Unexpected genomic relationships between *Bacillus anthracis* strains from Bangladesh and Central Europe

Farzana Islam Rume^{a,c}, Chowdhury Rafiqul Ahsan^a, Paritosh Kumar Biswas^b, Mahmuda Yasmin^a, Peter Braun^d, Mathias C. Walter^d, Markus Antwerpen^d, Gregor Grass^{d*}, Matthias Hanczaruk^d

^aDepartment of Microbiology, University of Dhaka, Dhaka-1000, Bangladesh

^bDepartment of Microbiology & Veterinary Public Health, Chittagong Veterinary and Animal Sciences University, Chittagong, Bangladesh

^cDepartment of Microbiology & Public Health, Patuakhali Science and Technology University, Khanpura, Babugonj, Barisal, Bangladesh

^dBundeswehr Institute of Microbiology, Munich, Germany

*Corresponding author:

Gregor Grass

Bundeswehr Institute of Microbiology

Neuherbergstr.11

80937 Munich, Germany

Email: gregorgrass@bundeswehr.org

ABSTRACT

The zoonosis anthrax caused by the bacterium *Bacillus anthracis* has a broad geographical distribution. Active enzootic areas are typically located away from central and northern Europe where cases of the disease occur only sporadically and in limited numbers. In contrast, a few out of the 64 districts of Bangladesh are hyper-endemic for anthrax and there the disease causes major losses in live-stock. In this study we genotyped eight strains of *B. anthracis* collected from the districts of Sirajganj and Tangail in 2013. All these strains belonged to canSNP group A.Br.001/002 Sterne differing only in a few of 31 tandem-repeat (MLVA)-markers. Whole genome sequences were obtained from five of these strains and compared with genomic information of *B. anthracis* strains originating from various geographical locations. Characteristic signatures were detected defining two “Bangladesh” clusters potentially useful for rapid molecular epidemiology. From this data high-resolution PCR assays were developed and subsequently tested on additional isolates from Bangladesh and Central Europe. Remarkably, this comparative genomic analysis focusing on SNP-discovery revealed a close genetic relationship between these strains from Bangladesh and historic strains collected between 1991 and 2008 in The Netherlands and Germany, respectively. Possible explanations for these phylogenetic relationships are discussed.

Keywords: *Bacillus anthracis*; Bangladesh; whole genome sequencing; comparative genomics; Single Nucleotide Polymorphism (SNP); high-resolution melting SNP analysis (HRM-SNP)

1. Introduction

Today, the zoonotic disease anthrax is rare in most European countries especially in central and northern parts of the continent. Historically, however, infections with *Bacillus anthracis*, the causative agent of the disease, were more common in animals and humans in these regions. Besides relatively rare importations of acute human infections from endemic areas (Anaraki et al., 2008), recent outbreaks have typically originated from infections through contaminated recreational drug use related to injectional anthrax in patients from Norway, the United Kingdom, Germany, France and Denmark (Hanczaruk et al., 2014; Ringertz et al., 2000) or livestock infections from recalcitrant spores by presumably disturbed old burial grounds of animals diseased by anthrax (Ågren et al., 2014). In contrast, *B. anthracis* is endemic in parts of Italy (Fasanella et al., 2005; Fasanella et al., 2010) and other Mediterranean countries (Durmaz et al., 2012) and hyperendemic with numerous annual outbreaks in parts of Africa and Asia including Bangladesh (Chakraborty et al., 2012). In 2014 this tropical country on the eastern edge of the Indian sub-continent has faced 225 reported human cutaneous anthrax cases alone (according to the Institute of Epidemiology, Disease Control & Research (IEDCR), Dhaka, Bangladesh; <http://www.iedcr.gov.bd>) and a history of earlier outbreaks (Be-Nazir et al., 2010; Chakraborty et al., 2012; Fasanella et al., 2013b; Siddiqui et al., 2012). Human anthrax in Bangladesh is intrinsically related to enzootic anthrax afflicting livestock. This is because of very low vaccination coverage (Chakraborty et al., 2012) despite the country's routine anthrax vaccination program for livestock (Siddiqui et al., 2012). On the country-side, people and their farm animals live in close vicinity and direct or indirect human contact with sick and diseased animals is frequent. As a result, humans may become infected during all phases of the

1 life-cycle of the anthrax bacterium, from contaminated soil or via animal products
2 (Chakraborty et al., 2012). Expectedly, the disease is more prevalent in the rural districts of
3 Pabna, Sirajganj and Tangail which support the country's largest cattle populations (Be-Nazir
4 et al., 2010).

5 Despite the country's history of anthrax outbreaks, Fasanella and co-workers (Fasanella et
6 al., 2013b) were the first to genotype *B. anthracis* strains from Bangladesh into the
7 canSNP-typing system (Van Ert et al., 2007) and clustered all 99 strains analyzed to the
8 A.Br.001/002 Sterne canSNP sub-lineage (Fasanella et al., 2013b). Taken together with
9 MLVA-typing this data suggested that Bangladeshi strains were most closely related to strains
10 from China (Simonson et al., 2009) and other South Asian countries (Fasanella et al., 2013b).
11 We noticed that published Bangladeshi strains (Chakraborty et al., 2012; Fasanella et al.,
12 2013b; Rume et al., 2016) had identical canSNP-profiles and also clustered according to
13 MLVA with certain strains from the *B. anthracis* collection of the Bundeswehr Institute of
14 Microbiology comprising older strains from German and Dutch collections, respectively. Here
15 we report the genotyping including whole genome sequencing of recently isolated strains from
16 Bangladesh and the results of comparative genomics using representative strains from
17 Germany, The Netherlands and published sequences to get insight into the phylogenetic
18 diversity of Bangladeshi strains and their relationships with strains from Central Europe. Novel
19 SNPs specifically discriminating these strains were identified. Using this information new
20 PCR-based high-resolution melting SNP-assays were developed and a variety of additional
21 strains from Bangladesh and Central Europe genotyped as well. Finally, we delineated genetic
22 relationships between these closely related strains from distant geographical origins.

