Using genomics to improve *Bacillus anthracis* diagnostics and outbreak investigations

Joakim Ågren

Faculty of Veterinary Medicine and Animal Science
Department of Biomedical Sciences and Veterinary Public Health
Uppsala

and

Department of Bacteriology, National Veterinary Institute, Uppsala

Doctoral Thesis
Swedish University of Agricultural Sciences
Uppsala 2014

Acta Universitatis agriculturae Sueciae
2014:60

Cover: Scanning electron microscopy image of *Bacillus anthracis* spores.

(photo: Joakim Ågren, SVA, and Leif Ljung, Uppsala University)

ISSN 1652-6880
ISBN (print version) 978-91-576-8068-6
ISBN (electronic version) 978-91-576-8069-3

Print: SLU Service/Repro, Uppsala 2014

Using genomics to improve *Bacillus anthracis* diagnostics and outbreak investigations

Abstract

The bacterium *Bacillus anthracis* causes the disease anthrax, primarily in herbivores but many mammals are susceptible to the disease. Its infective form is as a dormant spore that can lie in the soil for decades. Thus, in its cycle of infection, it spends most of the time in an inactive state and replication-induced DNA-mutations are therefore kept at a minimum. Partly due to these long periods of inactivity, all *B. anthracis* isolates found in the world are genetically very similar. This makes strain characterization difficult and requires high-resolution technologies. *Bacillus anthracis* also has similar DNA-content as other *Bacillus* spp. and therefore diagnostic cross-reactions are not uncommon.

Anthrax incidence has steadily declined in the world during the last century but there are still endemic areas. In 2008 and in 2011 Sweden suffered two large and costly outbreaks, most likely caused by the disturbance of old anthrax epizootic graves from the 1940s and 1950s. Several studies emanated from these outbreaks including how the bacteria in cows treated with penicillin developed penicillin resistance.

Next-generation sequencing (NGS) has revolutionized the way DNA is sequenced and the whole genome (i.e., all the DNA) of a bacterium can now be sequenced in only a few days. In this doctoral thesis, NGS and genomics were used to improve our capability to deal with anthrax outbreaks.

Genomic and genetic studies were applied to identify the non-anthrax *Bacillus* spp. available that were most closely related to *B. anthracis*. How well the strains could mimic *B. anthracis* in a model system for *B. anthracis* spores was evaluated and the best model strains found have since been used in exercises and as controls in real samples. For genome comparisons, a software was created that can compare the genetic content of several hundreds of bacterial genomes. This ensures a rapid characterization of an outbreak pathogen's genome. The software was also used to *in silico*-compare all published anthrax PCR-assays to determine which assays that had the highest specificity. This workflow ensures that our molecular diagnostics are as specific for *B. anthracis* as possible.

By using isolates from an anthrax outbreak, the mechanisms of beta-lactam resistance in *B. anthracis* were studied and chromosomal mutations in a negative sigma factor were found to be the cause. The genomic divergence of a strain during an outbreak was also studied to gain knowledge of the strengths and limitations of using NGS for epidemiological investigations.

In summary, this thesis describes different genomic approaches that have improved diagnostic methods, explained ambiguous antimicrobial resistance findings and enhanced the resolution of genomic epidemiological investigation to ensure a robust handling of future anthrax outbreaks.

Keywords: Bacillus anthracis, anthrax, genome, genomics, outbreak, diagnostics, PCR, biomarkers, penicillin resistance

Author's address: Joakim Ågren, SVA, Department of Bacteriology,
SE-751 89 Uppsala, Sweden *E-mail:* joakim.agren@sva.se

The capacity to blunder slightly is the real marvel of DNA. Without this special attribute, we would still be anaerobic bacteria and there would be no music.
- Lewis Thomas

Contents

List of Publications

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

I Fricker M*, **Ågren J***, Segerman B, Knutsson R, Ehling-Schulz M (2011). Evaluation of *Bacillus* strains as model systems for the work on *Bacillus anthracis* spores. *Int J Food Microbiol.* Mar 1;145 Suppl 1:S129-36.

II **Ågren J***, Sundström A*, Håfström T, Segerman B (2012). Gegenees: Fragmented Alignment of Multiple Genomes for Determining Phylogenomic Distances and Genetic Signatures Unique for Specified Target Groups. *PLoS ONE* 7(6): e39107. doi:10.1371/journal.pone.0039107

III **Ågren J**, Hamidjaja RA, Hansen T, Ruuls R, Thierry S, Vigre H, Janse I, Sundström A, Segerman B, Koene M, Löfström C, Van Rotterdam B, Derzelle S (2013). *In silico* and *in vitro* evaluation of PCR-based assays for the detection of *Bacillus anthracis* chromosomal signature sequences. *Virulence* 2013; 4:671 - 685.

IV **Ågren J**, Finn M, Bengtsson B, Segerman B (2014). Microevolution during an Anthrax Outbreak Leading to Clonal Heterogeneity and Penicillin Resistance. *PLoS ONE* 9(2): e89112. doi:10.1371/journal.pone.0089112

(* contributed equally)

The contributions to the papers included in this thesis were as follows:

I Martina Fricker, Joakim Ågren and Bo Segerman performed the experiments and analyzed the results. Martina Fricker wrote the first draft of the manuscript with input from Monika Ehling-Schulz, Rickard Knutsson, Bo Segerman and Joakim Ågren.

II Bo Segerman and Joakim Ågren conceived and designed the experiments. Joakim Ågren, Anders Sundström and Therese Håfström performed the experiments. Joakim Ågren and Bo Segerman wrote the paper. Anders Sundström and Bo Segerman designed and implemented the software.

III All authors gave suggestions regarding study design. Joakim Ågren and Anders Sundström performed the in silico experiments. Ingmar Janse and Raditijo A Hamidjaja designed and distributed the laboratory trial. Håkan Vigre performed the statistical analyses. Joakim Ågren, Raditijo Hamidjaja, Trine Hansen, Robin Ruuls, Simon Thierry, Håkan Vigre, Ingmar Janse, Anders Sundström, Bo Segerman, Miriam Koene, Charlotta Löfström and Bart Van Rotterdam contributed to the writing of the paper that was first drafted by Sylviane Derzelle.

IV Joakim Ågren and Bo Segerman conceived and designed the experiments. Joakim Ågren and Maria Finn performed the experiments. Joakim Ågren analyzed the data. Joakim Ågren, Bo Segerman and Björn Bengtsson wrote the paper.

Abbreviations

APHIS	US Animal and Plant Health Inspection Service
bp	base pair
BSL-3	biosafety level 3
canSNP	canonical single nucleotide polymorphism
ECF	extracytoplasmic sigma factor
FPKM	fragments per kilobase of exon per million fragments mapped
FRET	fluorescence resonance energy transfer
GABRI	Ground Anthrax Bacillus Refined Isolation
HEPA	high-efficiency particulate air
LD_{50}	median lethal dose
MATH	microbial adhesion to hydrocarbon
MIC	minimum inhibitory concentration
MLST	multilocus sequence typing
MLVA	multiple locus variable-number tandem repeat analysis
MRSA	methicillin-resistant *Staphylococcus aureus*
NGS	next-generation sequencing
PBP	penicillin-binding protein
PCR	polymerase chain reaction
PLET	polymyxin, lysozyme, ethylenediaminetetraacetic acid, thallous acetate
rsiP	anti-sigma factor P gene
sigP	sigma factor P gene
SNP	single nucleotide polymorphism
SNR	single nucleotide repeats
SNV	single nucleotide variant
spoIIIAB	sporulation stage III AB gene
SVA	Swedish National Veterinary Institute
VNTR	variable number tandem repeat
WGS	whole genome sequencing

1 Background

1.1 Bacillus anthracis

Bacillus anthracis is a rod-shaped, Gram-positive, spore-forming aerobic bacterium that causes the disease anthrax (Mock & Fouet, 2001). The bacterium was intensively studied by Robert Koch and Louis Pasteur and in 1876 Koch determined the full life cycle of the bacterium and it was the model for Koch's postulates on the transmission of infectious disease (Sternbach, 2003; Koch, 1876).

The bacterium belongs to the large genus of *Bacillus* and it is often placed in the so-called *Bacillus cereus* group (*B. cereus* sensu lato), which is a sub-group of closely related species (Daffonchio *et al.*, 2000). The group comprises *B. anthracis, B. cereus, B. mycoides, B. pseudomycoides, B. thuringiensis* and *B. weihenstephanensis*. It has been proposed that *B. anthracis*, *B. cereus* and *B. thuringiensis* should form a single species, in which *B. anthracis* would be one lineage (Helgason *et al.*, 2000) due to their close relatedness and the inability to fully differentiate all of them using 16S ribosomal RNA sequencing (Bavykin *et al.*, 2004) or multilocus sequence typing (MLST) (Priest *et al.*, 2004).

In contact with free oxygen, *B. anthracis* forms an endospore (Roth *et al.*, 1955), which has high levels of resistance to heat, drought, chemicals and UV-light (Kim *et al.*, 2004; Nicholson *et al.*, 2000; Watson & Keir, 1994). These endospores are metabolically inactive and designed to survive for long periods of time until conditions are suitable and they can germinate into vegetative cells again.

1.1.1 Closely related *Bacillus* spp.

Genetically, the major difference between *B. anthracis* and some of the closely related *B. cereus*-group strains are the two big virulence plasmids pXO1 and

pXO2. The pXO1 plasmid (182 kb) encodes toxins (Okinaka *et al.*, 1999) and the pXO2 plasmid (95 kb) encodes a protective poly-D glutamate capsule that is essential for virulence (Makino *et al.*, 1989; Green *et al.*, 1985; Uchida *et al.*, 1985). The chromosome of *B. anthracis* is 5.2 Mb and the main features discriminating it from the chromosomes of related *Bacillus* spp. are a nonsense mutation in the *plcR* gene (Agaisse *et al.*, 1999) together with four prophages of which only one of them has been found in other *Bacillus* spp. (Sozhamannan *et al.*, 2006; Read *et al.*, 2003). The transcriptional activator PlcR is responsible for the expression of extracellular proteins such as hemolysins, enterotoxins and proteases in *B. cereus* and *B. thuringiensis* (Slamti *et al.*, 2004; Gohar *et al.*, 2002). In *B. anthracis* the *plcR* gene product is truncated and non-functioning, which gives *B. anthracis* a different secretome than *B. cereus* (Sastalla *et al.*, 2010).

The terms ‘close neighbors’ or ‘near-neighbors’ are often used when describing strains of *Bacillus* spp. that are genetically very closely related to *B. anthracis* (Greenberg *et al.*, 2010; Kolsto *et al.*, 2009; Taitt *et al.*, 2008). These strains have chromosomes which are highly similar to the *B. anthracis* chromosome and some of them also have virulence plasmids, which resembles the plasmids pXO1 and pXO2 (Klee *et al.*, 2010; Hoffmaster *et al.*, 2004). The close neighbors comprise strains of *B. cereus* and *B. thuringiensis*, which are species that can cause food-poisoning and have insecticidal properties, respectively (Stenfors Arnesen *et al.*, 2008; Aronson *et al.*, 1986). The genetic similarities of these strains are illustrated in Figure 1 where genomes from the *B. cereus*-group of differing relatedness have been compared using the Gegenees software (Ågren *et al.*, 2012).

