investigations vétérinaires

Une première autopsie a été réalisée le 2 août 2013 sur le cadavre de la quatrième vache morte au pré. Celui-ci présentait des lésions de pneumonie, une entérite légèrement hémorragique, une hématurie, de nombreuses pétéchies thoraciques et abdominales, et une putréfaction rapide. L'autopsie a conclu à une mort par septicémie de cause inconnue. Une revaccination contre les infections à Clostridies des animaux du pré où pâturent les vaches allaitantes a été effectuée.

Une seconde autopsie a été pratiquée sur le dixième cadavre (9 août 2013). Les lésions constatées étaient les suivantes: splénomégalie avec décoloration et rate boueuse, entérite hémorragique, pneumonie sévère (hépatisation et abcès), pétéchies importantes sur la paroi thoracique et moins importante sur la paroi abdominale, absence de caillot intracardiaque. Sur base de ces signes, une suspicion de FC et/ou salmonellose et/ou d'infection à Clostridies, est posée. La rate du bovin a été prélevée pour analyse, ainsi que les selles de l'intestin grêle. Ces prélèvements ont été transmis le 12 août au LVD42.

L'anamnèse du treizième et dernier animal, une vache laitière retrouvée morte dans un pré distinct (13 août 2013), était différente. La prise de sang effectuée s'est révélée négative pour la FC.

Tableau 2. Compte rendu et chronologie de l'épisode FC de 1998

Dates	Événement
20/07	1 ^{re} mortalité brutale vache Pie Noire morte en pâture Animal en lactation ne présentant aucun symptôme le matin même.
22/07	2 ^e mortalité brutale dans le même pré d'une Charolaise en attente de vêlage Examen post-mortem du cadavre (animal météorisé, muqueuses oculaire et vaginale cyanosées, écoulement sanguinolent des narines) Suspicion d'entérototoxicité
23/07	3 ^e mortalité brutale d'une vache Charolaise; 1 ^{re} autopsie à l'équarrissage (lésions internes au niveau intestinal et hépatique) Vaccination totale du cheptel contre les Clostridies
Début août	Mortalité brutale d'une vache Montbéliarde (paraplégie, mort en six minutes) et d'une jeune jument de selle Autopsie des deux animaux (tissus sous-cutanés hémorragiques, météorisation, muqueuse cyanosées, écoulement hémorragique nasal, rupture de l'estomac de la jument suite à des coliques violentes)
Août-septembre	Deux nouvelles mortalités de vaches laitières
20/10	Mortalité à l'étable d'un jeune taurillon croisé Pie Noire-Charolais Autopsie à la ferme (météorisation, écoulement sanguinolent, rate dilatée et très épaissie avec plage œdématisée, intestins violacés), prélèvements pour analyse FC (résultats +) Mortalité d'une vache laitière dans une autre pâture, ouverture du cadavre sur un chemin de remembrement (météorisation gazeuse importante, rate et foie normaux), prélèvements pour analyse FC négatifs
23/10	Foyer de FC confirmé; arrêté préfectoral portant déclaration d'infection d'une exploitation pour FC
Du 27 au 31/10	Vaccination de 280 animaux

Investigations épidémiologiques

Le diagnostic de FC ayant tardé, les cadavres des animaux du foyer ont été livrés à l'équarisseur et détruits sans précaution particulière. Un suivi des personnes à risque (éleveurs, vétérinaires, personnels de l'équarrissage et laborantins) a été mis en place. Une dizaine de personnes ont été placées sous traitement antibiotique. Aucun cas humain de FC n'a été associé à ce foyer. La Fédération départementale des chasseurs (FDC) et l'Office national de la chasse et de la faune sauvage (ONCFS) ont été alertés, mais n'ont pas fait part d'observation d'une mortalité de la faune sauvage augmentée dans le secteur concerné.

Un précédent foyer de FC avait été déclaré en 1998 dans une autre parcelle de la même exploitation. Durant cet épisode, qui s'était échelonné du 20 juillet au 23 octobre 1998, neuf animaux de différentes espèces et catégories (laitières, allaitantes, taurillon et cheval) étaient morts subitement (Tableau 2). Suite aux examens *post-mortem* et autopsies pratiquées en juillet 1998 sur les cadavres successifs d'animaux, un diagnostic d'entérototoxicité fut à plusieurs reprises posé et le cheptel bovin vacciné contre les Clostridies. La présence du germe de la FC ne fut suspectée que très tardivement (le 22 octobre 1998). Toutes les bêtes, à l'exception d'un taurillon de six mois mort à l'étable, avaient pâture dans un même pré, suspecté d'être contaminé.

À l'avenir, il a été proposé à l'éleveur que le cheptel soit vacciné régulièrement contre la FC avec une prise en charge financière partielle de la part du GDS.

Les facteurs pouvant expliquer la ré-émergence de la maladie dans cette exploitation ont été recherchés. Plusieurs conditions propices à la remontée en surface des spores ont été évoquées. La pâture concernée est une prairie au réseau aquifère superficiel présentant une zone humide en son centre où affleure de l'eau souterraine (foletière). Deux sources alimentent en eau l'abreuvoir. Le puits d'une de celles-ci a été ré-ouvert en 2013. Le piétinement important des bovins observés sur cette parcelle a pu aider à remuer et disséminer les spores de *B. anthracis*. La remontée en surface des spores aurait également pu être favorisée à l'occasion d'un retournement du sol. Des travaux de drainage (pose de drains) ayant été effectués dans un champ à proximité de la pâture deux ans auparavant, des prélèvements ont aussi été effectués à ce niveau, bien que le délai entre ces travaux et la survenue des cas en fasse une hypothèse très peu probable. Enfin, les conditions climatiques particulières rencontrées en 2013 ont également pu jouer un rôle. Suite à un printemps particulièrement pluvieux, le pré incriminé est resté inondé de nombreux mois, avant la mise en pâture des animaux, ce qui a pu favoriser la remontée de spores.

Investigations microbiologique et environnementale

La rate de bovin prélevée le 9 août 2013 a été adressée au LNR le 13 août pour recherche et identification de *B. anthracis* par culture et PCR multiplexe spécifique. La rate présentait un aspect « boueux » avec du sang noir non coagulé. L'analyse PCR de l'organe a confirmé la suspicion de FC (16 août). Un isolement de colonies caractéristiques à *B. anthracis* a été obtenu et l'espèce bactérienne confirmée le 19 août. Les deux souches de *B. anthracis* isolées ont été génotypées à l'Anses et fait ensuite l'objet d'une caractérisation génétique fine (cf. ci-dessous).

Le 21 août, une investigation conjointe des services vétérinaires de la DDecPP de Haute-Marne et du LNR a été menée avec la participation du vétérinaire sanitaire de l'élevage. Des prélèvements de terre et d'eau sur les sites d'exposition suspectés ont été réalisés afin de rechercher les sources potentielles de contamination. La présence de spores aux emplacements des cadavres des animaux a également été recherchée. Trois échantillons ont été prélevés sur huit sites différents: sur la plateforme bétonnée de l'exploitation où l'autopsie du 9 août a été faite; sur quatre sites de découverte de cadavres (dans le pré et dans le sous-bois attenants); en surface des travaux de drainage ayant eu lieu à proximité du pré incriminé; dans la zone d'affleurement de l'eau souterraine du pré (foletière); au niveau des deux tuyaux alimentant l'abreuvoir en eau de source (dont l'un des puits avait été ré-ouvert dans l'année).

Tableau 3. Résultats d'analyse des prélèvements de terre

	Quantité de spores de <i>B. anthracis</i> isolées dans 0,15 g de matrice		
Site de prélèvement	Échantillon 1	Échantillon 2	Échantillon 3
Plateforme d'autopsie	1-5	> 100	1000
Pré, cadavre 1	Négatif *	100 *	Négatif *
Pré, cadavre 2	0-3	Négatif *	Négatif *
Orée du bois, cadavre 3	100	10	1000
Sous-bois, cadavre 4	0-3	Négatif	Négatif
Drain	Négatif	Négatif	Négatif
Eau de source (abreuvoir)	Négatif	Négatif	
Foletière	Négatif *	1	Négatif

*présence de spores caractéristiques du groupe *Bacillus cereus* (colonies hémolytiques)

La plupart des sites de prélèvement se sont avérés contenir des spores de *B. anthracis* dans des proportions plus ou moins variables d'un échantillon à l'autre (Tableau 3). Seules l'eau de source alimentant l'abreuvoir et la terre prélevée au niveau du drain sont restées négatives. Il faut néanmoins garder à l'esprit que l'échantillonnage réalisé étant très limité, il ne permet pas de conclure avec certitude à l'absence de contamination en spores de ces deux sites. Le lieu le plus chargé en spores, et ce malgré l'épandage de chaux effectué par l'éleveur les jours précédents, était la plateforme bétonnée sur laquelle avait été réalisée une des autopsies.

L'expertise menée est en faveur d'une résurgence de spores liée aux conditions météorologiques et à un sol déjà contaminé et favorable à la conservation des spores. La foletière est la seule source environnementale dans laquelle la présence de spores a pu être mise en évidence. La contamination détectée y était faible, probablement très hétérogène, mais représente néanmoins une source d'exposition très probable.

Typage des souches de Haute-Marne

L'Anses a récemment réalisé une cartographie à très haute résolution des génotypes de *B. anthracis* présents sur le territoire français. Un outil de typage moléculaire par PCR permettant une identification fine des isolats français en a découlé (Girault *et al.*, 2014).

