Infection, Genetics and Evolution 46 (2016) 50–58

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Research paper

Comparative genomics of *Bacillus anthracis* from the wool industry highlights polymorphisms of lineage A.Br.Vollum

Sylviane Derzelle [a,1], Lisandra Aguilar-Bultet [a,b], Joachim Frey [a,*]

[a] Institute of Veterinary Bacteriology, Vetsuisse, University of Bern, Laenggasstrasse 122, 3001 Bern, Switzerland
[b] Graduate School for Cellular and Biomedical Sciences, University of Bern, Bern CH-3012, Switzerland

ARTICLE INFO

Article history:
Received 15 March 2016
Received in revised form 21 October 2016
Accepted 22 October 2016
Available online 25 October 2016

Keywords:
Bacillus anthracis
Whole genome sequencing
Single nucleotide polymorphism
Phylogeny
Comparative genomics
High resolution melting

ABSTRACT

Background: With the advent of affordable next-generation sequencing (NGS) technologies, major progress has been made in the understanding of the population structure and evolution of the *B. anthracis* species. Here we report the use of whole genome sequencing and computer-based comparative analyses to characterize six strains belonging to the A.Br.Vollum lineage. These strains were isolated in Switzerland, in 1981, during iterative cases of anthrax involving workers in a textile plant processing cashmere wool from the Indian subcontinent.
Results: We took advantage of the hundreds of currently available *B. anthracis* genomes in public databases, to investigate the genetic diversity existing within the A.Br.Vollum lineage and to position the six Swiss isolates into the worldwide *B. anthracis* phylogeny. Thirty additional genomes related to the A.Br.Vollum group were identified by whole-genome single nucleotide polymorphism (SNP) analysis, including two strains forming a new evolutionary branch at the basis of the A.Br.Vollum lineage. This new phylogenetic lineage (termed A.Br.H9401) splits off the branch leading to the A.Br.Vollum group soon after its divergence to the other lineages of the major A clade (i.e. 6 SNPs). The available dataset of A.Br.Vollum genomes were resolved into 2 distinct groups. Isolates from the Swiss wool processing facility clustered together with two strains from Pakistan and one strain of unknown origin isolated from yarn. They were clearly differentiated (69 SNPs) from the twenty-five other A.Br.Vollum strains located on the branch leading to the terminal reference strain A0488 of the lineage. Novel analytic assays specific to these new subgroups were developed for the purpose of rapid molecular epidemiology.
Conclusions: Whole genome SNP surveys greatly expand upon our knowledge on the sub-structure of the A.Br.Vollum lineage. Possible origin and route of spread of this lineage worldwide are discussed.

1. Introduction

Bacillus anthracis, the causative agent of anthrax, forms endospores that can remain dormant for years in many types of soil all over the world. *B. anthracis* commonly infects wild and domesticated herbivorous mammals that ingest or inhale spores while grazing (Hugh-Jones and Blackburn, 2009). Once the endospores enter the host, they germinate and then spread by the circulation to the lymphatics. The poly-γ-D-glutamic acid capsule produced by the pathogen disguises *B. anthracis* from immune surveillance and allows its unimpeded growth in the host. When the bacteria gain access to the circulating blood, they multiply and typically kill the animal within a few days or weeks. Production of the powerful tripartite protein toxins by *B. anthracis*, lethal toxin and edema toxin in cooperation with the shared protective antigen that serves as the receptor binding subunit, causes cell death and tissue swelling. Anthrax is often fatal in cattle, sheep and goats. 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 (Hugh-Jones and Blackburn, 2009).

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 “woolsorter's” or “ragpicker's” disease. It was

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; Tm, melting temperature; NCBI, National Center for Biotechnology Information; SRA, Sequence Read Archive; SPR, Subtree-Pruning-Regrafting; ML, Maximum Likelihood; NL, Neighbor-Joining; MP, Maximum Parsimony.

* Corresponding author.
E-mail addresses: sylviane.derzelle@lncr.org (S. Derzelle), lisandra.aguilar@vetsuisse.unibe.ch (L. Aguilar-Bultet), joachim.frey@vetsuisse.unibe.ch (J. Frey).
[1] Present address: LNCR (Laboratoire National de Contrôle des Reproducteurs), Maisons Alfort, France

http://dx.doi.org/10.1016/j.meegid.2016.10.019

a serious threat after the 1830s, when various Eastern wools such as mohair or Cashmere goat hair were imported for use as textile fibers (Brachman et al., 1966; Carter, 2004).

Today, reports of anthrax cases related to handling of animal products in industrial settings are extremely rare in Europe, largely due to widespread animal vaccination during the 20th century, improved 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 remain at high biological risk. In Switzerland, between 1978 and 1981, small outbreaks of human anthrax occurred in a plant that processed synthetic fibers and cashmere wool from Pakistan. Within 3 years, 25 workers had contracted anthrax. Twenty-four cases had the cutaneous form and one had inhalation anthrax from which all the patients recovered (Pfisterer, 1991; Winter and Pfisterer, 1991). We have previously reported the MLVA-8 and canSNPs typing of some of the strains isolated from clinical cases, goat hairs, and air filters from this factory (Pilo et al., 2008). All isolates were found to cluster with the A4 genotype and to belong to the sublineage A.Br.Vollum. The minor variations found between some of these strains (strains could be differentiated by two VNTR allelic difference on the pXO1 and pXO2 markers) raised questions about the epidemiological significance of the differentiation (Pilo et al., 2008).