There are also reports of close neighbors causing anthrax-like disease. In 1987, the strain *B. cereus* G9241 caused anthrax-like pneumonia in welders in southern USA (Miller *et al.*, 1997). The strain had a plasmid, which shared 99.6 % similarity to the toxin-producing pXO1 plasmid and was confirmed to be virulent in a mouse model (Hoffmaster *et al.*, 2004). In 2001-2002 in Côte d’Ivoire and in Cameroon in 2004, closely related isolates caused anthrax-like disease leading to death in four infected chimpanzees and in one gorilla. The isolates were clearly not typical *B. anthracis* as they were motile and resistant to the gamma phage (Klee *et al.*, 2006), which *B. anthracis* is not. Genome sequencing of one of the isolates revealed two plasmids almost identical to pXO1 and pXO2 and also showed that the sequenced isolate was one of the most closely related to *B. anthracis* ever described (Klee *et al.*, 2010).

Organism	1	2	3	4	5	6	7	8	9	10	11	12	13
1: B. anthracis Vollum	100	100	95	93	93	83	81	74	73	70	69	68	56
2: B. anthracis Sterne	100	100	95	93	93	84	82	74	74	71	69	68	57
3: B. thuringiensis 4AJ1	94	95	100	93	92	83	81	74	73	70	68	68	55
4: B. thuringiensis 97-27	92	93	93	100	92	83	82	74	74	71	69	68	57
5: B. cereus NVH0597-99	92	92	92	92	100	83	81	74	73	71	69	69	57
6: B. thuringiensis 4Y1	83	83	83	83	83	100	83	75	74	71	68	68	56
7: B. cereus ATCC 10987	81	81	81	82	81	83	100	74	73	70	68	68	57
8: B. cereus ATCC 14579	74	74	74	74	74	75	74	100	89	83	69	69	57
9: B. thuringiensis 10792	73	73	73	73	73	74	73	88	100	83	68	68	56
10: B. thuringiensis 35646	71	71	71	71	71	72	71	84	85	100	67	67	54
11: B. weihenstephanensis	68	69	68	68	68	68	68	69	68	66	100	91	58
12: B. mycoides ATCC 6462	68	68	68	68	68	68	68	68	68	66	91	100	56
13: B. pseudomycoides 12442	55	56	55	56	55	56	56	56	56	53	57	56	100

Figure 1. Phylogenomic overview matrix created in Gegenees showing the average genomic core similarities between genomes. Final scores were calculated using a threshold value of 20 % for the normalized BLAST-scores to exclude genetic material not included in the core genome. Green represents a high similarity and yellow/orange represents a lower similarity.

The majority of the *Bacillus* spp. genomes in the genome databases are pathogenic strains of which many represent these close-neighbor strains (Ehling-Schulz & Messelhausser, 2013; Segerman, 2012). This means that much has been divulged on the genomic similarities of the *B. cereus* group. However, there are still most likely many more variants of anthrax-like bacteria in the environment that we know nothing about. These closely related bacteria can cause molecular anthrax-diagnostics to fail or produce ambiguous results because of the similarity of the genomes. This is especially a problem when analyzing environmental samples such as soil (Kuske *et al.*, 2006; Marston *et al.*, 2006) and the molecular assays used must be as specific for *B. anthracis* as possible.

1.1.2 Penicillin resistance

Bacillus anthracis is considered susceptible to beta-lactam antibiotics such as penicillin but a few contradictory findings have been made. In 1947, Barnes et al. showed that the bacterium could produce penicillinase (Barnes, 1947). In 2004, Athamna et al. showed that *B. anthracis* could be made resistant to a variety of antibiotics (e.g., ciprofloxacin, penicillin G and tetracycline) *in vitro* and suggested that monitoring of resistance in clinical cases should be made mandatory (Athamna *et al.*, 2004). Fortunately though, reports of resistant isolates from fresh clinical samples are rare and one review article states that the total number of reported isolates was only five by 2004 (Turnbull *et al.*, 2004).

It has been shown that *B. anthracis* has two beta-lactamase genes but that their transcription is very low and was not induced by beta-lactam presence (Chen *et al.*, 2003). The transcription of the beta-lactamases was shown to be controlled by an extracytoplasmic sigma factor (ECF) SigP and its anti-sigma factor RsiP (Ross *et al.*, 2009) and it was further shown that this beta-lactam feedback system is non-functioning in prototypical *B. anthracis*-isolates. However, in a penicillin resistant strain, the *rsiP* gene was nonsense mutated and the strain constitutively produced beta-lactamases (Ross *et al.*, 2009).

During a Swedish anthrax outbreak in cattle in 2011, several resistant clones were isolated from different animals that succumbed to the disease during or after penicillin treatment (Paper **IV**). Due to its low cost and availability, penicillin is the recommended treatment of animal anthrax (WHO, 2008) and it should probably stay that way due to its success in curing the disease in many outbreaks for many years. But an in-depth genomic analysis of these resistant isolates is still motivated to gain understanding into how resistance can arise.

1.1.3 Life cycle and pathogenesis

The bacterium spends the majority of its life as a dormant endospore in the soil. Exactly how an animal contracts the disease is not fully understood but ingestion and also inhalation of spores are considered the most likely routes of infection in animals (WHO, 2008). Insects have also been incriminated as being a vector for the disease (Fasanella *et al.*, 2013b; Braack & De Vos, 1990; Turell & Knudson, 1987). In an animal model it has been shown that when spores enter the host they are phagocytosed by macrophages, but instead of being lysed they germinate inside the macrophage to their vegetative state (Guidi-Rontani *et al.*, 1999). They then multiply in the lymphatic system followed by their release into the bloodstream, and as many as 10^7 to 10^8 organisms per milliliter of blood can be present in the late stages of the disease (Dixon *et al.*, 1999). The pXO2-encoded capsule substance is here essential for evading phagocytosis of the vegetative *B. anthracis* cells (Drysdale *et al.*, 2005). The pXO1 plasmid is responsible for a tripartite exotoxin system consisting of 1) an edema factor that causes severe edema, 2) a lethal factor that causes necrosis of tissue and 3) a protective antigen that couples with either of the two toxin factors to produce one of the two binary active toxins (Young & Collier, 2007). The toxemia leads to systemic shock and death (Dixon *et al.*, 1999). Due to scavenging or terminal hemorrhaging, blood can pour out of the animal after death (Bellan *et al.*, 2013; Shafazand *et al.*, 1999). When bacteria in the blood come into contact with free oxygen, they start to

sporulate again (Hugh-Jones & Blackburn, 2009) and complete the bacterium's life cycle.

1.1.4 Biosafety and anthrax in humans

Anthrax is a zoonotic disease, i.e., it can be transmitted between animals and humans. However, it is primarily a disease of herbivores. Human infections are rare and the major sources are contact with infected animals and working with contaminated animal products (WHO, 2008). *Bacillus anthracis* is classified as a risk group 3 organism because of its pathogenic properties and work with this organism must be conducted under biosafety level 3 (BSL-3) conditions (AFS.2005:1) due to the risks involved. Some Swedish requirements for BSL-3 work are: restricted entry to the lab and only via an airlock, separate HEPA-filtered ventilation with negative pressure, solid and liquid waste must be autoclaved or, in another approved way, decontaminated before leaving the laboratory and all work with infected material must take place in a biological safety cabinet class I or II. Personal protective equipment according to risk assessment must be used and in the case of *Bacillus anthracis* full protective clothing including breathing protection must be used. Figure 2 shows work inside a class II biological safety cabinet in the BSL-3 laboratory and the personnel is wearing the protective clothing and an over-pressure respirator commonly used in BSL-3 laboratories.

Figure 2. Laboratory work inside a BSL-3 facility (Photo: Anna Sollén/SVA)

Laboratory-acquired infections have been reported (Pike *et al.*, 1965) and even under modern laboratory standards, infections have occurred (CDC, 2002).

There are three different well-known manifestations of human anthrax disease and a fourth one recently discovered. The cutaneous form is the most common one, in which spores come in contact with a lesion in the skin (Pile *et al.*, 1998). Black necrotic lesions are formed and the name anthrax comes from these black ulcers as 'anthrax' is derived from the Greek word for coal. The prognosis for cutaneous infections if treated with antibiotics is good (Pile *et al.*, 1998). The second form is gastrointestinal that arises from eating meat contaminated with *B. anthracis* spores. This is a rare type of infection and it is believed that the mortality rate without treatment is high (Pile *et al.*, 1998). The most severe form of infection is caused by inhalation of spores into the lungs. The infectious dose is relatively high with an estimated median lethal dose (LD_{50}) of 2,500-55,000 spores (Inglesby *et al.*, 2002). Historically, the fatality rate has been reported to be around 92 %, but during the anthrax terrorism attack in 2001 the rate was instead 45 % (Holty *et al.*, 2006). The difference stems from advances in health care where rapid diagnosis leads to treatment in

the earlier phases of the disease. If the disease progresses to the fulminant septic phase, the mortality, as determined from case reports, is 97 % regardless of treatment (Holty *et al.*, 2006). In the last ten years in Europe there have been several cases of a new form of anthrax referred to as injectional anthrax (Ramsay *et al.*, 2010; Ringertz *et al.*, 2000). The form has been seen among heroin addicts who have injected heroin contaminated with *B. anthracis* spores. This form has an estimated mortality rate of over 30 % (Grunow *et al.*, 2013).

1.1.5 Biological warfare and bioterrorism

A 1993 report predicted that the release of 100 kg of anthrax spores upwind of Washington, D.C. from a single airplane would result in the death of 130,000 to 3 million humans (U.S. Congress, 1993). *Bacillus anthracis* is one of the four bacterial category A bioterrorism agents listed by the US Centers for Disease Control and Prevention. Category A agents are by definition easy to disseminate and/or easily transmitted person-to-person, result in high mortality rates, can cause public panic and require special action for public health preparedness (CDC, 2014). There is however, a discussion on whether non-experts can actually produce "weapon-grade" spores of *B. anthracis*, i.e., of uniform particle size, low electrostatic charge etc., to optimize aerosolization (Inglesby *et al.*, 2002). *Bacillus anthracis* is also specified as a select agent by the Animal and Plant Health Inspection Service (APHIS) of the Unites States Department of Agriculture due to its risk to animal health (US Department of Agriculture, 2014).

Although not believed to have an antagonistic cause, the outbreak of foot-and-mouth disease in the UK in 2001 has been estimated to have cost £2.4–4.1 billion (Kao, 2002) and has highlighted the impact an agricultural disease can have on society. Since anthrax is primarily an animal disease, there is of course a risk of a terrorist attack targeting animals instead of humans. Since many animals are more susceptible to the disease than humans are, the spores used in an animal attack would probably not have to be prepared to the weapon-grade form needed for a successful attack against humans.

During World War I, *B. anthracis* was weaponized when vials containing spores had been placed inside sugar lumps for potential use against animals (Redmond *et al.*, 1998). Since the 1940's, *B. anthracis* has been used in the biological warfare programs of UK, US and the Soviet Union (Inglesby *et al.*, 2002). In 1942-43, aerosolization of anthrax spores with explosive devices was tested on the island of Gruinard in Scotland. Animals placed on the island, as an indicator of a successful aerosolization, did die of anthrax and the spores persisted in the soil for four decades until the island was finally decontaminated using formaldehyde (Manchee *et al.*, 1994).