L'ensemble des souches isolées sur l'exploitation de Haute-Marne en 2013 (n = 8) et 1998 (n = 2) appartiennent à la lignée trans-eurasienne A.Br.011/009 incluant 30 % des souches françaises de *B. anthracis*. Elles se positionnent au sein de la branche phylogénétique française 2 (clade NE) qui regroupe sept souches de la base de données de séquence génomique de l'Anses isolées dans l'extrême Nord-Est de la France (Girault *et al.*, 2014). Sa localisation territoriale, incluant les départements de Moselle (57), Meurthe-et-Moselle (54), Marne (51), Haute-Saône (70) et du Bas-Rhin (67). Tous les isolats de Haute-Marne possèdent la mutation ponctuelle du marqueur SNP BA22 qui définit cette branche, en accord avec l'origine géographique du département.

L'analyse génétique comparative avec la souche séquencée de *B. anthracis* isolée dans la Marne en 1990 indique que les souches isolées de Haute-Marne sont distinctes. Plusieurs marqueurs spécifiques de la souche de la Marne ne se retrouvent pas dans les isolats des foyers de Haute-Marne. Le séquençage global d'une de ces souches est en cours. Il devrait contribuer à documenter le caractère supposé clonal des souches impliquées dans ces deux foyers (1998, 2013) et permettre l'identification de signatures ADN (SNP) discriminantes pour ces souches. Un examen approfondi des dix souches isolées de ces deux foyers devraient également nous fournir des informations importantes quant à l'hétérogénéité des spores présentes sur une exploitation et leur évolution génétique entre deux épisodes espacés de quinze ans.

Conclusions et recommandations

La cause de la résurgence de la FC en Haute-Marne en 2013 semble devoir être attribuée aux conditions climatiques particulières de l'année et aux caractéristiques hydro-géologiques propices du terrain. Les épisodes pluvieux importants associés à l'inondation prolongée des terres de pâturage permettent la prolifération de parasites (paramphistomes) et d'insectes (taons) qui peuvent jouer un rôle dans la déclaration ou la propagation de la maladie (Vaissaire *et al.*, 2001). En cas d'ingestion de spores de *B. anthracis* chez les ruminants (une hypothèse plausible pour un cheptel pâturant en « champs maudits »), des facteurs extérieurs, comme des affections digestives ou respiratoires ou du parasitisme qui pourraient provoquer des irritations ou des lésions des muqueuses, sont nécessaires à l'expression de la maladie.

Ce cas de FC nous rappelle que la maladie est de nature à resurgir périodiquement dans les zones contaminées. Or, force est de constater qu'avec le temps, la FC est devenu pour certains une maladie « du passé » tombée dans l'oubli. La FC ne fait souvent pas

partie des hypothèses diagnostiques classiques envisagées face à une hyperthermie, une affection respiratoire, digestive ou une mortalité brutale inexpliquée ou faisant penser à de l'entérototoxicité et n'est quasiment jamais envisagée face à des troubles nerveux (Vaissaire *et al.*, 2001). Cette perte de connaissance et d'expérience conduit à une sous-identification de la maladie et à un diagnostic tardif, avec les conséquences que cela implique. Rappelons qu'il peut y avoir suspicion de FC dans les cas suivants : mort brutale même en absence de lésions macroscopiques à l'autopsie, forte fièvre, hémorragies quelle qu'en soit l'origine (avec ou sans absence de coagulation), troubles respiratoires, réactions oedémateuses de taille importante. Face à deux cas de morts subites inexpliquées, la suspicion FC devrait être automatique, y compris en bâtiment (cas de FC lié à l'ingestion de foin contaminé). Conserver la mémoire des épisodes survenus dans une région et cartographier les terrains sur lesquels des animaux sont morts de FC sont également très importants (Anses, 2010).

En cas de suspicion de FC, un prélèvement de sang sur tube sec (sans anticoagulant) sur l'animal malade ou mort est recommandé. Afin d'améliorer la sensibilité du diagnostic, les prélèvements doivent être réalisés avant la mise en œuvre de l'antibiothérapie et de la vaccination. Les animaux morts ne doivent en aucun cas être autopsiés sur place. La DDéCP doit être prévenue immédiatement, ainsi que le laboratoire départemental et le service de l'équarrissage (protection du personnel et traitement des cadavres). Dans tous les cas, le prélèvement doit parvenir au plus vite au laboratoire car la qualité des résultats en dépend. Le transport doit être effectué par un transporteur spécialisé et en triple emballage. L'envoi est pris en charge par l'État dans le cadre de la police sanitaire.

Remerciements

Les auteurs remercient l'éleveur et les agents de la DDASS qui ont contribué à cette étude, ainsi que R. Aaziz (Anses) pour son assistance technique. Guillaume Girault est un doctorant co-financé par l'Anses et la Direction générale de l'armement (DGA).

Références

- Anses, 2010. Avis de l'Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail relatif à une demande d'avis sur les mesures de gestion en santé animale et en sécurité sanitaire des aliments lors de suspicions et de confirmations de cas de fièvre charbonneuse. SA-007.
- Calavas, D., Sala, C., Vaissaire, J., Condé, J., Thien-Aubert, H., Hessemann, M., Woronoff-Rehn, N., 2008. Retour d'expérience sur un épisode de fièvre charbonneuse chez les bovins dans le Doubs au cours de l'été 2008. Bull. Epid. Santé Anim. Alim. 32 32, 1-6.
- Girault, G., Blouin, Y., Vergnaud, G., Derzelle, S., 2014. High-throughput sequencing of *Bacillus anthracis* in France: investigating genome diversity and population structure using whole-genome SNP discovery. BMC genomics 15.
- Madani, N., Mendy, C., Moutou, F., Garin-Bastuji, B., 2009. La fièvre charbonneuse en France. Épidémiologie de l'été 2009 et foyers enregistrés sur la dernière décennie (1999-2009). Bull. Epid. Santé Anim. Alim 38, 1-3.
- OIE, 2008. Anthrax in humans and animals.
- Vaissaire, J., Mock, M., Le Doujet, C., Levy, M., 2001. Le charbon bactérien. Épidémiologie de la maladie en France. Médecine et maladies infectieuses 31, 257-271.

Annexe 3 :

Liste des souches et génomes utilisés au cours de l'étude

Name	Country	Year of isolation	Host	canSNP
2861	Germany	NR	NR	A.Br.001/002
2862	Germany	NR	NR	A.Br.001/002
2865	Germany	NR	NR	A.Br.001/002
11_4	France	2011	Bovine	A.Br.001/002
8_07	France	2008	Bovine	A.Br.001/002
8_09	France	2008	Bovine	A.Br.001/002
8_10	France	2008	Bovine	A.Br.001/002
8_11	France	2008	Bovine	A.Br.001/002
8_13	France	2008	Bovine	A.Br.001/002
8_14	France	2008	Bovine	A.Br.001/002
8_15	France	2008	Bovine	A.Br.001/002
8_16	France	2008	Bovine	A.Br.001/002
8_19	France	2008	Bovine	A.Br.001/002
8_20	France	2008	Bovine	A.Br.001/002
8_21-1	France	2008	Bovine	A.Br.001/002
8_21-2	France	2008	Bovine	A.Br.001/002
8_21-3	France	2008	Bovine	A.Br.001/002
8_22/2	France	2008	Bovine	A.Br.001/002
8_24	France	2008	Bovine	A.Br.001/002
8_25	France	2008	Bovine	A.Br.001/002
8_26	France	2008	Bovine	A.Br.001/002
8_27	France	2008	Bovine	A.Br.001/002
8_28	France	2008	Bovine	A.Br.001/002
8_29	France	2008	Bovine	A.Br.001/002
A0389	Indonesia	1989	NR	A.Br.001/002
A16	China	NR	NR	A.Br.001/002
VAR 85/1665	Belgium	1978	Bovine	A.Br.001/002
VAR 83/1653	Belgium	1983	Bovine	A.Br.001/002
Ba190	Bangladesh	2011	Soil	A.Br.001/002
Ba316	Bangladesh	2011	Soil	A.Br.001/002
Ba425	Bangladesh	2011	Soil	A.Br.001/002
Ba400	Bangladesh	2011	Soil	A.Br.001/002
Ba321	Bangladesh	2011	Soil	A.Br.001/002
Ba325	Bangladesh	2011	Soil	A.Br.001/002
Ba333	Bangladesh	2011	Soil	A.Br.001/002
Ba351	Bangladesh	2011	Soil	A.Br.001/002
Ba385	Bangladesh	2011	Soil	A.Br.001/002
Ba395	Bangladesh	2011	Soil	A.Br.001/002
Ba415	Bangladesh	2011	Soil	A.Br.001/002
BL8	Poland	NR	NR	A.Br.001/002
BL9	Poland	NR	NR	A.Br.001/002
BL9 hem	Poland	NR	NR	A.Br.001/002
CIP 53.169	France	1953	NR	A.Br.001/002
CIP 81.89	France	1981	NR	A.Br.001/002
CIP A211	France	1954	NR	A.Br.001/002
K929	Denmark	1966	NR	A.Br.001/002
A	Denmark	1935	NR	A.Br.001/002
B	Denmark	1935	NR	A.Br.001/002
C	Denmark	1960	NR	A.Br.001/002
Ba103	Japan	1991	NR	A.Br.001/002
NCTC 00109	UK	1920	NR	A.Br.001/002
NCTC 02620	UK	1928	NR	A.Br.001/002
NCTC 05444	UK	1938	NR	A.Br.001/002
CVI-128268	Netherlands	1989	Outbreak	A.Br.001/002
CVI-131959-5	Netherlands	1991	Outbreak	A.Br.001/002
CVI-13185	Netherlands	1991	Outbreak	A.Br.001/002
CVI-unknown-1	Netherlands	1968	NR	A.Br.001/002
CVI-132064-1	Netherlands	1991	Outbreak	A.Br.001/002
CVI-127491-V	Netherlands	NR	NR	A.Br.001/002
Sterne	South Africa	NR	NR	A.Br.001/002