2. Material and Methods

2.1. Collection of soil samples

Following two reports on anthrax outbreaks in humans and bovine animals, as published in ProMED-mail on May 17 (ProMED-mail archive no. 20130517.1720541) and September 7 2013 (ProMED-mail archive no. 20130907.1929746), we visited the reported areas located at two sub-districts (locally known as upazilas), namely Tangail Sadar (Thangail district, Dhaka division) near Tangail and Shahajadpur (Sirajganj district, Rajshahi division) in the north of Sirajganj district, respectively, between July 6 and September 27, 2013 and collected samples from different cattle farms in the reported areas. These sites had a history of sudden death of at least one bovine animal in the recent past with onset of convulsions or collapsing, with or without previous reported fever similar to what has been described earlier (Biswas et al., 2012). The cattle farms from which environmental samples were collected are referred to as case farms and all of these have not been previously reported for anthrax cases by mass-media in the country or by any global electronic based surveillance. Additional samples were collected from diverse sampling sites in the districts of Tangail (near the village of Katuli) and Sirajganj (near the village of Char Kaijuri) (Fig. 1) including soil, water, food and other environmental samples as well as swabs from animal bones or carcasses. The characteristics of the sampling sites are denoted in Table 1. In the Tangail district samples from an animal burial site in a watered forest and from two stables of previously diseased cows were collected, in the Sirajganj district samples from a carcass-containing water pit and from a stable of a previously diseased cow were drawn, respectively.

2.2. Isolation of *B. anthracis* from soil samples and inactivation

To detect *B. anthracis* from samples, the Ground Anthrax Bacillus Refined Isolation (GABRI) method was used (Fasanella et al., 2013a). The GABRI method is a sensitive method for isolation of *B. anthracis* spores from soil particles as it disrupts hydrophobic interactions between spores and the adhering solid matrices. After incubation at 37°C for 24 h on selective solid media (Tomaso et al., 2006) suspected colonies lacking hemolysis were picked and streaked on Columbia blood agar (Merck KGaA). In general, bacteria were inactivated (for DNA extraction) after a single passage on blood agar in order to minimize in vitro genetic changes. For each isolate a few colonies were scraped off the agar plate and transferred into a 1.5 ml reaction tube filled with 500 µl 2% (v/v) Terralin PAA (Schülke & Mayr) and the cell pellet carefully resuspended by pipetting up and down. The reaction tube was completely filled up with 2% (v/v) Terralin PAA and incubated for 30 min for inactivation of *B. anthracis* endospores which are vulnerable to peracetic acid, the active agent of Terralin PAA (Jacob et al., 2012). Following centrifugation at 6000 × g for 2 minutes the supernatant was removed and the pellet resuspended in 1 ml phosphate-buffered saline (PBS). After two more washing steps with 1 ml PBS the pellet was stored at -20°C until further use. Work involving live *B. anthracis* was performed in a biosafety level 3 laboratory (BSL-3) within a class III safety cabinet (Glove box). Inactivated preparations were transferred to a BSL-2 laboratory for further analysis.

2.3. Extraction of DNA from inactivated culture material

For isolation of DNA from inactivated bacterial suspensions, the DNeasy Blood and Tissue Kit (Qiagen) was used as described in the manufacturer's manual for Gram-positive bacteria with

the following minor changes. After cell wall lysis with lysozyme, 1 µl of RNase A (Qiagen) was added to each reaction tube and the mix was incubated at 37°C for another 10 min. Time for lysis with proteinase K was doubled to 60 min and elution was done twice (150 µl and 100 µl, respectively) into the same tube using sterile nuclease free Milli-Q water instead of Buffer AE. The DNA concentrations of each eluate were quantified using the Qubit dsDNA HS Assay Kit (Thermo Fisher) according to the supplier's protocol. DNA solutions were stored at -20°C until further use.

2.4. Diagnostic real-time PCR for chromosomal and plasmid markers of *B. anthracis*

The specific chromosomal marker *dhp61* (BA_5345) was used for identification of *B. anthracis* as described by (Antwerpen et al., 2008). Real-time PCR for detection of *capC* (pXO2) and *pagA* (pXO1) were performed according to (Turnbull, 2008). Assays were conducted using either LightCycler 2.0, LightCycler 480 II (Roche) or the Stratagene MX 3000P (Agilent Technologies). Data analysis was performed with the respective associated instrument software.

2.5. Analysis of canonical Single Nucleotide Polymorphisms (canSNPs) by Melt Mismatch

Amplification Mutation Assay (Melt-MAMA)

CanSNP assay inquiry is a phylogenetic approach to identify SNPs that efficiently partition bacterial strains into genetic groups consistent with their recognized population structure (Keim 2004, Pearson 2004, Van Ert 2007). For canSNP typing of the *B. anthracis*-isolates Mismatch Amplification Mutation Assays (Melt-MAMA) were performed (Birdsell et al.,

2012). In short, for every canSNP assay two previously genotyped *B. anthracis* isolates were used as controls carrying either the ancestral or derived allele, respectively. The 20 µl reactions contained 1 × LightCycler FastStart DNA Master SYBR Green-I, 4 mM MgCl₂, about 5 ng of template DNA and primer mixes specific for each assay (Birdsell et al., 2012). Real-time PCR was run on a LightCycler 2.0 or LightCycler 480 II (Roche). Thermal profiles used for amplification and subsequent melting curve analysis were as published previously (Birdsell et al., 2012).

2.6. Multi locus VNTR (variable number of tandem repeats) analysis (MLVA)

MLVA is a standard genotyping method for *B. anthracis* (Van Ert et al., 2007). For the time being MLVA is still valuable because MLVA provides an additional data set amending next generation sequencing approaches and because most of the long and highly-repetitive VNTRs cannot be determined by whole genome sequencing approaches relying on DNA-libraries yielding relatively short (<600 bases) read lengths. Therefore, DNA fragment lengths of fluorescently labeled amplicons covering complete VNTR loci were analyzed by capillary gel electrophoresis. In order to amplify the 31 different VNTR loci, five conventional multiplex PCRs were carried out as described previously (Beyer et al., 2012; Keim et al., 2000; Lista et al., 2006) with minor changes in primer sequences and primer concentrations.

PCR products were diluted 1:10 in water. A volume of 2 µl of each dilution and 0.5 µl of size standard Genescan 1200 LIZ (Applied Biosystems) were added to 17.5 µl HiDi formamide (Applied Biosystems) and incubated at 95°C for 5 min to denature DNA. Fragment lengths

were determined on a 3130 Genetic Analyzer (Applied Biosystems) and the associated GeneMapper software.