Three major anthrax-incidences have been described and they emphasize the risk of aerosolized spores. In 1979, there was an outbreak of inhalational anthrax in Sverdlovsk (former Soviet Union) due to an accidental release of spores from a military facility. Ninety-six people were reported to be infected and 64 died. The victims worked or lived in the wind direction from the facility. In the same direction, up to 50 km from the facility, animals also contracted anthrax (Meselson *et al.*, 1994). A large-scale anthrax attack was performed in July 1993 when the religious group Aum Shinrikyo aerosolized *B. anthracis* spores from the roof of an eight-story building in Tokyo, Japan. Although successful in aerosolization of viable spores, they had used the Sterne 34F2 strain, which is used in Japan for animal vaccination (Takahashi *et al.*, 2004). The strain lacks the pXO2 plasmid and no human cases were reported. The most famous bioterrorism incident is the anthrax letter attack in the USA in 2001. The letters contained fully virulent anthrax spores and were sent to news media offices and to two senators. The result of this terrorist action was 11 cases of cutaneous anthrax and 11 cases of inhalational anthrax and 5 of these cases were fatal (Jernigan *et al.*, 2002).

The incidence with the deliberate spread of anthrax in the US has raised the question of the possible use of anthrax for agroterrorism and APHIS has created a template for early-responders. The template is based on approximately 20 ecological and epizootiological criteria that can assist in the differentiation of a natural outbreak to a non-natural outbreak (US Department of Agriculture, 2007).

The history of anthrax, and especially the anthrax letters, resulted in a fear-potential that can have implications when handling a natural animal outbreak. However, it should be stressed that the massive amount of purified spores released in an antagonistic scenario and the high risk of infection in humans that follows is an extreme situation. It resembles nothing that can happen during a natural animal case (WHO, 2008).

1.2 Outbreaks and disease control

Anthrax is a notifiable disease under the Swedish Epizootic Act (SFS.1999:657, 1999), which means that any suspicion and/or occurrence of the disease must be reported to the Swedish veterinary authorities. Anthrax is also included in the disease list created by the World Organisation for Animal Health (OIE), which means that under the OIE Terrestrial Animal Health Code occurrences must be reported to the OIE (OIE, 2013).

Disease control

The following brief summary on disease control of anthrax is based on the recommendations from the WHO (WHO, 2008) and also on how the Swedish anthrax cases in 2008, 2011 and 2013 were handled.

After the diagnosis has been confirmed the remaining animals in the herd should as soon as possible be moved from the area of the first incident and be kept under strict observation. The first area should then as quickly as possible be decontaminated to prevent spread to wildlife. The area should also be placed under visiting restrictions and people and machines leaving the area should be properly decontaminated. If no other animals are showing symptoms there is a chance of the index case being an isolated infection and in this case a vaccination of the remaining animals can be performed. However, if more animals are showing symptoms or are dying, it is likely that many animals might have been infected and the disease progression should be halted by the use of penicillin. This will most likely save the majority of animals and if some animals still die from the toxins, the bacterial load has still been reduced and the carcasses will not contribute to further spread or contaminations (WHO, 2008). Culling of the remaining animals was performed during the 2008-outbreak due to practical issues of the decontamination of the farm and subsequent destruction (Lewerin *et al.*, 2010). If it is possible to avoid, culling is not necessary as anthrax is not a chronic infection and the animals usually respond well to penicillin treatment.

Anthrax worldwide

In Europe and in parts of North America, sporadic outbreaks can follow the disturbance of old anthrax epizootic graves, such as at the creation of a ditch or a road (WHO, 2008). In some Mediterranean countries like Italy (Garofolo *et al.*, 2011; Garofolo *et al.*, 2010) and Turkey (Durmaz *et al.*, 2012) there are endemic areas in which cattle, humans and wild animals have been infected. In Canada, there have been several reports on anthrax in wild Bison (Shury *et al.*, 2009; Nishi *et al.*, 2007). In Texas there is an area with almost enzootic anthrax in white-tailed deer (Blackburn & Goodin, 2013). In several Asian countries the disease is still a large problem and in Cambodia and Vietnam it is even considered enzootic. In Africa, outbreaks occur in national parks e.g., in South Africa, Botswana, Namibia and Uganda. The situation is considered hyperendemic in Zambia and Zimbabwe (WHO, 2008).

Anthrax outbreaks in Sweden

In the first half of the 20th century, anthrax outbreaks were common in Sweden with up to hundreds of reported outbreaks per year. One cause of these

outbreaks was shown to be imported contaminated meat-and-bone meal (Rutqvist L, 1957). In a study performed by the Swedish National Veterinary Institute (SVA), old records of anthrax cases between 1916 and 1961 were reviewed and summarized (Elvander, 2011). Over 3000 farms were identified, but the exact locations of the graves were not disclosed in these old records. However, the fact that a certain farm has had anthrax in the past can be of use in outbreak investigations.

In 1981 there was a confirmed case of anthrax at a farm in the northern part of the province of Uppland. During the following 27 years, samples were sent to SVA, especially in summer-time, but no sample tested positive. The outbreak in 2008 was unexpected and without classical clinical signs it was not suspected to be anthrax at first. The outbreak in 2011 was more efficiently dealt with after method developments and lessons learned from 2008.

The three latest outbreaks in Sweden are described below as they were the incitement of this thesis and they also provided good opportunities to evaluate developed methods.

1.2.1 Veddige, province of Halland 2008

The anthrax outbreak in 2008 occurred in a cattle herd with about 90 animals. The first seven dead animals had shown no clinical signs of anthrax except for fever. Eventually, at necropsy of three new dead animals, anthrax was suspected and the diagnosis could be made by cultivation and real-time polymerase chain reaction (PCR) (Lewerin *et al.*, 2010). This outbreak occurred during winter-time and the animals were stabled and given feed in the form of roughage, which suggested that the feed was the carrier of the spores. An investigation was started where feed- and environment samples were collected in and around the barn and also at locations where the feed had been harvested. The samples were analyzed with methods that were not validated for finding a low number of spores in large amounts of matrix. No samples gave positive results and the source of the outbreak has not been determined. The theory that the feed was the source is supported by the extreme drought the previous summer that had enabled harvesting of feed at a previously unavailable, flooded, area. The area laid next to the bank of a big river that people, allegedly, dumped dead animals into during the early 1900's (Lewerin *et al.*, 2010). The economic consequences of the outbreak were substantial as 60 million SEK was put into the sanitation of the farm. One report states that the high costs might have been due to an overestimation of the risks involved (Knutsson *et al.*, 2012).

1.2.2 Kvismaren nature reserve 2011 and 2013

In the summer of 2011 an outbreak in cattle occurred in a nature reserve called Kvismaren not far from the city of Örebro. Poisoning with cowbane (*Cicuta virosa*) was first suspected when grazing cattle abruptly died on a large pasture. However, when anthrax was confirmed the remaining cattle were moved to another area and treated with penicillin. As more animals died, a second treatment with antibiotics was started using procaine penicillin. Despite this, some fetuses were aborted. A total of 24 animals died and three fetuses were aborted. *Bacillus anthracis* isolated from animals and fetuses that died despite treatment were all confirmed resistant to penicillin by minimum inhibitory concentration (MIC)-tests.

The source of the outbreak was suspected to be an anthrax epizootic grave from 1943 (Elvander, 2011). Ditch-clearing had been performed in the spring 2011, which was suspected to have released spores from the old anthrax grave. A complication during the outbreak was that the spores were spread via the large ditches and the large Kvismare canal to a nearby pasture where yet another animal died. Soil-, sediment- and water samples were analyzed using PCR-analysis, which indicated where contamination was present and action was taken to prevent further disease from this source. All remaining animals were vaccinated and were allowed to return to the outbreak area the following year.

Surprisingly, in 2013, a heifer died of anthrax on a farm not far from the 2011-outbreak. How the anthrax spores got there could not be explained as water flows in the opposite direction and the animals had no other interaction with the area contaminated in 2011. Wildlife seems to be the only reasonable explanation for the spread to this new area. All remaining animals were vaccinated. The 2013-isolate was whole-genome sequenced and found to be nearly identical with the 2011-isolates as it showed only 3 unique chromosomal base pair mutations (unpublished).

1.3 Diagnostic methods

There are many methods published on detection and identification of *B. anthracis*, but in this thesis only the methods commonly used in diagnostic labs in Sweden will be discussed. They comprise the classical bacteriology, e.g., cultivation and microscopy coupled with DNA-extraction and real-time PCR. Genotyping methods are also discussed as they are important for inferring relationships between strains.

1.3.1 Bacteriology

Bacillus anthracis is readily isolated from, for instance, a fresh blood sample as it grows very well on blood agar. The colonies are large, white, non-hemolytic and irregular. Colonies on a horse-blood agar plate are shown in Figure 3.

Figure 3. Colonies of *B. anthracis* strain SVA11, from the outbreak of 2011, on a blood-agar plate. The insert shows a single colony of the same strain.

The Gram-positive rods can, under the microscope, be seen as long chains of cells. This is more often seen from *in vitro* cultivated material as the chains are often not as long from fresh clinical material (WHO, 2008). If cultivated under elevated CO_2-levels and on bicarbonate-serum agar the bacterium produces its capsule substance (Meynell & Meynell, 1964). The capsule can be stained using McFadyean's polychrome methylene blue staining and viewed under microscope (Spencer, 2003) but the change in colony morphology alone

may also be indicative enough. Separating *B. anthracis* from other closely related *Bacillus* spp. is not always trivial but there are some classical distinguishing traits that when put together should differentiate it from the other species. They include i) colony morphology ii) no hemolysis and no motility iii) sensitivity to the 'gamma' phage and penicillin iv) capsule formation (WHO, 2008). However, in unusual cases, a close-neighbor strain might be indistinguishable from *B. anthracis* using phenotypical traits (Marston *et al.*, 2006) and a combination of traditional bacteriology and molecular methods is therefore preferably used.

Isolation of an outbreak strain is of highest importance as a pure isolate is required for whole-genome sequencing, antimicrobial susceptibility testing and other follow-up analyses. Isolating the outbreak strain is preferably done from fresh animal blood samples. However, in order to connect a potential source to a clinical case, isolation of the strain from the potential source is required. Cultivating the bacteria from soil- or environmental samples can be difficult due to the background flora in them. In these cases, cultivation on semi-selective medium and/or treatments with heat or ethanol should be attempted. Since spores are resilient to heat and ethanol, the background flora can be decreased using either one of the treatments before plating the sample (Dragon & Rennie, 2001). This would enable growth from samples containing anthrax spores and various vegetative bacteria since theoretically only the spores would survive. However, soil samples will most likely contain other spore-forming *Bacillus*-species, which also can survive the pre-treatments.