BFV	NR	NR	NR	A.Br.001/002
Han	China	2012	Bovine	A.Br.001/002
V770-NP1-R	Israel	1954	NR	A.Br.003/004
2054 HB2	Belgium	NR	NR	A.Br.003/004
SVA 2008	Sweden	2008	NR	A.Br.003/004
IEMVT89	Cameroon	1989	NR	A.Br.005/006
ULG 197 A2/19	Belgium	NR	NR	A.Br.005/006
CZC5	Zambia	NR	NR	A.Br.005/006
E	Denmark	1937	NR	A.Br.005/006
K836	Denmark	1968	NR	A.Br.005/006
Ba050	Italy	1995	Animal	A.Br.005/006
JF3966	Cameroon	2007	Cattle	A.Br.005/006
JF4685	Chad	1996-2003	NR	A.Br.005/006
JF4688	Chad	1996-2003	NR	A.Br.005/006
JF4691	Chad	1996-2003	NR	A.Br.005/006
ULG 237 A2/16	Belgium	NR	NR	A.Br.008/011
ULG 290 A2/17	Belgium	NR	NR	A.Br.008/011
ULG 239 A2/18	Belgium	NR	NR	A.Br.008/011
ULG 238 A2/20	Belgium	NR	NR	A.Br.008/011
ULG 241 A2/9	Belgium	NR	NR	A.Br.008/011
ULG 236 A2/10	Belgium	NR	NR	A.Br.008/011
ULG 238 A2/11	Belgium	NR	NR	A.Br.008/011
BL6	Poland	1996	NR	A.Br.008/011
Ba 3154	Bulgaria	NR	Soil	A.Br.008/011
Ba 3166	Bulgaria	NR	Soil	A.Br.008/011
K169/66	Denmark	1966	NR	A.Br.008/011
D	Denmark	1973	NR	A.Br.008/011
H18	Hungary	NR	NR	A.Br.008/011
H20	Hungary	NR	NR	A.Br.008/011
H28	Hungary	NR	NR	A.Br.008/011
H55	Hungary	NR	NR	A.Br.008/011
H65	Hungary	NR	NR	A.Br.008/011
Ba4599	Scotland	2009	Human	A.Br.008/011
UR-1	Germany	2012	Human	A.Br.008/011
Ba152	Italy	2009	Animal	A.Br.008/011
Ba180	Italy	2009	Animal	A.Br.008/011
Ba007	Italy	1996/1997	Animal	A.Br.008/011
NCTC 07753	UK	1948	NR	A.Br.008/011
CVI-unknown2	Netherlands	1976	NR	A.Br.008/011
Tsiankovskii-1	USSR	NR	NR	A.Br.008/011
Smith 1013	NR	NR	NR	A.Br.008/011
A0897	Ethiopia	2000	Donkey	A.Br.008/011
A0149	Turkey	1991	Human	A.Br.008/011
Cvac02	China	2012	NR	A.Br.008/011
30	France	2007	Bovine	A.Br.011/009
32	France	2007	Bovine	A.Br.011/009
34	France	2007	Bovine	A.Br.011/009
86	France	1982	Sheep	A.Br.011/009
87	France	1982	NR	A.Br.011/009
88	France	1990	Bovine	A.Br.011/009
89	France	1984	NR	A.Br.011/009
90	France	1986	NR	A.Br.011/009
92	France	1989	Bovine	A.Br.011/009
93	France	1990	NR	A.Br.011/009
94	France	1991	Bovine	A.Br.011/009
95	France	1991	Bovine	A.Br.011/009
99	France	2000	Bovine	A.Br.011/009
100	France	1999	Bovine	A.Br.011/009
101	France	1999	Ovine	A.Br.011/009
102	France	1999	Bovine	A.Br.011/009
103	France	1999	NR	A.Br.011/009

104	France	1999	NR	A.Br.011/009
116	France	1996	Bovine	A.Br.011/009
118	France	1996	NR	A.Br.011/009
122	France	1992	NR	A.Br.011/009
124	France	2008	Bovine	A.Br.011/009
136	France	2003	Bovine	A.Br.011/009
143	France	2000	Bovine	A.Br.011/009
11_5	France	2011	Goat	A.Br.011/009
17 JB	France	NR	NR	A.Br.011/009
9_08	France	2009	Bovine	A.Br.011/009
9_13	France	2009	Bovine	A.Br.011/009
Carbosap	Italy	2013	Vaccine	A.Br.011/009
CIP 74.12	France	1974	NR	A.Br.011/009
CIP A204	France	NR	Cobaye	A.Br.011/009
CIP A205	France	NR	NR	A.Br.011/009
CIP A206	France	1954	Pustule	A.Br.011/009
Gmb1	Gambia	NR	NR	A.Br.011/009
Ba005	Italy	NR	Animal	A.Br.011/009
Ba137	Italy	2006	Animal	A.Br.011/009
Ba157	Italy	2009	Animal	A.Br.011/009
Ba297	Italy	2011	Animal	A.Br.011/009
Ba315	Italy	2011	Animal	A.Br.011/009
Ba430	Italy	2012	Animal	A.Br.011/009
Ba010	Italy	1996-1997	Animal	A.Br.011/009
Ba031	Italy	1984	Animal	A.Br.011/009
Ba039	Italy	1972	Animal	A.Br.011/009
Ba019	Italy	1999	Animal	A.Br.011/009
Ba040	Italy	1984	Animal	A.Br.011/009
Ba048	Italy	2000	Animal	A.Br.011/009
Ba051	Italy	NR	Animal	A.Br.011/009
Ba053	Italy	NR	Animal	A.Br.011/009
Ba079	Italy	2004	Animal	A.Br.011/009
NCTC 01328	UK	1922	NR	A.Br.011/009
CVI-260187	Netherlands	1987	NR	A.Br.011/009
Butare	Congo	NR	NR	A.Br.011/009
Sen2 Col2	Senegal	NR	NR	A.Br.011/009
Sen3	Senegal	NR	NR	A.Br.011/009
95014	NR	NR	NR	A.Br.011/009
A0248	USA	NR	Human	A.Br.Ames
A2012	USA	2001	Human	A.Br.Ames
Ames	USA	2003	NE	A.Br.Ames
Ames Ancestor	USA	1981	Animal	A.Br.Ames
K670/88	Denmark	1988	Spleen	A.Br.Ames
K35/88	Denmark	1988	Spleen	A.Br.Ames
STI-1	Georgia	NR	NR	A.Br.Ames
2863	Germany	NR	NR	A.Br.Aust94
2864	Germany	NR	NR	A.Br.Aust94
2866	Germany	NR	NR	A.Br.Aust94
52_G	Georgia	NR	NR	A.Br.Aust94
8903_G	Georgia	NR	NR	A.Br.Aust94
9080_G	Georgia	NR	NR	A.Br.Aust94
Australia 94	Australia	NR	NR	A.Br.Aust94
1409	Denmark	1974	NR	A.Br.Aust94
JF3853	Switzerland	1952	Cattle	A.Br.Aust94
CVI-56430	Netherlands	1987	Bovine	A.Br.Aust94
CVI-188678-1	Netherlands	1993	Outbreak	A.Br.Aust94
A.Br003	NR	NR	NR	A.Br.Aust94
H9401	Korea	1994	Human	A.Br.H9401
A0488	UK	1935	Cattle	A.Br.Vollum
AF1	Afghanistan	NR	Cashmeer	A.Br.Vollum
AF2	Afghanistan	NR	Cashmeer	A.Br.Vollum