2.7. Ion Torrent next generation whole genome sequencing

DNA-Library preparation for whole genome sequencing with the Personal Genome Machine (PGM, Thermo Fisher) required 100 – 1000 ng of high quality DNA in 50 µl nuclease free water. Where required, eluates from genomic extractions were concentrated to 2 -20 ng/µl. For this a Concentrator Plus (Eppendorf) was used to evaporate water from aqueous solutions.

Whole genome sequencing of DNA of *B. anthracis* isolates was carried out with the Ion PGM (Thermo Fisher). To this end, library preparations were made according to the Ion Xpress Plus Fragment Library Preparation Guide (Thermo Fisher) using a sonication method using the Bioruptor system UCD-200 (Thermo Fisher) and subsequent end-repair (Thermo Fisher). The following settings of the Bioruptor were used to obtain fragment sizes of approx. 400 bp: time On/OFF: 3 × 15 min, 0.5 min/µM, min, Power level L (low). The 2100 Bioanalyzer (Agilent Technologies) was used for quality control after end-repair. Size selection and adapter ligation was achieved following the instructions of the Agilent High Sensitivity DNA Quick Start Guide (Agilent Technologies). Final DNA concentrations of the libraries were measured using the Qubit dsDNA HS Assay Kit (Thermo Fisher) and the DNAs were diluted to 44.2 pM. Emulsion PCR and enrichment of template positive Ion Sphere Particles were performed according to the Ion PGM Template OT2 400 Kit (Thermo Fisher). Chip loading was carried out according to the manual for the Ion 318 Chip Kit v2 (Thermo Fisher) in combination with Ion PGM Sequencing 400 Kit-chemistry (Thermo Fisher) producing reads of up to 400 bases.

2.8. Analysis of whole genome sequencing data – SNP calling

Sequencing output files were mapped to the reference genome *B. anthracis* Ames Ancestor (GenBank accessions: chromosome: AE017334.2; pXO1 AE017336.2; pXO2: AE017335.3) using bowtie2 (Langmead and Salzberg, 2012). A Variant Call File (VCF) was generated by SAMtools (Li et al., 2009) for every isolate containing all variations compared to the reference genome including quality and coverage information. A python script was used to identify all single base variations that were both present in more than 75% of the reads and marked as high-quality bases. This analysis was carried out twice, once with a minimum coverage depth for variations of 15 and once with a minimum of 3. In order to avoid false negative as well as false positive results, the output data of the two approaches were compared. All results that indicated a SNP in one isolate compared to others were re-checked by manual inspecting the respective BAM files using Tablet software (Milne et al., 2013).

Next progressiveMauve (Darling et al., 2010) was used to align genomes to *B. anthracis* Ames Ancestor where only contig sequences were available (Table 2) and observed SNPs were exported. Both SNP data sets were combined and filtered excluding gaps and ambiguous bases. SNPs located in annotated regions of rRNAs, tRNAs, repeat regions and tandem repeats (predicted by TRF (Benson, 1999)) were also removed using bedtools (Quinlan and Hall, 2010). Closely adjacent SNPs with a distance of less than 10 bp were also removed. From the concatenated SNP sequences, a UPGMA tree was computed and optimized according to the maximum parsimony approach of Bionumerics 6.6 (Applied Maths). For validation, concatenated SNP sequences were used to compute a tree with the phangorn package (Schliep, 2011) and the R statistical framework.

The draft genome of isolate *B. anthracis* strain “Tangail-1” can be accessed under accession number CP015779 (chromosome), CP015777 (pXO1) and CP015778 (pXO2). Strains included in phylogenetic analysis are summarized in table 2.

2.9. Interrogation for branch-specific SNPs by high resolution melting analysis (HRM-SNP)

In order to validate the distribution of branch-specific SNPs obtained by whole genome sequencing, HRM-SNP assays were designed surrounding representative SNP positions, since base pair exchange from A + T to G + C or vice versa result in differences in melting temperatures of short amplicons. HRM-SNP primer sequences for realtime PCR-assays are listed in Table 3. Primer oligonucleotides were designed with the Primer-BLAST tool of NCBI (Ye et al., 2012) using the *B. anthracis* Ames Ancestor genome as a reference template.

All SNPs were tested in 10 µl singleplex reactions containing 0.4 µM of the respective primer pair (B1, B2, NL/B1, A02/B1-NL and A02/A), 3 mM MgCl₂, approx. 1 ng of template DNA and 1 × LightCycler 480 High Resolution Melting Master Mix (Roche). Amplification and melting curve analysis was carried out on the LightCycler 480 II (Roche) using the following thermal profile for amplification: 95°C for 10 min, 45 cycles at 95°C for 10 s, 55°C (HRM-SNP B1: 52°C) for 20 s, 72°C for 10 s. The thermal profile used for melting curve analysis was for all samples: 95°C for 1 min, 60°C for 1 min and a final heating step to 95°C with a ramp rate of 0.02°C/s and continuous quantification at 530 nm.

3. Results and discussion

3.1. *B. anthracis* isolation from environmental samples and initial genotyping

All non-hemolytic candidate strains retrieved by the modified GABRI method (Fasanella et al., 2013a) from Bangladesh samples were identified as *B. anthracis* by diagnostic real time PCR on three markers *dhp61* (chromosomal), *pagA* (pXO1) and *capC* (pXO2), harboring both virulence plasmids (pXO1 and pXO2) (data not shown). The geographic attribution of these isolates within Bangladesh is presented in Table 1. All isolates were classified within the 12 original canonical SNP-groups of *B. anthracis* comprising the three major branches A, B and C for initial sub-typing of the isolates on a global phylogenetic background (Van Ert et al., 2007). As expected from published work (Fasanella et al., 2013b; Rume et al., 2016) all isolates belonged to the A. Br. 001/002 canSNP lineage indicating homogeneity of Bangladeshi strains in terms of canSNP-designation. Notably, previously described strains from very closely located sampling sites in the upazilas Tangail Sadar and Shahjadpur, respectively, also belonged to this lineage (Rume et al., 2016). This initial canSNP-typing agreed well with the previous finding that the A. Br. 001/002 canSNP lineage has a major presence in most of China and other south-eastern Asian countries, where this lineage is thought to be old and well established (Simonson et al., 2009).