The polymyxin, lysozyme, ethylenediaminetetraacetic acid, thallous acetate (PLET) agar has since its development in 1966 (Knisely, 1966) been of great use. It is regarded to be semi-selective as some other *Bacillus* spp. have been shown to grow on PLET (Marston *et al.*, 2008; Dragon & Rennie, 2001). Dragon et al. saw indications that not all of the *B. anthracis* spores added to the PLET-agar plate germinated which can be detrimental to recovery from samples with a low number of spores (Dragon & Rennie, 2001). The PLET-agar has recently been modified by Luna et al. by adding trimethoprim and sulfamethoxazole to further enhance its selectivity for *B. anthracis* (Luna *et al.*, 2009). An isolation method called Ground Anthrax Bacillus Refined Isolation (GABRI) has shown good results as it effectively reduced the number of non-anthrax bacteria in the sample (Fasanella *et al.*, 2013a). The method employs washing of relatively large soil samples followed by centrifugation to pellet sample debris. The supernatant is then heat-treated, subjected to fosfomycin and then plated on a semi-selective medium similar to PLET. GABRI was also tested, with success, on naturally contaminated soil samples from an outbreak area in Bangladesh Fasanella (Fasanella *et al.*, 2013a).

1.3.2 DNA-extraction and PCR

Molecular detection of *B. anthracis* is advantageous as the bacteria do not need to be viable, in contrast to the cultivation analyses. In some cases this strength is also a weakness as no information is given on viability of the bacteria in the sample. Cultivation and molecular methods should be used as complements to each other as no method answers all the questions. As an example, an animal that has been dead for some time usually does not contain viable bacteria so PCR is needed to confirm anthrax. However, the spores around a carcass are probably easier to find using selective cultivation media.

The difficult part of molecular detection of *B. anthracis* is that the bacteria form endospores, which are very difficult to extract DNA from. Spores are resilient to many lysis buffers and chemicals as well as to heat and mechanical force (Nicholson *et al.*, 2000). Spores that have germinated into its vegetative state are easier to lyse using lysis buffers but the germination step would of course increase the analysis time. Germination of spores occurs when nutrients (e.g., broth) are added to the sample and the spores come in contact with certain molecules (e.g., L-alanine and Inosine) that trigger germination (Fisher & Hanna, 2005). Enrichment is one way of increasing both the chance of extracting DNA from the bacterium (since it is no longer a spore) and also the number of bacteria to extract DNA from (since the bacteria have multiplied). Enrichment overnight in PLET-broth significantly improved DNA extraction and detection from *B. anthracis* spores in soil (Gulledge *et al.*, 2010). Enrichment for 3-4 hours in brain-heart infusion broth has showed good results in subsequent DNA-extraction and PCR with detection levels down to 10 viable spores/0.25 g of soil (Ågren, unpublished).

A direct DNA extraction from a sample can be attempted when analysis time is of utmost importance. Wielinga et al. spiked feed and food matrices with *B. anthracis* spores and used a kit that contains guanidine thiocyanate as lysating substance, which had earlier shown good results (Wielinga *et al.*, 2011a; Dauphin *et al.*, 2009). The study evaluated the viability of two different *Bacillus*-strains after lysis buffer treatment with different concentrations of lysis buffer, and it showed that a fraction (15-30 %) of the spores survived lysis treatment. The study also took into account the adaptations and optimizations that were needed to use a kit or protocol inside a BSL-3 laboratory. This is often overlooked or not mentioned in reports on DNA-extraction. Another finding was that the different matrices tested produced different levels of DNA-recoveries and subsequent PCR-inhibition (Wielinga *et al.*, 2011a).

There are several commercially available DNA-extraction kits and they typically contain some kind of lysis buffer that, sometimes in conjunction with mechanical force, lyses the cells. Silica columns, magnetic beads or DNA-

precipitation are often used to purify the DNA from the remaining sample. A recent comparison of nine commercial kits showed that the choice of kit is crucial as there were large differences in DNA-extraction efficiencies (Thomas *et al.*, 2013). For automation and higher sample throughput the use of DNA extraction robots has become the standard way of extracting DNA for many pathogens. The Qiagen EZ1 (Qiagen, Hilden, Germany) for instance has been extensively used at SVA for DNA-extraction of *B. anthracis* from blood and tissue samples. However, soil, feed and environmental samples are often analyzed for anthrax outbreak investigations and then DNA could be extracted using an automated solution but there would be a need for some kind of pre-treatment. Straws of hay, soil particles, small rocks etc. are components in these samples that cannot go into the automated DNA extraction robots for obvious reasons. Manual kits and/or several pre-treatment steps are needed for such samples due to their complexity. These sample matrices can give rise to problems for the extraction kits and can contain inhibitors that also inhibit the PCR analysis of the extracted DNA. The inhibition comes mostly from organic substances that are co-extracted with the DNA such as humic substances from soil that are strong PCR inhibitors (Tebbe & Vahjen, 1993). The method used for DNA-extraction must both be efficient at extracting DNA but also at purifying the DNA so that inhibitor substances are washed away.

Once the DNA has been extracted from the sample, the PCR is the most commonly used DNA-detection method. A problem in *B. anthracis* detection by PCR is that the close-neighbor strains (described under 1.1.1) can contain DNA-sequences that are identical, or almost identical. This can enable amplification of DNA although no *B. anthracis* bacteria were present in the sample. A thorough primer design and PCR optimization is therefore of greatest importance.

1.3.3 Molecular genotyping

Molecular strain characterization, so called genotyping, is used to determine how a strain is genetically related to other strains. *Bacillus anthracis* is one of the most clonal species known, meaning there is extremely little genetic heterogeneity between strains found all over the world (Van Ert *et al.*, 2007). This clonality makes molecular typing difficult and high-resolution technologies are needed to separate the strains from each other (Keim *et al.*, 2004).

A technique that has been widely used since its introduction in 2000 is the multiple locus variable-number tandem repeat analysis (MLVA) for *B. anthracis*, which basically measures the amount of repeats in eight repetitive regions in the genome (Keim *et al.*, 2000). This method has been extended in

steps from the 8 loci looked at in 2000 to 31 loci in 2012 (Beyer *et al.*, 2012). A variant of a variable-number tandem repeat (VNTR) is the single nucleotide repeats (SNR) which possesses an extreme variability thus being suitable for studies looking at a short time frame (Derzelle & Thierry, 2013). VNTR and SNR loci diverge faster than the rest of the genome due to their proneness to accumulate errors during replication (Keim *et al.*, 2004), but they are also subject to reversals and homoplasy, which sometimes limit their phylogenetic values (Derzelle & Thierry, 2013). For elucidating deeper phylogenetic relationships, single nucleotide polymorphisms (SNPs) are very stable over time (Keim *et al.*, 2004). SNPs are single base mutations that can occur at most base positions in the genome and they are not prone to homoplasy, which makes their phylogenetic value high (Pearson, 2004). Van Ert et al. developed a genotyping scheme based on a set of so-called canonical SNPs (canSNPs) using more than 1000 isolates (Van Ert *et al.*, 2007). These twelve canonical SNPs represent the big evolutionary lineages for *B. anthracis* and can replace a genome-wide SNP analysis when placing a new isolate phylogenetically. If strains of different canSNP-types cause outbreaks the canSNPs method could separate the outbreak strains. However, the canSNPs are evolutionary markers and they might not be useful for epidemiological analyses during a single outbreak (not even if the outbreak occur over an extended period of time). In these cases, either MLVA or a whole genome sequencing approach with genome-wide SNP-analysis (Kuroda *et al.*, 2010) could be performed to find the few mutations acquired during this short time frame.

It should also be noted that since anthrax is not commonly transmitted animal-to-animal, epidemiological studies on anthrax outbreaks are more focused towards finding a source for the outbreak or a point of introduction. Determining a genotype can say something about the origin of an outbreak strain (Grunow *et al.*, 2013; Price *et al.*, 2012), whether it may come from a laboratory (Hoffmaster *et al.*, 2002) or if it is commonly found in an area of previous natural outbreaks (Fasanella *et al.*, 2005; Fouet *et al.*, 2002).

1.4 Genomics and next-generation sequencing

The genome of an organism refers to the total amount of DNA within that organism. In genetics, one or several genes are studied whereas in genomics the whole genome is studied. This includes comparative genomics where genomes from several organisms are compared to each other. Technologies used in genomic studies include whole-genome sequencing (WGS) and assembly of the DNA-sequence to facilitate further genomic analysis. To have the genome-sequence of a pathogen can answer many questions e.g., why a

strain has become resistant to an antimicrobial substance (Ågren *et al.*, 2014), why a strain is extremely pathogenic and causes severe complications (Mellmann *et al.*, 2011) or how to find biomarkers for a new emerging pathogen (Hoffmann *et al.*, 2012). Genomic studies have been the basis for the results of this thesis and some background to that field is presented here.

When sequencing DNA, a sequencing technology is used to 'read' the sequence of A, C, T and G of a small part of the DNA, usually some hundred base pairs (bp). The most commonly used method has been the chain-terminating approach developed by Sanger et al. in 1977 (Sanger *et al.*, 1977). This technology has been further developed and has been used to sequence whole genomes of organisms. The sequence gathered using Sanger sequencing can be up to 900 bp long (Liu *et al.*, 2012). The first bacterial genome to be fully sequenced was that of *Haemophilus influenza* in 1995 (Fleischmann *et al.*, 1995). The first complete genome sequence of *B. anthracis* was published in 2003 by Read et al. and it was sequenced using Sanger sequencing (Read *et al.*, 2003). This was both costly and time-consuming as the sequencing work took 18 months and the genome finishing (see below) work took over 2 years (Bergman, 2011).

The advent of different next-generation sequencing (NGS) technologies has revolutionized the way DNA is sequenced. Their common denominator is that they sequence numerous randomly located DNA-sequences at the same time (i.e., parallel sequencing), up to billions of sequences simultaneously (Liu *et al.*, 2012). Each sequence gathered is called a 'read' and is usually some hundred bp long for NGS (Quail *et al.*, 2012) although the NGS technology by Pacific Biosciences has produced average read lengths of over 4,000 bp and maximum read lengths of 27 kb (Koren *et al.*, 2013).

The *B. anthracis* genome is 5.4 million bp (Read *et al.*, 2003), which means that 6,000 of 900-bp Sanger-reads are needed to sequence the whole genome to a genome-coverage of 1x. However, in a shotgun approach with random sequence start points, to create several overlapping reads with several different starting points, a high coverage of sequencing reads is needed. Overlapping reads, that are covering a longer stretch of DNA, can be put together to create long, so called, contigs of DNA sequence. Putting short DNA-reads together to form longer contigs is called assembling. If there are no reads covering an area, or if there is uncertainty as to where the reads should be (usually because of repetitiveness), the contig cannot be elongated further and a gap is formed. If all of these gaps have been solved and genome finishing work has ensured that the single contig left is the correct DNA-sequence, it is called a 'complete genome' sequence. If the sequence has some gaps left or there are possible misassemblies or poorly covered areas, it is called a 'draft genome'. An

illustration of how sequence reads are assembled into longer sequences is shown in Figure 4.