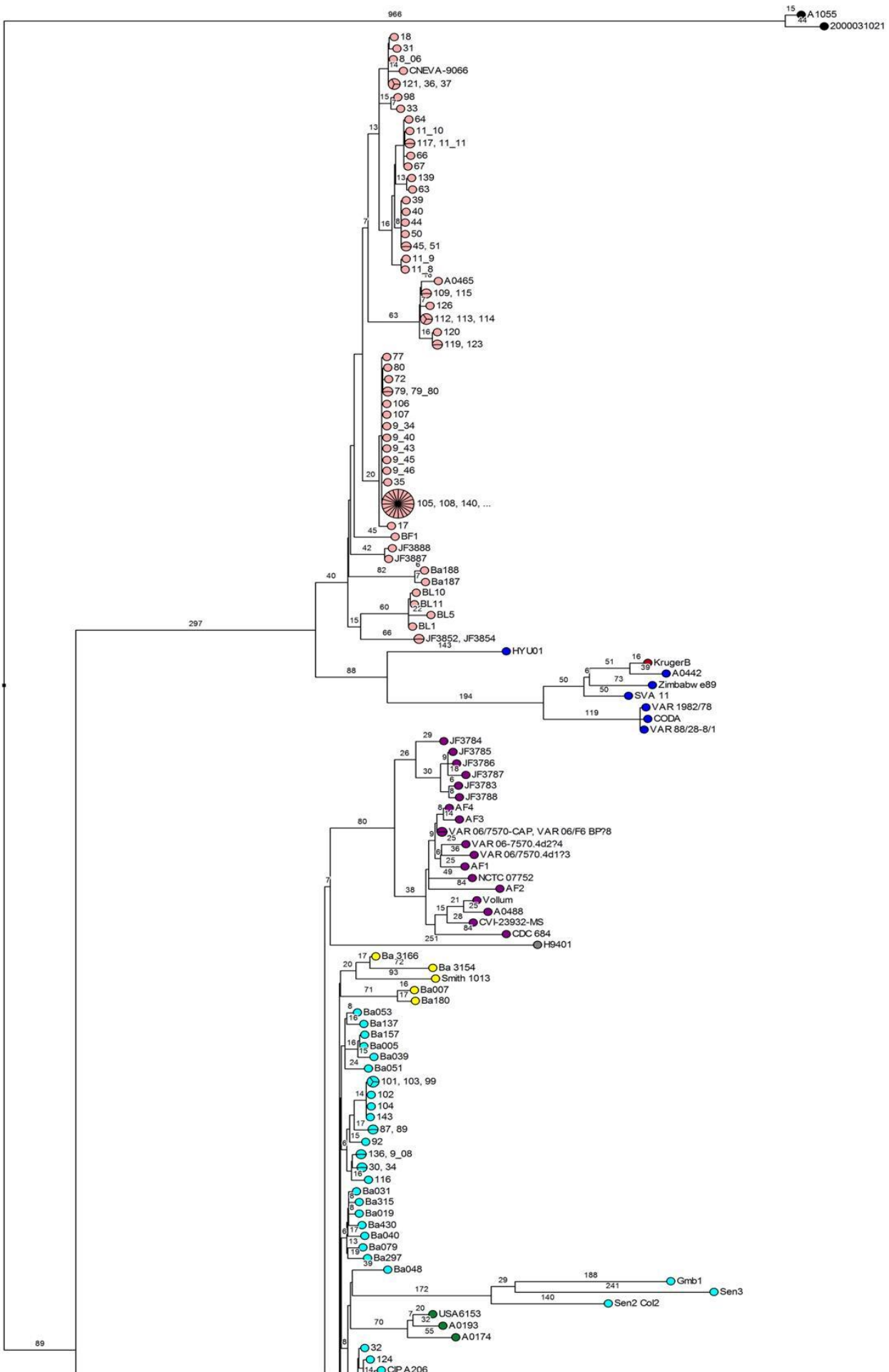
AF3	Afghanistan	NR	Cashmeer	A.Br.Vollum
AF4	Afghanistan	NR	Cashmeer	A.Br.Vollum
CDC 684	USA	NR	NR	A.Br.Vollum
JF3783	Pakistan	1981	Human	A.Br.Vollum
JF3784	Pakistan	1981	Air	A.Br.Vollum
JF3785	Pakistan	1981	Cashmeer	A.Br.Vollum
JF3786	Pakistan	1981	Cashmeer	A.Br.Vollum
JF3787	Pakistan	1981	Cashmeer	A.Br.Vollum
JF3788	Pakistan	1981	Human	A.Br.Vollum
NCTC 07752	UK	1948	NR	A.Br.Vollum
CVI-23932-MS	Netherlands	1989	Canine	A.Br.Vollum
VAR 06/F6 BP#8	Russia-Afghanistan-Iran	NR	Cashmeer	A.Br.Vollum
VAR 06/7570-CAP	Russia-Kazakhstan	NR	Cashmeer	A.Br.Vollum
VAR 06-7570.4d2#4	Russia-Kazakhstan	NR	Cashmeer	A.Br.Vollum
VAR 06/7570.4d1#3	Russia-Kazakhstan	NR	Cashmeer	A.Br.Vollum
Vollum	USA	NR	NR	A.Br.Vollum
A0174	Canada	NR	NR	A.Br.WNA
A0193	USA	NR	Bovine	A.Br.WNA
USA6153	USA	NR	NR	A.Br.WNA
A0442	South Africa	NR	Antelope	B.Br.001/002
VAR 1982/78	Belgium	1978	NR	B.Br.001/002
VAR 88/28-8/1	Belgium	1988	Bovine	B.Br.001/002
CODA	Belgium	NR	NR	B.Br.001/002
SVA 2011	Sweden	2011	NR	B.Br.001/002
HYU01	Korea	NR	Soil	B.Br.001/002
Zimbabwe89	NR	NR	NR	B.Br.001/002
17	France	2006	Bovine	B.Br.CNEVA
18	France	2007	Bovine	B.Br.CNEVA
31	France	2007	Bovine	B.Br.CNEVA
33	France	2007	NR	B.Br.CNEVA
35	France	2007	Bovine	B.Br.CNEVA
36	France	2006	Bovine	B.Br.CNEVA
37	France	2006	Bovine	B.Br.CNEVA
39	France	2005	Bovine	B.Br.CNEVA
40	France	2005	Bovine	B.Br.CNEVA
41	France	2005	Bovine	B.Br.CNEVA
44	France	2004	Bovine	B.Br.CNEVA
45	France	2004	Bovine	B.Br.CNEVA
50	France	2004	Bovine	B.Br.CNEVA
51	France	2004	Bovine	B.Br.CNEVA
63	France	2003	Bovine	B.Br.CNEVA
64	France	2003	Bovine	B.Br.CNEVA
66	France	2003	Bovine	B.Br.CNEVA
67	France	2003	NR	B.Br.CNEVA
72	France	2000	NR	B.Br.CNEVA
73	France	2000	Bovine	B.Br.CNEVA
75	France	2000	Bovine	B.Br.CNEVA
77	France	2000	Bovine	B.Br.CNEVA
79	France	2000	Bovine	B.Br.CNEVA
80	France	2000	Bovine	B.Br.CNEVA
82	France	2000	Bovine	B.Br.CNEVA
97	France	2000	Animal	B.Br.CNEVA
98	France	2000	Animal	B.Br.CNEVA
105	France	1997	Bovine	B.Br.CNEVA
106	France	1997	Bovine	B.Br.CNEVA
107	France	1997	Bovine	B.Br.CNEVA
108	France	1997	Bovine	B.Br.CNEVA
109	France	1997	Bovine	B.Br.CNEVA
112	France	1997	Bovine	B.Br.CNEVA
113	France	1997	Bovine	B.Br.CNEVA
114	France	1997	Bovine	B.Br.CNEVA