Next, we performed Multiple Loci Variable Number of Tandem Repeat (VNTR) analysis (MLVA). Currently, MLVA data are notoriously difficult to extract from whole genome projects sequencing since assemblies often fail because of their partially long repetitive DNA-structures. Typically, MLVA is used in a series of progressive hierarchical resolving assays for sub-typing *B. anthracis* following canSNP-typing because of MLVA's higher

1 resolution power (Keim et al., 2004). All isolates were screened for variations in their 31
2 MLVA markers (Table 4) and assigned to five similar MLVA genotypes. The majority of
3 fragment length variations were observed in the markers bams5, bams15, bams21, bams24 and
4 bams31 which are part of the MLVA-31 typing scheme (Beyer et al., 2012). Additional
5 fragment length variations occurred in markers VNTR16, VNTR19 and VNTR35 which are
6 part of MLVA-15 (Van Ert et al., 2007). These results agree with the findings of (Chakraborty
7 et al., 2012; Fasanella et al., 2013b; Rume et al., 2016) who reported similar data sets for
8 Bangladeshi strains. Notably, much of the diversity seen in our study using MLVA-31 has not
9 been seen using the smaller MLVA-8 (Chakraborty et al. 2012) or MLVA-15 (Chakraborty et
10 al., 2012; Fasanella et al., 2013b) panels. Thus, data from various MLVA-typing studies
11 confirm the close genetic relatedness of strains from the regions near Tangail and Sirajganj and
12 provide strong evidence for the ecological establishment of the A.Br.001/002 sub-lineage in
13 Bangladesh. We did not include MLVA data from previous genotyping studies of Bangladeshi
14 strains because no reference strains such as Ames had been published along these MLVA-8
15 (Chakraborty et al. 2012) or MLVA-15 (Fasanella et al., 2013b; Rume et al., 2016) datasets
16 required for data calibration. Next, a panel of diverse *B. anthracis* A.Br.001/002 branch strains
17 was analyzed alongside the new Bangladesh strains (Table 4). Interestingly, MLVA-31 typing
18 of Chinese isolates of the A.Br.001/002 did not provide further clues on the potential origin of
19 Bangladeshi strains, though a close relatedness, caused by imported contaminated bone meal,
20 between strains from these southeastern Asian regions has been postulated previously
21 (Fasanella et al., 2013b). Also genotypes of strains from Thailand and France were clearly
22 distinct from the Bangladeshi isolates (Table 4). Conversely, we detected suspicious

1 similarities between European strains from The Netherlands or Germany and the new strains
2 from Bangladesh. Thus, MLVA-genotype data for the most similar Central European strains
3 were also included in Table 4 to demonstrate the strains' potential close relationship with
4 strains from Tangail and Sirajganj districts.

5 6 *3.2. Whole genome sequencing and assembly*

7 Though MLVA genotyping has rather limited resolution and does not provide phylogenetically
8 reliable information, MLVA data for new strains will allow backward compatibility with
9 earlier studies in which MLVA has been a central part of genotyping. For high-resolution
10 genotyping, three strains from our collection isolated from Germany (A163) and The
11 Netherlands (CVI 131959, CVI 132064-1) along with the five new strains from Bangladesh
12 (Table 1) were genome sequenced using the Ion Torrent personal genome machine.

13 For the WGS library, at least 1,745,000 reads with a total of 470 Mbases were generated.
14 Using bowtie2-software (Langmead and Salzberg, 2012) high-quality reads were mapped to
15 Ames Ancestor chromosome, plasmid pXO1 and pXO2, respectively. This resulted in an
16 average sequencing depth of 75-fold for the chromosome and a genome coverage in relation to
17 the reference genome of higher than 98%.

18 19 *3.3. SNP discovery and analysis suggest two Bangladeshi clusters related to strains from* 20 *Central Europe*

21 Comparative analysis of the genomic sequences of five Bangladeshi, one German, and two
22 additional published Dutch *B. anthracis* isolates was carried out with *B. anthracis* Ames

1 Ancestor as reference and strain Australia_94 as outlier (SNPs of newly sequenced strains are
2 summarized in Suppl. Table 1). Next, this analysis was amended with the genomes
3 summarized in Table 2. From this complete dataset of 1,185 SNPs a group of 231 were
4 parsimony informative and 218 non-homoplasious. The resulting 218 SNPs were sufficient to
5 infer the phylogeny of these strains, especially of the Bangladeshi isolates in comparison with
6 Central European strains of the A.Br.001/002 canSNP sub-lineage as pre-selected by
7 MLVA-31. For this, a UPGMA tree was calculated according to (Kimura, 1980) and optimized
8 for maximum parsimony. The overall consistency of the optimized maximum parsimony tree
9 was > 0.97 , thus, 2.7 % of the initial 1,185 SNPs were homoplasious. Figure 2 shows a
10 maximum parsimony tree generated from these chromosomal SNPs. The Bangladeshi strains
11 fall into two distinct sub-groups B1 and B2 which are separated by 65 SNPs.

12 Notably, the Dutch isolates (CVI 131959 and CVI 132064-1) are clearly separated from
13 B2 (71 SNPs), whereas the distance between B1 and the Dutch strains is 46 SNPs. German
14 strain A163 is separated by 22 SNPs from the Bangladeshi B1/Dutch branching node (Fig. 2).
15 A sub-dataset of SNPs located on both virulence plasmids of our in-house sequenced strains
16 was also analyzed yielding 16 and 10 SNPs for pXO1 and pXO2, respectively) (Suppl. Table
17 1). However, SNP differences in this plasmid data-set lack sufficient resolution for meaningful
18 phylogenetic analysis.

19 The Bangladesh-Central Europe Cluster is clearly separated from the cluster comprising
20 the Sterne strain (55 SNPs). To this Sterne cluster also belonged K929 and C, two Danish
21 strains (Derzelle et al., 2015a), and the French isolate ANSES_08-8_20 (Girault et al., 2014b).
22 The other strains of the A.Br. 001/002 Sterne/Ames polytomy (Vergnaud et al., 2016) isolated

1 in the U.S.A. (Ames) China (A16, Han), Indonesia (A0389), Japan (52-40-NIAH, 44-NIAH)
2 or Denmark (A, B) were clearly distinct from both the Bangladesh-Central Europe cluster and
3 the Sterne cluster, respectively (Fig.2).