Classically, the lifecycle of *B. anthracis* is described by short vegetative bursts in infected hosts, which is estimated to last 20–40 bacterial generations (Keim et al., 2004), alternating with long periods of dormancy as an environmental spore until disease is re-established. The capacity of *B. anthracis* spores to germinate and establish populations of vegetative cells in the environment is a matter of discussion in the scientific community. Recent laboratory findings tend to indicate that *B. anthracis* may have some ability to interact with members of the grassland-soil community, including plants, invertebrates and protists (Dey et al., 2012; Saile and Koehler, 2006; Schuch and Fischetti, 2009). In any case, *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. The global structure of the *B. anthracis* population is divided into three major genetic clades (A, B and C), with further subdivisions into 12 main lineages (Pilo and Frey, 2011; Van Ert et al., 2007). The type strain of the species, strain Vollum (NCTC10340^T; ATCC14578^T), is affiliated to the A.Br.Vollum lineage (corresponding to MLVA cluster A4). The geographically broad distribution of this lineage throughout the world is thought to be a consequence of human activities in commerce and industrialization. Indeed, the occurrence of A.Br.Vollum strains in Western industrialized regions has often been tied to the international trade of spore-infected items, in particular wool or goat hair from the Middle-East and central Asian area (Pilo et al., 2008; Van Ert et al., 2007; Wattiau et al., 2008).

To infer the epidemiological relationships existing between strains and gain further resolution into the population sub-structure of the A.Br.Vollum lineage, six strains associated with the wool-processing factory outbreaks were sequenced. Genetic characterization was performed by whole-genome SNP analysis and canonical SNP genotyping (Keim et al., 2004; Van Ert et al., 2007), which best roots the phylogenetic relationship of *B. anthracis* strains. Comparative genomics were conducted using twenty-eight additional strains belonging to lineage A.Br.Vollum and 47 diverse strains identified among the hundreds of *B. anthracis* genome sequences available in public databases.

2. Methods

2.1. Bacterial strains, DNA extraction and biosafety procedures

Six *B. anthracis* strains belonging to A.Br.Vollum were sequenced in this study (Table 1). They were collected during anthrax outbreaks that have occurred in northern Switzerland over three years (1978–1981) in a textile factory (Pfisterer, 1991; Pilo et al., 2008; Winter and Pfisterer, 1991). All *B. anthracis* manipulations were performed in a biosafety level 3 laboratory using class II type A2 biosafety cabinet.

DNA was obtained from vegetative cells grown 16 h at 37 °C on 5% horse blood agar plates by scraping the agar surfaces to remove bacterial colonies. DNA was purified using the QIAGEN® Genomic-tip 100/G columns and QIAGEN® Genomic DNA Buffer Set. After isopropanol precipitation, genomic DNA was suspended in 400 μl of 10 mM Tris HCl (pH 8) for 2 h at 50 °C. Subsequently DNA preparations were filtered through a 0.2 μm Acrodisc® syringe filter (Pall Corporation, Ann Arbor, MI, USA) to warrant full removal of remaining *B. anthracis* cells or spores. Viability testing was systematically performed before DNA was taken out of the BSL-3 facility.

2.2. Whole genome sequencing and SNPs discovery

DNA was subjected to paired-end whole genome sequencing on the Illumina MiSeq platform (Illumina Inc., San Diego, CA, USA). The number of reads that passed Illumina quality filters varied from 0.5 to 1.7 million. Chromosomal, pXO1 and pXO2 sequences of the Ames Ancestor strain were used as reference for genome and plasmids assembly [RefSeq: NC_007530.2, 129 NC_007322.2 and NC_007323.3, respectively]. Paired-end reads were mapped to the Ames Ancestor genome using the Bowtie 2 tool and its default parameters (Langmead and Salzberg, 2012). Whole-genome SNPs discovery and variant calling were performed for each alignment using the SAMtools package (Li et al., 2009). Individual lists of SNPs were compiled and data filtered to remove SNP positions at ribosomal operons and 31 VNTR loci. Contiguous SNPs were also excluded from the analysis.

2.3. Comparative genomics and in silico canSNP genotyping

In order to identify additional strains affiliated to the A.Br.Vollum lineages, all publicly available *B. anthracis* genomes were retrieved from database and in silico genotyped. Briefly, the complete or draft genome assemblies from a hundred strains were used to generate sets of artificial, overlapping, 200 bp-long reads (with a 4-fold genome coverage). Each set of reads was aligned to the Ames Ancestor sequences using the same pipeline as described above for the six Swiss isolates. Genotypes were next determined based on 13 published canonical SNPs (Van Ert et al., 2007) and a few additional SNPs. The locations of these SNPs along the Ames Ancestor genome are listed in Table 3. Seventy-five strains, representing the genetic and geographic diversity observed in *B. anthracis*, were further selected for phylogenetic analysis and comparison (Tables 1 and 2). Besides the 28 A.Br.Vollum isolates we identified, the sample includes four genomes belonging to each of the eleven other canSNP lineages or groups described so far. Strains harbouring both virulence plasmids (pXO1 and pXO2) were preferred.

2.4. Phylogenetic analysis

Evolutionary analyses were conducted in MEGA7 (Kumar et al., 2016) using the *B. cereus* AH820 strain as outgroup and the whole genome SNP data set as input. Phylogenetic relationships were inferred using three distinct methods, i.e. Maximum Likelihood, Maximum Parsimony and Neighbor-Joining. The Tamura 3-parameter model was statistically selected as the best-fit nucleotide substitution model using MEGA7 tools, and applied for modelling how DNA sequences change over evolutionary time. The Subtree-Pruning-Regrafting (SPR) algorithm was used as heuristic model for tree inference. Bootstrap scores (250 replicates) were computed for each phylogenetic reconstruction.

The global evolutionary history of the species was inferred by using the Maximum Likelihood (ML) method. Initial ML trees were obtained by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimates using the Maximum Composite Likelihood