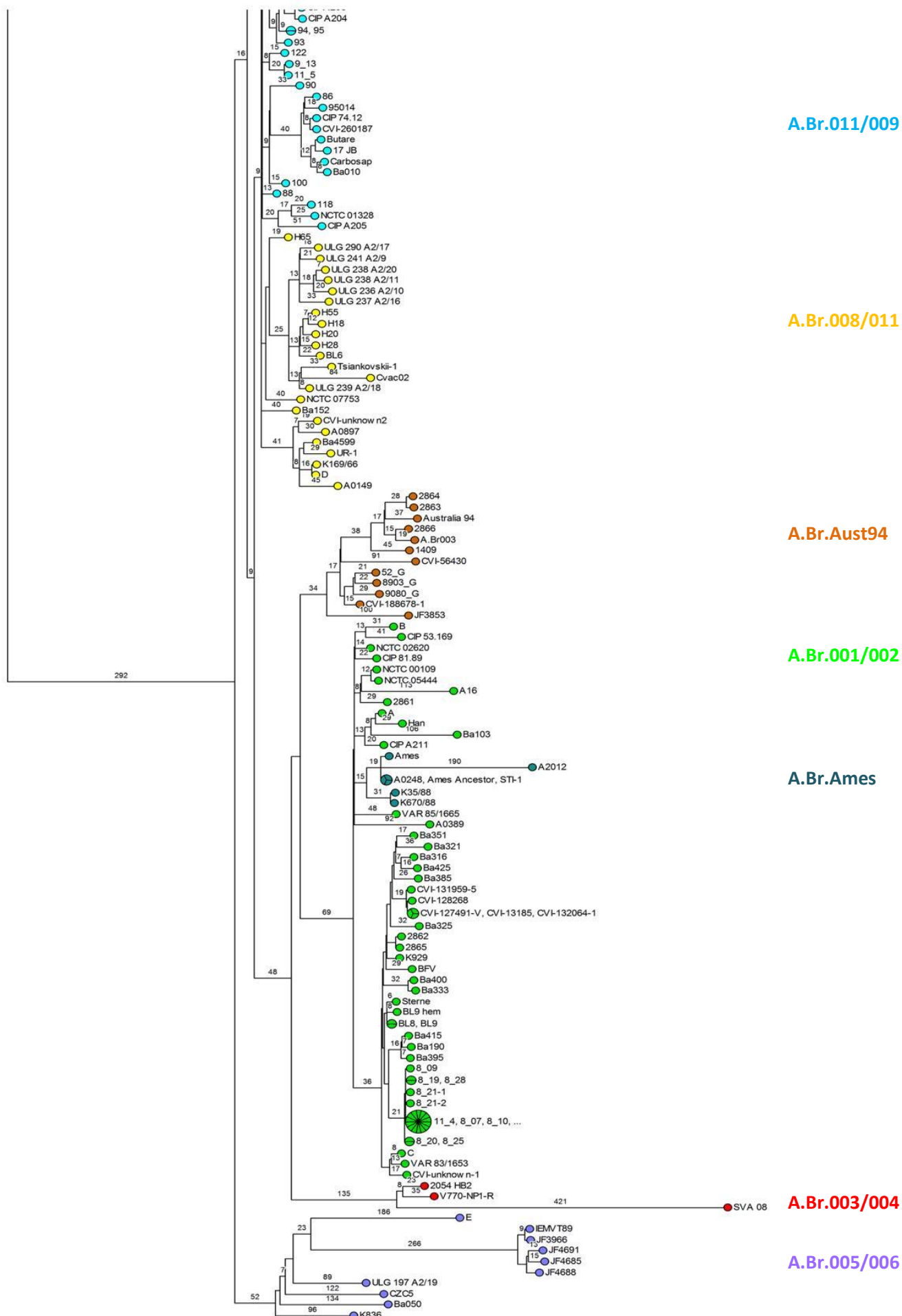
115	France	1997	Bovine	B.Br.CNEVA
117	France	2003	Bovine	B.Br.CNEVA
119	France	1997	Bovine	B.Br.CNEVA
120	France	1994	Bovine	B.Br.CNEVA
121	France	1992	NR	B.Br.CNEVA
123	France	1992	Bovine	B.Br.CNEVA
126	France	1997	Soil	B.Br.CNEVA
139	France	2001	Bovine	B.Br.CNEVA
140	France	2000	Animal	B.Br.CNEVA
11_10	France	2011	Bovine	B.Br.CNEVA
11_11	France	2011	Bovine	B.Br.CNEVA
11_8	France	2011	Bovine	B.Br.CNEVA
11_9	France	2011	Bovine	B.Br.CNEVA
79_80	France	2000	Bovine	B.Br.CNEVA
8_06	France	2008	Bovine	B.Br.CNEVA
9_28	France	2009	Bovine	B.Br.CNEVA
9_29	France	2009	Bovine	B.Br.CNEVA
9_30	France	2009	Bovine	B.Br.CNEVA
9_32	France	2009	Bovine	B.Br.CNEVA
9_33	France	2009	Equine	B.Br.CNEVA
9_34	France	2009	Bovine	B.Br.CNEVA
9_35	France	2009	Equine	B.Br.CNEVA
9_36	France	2009	Bovine	B.Br.CNEVA
9_38	France	2009	Bovine	B.Br.CNEVA
9_39	France	2009	Bovine	B.Br.CNEVA
9_40	France	2009	Bovine	B.Br.CNEVA
9_41	France	2009	Bovine	B.Br.CNEVA
9_42	France	2009	Bovine	B.Br.CNEVA
9_43	France	2009	Bovine	B.Br.CNEVA
9_44	France	2009	Bovine	B.Br.CNEVA
9_45	France	2009	Bovine	B.Br.CNEVA
9_46	France	2009	Bovine	B.Br.CNEVA
A0465	France	NR	Bovine	B.Br.CNEVA
BF1	Germany	2009	Bovine	B.Br.CNEVA
BL1	Poland	1993	NR	B.Br.CNEVA
BL10	Poland	1993	NR	B.Br.CNEVA
BL11	Poland	1993	NR	B.Br.CNEVA
BL5	Poland	1988	NR	B.Br.CNEVA
CNEVA-9066	France	1992	NR	B.Br.CNEVA
Ba187	Italy	2006	Animal	B.Br.CNEVA
Ba188	Italy	2007	Animal	B.Br.CNEVA
JF3852	Switzerland	1953	Bovine	B.Br.CNEVA
JF3854	Switzerland	1957	Bovine	B.Br.CNEVA
JF3887	Switzerland	1960	Cattle	B.Br.CNEVA
JF3888	Switzerland	1962	Cattle	B.Br.CNEVA
KrugerB	South Africa	NR	NR	B.Br.Kruger
A1055	USA	NR	NR	C.Br.A1055
2000031021	NR	NR	NR	C.Br.A1055

Annexe 4 :

Arbre phylogénétique de *B. anthracis* basé sur 292 génomes (échelle non logarithmique)

Minimum Spanning Tree avec nœuds hypothétiques généré à partir d'une matrice SNP (10583 SNP) et avec une approche parcimonieuse. L'indice d'homoplasie est ici de 1.5%.

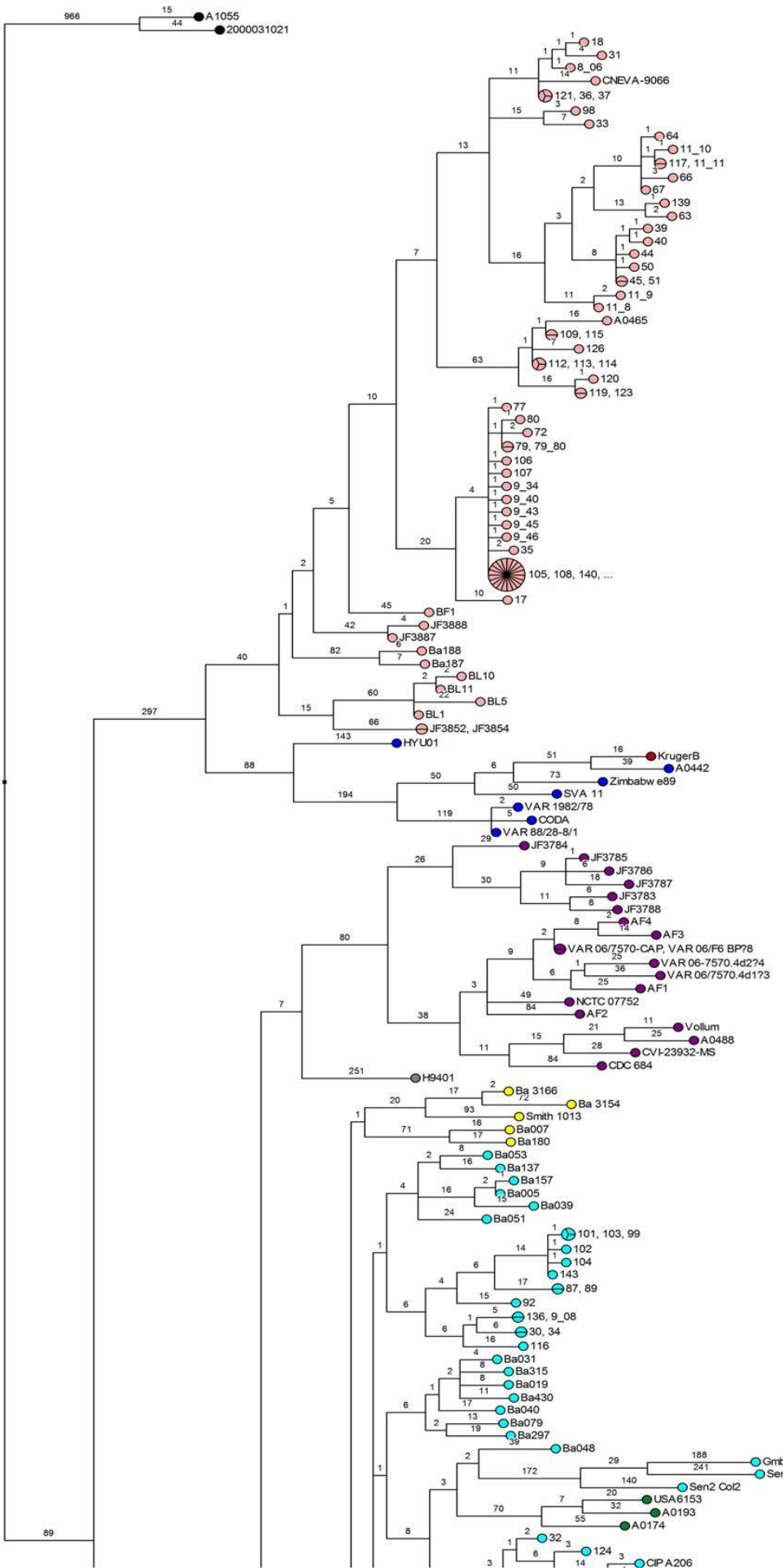




Annexe 5 :

Arbre phylogénétique de *B. anthracis* basé sur 292 génomes (échelle logarithmique)

Minimum Spanning Tree avec nœuds hypothétiques généré à partir d'une matrice SNP (10583 SNP) et avec une approche parcimonieuse. L'indice d'homoplasie est ici de 1.5%.