4 Previously, the molecular diversity of *B. anthracis* from the Netherlands has been
5 described by analyzing historic strains from this country which is nowadays rarely affected by
6 anthrax (Derzelle et al., 2015b). Further, the A.Br.001/002 sub-population can be further
7 resolved into two distinct phylogenetic sub-groups termed A01 and A02 (Girault et al., 2014c).
8 The first one is defined by the Ames strain and relatives (thus named “Ames sub-group”), the
9 other, A02 sub-group (“Sterne sub-group”), includes reference vaccine strain Sterne and the
10 six Dutch A.Br.001/002 isolates (Derzelle et al., 2015b). Also clustered in the A02 sub-group
11 are strains isolated from Northeast France. Unfortunately, little genomic data has been
12 published alongside for comparison (Girault et al., 2014c). It is unlikely that these strains also
13 constitute a link to Bangladeshi strains described here because French A.Br.001/002 strain
14 ANSES_08-8_20 described by (Girault et al., 2014b) did not closely cluster with the
15 Bangladeshi or Dutch strains (compare (Derzelle et al., 2015b) and Fig. 2).

16 The five Dutch A.Br.001/002 strains from (Derzelle et al., 2015b) were isolated during a
17 series of anthrax outbreaks on a farm in the Netherlands. The history of that outbreak, which
18 had started in July 1991 when the carcass of a bull diseased from anthrax was opened on the
19 livestock farm in Winsum (Groningen), is well investigated. Consecutively, this led to
20 contamination of the premises and to at least four additional cases occurred among dairy and
21 beef cattle on the same farm until November 1991. Ultimately, this resulted in nine animal
22 casualties. These five outbreak strains are relatively closely related to each other. Three SNPs

were detected among the five outbreak strains (Derzelle et al., 2015b), which is a typical finding (Girault et al., 2014a). One additional Dutch strain 61 SNPs away is clearly distinct and may represent another lineage within the A.Br.001/002 Sterne sub-group (Derzelle et al., 2015b).

3.4. Epidemiological analysis suggests historic importation of B. anthracis from Southeast Asia to Central Europe

It has been suggested that the A.Br.001/002 clade might have its origin in China (Simonson et al., 2009). Strains from this clade, however, have also spread to higher latitudes in Central and Western Europe and contribute to a significant part of the European *B. anthracis* diversity in Denmark (Derzelle et al., 2015b), Germany (Antwerpen et al., 2016), Belgium, the United Kingdom, the Netherlands as well as Northeastern France (Derzelle and Thierry, 2013; Girault et al., 2014a). Such a spread of related strains over vast geographic distances seems not to be uncommon for *B. anthracis* as evidenced by the A.Br.013/015 node which is basal to the Aust94 lineage. This A.Br.013/015 node and its expansion comprise isolates from the countries of Georgia, Europe and the U.S.A. (Khmaladze et al., 2014). The authors deduced from their data that the A.Br.013/015 node is not ecologically established in Europe and the USA.

Similarly, we also suggest that the sole identification of isolates from the A.Br.001/002 lineage in Central Europe does not mean that this genetic group is fully ecologically established on this continent. Conversely, the exclusive isolation of A.Br.001/002 branch strains from Bangladesh in this work and in previous studies (Chakraborty et al., 2012; Fasanella et al., 2013b; Rume et

al., 2016) clearly suggests that this clade is the predominant lineage in this Southeastern Asian country.

From the rare isolation of such A.Br.001/002 strains in Europe it may be hypothesized that their origin might have been the repeated importation of contaminated animal products such as wool or animal hides, from Southeastern Asian countries. The exact sources, however, remain vague because these geographic regions are currently still heavily undersampled. Thus, the probability that Central European strains were the source of an introduction into Southeastern Asian countries including Bangladesh is low but difficult to validate. This is due in part, because meta-data on the origins of older European strains are often fragmentary and MLVA-data for individual markers (MLVA-31) of many strains in collections are largely not published (Derzelle and Thierry, 2013; Girault et al., 2014a; Girault et al., 2014c). Similarly, it has been proposed that the presence of the A.Br.013/015 node which leads to the Aust94 branch, in Europe, the U.S.A. and possibly Georgia and Turkey actually originated from the importation of contaminated animal products (Khmaladze et al., 2014).

HRM SNP-PCR derived from our new genome data (Table 3) and from (Derzelle et al., 2015b) were used to interrogate and confirm representative SNP positions (Table 5) on branches separating the Bangladesh B1 and B2 lineage as well that leading to the Dutch strains (Derzelle et al., 2015b). As expected, the Dutch strains exhibited the ancestral allele for SNP A02/A and were derived for A02/B1-NL. The Bangladesh strains showed the ancestral alleles for these both SNPs regardless if they belonged to either group B1 or B2. Interestingly, German strain A40 which was not genome sequenced exhibited the derived state for SNP A02/A, similar to strain CVI-un1 from (Derzelle et al., 2015b), whereas genome sequenced

strain A163 had the ancestral allele. Analysis of SNP B1+NL confirmed that the Dutch strains (Derzelle et al., 2015b) and our strains from Bangladesh group B1 (Table 5) share the derived allele, whereas the German and Bangladesh B2 strains were ancestral for this marker. Finally, specific SNPs for each Bangladeshi group B1 (SNP A02/B1) and B2 (SNP A02/B2) confirm positions of all analyzed B1 and B2 strains to their respective clade and separated the German and Dutch strains which exhibited the ancestral allele for these both markers. Thus, HRM-SNP analysis verified the relationships of the genome sequenced strains and amended this with information from the MLVA-typed but not genome-sequenced strains from this study.

Conflict of interests

None of the authors have a conflict of interest to declare.

Acknowledgements

Miriam Koene is acknowledged for the generous gift of the Dutch A.Br.001/002 strains and Wolfgang Beyer for strains A163 and A40. Thanks are due to Linda Dobrzykowski and Phillip Vette for skillful technical assistance. This work was supported by funds from the German Ministry of Defense (Sonderforschungsprojekt 25Z1-S-431214). Work in Bangladesh was supported by the Bangladesh Academy of Sciences under the BAS-USDA Program in Agriculture and Life Science.