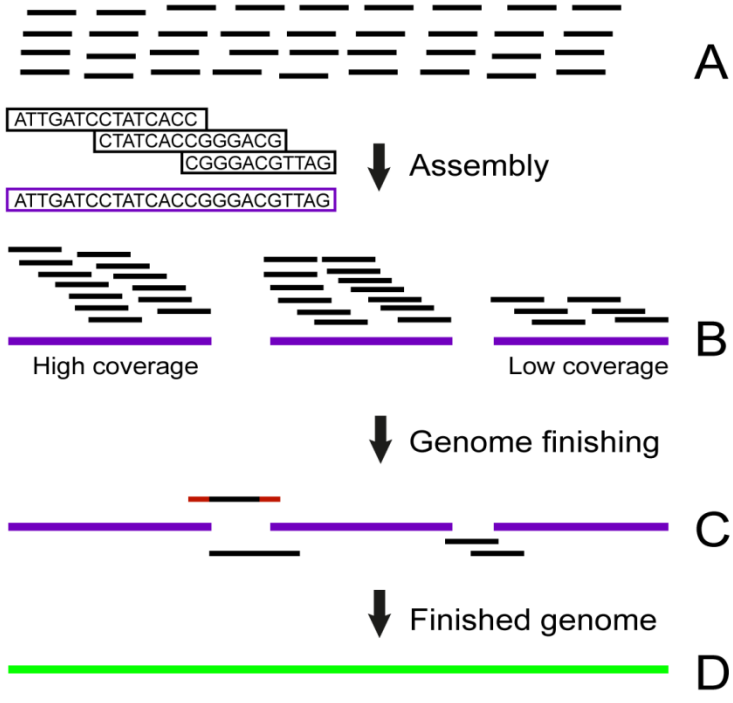

Figure 4. Overview of the sequence assembly, from sequence reads to the finished genome. A) DNA sequence reads from the sequencing machine. B) Assembled contigs. The assembly of reads is done by assembly software that tries to elongate the consensus sequence as long as possible. See inserted example. A high coverage generally gives a better assembly with higher accuracy. C) Genome finishing work requires hands-on work as the contigs must be elongated manually as far as possible to get them to overlap. The use of extra Sanger-sequencing might be needed. Paired-end read and/or mate-pair read information (red-and-black read) can be used to determine order of contigs. D) The finished genome sequence that should accurately depict the original genome sequence.

NGS has opened up new areas of research and also the speed at which researchers can obtain genomic DNA sequence data. Bacteria and viruses can be sequenced during the disease outbreak they are causing (Baize *et al.*, 2014; Köser *et al.*, 2012) and investigations of a pathogen's origin or transmission history can be performed with high resolution (Harris *et al.*, 2013). Publically available sequence databases contain these growing numbers of DNA-sequences and this creates valuable reference data. Good reference databases with sequences together with high-quality metadata (e.g., epidemiological data

such as geographical location of a confirmed disease outbreak together with information about time of isolation of outbreak strain) are essential for investigations on pathogen outbreaks as the pathogen can then be related to previous outbreaks or other sequenced pathogens (Sjödin *et al.*, 2013).

In 2011, NGS machines were released from Life Technologies and Illumina (IonTorrent PGM and the MiSeq) that had lower output than the previously available machines but came at a lower cost and had faster turnaround times (Quail *et al.*, 2012). Where the larger machines (e.g., Illumina HiSeq 2000) are aimed at large scale sequencing producing facilities and at large eukaryote genome sequencing projects, these smaller machines are more suited for rapid local sequencing of small bacterial genomes. As they are fast and do not require a specialized core-facility they are suitable for bacterial or viral outbreak investigations (Quail *et al.*, 2012).

2 Aims of the thesis

The overall aim of this project was to facilitate the handling of anthrax outbreaks by genomic studies and improved diagnostics. More specifically, the aims were:

- To investigate whether genetically close *Bacillus*-species could mimic *B. anthracis*-spores phenotypically and thus be used as a model system for anthrax spores. The model system could be used in developing diagnostic methods and be a simulant in realistic exercises.

- To develop methodology for comparing several hundreds of bacterial genomes to each other on a normal workstation computer. If this would be possible the genomic characterization of an outbreak pathogen's genome would give genomic support to real-time epidemiological investigations.

- To evaluate the capacity of the developed methodology (bullet point above) to predict PCR-cross reactions *in silico*.

- To investigate how the antibiotic resistance arose during an anthrax outbreak in 2011 and also if this has implications for the current treatment procedures.

- To investigate what support whole-genome sequencing could bring to an epidemiological investigation.

3 Comments on materials and methods

3.1 Spore hydrophobicity

An important phenotypic trait that is important for extracting a spore from a sample matrix is the hydrophobicity of the spore. It determines which buffers should be chosen and how the spore will behave in different environments. A commonly used technique to determine the hydrophobicity is the microbial adhesion to hydrocarbon (MATH) test, which was developed in 1980 (Rosenberg *et al.*, 1980). Although slight modifications of the method have been reported, it is based on the separation of cells or spores between an aqueous phase and a hydrocarbon phase. Briefly, the MATH-assay is commonly performed in glass vials where the spores in the aqueous phase are mixed or vortexed with the hydrocarbon phase. After a settling time the absorbance of the aqueous phase is measured and compared with its original absorbance value.

Although the MATH-assay was originally supposed to be included in Paper **I** there are no hydrophobicity results reported and there are several reasons for this. Inside the BSL-3 facility glass tubes or vials were not allowed, due to risk of breakage. Several different plastic tubes were evaluated but they all produced ambiguous results with little or no reproducibility. The spores seemed to adhere to the surface of the plastic tubes, which was somewhat remedied by using Protein LoBind tubes (Eppendorf) but reproducible results were still not achievable.

Other studies have reported different setbacks of the method. One research group found that microdroplets of hexadecane were formed that interfered with the absorbance measuring (Zoueki *et al.*, 2010). This is in accordance with a paper by Leishman et al. where the recovery of the spores from the hexadecane phase was not correlated to the change in absorbance. In that paper it was also stated that heating of the spores, the amount of hexadecane, the mixing times,

the settling times and also strain-dependant sedimentation rates could interfere with the final MATH-results (Leishman *et al.*, 2010).

3.2 Whole genome sequencing

For the initial assembly of the 2011-outbreak strain, the isolate from the first sample sent to SVA was chosen. DNA was prepared from an overnight culture using the MasterPure Gram positive kit (Epicentre Biotechnologies, Madison, WI, USA) and filtered through an Ultrafree-MC 0.22 µm sterile filter (Merck Millipore, Billerica, MA, USA). Sequencing on both the Roche 454 GS FLX+ and the Illumina HiSeq 2000 was performed by Science for Life Laboratory (Stockholm, Sweden). The 454-run used chemistry that enabled long read-lengths and it produced an average genome coverage of 21x. The reason for using both Roche and Illumina techniques was to achieve long 454-reads for the initial assembly and using the shorter, more accurate, Illumina-reads for assembly-correction. The 454-technology suffers from poor sequencing accuracies across homopolymeric regions, which the Illumina-technology does not (Luo *et al.*, 2012) making them a good complement. The Illumina-reads yielded a coverage of 112x and were used to correct several hundreds of mostly minor assembly errors. A few weakly resolved areas still remained and these were solved by Sanger sequencing. The remaining genomes were sequenced using an Illumina MiSeq with 2x250 bp paired-end reads and the reads were mapped to the finished outbreak genome. Differences were located using Consed/cross-match and MUMmer.

For the RNA-sequencing the MiSeq was used but the input RNA was first depleted of ribosomal RNA before prepared for sequencing. The ribosomal RNA is the most abundant RNA and would make up the majority of the input material and the output reads if not removed. The RNA was sequenced using short 1x50 bp settings since the transcriptome was not to be determined but rather quantified. Reads were mapped to the coding regions of the genome using Bowtie2 with the most sensitive settings. The output of aligned reads was analyzed using the software Cufflinks and Cuffdiff 2. Since longer genes would get more “hits” when mapping the reads to the transcriptome, they are normalized and the transcript abundances are given as fragments per kilobase of exon per million fragments mapped (FPKM). Cuffdiff 2 tests each transcript for significant difference of abundance between two samples.

The outbreak-isolate of 2013 was sequenced using the fastest possible methods to provide real-time support to the investigation. DNA from two colonies was extracted using the EZ1 Advanced (Qiagen), sterilized using Millipore Ultrafree-MC 0.22 µm sterile filter and taken from the BSL-3

laboratory. After concentration measurement, the DNA was prepared using the Nextera XT-kit (Illumina) and sequenced on a MiSeq using 1x50 bp settings. For quality assurance, the FastQC software (Andrews, 2010) analyzed the reads and, due to the short read length, the majority had quality values close to Q40. By aligning the reads using Bowtie2, the fraction of reads not aligned was determined to be less than 0.3 %. This showed that 99.7 % of the reads aligned perfectly to the outbreak genome of 2011. By mapping the reads to the 2011-outbreak-genome using Consed/cross-match, the polymorphisms could be determined.

3.3 Creation of the phylogenomic overviews in Figure 1 and Paper III

To illustrate the genome similarities of the *B. cereus* group a few strains have been included in a phylogenomic overview created in Gegenees (Paper **II**), both in this thesis with fewer strains and as a table in Paper **III**. The genomes were compared using the accurate-setting of Gegenees and the phylogeny is based on score values with a threshold of 20 %. This means that the average BLASTN-scores of the core genomes are used, as genetic material not included in the core genome is sorted out with the threshold.

In Paper **III**, we considered genomes of 80 % average core genome similarity to *B. anthracis* to be called near neighbors, as to our knowledge, no real definition of the close neighbor-concept exist. This might seem arbitrary but genomes of strains that have been called near neighbor or close-neighbor in the literature (although they are not all included in either of the illustrations) all pass this criterion. Strains with 80 % similarity or above are also part of one of the three main phylogenetic clusters of the *B. cereus* group population, along with *B. anthracis* (Kolsto *et al.*, 2009). Strains with less than 80 % similarity in such a phylogenomic overview created in Gegenees are part of the other two clusters. In Figure 1 strains number 6 and 7 represent strains that are relatively closely related to *B. anthracis* but not as much as strains number 3, 4 and 5 and might not be traditionally viewed as close neighbors.

3.4 Growth rate experiment

A growth rate test was performed on one resistant isolate and one previously resistant that had acquired a counter-acting mutation. Five 1.5 ml micro tubes with 1.0 ml of Luria-Bertani medium were inoculated for each of the two variants of *B. anthracis*. The tubes were incubated at 37°C in a Thermomixer (Eppendorf) with a mixing frequency of 400 rpm. During five hours of

incubation, at every 20 minutes, the growth was measured by placing part of the medium in plastic cuvettes in a spectrophotometer and then transferred back to the micro tube. The growth rates were estimated by fitting exponential growth curves to the experimental data using non linear regression but no significant differences in growth rate were seen.

Since these isolates with counter-acting mutations were clearly enriched somehow in our standard laboratory procedures (Paper **IV**) we suspected that the growth rate difference must be substantial but this growth test showed it was not. However, this study should be complemented with a more high-resolution method such as controlled head-to-head competition (Elena & Lenski, 2003) to conclusively determine the fitness of the two variants of the strain.

3.5 DNA-extraction from environmental samples

During the outbreak investigation of 2011, over 200 soil- and sediment samples were processed and a DNA-extraction method was developed that produced robust extractions with little or no subsequent PCR-inhibition. A problem with direct DNA-extraction from such samples is that the spores are hard to lyse and that they are usually found in very low numbers (Dragon *et al.*, 2005; Manchee *et al.*, 1994). Therefore a short enrichment step was evaluated as this would germinate the spores and increase their number slightly.