C.Br.A1055

B.Br.CNEVA

B.Br.Kruger

B.Br.001/002

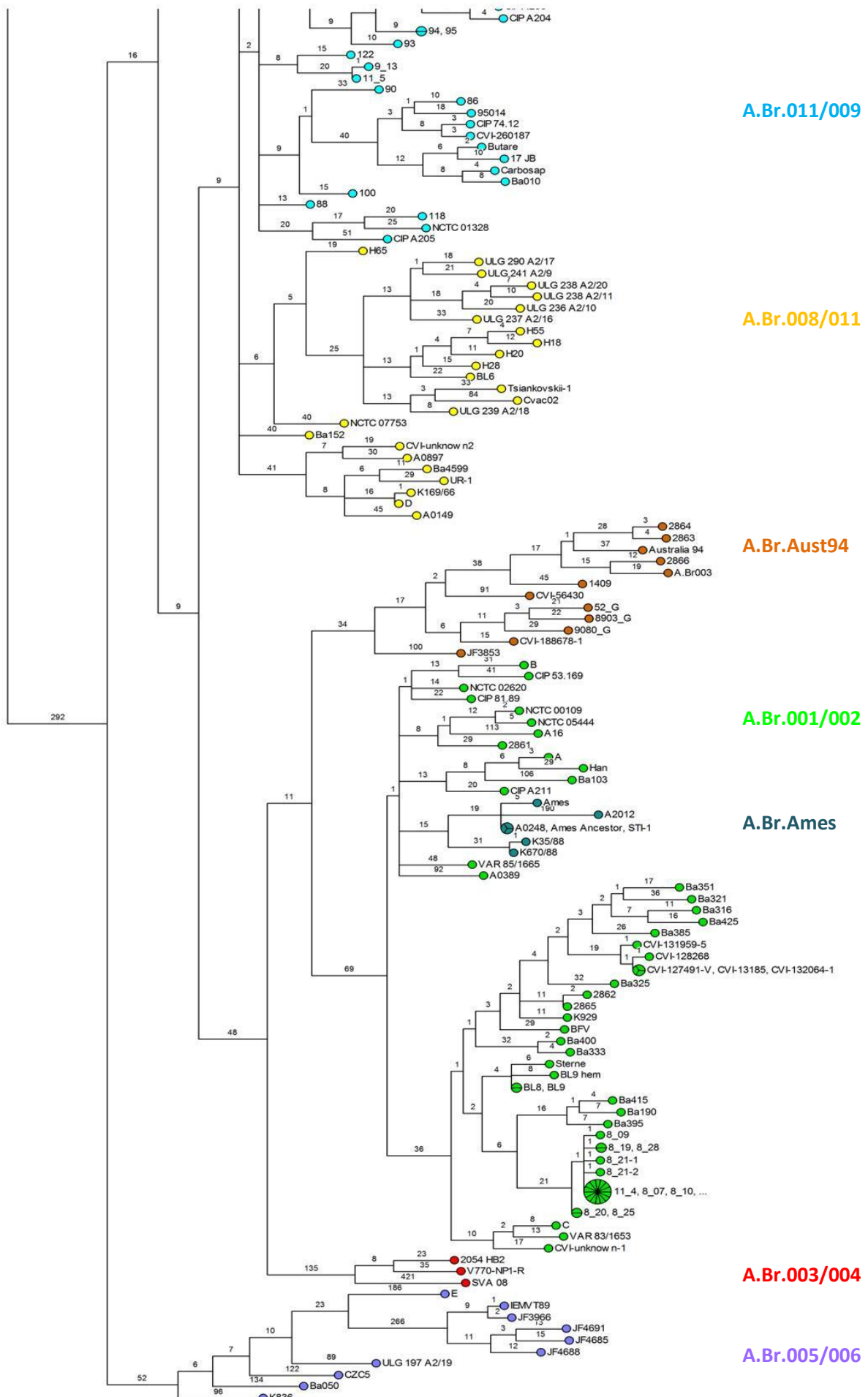
A.Br.Vollum

A.Br.H9401

A.Br.008/011

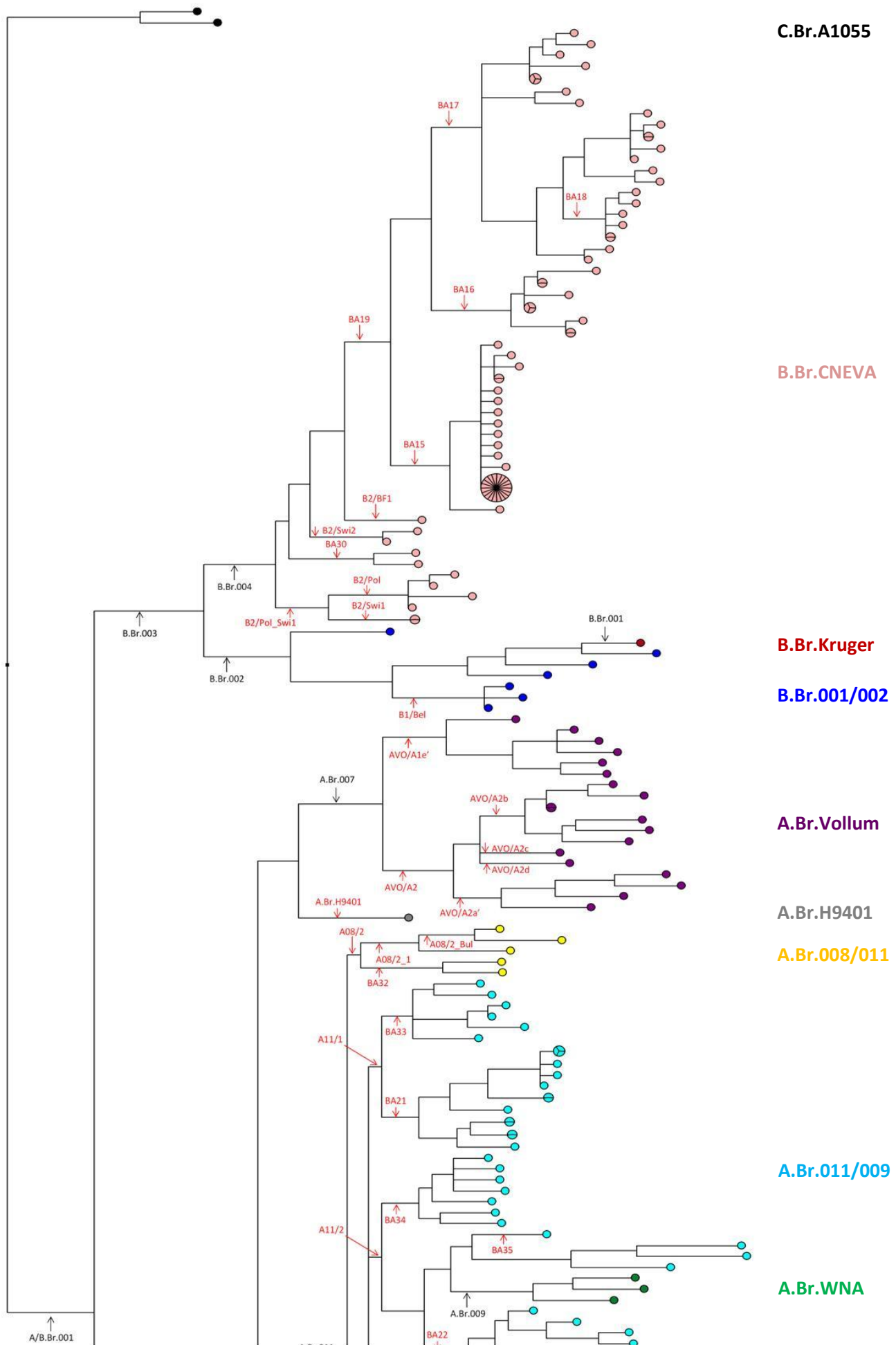
A.Br.011/009

A.Br.WNA



Annexe 6 :

Position des 14 canSNP (en noir) (Marston *et al.*, 2011; Van Ert *et al.*, 2007a) et des 60 nouveaux marqueurs SNP (en rouge)



Annexe 7:

Liste des communications effectuées durant le doctorat

Communications orales :

GIRAULT G, BLOUIN Y, DERZELLE S. « **Etude pangénomique des souches de *Bacillus anthracis* isolées en France** », 9^{ème} congrès national de la Société Française de Microbiologie, 08-09 février 2013, Lille

GIRAULT G, MADANI N, DERZELLE S., « **Anthrax in France: a pangenomic study** », Bacillus ACT 2013, 01-05 septembre 2013, Victoria (British Columbia, Canada)

GIRAULT G, BLOUIN Y, DERZELLE S., « **Génotypage de *Bacillus anthracis* en France : une approche par séquençage à haut débit** », Journée des Doctorants et post-Doctorants ANSES, 28-29 novembre 2013, Maisons-Alfort

GIRAULT G, FASANELLA A, DERZELLE S., « **High-throughput sequencing of *Bacillus anthracis* in France and Italy: investigating genome diversity and population structure** », Workshop on the Biology of Anthrax, 11-12 Mars 2014 Cardiff (Pays de Galles)

Communications écrites :

GIRAULT G, BLOUIN Y, DERZELLE S., « **A pangenomic study of French *Bacillus anthracis* strains** », Journées ABIES, 26-27 février 2013, Paris

GIRAULT G, BLOUIN Y, VERGNAUD G, MADANI N, DERZELLE S., « **Génotypage de *B. anthracis* par séquençage du génome complet** », Journée des Doctorants et post-Doctorants ANSES, 27-28 novembre 2014, Maisons-Alfort

➔ PRIX DU MEILLEUR POSTER

A pangenomic study of French *Bacillus anthracis* strains

Guillaume GIRAULT¹, Yann BLOUIN², Sylviane DERZELLE¹

1. ANSES, Laboratoire de santé animale, Unité des Zoonoses Bactériennes, Maisons-Alfort, France
2. Université Paris Sud, Institut de Génétique et Microbiologie, Equipe Génomes et Polymorphismes, Orsay, France

Introduction

Bacillus anthracis, a spore-forming Gram-positive bacteria, is the etiological agent of anthrax, a zoonosis with a worldwide distribution. Anthrax is an important infection in both livestock and humans. Ruminants are the more susceptible animals to this disease because of the presence of spores of *B. anthracis* in soils. Spores, the infectant form of the bacteria, are highly resistant and can persist for long period of times. Humans can also be contaminated. The most severe form of the disease (inhalation anthrax) is lethal in absence of rapid antibiotic treatment.