References

- Ågren, J., Finn, M., Bengtsson, B., Segerman, B., 2014. Microevolution during an anthrax outbreak leading to clonal heterogeneity and penicillin resistance. *PLoS One*. 9, e89112.
- Anaraki, S., Addiman, S., Nixon, G., Krahe, D., Ghosh, R., Brooks, T., Lloyd, G., Spencer, R., Walsh, A., McCloskey, B., Lightfoot, N., 2008. Investigations and control measures following a case of inhalation anthrax in East London in a drum maker and drummer, October 2008. *Euro Surveill*. 13.
- Antwerpen, M., Elschner, M., Gaede, W., Schliephake, A., Grass, G., Tomaso, H., 2016. Genome sequence of *Bacillus anthracis* strain Stendal, isolated from an anthrax outbreak in cattle in Germany. *Genome Announc*. 4.
- Antwerpen, M.H., Zimmermann, P., Bewley, K., Frangoulidis, D., Meyer, H., 2008. Real-time PCR system targeting a chromosomal marker specific for *Bacillus anthracis*. *Mol Cell Probes*. 22, 313-5.
- Be-Nazir, A., Sultana, Y., Fatema, D., Ara, K., Begum, N., Mostanzid, S.M., Jubayer, S., 2010. Anthrax: an emerging zoonotic disease in Bangladesh. *Bangladesh J Med Microbiol*. 04, 46-50.
- Benson, G., 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res*. 27, 573-80.
- 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 Negl Trop Dis*. 6, e1534.
- 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.
- Biswas, P.K., Islam, M.Z., Shil, S.K., Chakraborty, R.K., Ahmed, S.S., Christensen, J.P., 2012. Risk factors associated with anthrax in cattle on smallholdings. *Epidemiol Infect*. 140, 1888-95.
- Chakraborty, A., Khan, S.U., Hasnat, M.A., Parveen, S., Islam, M.S., Mikolon, A., Chakraborty, R.K., Ahmed, B.N., Ara, K., Haider, N., Zaki, S.R., Hoffmaster, A.R., Rahman, M., Luby, S.P., Hossain, M.J., 2012. Anthrax outbreaks in Bangladesh, 2009-2010. *Am J Trop Med Hyg*. 86, 703-10.
- Darling, A.E., Mau, B., Perna, N.T., 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One*. 5, e11147.
- Derzelle, S., Thierry, S., 2013. Genetic diversity of *Bacillus anthracis* in Europe: genotyping methods in forensic and epidemiologic investigations. *Biosecur Bioterror*. 11 Suppl 1, S166-76.
- Derzelle, S., Girault, G., Kokotovic, B., Angen, O., 2015a. Whole genome-sequencing and phylogenetic analysis of a historical collection of *Bacillus anthracis* strains from Danish cattle. *PLoS One*. 10, e0134699.

- 1 Derzelle, S., Girault, G., Roest, H.I., Koene, M., 2015b. Molecular diversity of *Bacillus*
2 *anthracis* in the Netherlands: Investigating the relationship to the worldwide
3 population using whole-genome SNP discovery. *Infect Genet Evol.* 32, 370-376.
- 4 Durmaz, R., Doganay, M., Sahin, M., Percin, D., Karahocagil, M.K., Kayabas, U., Otlu, B.,
5 Karagoz, A., Buyuk, F., Celebi, O., Ozturk, Z., Ertek, M., 2012. Molecular
6 epidemiology of the *Bacillus anthracis* isolates collected throughout Turkey from 1983
7 to 2011. *European Journal of Clinical Microbiology & Infectious Diseases.* 10,
8 2783-90.
- 9 Fasanella, A., Van Ert, M., Altamura, S.A., Garofolo, G., Buonavoglia, C., Leori, G., Huynh,
10 L., Zanecki, S., Keim, P., 2005. Molecular Diversity of *Bacillus anthracis* in Italy.
11 *Journal of Clinical Microbiology.* 43, 3398-3401.
- 12 Fasanella, A., Garofolo, G., Galante, D., Quaranta, V., Palazzo, L., Lista, F., Adone, R., Jones,
13 M.H., 2010. Severe anthrax outbreaks in Italy in 2004: considerations on factors
14 involved in the spread of infection. *New Microbiol.* 33, 83-6.
- 15 Fasanella, A., Di Taranto, P., Garofolo, G., Colao, V., Marino, L., Buonavoglia, D., Pedarra, C.,
16 Adone, R., Hugh-Jones, M., 2013a. Ground Anthrax *Bacillus* Refined Isolation
17 (GABRI) method for analyzing environmental samples with low levels of *Bacillus*
18 *anthracis* contamination. *BMC Microbiol.* 13, 167.
- 19 Fasanella, A., Garofolo, G., Hossain, M.J., Shamsuddin, M., Blackburn, J.K., Hugh-Jones, M.,
20 2013b. Bangladesh anthrax outbreaks are probably caused by contaminated livestock
21 feed. *Epidemiol Infect.* 141, 1021-8.
- 22 Girault, G., Blouin, Y., Vergnaud, G., Derzelle, S., 2014a. High-throughput sequencing of
23 *Bacillus anthracis* in France: investigating genome diversity and population structure
24 using whole-genome SNP discovery. *BMC Genomics.* 15, 288.
- 25 Girault, G., Parisot, N., Peyretailade, E., Peyret, P., Derzelle, S., 2014b. Draft genomes of
26 three strains representative of the *Bacillus anthracis* diversity found in France. *Genome*
27 *Announc.* 2.
- 28 Girault, G., Thierry, S., Cherchame, E., Derzelle, S., 2014c. Application of high-throughput
29 sequencing: discovery of informative SNPs to subtype *Bacillus anthracis*. *Advances in*
30 *Bioscience and Biotechnology.* 05, 669-677.
- 31 Hanczaruk, M., Reischl, U., Holzmann, T., Frangoulidis, D., Wagner, D.M., Keim, P.S.,
32 Antwerpen, M.H., Meyer, H., Grass, G., 2014. Injectional anthrax in heroin users,
33 Europe, 2000-2012. *Emerg Infect Dis.* 20, 322-3.
- 34 Jacob, D., Sauer, U., Housley, R., Washington, C., Sannes-Lowery, K., Ecker, D.J., Sampath,
35 R., Grunow, R., 2012. Rapid and high-throughput detection of highly pathogenic
36 bacteria by Ibis PLEX-ID technology. *PLoS ONE.* 7, e39928.
- 37 Keim, P., Price, L.B., Klevytska, A.M., Smith, K.L., Schupp, J.M., Okinaka, R., Jackson, P.J.,
38 Hugh-Jones, M.E., 2000. Multiple-locus variable-number tandem repeat analysis
39 reveals genetic relationships within *Bacillus anthracis*. *J Bacteriol.* 182, 2928-36.
- 40 Keim, P., Van Ert, M.N., Pearson, T., Vogler, A.J., Huynh, L.Y., Wagner, D.M., 2004. Anthrax
41 molecular epidemiology and forensics: using the appropriate marker for different
42 evolutionary scales. *Infect Genet Evol.* 4, 205-13.
- 43 Khmaladze, E., Birdsell, D.N., Naumann, A.A., Hochhalter, C.B., Seymour, M.L., Nottingham,
44 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, e102651.