Soils and sediments vary in physical characteristics and in composition which can affect the growth of certain bacteria (Ellis *et al.*, 2003). This means that positive control spores should be added to each sample as there is no other way of determining the success of the enrichment and the subsequent DNA-extraction. We solved this by using different model spores evaluated in Paper **I**. A possible problem with the *B. cereus* spores that mimicked *B. anthracis* spores the best, is that there is a slight risk of similar bacteria already being present in the soil sample. This had not been seen during development of the method where a variety of soils were tested, but the risk cannot be neglected. For the 2011-outbreak samples we used spores of *B. thuringiensis* var. *kurstaki* HD-1 that is commonly used as an insecticide. The crystal protein gene *cry1* found in *B. thuringiensis* strains was used as PCR-target. This control strain might be a poor choice if the samples come from an area where *B. thuringiensis* spores are used to decimate the mosquito population, e.g., around the Swedish river Dalälven where var. *israelensis* is used. Guidi et al. stated that *B. thuringiensis* spores could be found in very high numbers in an area that had been annually treated for mosquito control (Guidi *et al.*, 2011). We made

no such findings during the 2011-outbreak and the area has most likely never been treated since it is a nature reserve.

Briefly, 400 µl of brain-heart-infusion medium was added to roughly 0.25-0.5 g of sample in a 1.5 ml micro tube, which was then inoculated with 40 spores of *B. thuringiensis* var. *kurstaki* HD-1. The control spore stock consisted of circa 40 spores per 10 µl and stored in Protein LoBind-tubes (Eppendorf). The concentration was determined by plating once a week and the concentration was stable throughout the outbreak period. The sample was then incubated for four hours at 37°C with shaking in a thermomixer at 500 rpm, before extracted using a DNA-extraction kit.

Evaluation of suitable soil-extraction kits was made and the SoilMaster™ DNA Extraction kit (Epicentre Biotechnologies, Madison, WI, USA) gave high DNA-yields, but together with high levels of PCR-inhibition as not even a hundredfold dilution of the extracted DNA could circumvent the inhibition. Instead, the PowerSoil® DNA Isolation Kit (MO BIO Laboratories, Inc., Carlsbad, CA, USA) was deemed the best solution as it included a bead-beating step that could safely be used in the BSL-3 laboratory, and of the over 200 soil- and sediment samples analyzed with it, only one showed PCR-inhibition and that sample contained feces.

PCR was then performed that detected targets in *B. anthracis* and in the spore control to ensure the enrichment and DNA-extraction had been successful. The method proved very stable as the PCR-results were basically identical for the control spores across all samples in that batch. A separate PCR-reaction was performed on the same extracted DNA in which unrelated DNA was added to monitor PCR-inhibition. Optimally, these two reactions should have been performed in the same tube with an internal inhibition control but it was found that when multiplexing using three or four targets, the sensitivity and efficiency of the *B. anthracis* PCR was lowered.

4 Results and discussion

4.1 A model system for *B. anthracis*-spores (I and II)

Work with *B. anthracis* in Sweden is restricted to special BSL-3 facilities in which work is both time consuming and costly. In many countries, the BSL-3 restriction is also true for attenuated strains, which lack one or both virulence plasmids. Surrogate or model strains of non-virulent species are therefore desired when developing new diagnostic methods or during exercises or trainings.

Model strains for *B. anthracis* have been used frequently in studies in the past, and this has been reviewed by Greenberg et al. (Greenberg *et al.*, 2010). However, the review showed that the majority of the studies had used *B. atrophaeus* (formerly *B. globigii* and *B. subtilis* var. *niger*), which is not closely related to *B. anthracis* (Gibbons *et al.*, 2011; Xu & Cote, 2003). The benefit of using *B. atrophaeus* lies in its complete lack of pathogenicity, which enables large-scale aerosolization experiments and exercises (Edmonds *et al.*, 2009; Stratis-Cullum *et al.*, 2003).

In Paper **I** the assumption that spores of genetically related strains behaved similarly phenotypically was evaluated. A model system for *B. anthracis* spores was to be proposed that was intended to be used as model in detection studies in feed and food and in exercises. However, the model was not intended for aerosolization studies thus not having the non-pathogenicity demand since experiments could be carried out in normal BSL-2 laboratory settings.

For the genetic analysis we used the sporulation stage III AB gene (*spoIIIAB*) to screen over 50 strains of *Bacillus* spp. as described by Ehling-Schultz et al. (Ehling-Schulz *et al.*, 2005). This method utilizes a single gene instead of the seven used for MLST, thus being cheaper to use but it still yielded the same tree topology. A whole genome sequence comparison was also performed for the strains with draft or completed genome sequences

deposited at Genbank, using the ASC method (Segerman *et al.*, 2011). It was found that the clustering of the *spoIIIAB* phylogenetic tree was supported by the whole genome analysis (Paper **I**).

Six *Bacillus* spp. from different clusters in the *spoIIIAB* analysis were chosen and were together with two *B. anthracis* strains sporulated and phenotypically analyzed. It was shown that the two *B. cereus* strains NVH0597-99 and F2085/98 were highly related to *B. anthracis* genotypically and they also showed the most similar phenotypical properties (Paper **I**). They showed similar heat-inactivation patterns, similar germination patterns and they even grew on the modified PLET agar, which is valuable for realistic exercises.

The two model strains proposed are suitable if parameters like germination, growth or heat-deactivation are important in an experiment. Their close genetic relationship with *B. anthracis* suggests they are also suitable for other types of experiments but this would of course have to be evaluated first. When performing studies of heat deactivation or decontamination it might be wise to choose a more conservative, fail-safe model to introduce some margin of error. In our case this would be the *B. thuringiensis* strain, which showed a higher tolerance to heat than *B. anthracis* or the two suggested *B. cereus* models (Paper **I**). It should also be emphasized that the two model strains suggested are isolated from human infections and must only be handled in a BSL-2 laboratory.

A topic not discussed in Paper **I** is the ability to trace the model organisms in a complex experiment with perhaps several other *Bacillus* spp. in the sample matrix, e.g., a spiked soil sample. A solution to this problem has recently been proposed where genetic signatures, also called 'barcodes', were introduced into the chromosome of a *B. thuringiensis* subsp. *kurstaki* strain (Buckley *et al.*, 2012). The inserted barcodes then worked as targets for specific PCRs and the model organism could be traced in a mix of near-neighboring strains. The two strains we proposed as models have been whole genome sequenced; the NVH0597-99 is available from Genbank (NCBI Acc. no. ABDK00000000) and the F2085/98 genome has been sequenced at SVA to a draft state but not yet made publically available. Our solution to the tracing issue was to use the Gegenees software (Paper **II**) to compare the models' genomes to all *Bacillus*-genomes publically available, thus locating possibly unique signature sequences. Real-time PCRs were then designed to these regions, evaluated *in vitro*, and then used to trace the model in complex samples in different experiments (unpublished).

Strains identified in Paper **I** have been successfully used during the Swedish anthrax outbreak of 2011 as controls for the DNA-extractions and detection by

PCR. This ensured a reliable interpretation of results from environmental samples. They have also been used in exercises and in several method developments.

4.2 Characterization of a bacterial genome during an outbreak (II)

When a bacterial disease outbreak occurs, the outbreak strain is often genotyped using MLST, MLVA or other genetic methods. This can provide information about the origin of the strain and if the pathogenic strain, or at least if a related strain, has been isolated and genotyped in the past. With the improvement of NGS technology and the large number of sequenced strains in reference databases, it is becoming more and more common to whole-genome sequence a pathogen and compare it to previously sequenced strains. However, the technological advancements in NGS have not been matched by analysis software that can handle the large amount of genomes.

For this purpose, the software called Gegenees was created (Paper **II**). It can compare several hundred bacterial genomes on a normal workstation which, together with the graphical user interface, enables whole genome comparisons for a bigger community of researchers and clinicians. Whole genome alignment in the traditional sense is very computationally intensive and time-consuming. In Paper **II** it was shown that comparing 30 *Bacillus*-genomes in the software progressive Mauve (Darling *et al.*, 2010) took almost 70 hours and that no more than 30 genomes could be aligned due to memory error issues. Gegenees is highly scalable, which means that the amount of computer memory used stays at the same low level regardless of the number of genomes aligned. This is due to the fragmented approach where the genomes are fragmented into smaller pieces (usually 100-500 bp), which are then sequentially aligned to the whole genomes using Basic Local Alignment Search Tool (BLAST) (Camacho *et al.*, 2009; Altschul *et al.*, 1990).

Although Gegenees uses alignments, it does not produce a traditional alignment file but instead produces conservation patterns. The default setting is BLASTN which aligns nucleotide sequences but the translated TBLASTX can also be used if genomes are more distantly related to each other. Aligning amino acids to each other can yield valid alignments even though the DNA-sequence lacks similarity due to divergence. This has been used in a recent publication (Rokicki *et al.*, 2014) where a developed software (CodaChrome) was used to visualize proteome conservation across thousands of genomes. However, that software cannot work with draft genomes and that is crucial during an outbreak as there will be no time to fully complete the DNA-

sequence of the pathogen and the majority of whole-genome sequences available are in a draft state (Koren *et al.*, 2013; Segerman, 2012).

During the German enterohemorrhagic *Escherichia coli* O104:H4 outbreak in 2011, Mellmann et al. (Mellmann *et al.*, 2011) and Rasko et al. (Rasko *et al.*, 2011) used NGS to characterize the outbreak strain. Both studies found that it was most related to the enteroaggregative strain 55989 and that the strain contained an interesting set of plasmids and virulence genes. In Paper **II** it was demonstrated that all of these findings could be made in Gegenees by simply comparing the strain to all previously sequenced *E. coli* strains. This exemplifies how Gegenees can be used for conservation analysis.

To investigate which genomic areas that were conserved and also unique to *B. anthracis*, all whole-genome sequenced *Bacillus* spp. genomes were compared in Gegenees. The four prophages and three other shorter sequences were highlighted as unique for *B. anthracis* (Paper **II**). Gegenees showed that the prophage lambdaBa03 was basically fully conserved and unique to *B. anthracis*. This supported the PCR-design work published by Wielinga et al. in which the PCR was placed in this prophage (Wielinga *et al.*, 2011b). Having a rapid method of assessing an outbreak pathogen is crucial for developing new diagnostics, for source attribution and for answering questions about peculiar traits (e.g., regarding antimicrobial resistances or virulence factors). For anthrax outbreaks, software like Gegenees can quickly indicate whether the outbreak strain has any peculiar genomic traits that are missing in other sequenced strains, or vice versa, and this brings further support into the investigation.

Another useful application of Gegenees is the creation of collections of biomarkers instead of genomes that can contain any information the user supplies. For instance a set of plasmids usually found in enteric bacteria can be aligned together with the newly sequenced genome thus creating a rapid plasmid profiling for an enteric pathogen (Paper **II**). A similar approach was used by Köser et al. where an artificial methicillin-resistant *Staphylococcus aureus* (MRSA) 'resistome' and also a 'toxome' were created that contained resistance genes and toxin genes, respectively. Depending on the gene-content in the outbreak- and non-outbreak strains it created a fingerprint of the strains and relationships could be indicated (Köser *et al.*, 2012). Once the toxome and resistome were created they could be treated as 'genomes' in Gegenees and the same graphical results would be acquired, without more manual analysis work.