B. anthracis is considered as an evolutionarily young species, recently emerged from the *B. cereus* group. It presents very low molecular diversity among strains. Due to this genetic homogeneity, only genotyping methods with high discrimination power have proved to be successful in differentiating strains.

To have a better knowledge of the diversity existing in France among *B. anthracis* isolates, whole-genome sequencing and comparative genomics has been applied to 115 French strains from different geographic areas. In order to develop a highly precise sub-typing system, key SNP signatures were next identified. SNPs are rare diagnostic markers that can provide the level of discrimination required for microbial forensics and precise strain identification. The sub-typing tool will be used to investigate suspect and natural anthrax cases occurring in France. Since the 2001 US postal anthrax attack, *B. anthracis* is considered as the main threat in what concerns bioterrorism.

Technical Background

B. anthracis is a very clonal species. To differentiate strains, only few methods have proved to be successful:

- Multiple-Locus Variable-number tandem repeat Analysis (MLVA)
- Single Nucleotide Repeat (SNR)
- Genome-wide Single Nucleotide Polymorphism (SNP) analysis
- Whole-genome sequencing

SNPs are important phylogenetic markers for diagnostic because of their stability. A set of 14 «canonical» SNPs (canSNPs) allows the classification of *B. anthracis* into 3 major lineages (A, B and C) and 13 sub-lineages spread throughout the world (Fig.1).

In France, 3 sub-lineages are naturally present : B.Br CNEVA (major sub-lineage), A.Br 011/009 and A.Br 001/002.

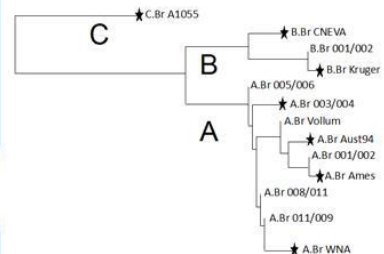


Fig.1 : *Bacillus anthracis* canSNPs tree (Van Ert et al., 2007, Price et al., 2012)

Materials & Methods

The aim of my thesis is to improve knowledge of anthrax diversity in France and to develop a molecular typing tool. Whole-genome sequencing of 115 French strains has been carried out for this purpose. The workflow is the following:

DNA extraction in NSB3 Lab

- Development of a genomic DNA extraction method for *B. anthracis*



- DNA extraction from 115 French strains



Whole-genome sequencing

On the IMAGiF sequencing platform (CNRS):



WGS by Illumina technology
Paired-end sequencing (2x75 & 2x100bp)

Bioinformatics analysis

- Genome assembly (homology versus reference sequence) → BioNumerics software



- Alignment and SNPs identification
- SNPs filters (VNTR, rRNA, incomplete data,...)

Fig.2: From DNA extraction to Bioinformatics analysis

Results

Filtered SNPs can be used for phylogenetic studies. After filtering, 3223 SNPs have been identified for the 115 French strains. As shown in Fig.3, whole genome SNPs analysis allows:

- distinction between A-branch and B-branch
- distinction between sublineages (defined by canSNPs)

The typing is extremely accurate : it can define strain-specific SNPs.

The major B.Br CNEVA sublineage in France was first investigated further. 561 SNPs specific to the sublineage were identified that differentiated 61 strains. A geographic discrimination in four clusters was observed: Alpes, Pyrenees, Massif Central and Saône et Loire. 5 informative SNPs were selected and validated by PCR-HRM (High Resolution Melting).

All the B.Br CNEVA affiliated strains were successfully tested (Fig.4). *In silico* results are in perfect correlation with *in vitro* results. Geographic clustering is possible for this lineage. Moreover, the 5 informative SNPs selected appear to be specific of French strains (not present in European strains from the B.Br CNEVA sublineage).

- This tool allows to determine the French origin of an unknown B.Br CNEVA strain, and its geographic localization within France

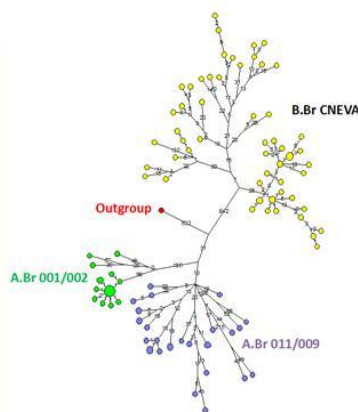


Fig.3 : Minimum Spanning Tree with hypothetical nodes based on 3223 chromosomal SNPs (115 strains)

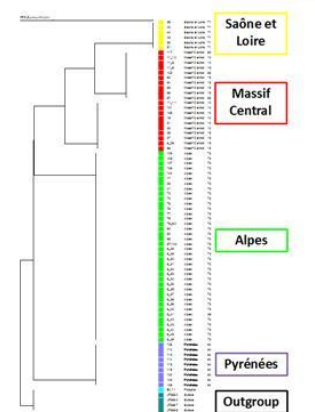


Fig.4 : Geographic clustering of the B.Br CNEVA lineage in France (NJ, based on 5 SNPs)

Conclusions

We have developed a high resolution molecular typing tool that allows to determine the French geographic origin of strains from the B.Br CNEVA sublineage.

The next step will be the identification of additional informative SNPs for the two others French sublineages and the development of a molecular typing tool able to discriminate all European isolates of *B. anthracis*. A definitive cartography of anthrax in France will be defined.

References

- Keim et al., Multiple-Locus Variable-Number Tandem Repeat Analysis Reveals Genetic Relationships within *Bacillus anthracis*, *Journal Of Bacteriology* **182**(10): 2928–2936 (2000)
- Van Ert et al., Global Genetic Population Structure of *Bacillus anthracis*, *PLoS ONE* **2**(5): e461 (2007)
- Kerecic et al., Pre-Columbian Origins for North American Anthrax, *PLoS ONE* **4**(3): e4813 (2009)
- Derzelle et al., Characterization of the genetic diversity of *Bacillus anthracis* in France using high resolution melting assays and multilocus variable-number tandem repeat analysis, *Journal Of Clinical Microbiology* (2011)
- Price et al., Molecular Epidemiologic investigation of an Anthrax Outbreak among Heroin Users, Europe, *Emerging Infectious Diseases* Vol. **18**, No. **8** (2012)

Acknowledgements

- ANSES (thesis co-financing)
- Direction Générale de l'Armement (thesis co-financing)
- IGM Orsay (bioinformatics formation)
- IMAGiF (DNA Sequencing)

Génotypage de *Bacillus anthracis* par séquençage du génome complet

Guillaume GIRAULT¹, Yann BLOUIN^{2,3}, Gilles VERGNAUD^{2,3,4}, Nora MADANI¹ et Sylviane DERZELLE¹



1. Université Paris-Est, ANSES, Laboratoire de Santé Animale, Unité des Zoonoses Bactériennes, Maisons-Alfort, France
2. Université Paris Sud, Institut de Génétique et Microbiologie, Orsay, France
3. Centre National de la Recherche Scientifique (CNRS), UMR 8621, Orsay, France
4. DGA/MRIS – Mission pour la Recherche et l'Innovation Scientifique, Bagnex, France



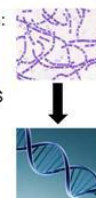
La Fièvre Charbonneuse

- ❖ Zoonose de répartition mondiale causée par la bactérie *B. anthracis* (Van Ert et al., 2007)
- ❖ *B. anthracis* = Arme Biologique potentielle (agent de classe 3)
- ❖ Maladie sporadique mais toujours d'actualité en France (Derzelle et al., 2011; Girault et al., 2014a)



Séquençage du génome

- ❖ Permet de « décoder » l'information génétique d'un organisme
- ❖ Plusieurs générations de méthodes de séquençage sont disponibles:
 - 1^{ère} génération (méthode Sanger)
 - 2^{ème} génération (454, Illumina, SOLID)
 - 3^{ème} génération (Pacific Bioscience, Oxford Nanopore,...)
- ❖ NGS (Next Generation Sequencing)
 - Séquençage massif en parallèle
 - Grand volume de données générées



Etat de l'Art & Approche Expérimentale

- ❖ *B. anthracis* = Pathogène d'intérêt (zoonose + arme biologique) **MAIS...** Pathogène hautement clonal = Difficultés pour différencier les souches entre elles
- ❖ Méthodes employées actuellement pour différencier les isolats:
 - MLVA (Multiple Locus Variable Number of Tandem Repeats)
 - SNR (Single Nucleotide Repeats)
 - canSNP (canonical Single Nucleotide Polymorphisms)
- ❖ Niveau de résolution encore insuffisant pour identifier avec précision une souche (Traçabilité, Filière de la preuve)
- ❖ En France, 3 sous-lignées coexistent à l'état naturel: A.Br.001/002, A.Br.011/009 et B.Br.CNEVA (Figure 1) (Derzelle et al., 2011)