Kimura, M., 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol. 16, 111-20.

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods. 9, 357-9.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The sequence alignment/map format and SAMtools. Bioinformatics. 25, 2078-9.

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 Microbiol. 6, 33.

Milne, I., Stephen, G., Bayer, M., Cock, P.J., Pritchard, L., Cardle, L., Shaw, P.D., Marshall, D., 2013. Using Tablet for visual exploration of second-generation sequencing data. Brief Bioinform. 14, 193-202.

Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 26, 841-2.

Ringertz, S.H., Hoiby, E.A., Jensenius, M., Maehlen, J., Caugant, D.A., Myklebust, A., Fossum, K., 2000. Injectional anthrax in a heroin skin-popper. Lancet. 356, 1574-5.

Rume, F.I., Affuso, A., Serrecchia, L., Rondinone, V., Manzulli, V., Campese, E., Di Taranto, P., Biswas, P.K., Ahsan, C.R., Yasmin, M., Fasanella, A., Hugh-Jones, M., 2016. Genotype analysis of *Bacillus anthracis* strains circulating in Bangladesh. PLoS One. 11, e0153548.

Schliep, K.P., 2011. phangorn: phylogenetic analysis in R. Bioinformatics. 27, 592-3.

Siddiqui, M.A., Khan, M.A., Ahmed, S.S., Anwar, K.S., Akhtaruzzaman, S.M., Salam, M.A., 2012. Recent outbreak of cutaneous anthrax in Bangladesh: clinico-demographic profile and treatment outcome of cases attended at Rajshahi Medical College Hospital. BMC Res Notes. 5, 464.

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 Microbiol. 9, 71.

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.

Tomaso, H., Bartling, C., Al Dahouk, S., Hagen, R.M., Scholz, H.C., Beyer, W., Neubauer, H., 2006. Growth characteristics of *Bacillus anthracis* compared to other *Bacillus* spp. on

the selective nutrient media Anthrax Blood Agar® and Cereus Ident Agar®.
 Systematic and Applied Microbiology. 29, 24-28.

Turnbull, P.C., 2008. World Health Organization. Anthrax in humans and animals. Vol., P.T. (Editor), ed.^eds. WHO Press, Geneva (CH).

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., Farlow, 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. 2, e461.

Vergnaud, G., Girault, G., Thierry, S., Pourcel, C., Madani, N., Blouin, Y., 2016. Comparison of French and worldwide *Bacillus anthracis* strains favors a recent, post-Columbian origin of the predominant North-American clade. PLoS One. 11, e0146216.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., Madden, T.L., 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics. 13, 134.

Zhou, W., Sun, Y., Zhu, L., Zhou, B., Liu, J., Ji, X., Wang, X., Wang, N., Gu, G., Feng, S., Qian, J., Guo, X., 2015. Investigation of anthrax cases in north-east China, 2010-2014. PLoS One. 10, e0135777.

Tables

Table 1: Strains of *B. anthracis* belonging to canSNP-group A.Br.001/002 Sterne from Bangladesh, Germany and the Netherlands used for analysis in this study.

Name	Origin	Region	Country	Sampling date	Additional Information
CVI127491 -V08551	Bull	Winsum	The Netherlands	01.07.1991	carcass opened on farm, contamination of premises
CVI128268	Bull			14.08.1991	Additional;
CVI131185	Beef Cattle			19.10.1991	cases
CVI131959**	Dairy Cattle			28.10.1991	from the
CVI132064-1**	Cattle			30.10.1991	same farm
A163	Cattle	Aulendorf	Germany	2008	
A40		Giessen		unknown	
Tangail-1*	Soil	Village: Katuli	Bangladesh	16.09.2013	same stable of
Tangail-2*		(24.2085 ⁰ N to 89.8376 ⁰ E), Upazila: Tangail		26.09.2013	one diseased cow;
Tangail-3*		Sadar,		26.09.2013	stable of a diseased cow;
Tangail-4/1*		District: Tangail.		26.09.2013	animal burial site in
Tangail-4/2*				26.09.2013	a flooded forest
Sirajganj-1/1*	Soil	Village: Char Kaijuri	Bangladesh	27.09.2013	sampled from a
Sirajganj-1/2*		(24.1731 ⁰ N to 89.6709 ⁰ E), Upazila: Shahjadpur,		27.09.2013	water pit with carcass;
Sirajganj-2*		District: Sirajganj.		09.07.2013	stable of a diseased cow

Strains isolated from Bangladeshi samples are labeled with an asterisk (*); whole genome sequences were performed for all strains marked in **bold** letters. Strains from (Derzelle et al., 2015b) were re-sequenced because no genomic data has been released as part of the original publication (**).

Table 2: Additional *B. anthracis* genome sequences from publically available databases used for analysis.

Strain designation	Country	Accession number
K929	Denmark	SAMEA3736003
A	Denmark	SAMEA3735716
B	Denmark	SAMEA3735734
C	Denmark	SAMEA3736006
A16	China	CP001970.1
Han	China	CP008854
52-40-NIAH	Japan	BBWY01000001 to BBWY01000036
44-NIAH	Japan	BBWX01000001 to BBWX01000124
A0389	Indonesia	ABLB00000000.1
Anses_08-8_20	France	JHCB00000000
Sterne	South Africa	AE017225.1
Australia_94	Australia	AAES00000000.1
Ames Ancestor	USA	AE017334.2

Table 3: SNPs and respective primer sequences used for HRM-SNP

SNP	Position *	Variation	Forward primer (5'-3')	Reverse primer (5'-3')	Reference
A02/B1	59522	C to T	GCAAACGATTCGTGTTCTATTATT	ACCTGCGATTGAACCTAGCA	This work
A02/B2	549428	G to A	AGCGGGAAGATTTGCGCTTA	GAGCGATACTCTGTTTCATTTCGT	This work
A02/B1+NL	3766679	G to A	ATCCGCGTAACGCATTTC	AGTCGGCAGTTGCAGATTAC	This work
A02/B1-NL	4248098	G to A	CGTAAACGATTCAACGATCATCT	GCGGAAATACATTACATGACTCC	Derzelle et al. (2015)
A02/A	4961938	G to A	TTGTTAACGCAACATCTCTACGC	CATTTAATACGCCACCAATTACG	Derzelle et al. (2015)