Lastly, biomarker discovery is an application that benefits from Gegenees' ability to compare vast numbers of genomes as the specificity of the biomarkers will be assessed using all available genomic material and not just a subset. This feature is useful for diagnostic assay development, for instance

when a newly sequenced pathogen needs to be specifically detected using PCR. This is further discussed in 4.3.

4.3 Evaluating PCR-assays for specificity (II and III)

When developing a novel PCR assay or evaluating an existing one, the analytical specificity is one of the crucial parameters to evaluate. The analytical specificity includes the capability to detect all the targets and at the same time not give false positive reactions for non-targets (Bustin *et al.*, 2009). This is often done by placing the primers and probes in a conserved gene that is not found in any other organisms. However, a gene that we think is unique to the target organism today might not be unique in the near future as more and more organisms are sequenced and it might be shown that the gene is found in a non-target organism. That is why it would be beneficial to test currently used assays for specificity on regular intervals.

In the case of *B. anthracis* detection, the need to have specific PCR-targets might not be that important when analyzing a fresh blood sample because the likelihood of finding a close-neighbor strain that causes anthrax-like disease is very low. But for more complex samples like feed and environmental samples, e.g., soil, the issue with cross-reactivity becomes important. During the outbreaks in Sweden in 2008 and 2011, PCR cross-reactions were observed on several occasions but only for the chromosomal marker as pXO1/pXO2-like plasmids are uncommon (Bergman, 2011). In contrast to this, a sample of organic material of animal origin from India intended as ecological fertilizer was found PCR-positive for the virulence plasmid pXO2 in the spring of 2014 at SVA. This case created a stressful situation as people had been in contact with the material and several confirmatory analyses had to be performed to clear the sample of suspicion (unpublished). The case stresses the importance of having as specific PCR-markers as possible.

The software Gegenees (Paper **II**) is complemented with a short sequence alignment setting. This enables the alignment of a list of primers and probes to a database containing whole genomes. The software can sort the primer alignments according to the hit score against the target genomes or to the outgroup genomes. Its functionality was demonstrated when PCR cross-reactions *in vitro* could be explained using Gegenees (Paper **II**).

In order to fully elucidate how these multiple primer alignments can be used for diagnostic development a large scale *in silico* study of PCRs was initiated. Using Gegenees, all publically available primer/probe systems targeting *B. anthracis* were aligned to all available *Bacillus* spp. genomes on Genbank. This *in silico* analysis showed that there was no plasmid marker sequence that

was specific and that the majority of the published PCR-assays were not specific for the *B. anthracis* chromosome at all (Paper **III**). Since no plasmid-marker was specific for *B. anthracis*, only the chromosomal targets are hereafter discussed. The four assays that showed 100 % specificity were all relatively recently published, which is not surprising as more and more sequence data produced give better understanding of the genomics in the *Bacillus* genus. The close neighbor strains (e.g., *B. cereus* G9241 (NCBI accession NZ_AAEK) and *B. cereus* var. *anthracis* CI (NCBI accession CP001746)) did, as expected, produce the majority of the cross-reactions *in silico*, at least for the assays that were relatively specific for *B. anthracis*.

To further evaluate the assays an inter-laboratory trial was performed on a collection of DNA from 90 *Bacillus* spp. including some strains that are close neighbors to *B. anthracis* and also were part of the *in silico* analysis. To evaluate the *in silico* methodology, PCRs that were not 100 % specific were also included in the laboratory trial. The results from this *in vitro* study were consistent with the *in silico* analysis. The close neighboring strains produced cross reactions *in vitro* but three assays performed adequately, i.e., produced no false positive and no false negative results (Paper **III**).

Interestingly, the three specific PCRs are all located within the prophage named 'LambdaBa03' in the Ames genome (NCBI accession NC_003997). Placing a diagnostic marker within a prophage might be considered a risk because of the instability of prophages (Canchaya *et al.*, 2004). However, the presence of these prophages in 192 isolates from different parts of the world analyzed in 2006 (Sozhamannan *et al.*, 2006) indicates that they are stable. The four prophages were also unique for *B. anthracis* when analyzed in Gegenees (Paper **II)**. Still, as more and more genomes are sequenced, the assumption that the phages are unique to *B. anthracis* might have to be revised in the future.

The chromosomal real-time PCR referred to by the WHO (WHO, 2008) was also included in the laboratory trial and it showed a relatively good specificity. However, it is based on fluorescence resonance energy transfer (FRET) probes and this requires certain features (i.e., decoupled excitation/emission filters) of real-time PCR machines that are not available on all machines. Hence, two of the five participating laboratories could not evaluate the WHO-assay. This was of course not optimal, but it does point out that the commonly used hydrolysis probes (e.g., Taqman) should be preferred since this is a more widely used technology.

To conclude, this study demonstrates a method of evaluating PCR assays. Once a diagnostic PCR is in use, its specificity can be controlled by downloading all newly sequenced relevant genomes to Gegenees on a regular basis, and aligning the primer/probes to them.

4.4 Penicillin resistance during an anthrax outbreak (IV)

Bacillus anthracis has two beta-lactamase genes on the chromosome but they are non-inducible thus making *B. anthracis* susceptible to penicillin (Chen *et al.*, 2003). To understand how several *B. anthracis* isolates became penicillin resistant during the Swedish anthrax outbreak in 2011, six isolates were whole-genome sequenced. They represented both penicillin susceptible and resistant phenotypes and by comparing their genome sequences to each other, relevant mutations could be determined. It was found that the reason for the penicillin resistance were mutations (one per resistant isolate) in the *rsiP* gene which codes for a negative extracytoplasmic sigma factor (Paper **IV**). Different frameshift mutations led to incomplete RsiP proteins. This enabled SigP, which is the positive sigma factor, usually sequestered by RsiP, to induce a constitutive expression of the two beta-lactamase genes (Paper **IV**). This means that the feedback control of this system was destroyed in the penicillin resistant isolates.

One of the isolates showed a hypermutator phenotype which is described in detail under heading 4.5. A hypermutator has an elevated mutation rate and this could theoretically lead to faster development of *rsiP*-mutations. It has been shown that hypermutators play a key role in the development of antimicrobial resistance in *Pseudomonas aeruginosa* causing human lung infections (Macia *et al.*, 2005). These hypermutators were found to be crucial for the development of a long-term chronic infection. However, hypermutators are associated with a loss of fitness. An *in vitro* study has shown that the relative fitness of hypermutator populations of *P. aeruginosa* was lowered as the number of generations increased and that the cause of this was only a few, deleterious, mutations of the many mutations accumulated (Heilbron *et al.*, 2014). The *B. anthracis* hypermutator isolate recovered in 2011 might have acquired its hypermutator phenotype before acquiring its resistance mutation thus increasing the rate of mutation and the chance of becoming resistant. However, only one of the resistant isolates was a hypermutator so it was not a prerequisite for resistance development.

The mutation rate was estimated for one of the anthrax outbreak isolates by whole-genome sequencing of the isolate before and after a serial passage study and it was shown to be similar to a previously reported value (Vogler *et al.*, 2002). This implies that the outbreak strain (except for the hypermutator isolate) had the normal mutation rate for *B. anthracis* which, if proven otherwise, could have been a factor in why this outbreak yielded several independent penicillin-resistant isolates.

By re-sequencing isolates some were found to have acquired nonsense and frameshift mutations in the *sigP* gene as well. This did, as expected, abolish the

penicillin resistance in these isolates. For one isolate, it was discovered that roughly 10 % of the frozen bacteria stock carried a "counter-acting" mutation. An amplicon sequencing experiment showed that the original blood sample that was sent to SVA during the outbreak contained less than 1 % of these mutated bacteria as it was not detected and the detection limit was about 1 % (Paper **IV**). This suggests there is some kind of enrichment of bacteria with counter-acting mutations during the isolation procedures used at SVA.

Follow-up studies have been performed to clarify what the selective advantage might be of the counter-acting mutations. A growth rate study in rich medium showed no significant difference between a penicillin-resistant isolate and one with a counter-acting mutation. However, this growth rate study should have been extended with a relative competitive fitness study that has the resolution to confidently say anything about the possibly small changes in fitness. They did not show any difference in survival in -80°C or storage at +4°C in different commonly used buffers and they both sporulated successfully.

A finding made in Paper **IV** was that the SigP induced transcription of a penicillin-binding protein (PBP) along with the beta-lactamases. This PBP is not expressed at all in prototypical *B. anthracis* strains and thus is not needed for normal growth. PBPs in bacteria synthesize cell-wall peptidoglycan and they have a strong affinity to beta-lactams (Brannigan *et al.*, 1990; Chambers & Sachdeva, 1990). For MRSA it has been shown that resistance to beta-lactams requires that a new form of PBP is acquired that has a low affinity to beta-lactams, thus surviving the penicillin-pressure (Pinho *et al.*, 2001). The same study did however also show that the interaction between the new PBP and the native PBPs is complex. Since there is a strong link between cell wall synthesis and PBP there is a possibility that the "new" PBP affects *B. anthracis* in a way that lowers the fitness of the bacterium. If this is the case, counter-acting mutations might also be found in this gene instead of in the *sigP* gene. Ross et al. also described a *B. anthracis* isolate that had an amino acid change in the sigP compared to all other sequenced *B. anthracis* strains (Ross *et al.*, 2009). This change presumably led to a reduced activity and the authors suggest that full activity of the sigma factor is detrimental to bacterial growth. It is tempting to assume that their isolate had acquired a missense mutation and the Swedish isolates acquired nonsense/frameshift mutations because the bacterium somehow benefits from having the penicillin resistance tuned down or completely shut off. This theory is slightly substantiated by the fact that all known sequenced *B. anthracis* strains have an identical DNA-sequence over the sigma factors thus presumably also have no expression of the beta-lactamases and the PBP.

The fact that penicillin resistance in *B. anthracis* only requires one random detrimental mutation in the *rsiP* gene means that penicillin resistance can occur at any time. However, it also means that a high number of bacteria are needed in order to have one bacterium with that kind of mutation. The fact that the bacterium can revert to a sensitive phenotype once the beta-lactam pressure is gone, can have implications when determining the resistance *in vitro*.

4.5 Heterogeneity among *B. anthracis* isolates during an outbreak (IV)

Bacillus anthracis isolates from the 2011 outbreak in Sweden were whole-genome sequenced to evaluate the discriminatory power needed to distinguish between isolates from the same outbreak. The outbreak strain's genome was completed and by looking at canonical SNPs the strain belonged to the B.Br.001/002 lineage. Repeat regions that could have been used for MLVA or SNR analysis did not differ between isolates for such a short time frame of evolution. A genome-wide single nucleotide variant (SNV) analysis was performed for several isolates to determine what a "normal" genomic divergence during an outbreak was. Penicillin susceptible isolates had acquired between 1-2 unique SNVs each from the theoretical progenitor sequence (Paper **IV**).