3 lignées majeures (A, B et C) et 13 sous-lignées identifiées dans le Monde
(Van Ert et al., 2007; Marston et al., 2011)

OBJECTIF = Etudier la diversité de *B. anthracis* en France et développer une méthode de typage à haut pouvoir résolutif

- Approche par NGS (Illumina, plateforme IMAGIF)
- Analyse Bioinformatique (BioNumerics):
 - Assemblage de génomes
 - Alignement multiple de génomes complets
 - Identification de SNP sur génome complet
 - Traitement des données
 - Construction d'arbres phylogénétiques
 - Identification de marqueurs canoniques
- Développement d'un outil de typage à haut pouvoir résolutif (PCR-HRM)

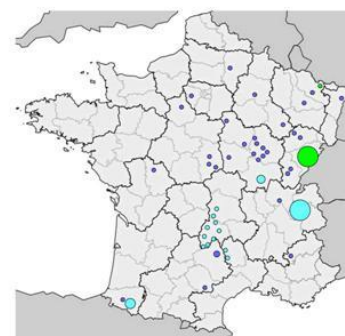


Figure1: Répartition de *B. anthracis* en France
(A.Br.001/002 en vert, A.Br.011/009 en violet et B.Br.CNEVA en bleu)
(Girault et al., 2014b)

Résultats

- ❖ Séquençage de 123 souches Françaises de *B. anthracis*
 - Identification de près de 4000 SNP chromosomiques
- ❖ Reconstruction de la structure de population de *B. anthracis* en France (Figure 2)
- ❖ Identification de 8 marqueurs canoniques Français (Girault et al., 2014c)
- Le niveau de résolution obtenu est sans précédent.
- Relation génotype-localisation géographique (B.Br.CNEVA): Alpes, Pyrénées, Massif Central et Saône-et-Loire sont définis.
- Polytomie remarquable observée pour la sous-lignée A.Br.011/009 (6 branches partant d'un même point).
- Complexe clonal observé pour la sous-lignée A.Br.001/002 (souches du Doubs), caractéristique d'un foyer.
- Développement de la PCR-HRM pour identifier les SNP choisis.
- Résultats par typage PCR-HRM en parfaite corrélation avec les résultats *in silico*.

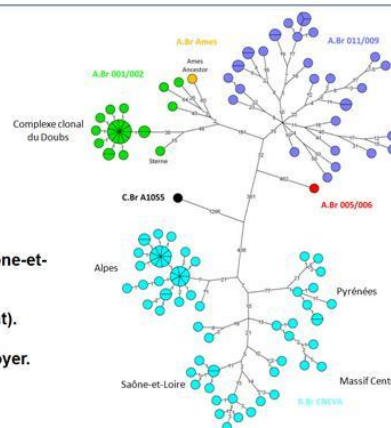


Figure2: Arbre phylogénétique de *B. anthracis* en France
généralisé à partir de 4018 SNP.
(Girault et al., 2014b; Girault et al., 2014c)

Conclusions

- ❖ Obtention d'une cartographie à très haute résolution de *B. anthracis* en France → **NGS = méthode de typage ultime**
- ❖ Phylogénie des souches Françaises permise grâce à cette approche → **Reconstruction de l'historique de la bactérie sur le territoire Français**
- ❖ Développement de la PCR-HRM (haut pouvoir résolutif, robuste, peu coûteuse) → **Outil testé en situation réelle (Cas de FC lors de l'été 2013)** (Girault et al., 2014a)

Références

- Van Ert et al., (2007) Global Genetic Population Structure of *Bacillus anthracis*. PLoS ONE 5
- Derzelle et al., (2011) Characterization of genetic diversity of *Bacillus anthracis* in France by using HRM assays and MLVA. Journal of clinical microbiology 49
- Marston et al., (2011) Molecular epidemiology of anthrax cases associated with recreational use of animal hides and yarn in the United States. PLoS ONE 6
- Girault et al., (2014a) Fièvre charbonneuse en 2013: un foyer en Haute-Marne appelle que cette zoonose est toujours d'actualité en France. BEP 63
- Girault et al., (2014b) Apport de la biologie moléculaire pour la connaissance de la fièvre charbonneuse en France. BEP 63
- Girault et al., (2014c) High-throughput sequencing of *Bacillus anthracis* in France: investigating genome diversity and population structure using whole-genome SNP discovery. BMC Genomics 15

Remerciements

- ANSES
- Direction Générale de l'Armement
- UPSUD – IGM
- ANR
- IMAGIF



RESUME

Bacillus anthracis est une bactérie pathogène de répartition mondiale qui est responsable d'une zoonose, le charbon bactérien. Appartenant au genre *Bacillus*, elle possède la capacité de sporuler : cette particularité lui permet de rester en dormance dans les sols et de résister à de nombreux stress (UV, chaleur...). Infectant principalement les mammifères, cette bactérie est responsable de foyers animaux et humains recensés annuellement à travers le monde. Certaines régions sont endémiques, d'autres, comme la France, ne connaissent que quelques cas sporadiques par an. Malgré l'incidence décroissante de la maladie à travers le monde, *B. anthracis* reste néanmoins d'un intérêt majeur pour de nombreux pays étant donné son usage potentiel en tant qu'arme biologique. L'étude de cette bactérie a une finalité duale, d'une part pour la mortalité qu'elle cause au sein des cheptels bovins et de la faune sauvage, et d'autre part, pour son potentiel en tant qu'arme de destruction massive. Cette espèce bactérienne est considérée comme très clonale : toutes les souches sont extrêmement proches d'un point de vue génétique. Afin de pouvoir identifier précisément les souches et remonter à la source d'infection, diverses méthodes de diagnostic et de typage moléculaire sont disponibles. Cependant, toutes ne possèdent pas le pouvoir de résolution souhaité, la robustesse ou la simplicité d'utilisation. Mon travail a consisté à étudier la diversité des souches de *B. anthracis* disponibles en Europe. Pour cela, un séquençage du génome complet a été réalisé sur un panel d'environ 250 souches (France, Europe) afin de disposer d'une très grande diversité. Une analyse bioinformatique comparative a permis de reconstruire les génomes des isolats européens et l'identification de polymorphismes de séquence spécifiques à ces souches (ici, les Single Nucleotide Polymorphisms = SNP). Ces marqueurs ont conduit à établir avec précision la phylogénie de 292 souches de *B. anthracis* au niveau mondial. Plusieurs hypothèses concernant les origines et l'évolution de ce pathogène ont été proposées. Une proposition d'une nouvelle sous-lignée a également été faite. Enfin, un panel de soixante nouveaux marqueurs de type SNP a été identifié et validé pour génotyper les groupes et lignées phylogénétiquement apparentés présents en Europe. Deux méthodes de biologie moléculaire basées sur l'amplification PCR de ces marqueurs SNP ont été mises au point. La première permet une analyse rapide à faible coût (PCR HRM) alors que la seconde permet un multiplexage des tests (Luminex). Mon travail a permis un accroissement significatif de la connaissance de ce pathogène qu'est *B. anthracis*. Les outils de typage mis en place permettront à la France de disposer d'un outil de diagnostic performant, assurant la traçabilité et l'identification rapide des souches impliquées lors de foyers de fièvre charbonneuse ou lors de cas importés.

ABSTRACT

Bacillus anthracis is a pathogenic bacterium with a worldwide repartition. It is the causative agent of a zoonosis named Anthrax. Belonging to the *Bacillus* genus, *B. anthracis* has the ability to sporulate: this particularity allows the bacterium to stay as quiescent spores into the soils and to resist against different kind of stresses (UV, heat treatment...). Mammals are principally infected and human or animal outbreaks are reported annually in the world. Several regions are endemic while some others, like France, report more sporadic cases. Even if Anthrax incidence is in constant decrease all over the world, *B. anthracis* is still a pathogen of interest for many countries because of its potential use as a biological weapon. The study of this bacterium has a two-tier purpose: first, it is the cause of a relative mortality in livestock and wildlife, and second it is potentially used as a weapon of mass destruction. This bacterial species is considered highly monomorphic and all strains are extremely closed genetically. In order to precisely identify strains during outbreaks or bioterrorism attempt and track the source of infection, several diagnosis and typing methods are available. However, not all of these methods have the discrimination power, the robustness or the ease of use that is required in the laboratory. During this work, I have studied the diversity of European isolates of *B. anthracis*. A whole genome sequencing approach for approximately 250 strains has been done with a great diversity into strains (France, Europe). A comparative bioinformatic analysis allowed genomes reconstruction and polymorphisms identification among European isolates (Single Nucleotide Polymorphism = SNP). These markers led to establish a precise phylogeny among 292 *B. anthracis* isolates at a world level. Several hypotheses concerning the origins and the evolution of this pathogen have been proposed. A new sublineage has been potentially discovered. A panel of sixty SNPs has been identified and confirmed to genotype the major groups and lineages phylogenetically related in Europe. Two molecular typing approaches based on PCR amplification have been developed. Using the identified SNP, they can be used to discriminate strains between each others. The first one is a quick and low cost approach (PCR HRM) whereas the second can be used to multiplex a lot of analyses (Luminex). My work allowed a significative increase into *B. anthracis* knowledge. The typing tools developed will allow traceability and quick identification of the strains involved in an Anthrax outbreak or in a suspected bioterrorist attempt in France.