*in relation to the Ames Ancestor reference genome (AE017334.2)

1 **Table 4: Results of MLVA-31 typing.**

Strain	Country	bams1	bams3	bams5	bams13	bams15	bams21	bams24	bams30	bams31	pX01	pX02	VNTR12	VNTR16	VNTR19	VNTR23	VNTR32	VNTR35
Ames		490	548	300	823	418	660	599	733	774	121	137	113	269	90	192	562	112
CVI 127491-V08551											124							
CVI 128268	The Netherlands	296	578	378	869	612	660	599	677	647	124	133	113	140	90	180	385	112
CVI 131959											124							
CVI 131185											127							
CVI 132064-1											127							
A163	Germany	296	578	378	878	612	660	599	677	647	127	133	113	140	90	180	385	112
A40											133		111					
Tangail-1	Bangladesh													140	90			112
Tangail-2														140	90			112
Tangail-3		296	578	378	869	612	660	599	677	647	127	133	113	140	90	180	385	112
Tangail-4/1														140	90			112
Tangail-4/2														269	93			106
Sirajganj-1/1	Bangladesh						660	599		647								
Sirajganj-1/2		296	578	339	869	576	400	599	677	647	127	133	113	140	90	180	385	112
Sirajganj-2							360	553		667								
Ba#112	France ^a				786				733	774	124	139		269		192		
Ba#109					434				492	728	127	135		269		192		
Ba#107		490	578	378	434	612	660	599	695	728	121	141	113	269	90	180	(562)	112
Ba#084					878				668	783	130	133		140		180		
(Anses08-8_20)																		
A0719CHI ^b	China ^c	490	578	378	841	612	660	599	733	774		141				192		
A0591CHI		“	“	“	“	“	“	“	“	“		133				192	562	112
A0551CHI		“	“	“	“	“	“	“	“	“	127	135	113	269	90	192		
A0587CHI		“	“	“	“	“	“	“	“	“		137				180		
SY (2009)	North-East China ^d	“	“	“	“	“	“	“	“	“	127	139						
LZ (2012)		“	“	“	“	“	“	“	“	“	130	139	117	277	93	192	562	112
NMG (2011)		“	“	“	“	“	“	“	“	“	130	133						
A0935THA	Thailand ^c	“	“	“	“	“	“	“	“	“		143		140		192		
A0936THA		“	“	“	“	“	“	“	“	“	127	143	113	140	90	216	562	112
A0937THA		“	“	“	“	“	“	“	“	“		137		269		204		

2 Markers identical in Bangladeshi strains and Ames reference strain are not shown in this table,
3 namely bams22 (711 bp), bams23 (633 bp), bams25 (392 bp), bams28 (498 bp), bams34 (509
4 bp), bams44 (420 bp), bams51 (498 bp), bams53 (234 bp), CG3 (153 bp), vrrA (307bp), vrrB1
5 (224 bp), vrrB2 (149 bp), vrrC1 (567 bp), vrrC2 (522 bp), and VNTR17 (386 bp). Alleles
6 differing from Bangladesh strain “Tangail-1” are indicated in **bold** letters. Value for VNTR32
7 in brackets was calculated from data of the identical MLVA marker bams1.

8 ^adata for the French strains is from (Thierry et al., 2014); ^bMLVA-31-data for A0719CHI was
9 amended in house; “data not part of the original publications (^cChina, Thailand (Van Ert et al.,
10 2007); ^dNorth-East China (Zhou et al., 2015)).

Table 5: Results of HRM-SNP typing.

Strain	Country	A02/A*	A02/B1-NL*	A02/B1+NL	A02/B1	A02/B2
Ames Ancestor		ancestral (G)	ancestral (G)	ancestral (G)	ancestral (C)	ancestral (G)
CVI 127491-V08551						
CVI 128268	The Netherlands	ancestral (G)	derived (A)	derived (A)	ancestral (C)	ancestral (G)
CVI 131959						
CVI 131185						
CVI 132064-1						
A163	Germany	ancestral (G)	ancestral (G)	ancestral (G)	ancestral (C)	ancestral (G)
A40		derived (A)				
Tangail-1						
Tangail-2						
Tangail-3	Bangladesh	ancestral (G)	ancestral (G)	ancestral (G)	ancestral (C)	derived (A)
Tangail-4/1						
Tangail-4/2						
Sirajganj-1/1	Bangladesh	ancestral (G)	ancestral (G)	derived (A)	derived (T)	ancestral (G)
Sirajganj-1/2						
Sirajganj-2						

* published as canSNP A02/A and A02/B1-NL by (Derzelle et al., 2015b). Indicated in **bold** are strains with genomes sequenced in this study.

Figure legends

Figure 1: Map of Bangladesh with locations of sampling sites. Stars indicate where samples were drawn in the regions of Tangail (A) and Sirajganj (B). Map of Bangladesh from <http://www.worldofmaps.net>.

Figure 2: Phylogenetic tree derived from chromosomal SNP analysis of Bangladeshi strains and their close relatives. A maximum parsimony minimum spanning tree was inferred from 218 non-homoplasious chromosomal SNPs (out of 1,185 SNPs in the complete dataset). Positions of five Bangladeshi, two Dutch and one German *B. anthracis* strains in relation to their relatives of the A.Br.001/002 Sterne and Ames canSNP-groups are shown. Two distinct Bangladeshi regional clusters for Sirajganj and Tangail are indicated as B1 and B2, respectively. The *B. anthracis* Australia_94 chromosome served as an outlier for rooting the A.Br.001/002/Ames polytomy. Arrows indicate the position of new SNPs (IV.: SNP “A02/B1+NL”, I.: SNP “A02/B1” and II.: SNP “A02/B2”) or previously published SNP (specific for canSNP A.Br.001/002 sub-group A02 III.: SNP A02/B1-NL (Derzelle et al., 2015b)). The length of each branch is proportional to the number of SNPs identified between strains (the number of SNPs specific for each branch is indicated).