The *in vitro* mutation rate (discussed in 4.4) indicated a typical mutation rate for *B. anthracis* but when we estimated the *in vivo* mutation rate needed to yield 1-2 SNVs per isolate, it was 10-20 times higher (Paper **IV**). This was of course only a rough estimate of the true *in vivo* rate but it was indicative enough to lead us to the theory that mutations are inherited from the source. The source in this case was thought to be an anthrax epizootic grave from 1943 and environmental isolates from the presumed source had also been sequenced in an attempt to connect the source isolates to the animal isolates. As these were also heterogeneous the inheritance-theory was supported. This means that mutations we saw 2011 might have been acquired during the 1943-outbreak or they may even have been inherited into the animals that died in 1943. Founder effects via the infectious dose can give rise to new a population which only inherits certain mutations from the original population.

This led us to believe that these different SNV-populations would be distinguishable within an animal and this was evaluated with an amplicon sequencing experiment. Briefly, PCRs were designed over genomic areas containing the mutations and they were amplified from the different DNA-extractions from the original blood samples sent to SVA. The amplicons were then sequenced and the number of reads supporting each mutation was

determined. This showed that the relative occurrence of the SNVs were between fully clonal (i.e., ~100 %) to subclonal (< 100 %) to below the detection limit, which was around 1 %. The detection limit was set because of the chance of finding a random mutation due to sequencing errors, which was below 1 %, although often much lower. Intra-animal spatial variation was also seen, which means that different clones are more or less likely to be isolated, depending on where the sample is extracted from the animal.

The occurrence of a hypermutating isolate (Paper **IV**) is notable as this is uncommon and might interfere with the interpretation of a SNV-based analysis. Köser et al. found a hypermutator during an investigation of a putative outbreak of MRSA and the cause was a nonsense mutation in the isolate's *mutS* gene (Köser *et al.*, 2012). MutS is part of the DNA proofreading mismatch repair pathway and the truncation of this protein led to an elevated mutation rate. A similar cause was identified for the *B. anthracis* isolate as the *mutL* gene product was truncated (Paper **IV**) and MutL is a crucial component in the mismatch repair system for *Bacillus* spp. (Ginetti *et al.*, 1996). Hypermutators might be discarded as not being part of the outbreak if only a set SNP cutoff value is used (Köser *et al.*, 2012) and this further highlights the need to sequence several isolates and make a judgment based on a whole population of outbreak isolates.

To conclude, a genome-wide SNV analysis was needed to differentiate the isolates obtained from the 2011 outbreak. There are however limitations to be aware of: 1) clonal heterogeneity within an animal that is in the same magnitude as between isolates from different animals, 2) intra-animal spatial heterogeneity, which has implications on sampling procedures, and 3) the fact that we cannot differentiate between inherited mutations and current-outbreak-specific mutations.

5 Conclusions

This thesis demonstrates different uses for genomic studies and NGS to improve diagnostics and facilitate outbreak investigations. The specific conclusions of the thesis work are:

- Spores of *Bacillus* spp. genetically closely related to *B. anthracis* behaved phenotypically similar to *B. anthracis* spores thus being suitable model organisms.
- Fragmented alignment was shown to be an efficient alternative to traditional whole-genome alignment for conservation analysis. The developed software, named Gegenees, can be used in an outbreak situation to rapidly characterize the pathogen's genome.
- Gegenees could be used to evaluate specificity of PCR-assays *in silico* against a large panel of genomes.
- The reason for the *B. anthracis* penicillin resistance development was a defect *rsiP* gene product and counter-acting mutations in the *sigP* gene could reverse a resistant isolate back to a penicillin-susceptible phenotype. The ease of which resistance can occur and also disappear can have implications for resistance determinations. Penicillin should probably stay the primary choice for animal anthrax treatment but the risk of resistance development should be considered.
- Several isolates from each animal from an outbreak should be sequenced to use SNVs for epidemiological investigations as there are risks of sequencing a hypermutator or a sub-clone that poorly represents the population in an animal. Spatial variations in the animals were found as the ratios of different sub-clones were different depending on where in the animal the sample was extracted. It was also found that inherited and acquired SNVs are not easily separated. NGS and genomic studies proved powerful tools for providing outbreak support but good reference data is needed to put new results in context.

6 Perspectives for the future

Genomics in epidemiological investigations – retrospectively and in real-time

Using WGS and genomics as tools for epidemiological investigations has been demonstrated for several pathogens. The *Vibrio cholera* outbreak on Haiti that followed the earthquake in 2010 was connected to Nepalese soldiers sent by the United Nations (Hendriksen *et al.*, 2011). Beres et al. described how three successive epidemics of invasive infections of group A *Streptococcus* consisted of different strains (Beres *et al.*, 2010). Much attention has been awarded the studies by Köser et al. and Harris et al. (Harris *et al.*, 2013; Köser *et al.*, 2012) in their work to explain transmissions that took place during MRSA outbreaks. The Harris et al. study is a retrospective descriptive study that explained the transmission events of the MRSA clone in the hospital. In contrast, Köser et al. attempted to differentiate between outbreak-associated isolates and non-outbreak isolates during the putative outbreak itself. Real-time studies like the latter one can provide information that can improve treatment schemes and outbreak countermeasures.

The Swedish anthrax outbreaks in 2011 and 2013 are examples of a retrospective study and a real-time study, respectively. After sending DNA from the 2011-isolates to a sequencing core-facility, the DNA-reads were delivered to SVA three months later. Genome finishing and comparative studies to fully elucidate on the SNV-heterogeneity and penicillin resistance then took several months. In contrast, the 2013-outbreak isolate was sequenced using an Illumina MiSeq available at SVA in less than 24 hours and the determination that the 2013-isolate was highly similar to the 2011-strain took 30 minutes. The difference in sequencing time between the two cases stem from the availability of a smaller benchtop NGS machine at SVA but the difference in analysis time is due to reference data. If the 2011-study had not been so in-depth, little could have been said about the 2013-isolate. To have

data to compare new data to is of greatest importance and much time-consuming work enabled the new genomic characterization in minutes.

The NGS technology is advancing at an astounding speed. What took years to perform seven years ago now takes hours and the scientific community is still trying to keep up with this rate of progression. In the near future, DNA sequencing will probably be made in the field next to a sick or deceased animal, but these technological gains must be matched by software that can handle the large amount of data and put it to use.

The primary goal for the field of *B. anthracis* research should be to create a reference database of sequencing and/or genotyping results that is shared between countries, starting with the countries in Europe. Secondly, improvements in genomic analysis software are needed as too much is dependent on user experience in command-line interface software that is developed *ad hoc* and has to be modified for each task. If these two goals were met it would enable source indications and genomic characterizations in a fraction of the current time and would strengthen our capacity to respond to future natural or antagonistic outbreaks of *B. anthracis*.

7 Populärvetenskaplig sammanfattning

7.1 Bakgrund

Bacillus anthracis är bakterien som orsakar sjukdomen mjältbrand, också kallad antrax. Sjukdomen är en zoonos, d.v.s. den kan överföras mellan djur och människa, men det är under naturliga förhållanden framförallt djur som drabbas. *B. anthracis* är en av få bakterier som kan omvandla sig själv till en överlevnadskraftig spor. Sporen är nästan helt inaktiv och kan därför ligga i marken och vänta på att infektera ett djur i flera decennier. Det finns till och med rapporter på runt 100 år gamla sporer som väckts till liv. Betande djur är väldigt känsliga för dessa sporer. När ett djur blir infekterat genom att andas in eller äta sporer som finns i fodret eller på marken så är sjukdomsförloppet oftast mycket snabbt och leder ofta till döden. Från det döda djuret kan det rinna okoagulerat blod ur kroppsöppningar och bakterierna i blodet (ofta över 100 miljoner per ml blod) som kommer ut i syrerik atmosfär sporulerar. På grund av att marken, och inte djuret, är den naturliga reservoaren för dessa sporer så kan det vara svårt att få bort sjukdomen från ett område.

Mjältbrand är än idag vanligt i vissa länder men i Europa sker endast sporadiska utbrott t.ex. när gamla mjältbrandsgravar störs, antingen genom grävarbete eller extrema väder. Mjältbrand var vanligt i Sverige fram till mitten av 1950-talet då importen av benmjöl (som användes i fodermedel) stoppades då detta identifierats som källan till utbrotten. Efter 27 mjältbrandsfria år bröt smittan ut 2008 i Veddige, Halland. Det blev en kostsam historia där över 60 MSEK gick åt för att sanera gården m.m. 2011 utbröt sjukdomen bland betande kor i Kvismare naturreservat. Efter antibiotikabehandling så fortsatte ändå djur att dö och från dessa djur isolerades antibiotikaresistenta bakterieisolat. Genom markprovtagning och molekylär diagnostik försökte SVA finna källan till utbrottet. Eftersom *B. anthracis* är väldigt lik andra naturligt förekommande

jordbakterier kan markprover ge falskt positiva signaler då diagnostiska markörer kan korsreagera med dessa närbesläktade bakterier.

Genomik kallas det när hela arvsmassan (s.k. genom) hos en organism eller flera studeras och jämförs. Eftersom hela arvsmassan används måste den först sekvenseras och detta har varit en väldigt kostsam och tidskrävande process men genom nya typer av sekvenseringsteknologier, som först blev tillgängliga 2007, så kan ett helt bakteriellt genom sekvenseras på bara dagar istället för månader eller år som förr var fallet. Genomikstudier har utförts i de fyra delstudierna i denna avhandling för att förbättra diagnostik, för att smittspåra och för att ta reda på hur antibiotikaresistensen uppkom.

7.2 Studier och resultat

7.2.1 Modellsporer för övningar och metodutveckling

Eftersom *B. anthracis* är patogen för människor så krävs hög nivå av skyddsutrustning och så kallade skyddsnivå-3-laboratorium för att få arbeta med bakterien. För att underlätta vid metodutveckling och för att ha ett ofarligt alternativ till *B. anthracis* vid övningar så har sporer av arten *B. globigii*, numera *B. atrophaeus* använts. De är dock inte alls närbesläktade med *B. anthracis* och hypotesen att mer närbesläktade arter skulle bete sig mer som *B. anthracis* testades därför.

Det som eftersöktes var relativt ofarliga *Bacillus*-stammar som liknade *B. anthracis* vad gäller egenskaper som till exempel känslighet för värme. Genom-analyser tillsammans med analyser av en gen som heter *spoIIIAB* utfördes för att se vilka stammar som var genetiskt sett mest lika *B. anthracis*. De som var mest lika, tillsammans med mer olika arter, valdes ut och sporulerades. Dessa sporer jämfördes med *B. anthracis*-sporer angående värmekänslighet, svar på olika germinerande ämnen, storlek och utseende och växt på agarplattor som är relativt specifika för *B. anthracis*.

Två stammar av arten *B. cereus* identifierades som extremt lika *B. anthracis* – både genomiskt och i de efterföljande experimenten. Ett oväntat resultat var att de även växte på agarplattor som är utvecklade för växt av *B. anthracis*.

Dessa två stammar har, sedan studien publicerades, använts i flertalet övningar där likheten med *B. anthracis* gjort övningarna mer realistiska. De har använts i metodutvecklingar där metoden kunnat utvecklas och optimeras på ofarliga sporer på ett vanligt skyddsnivå-2-laboratorium för att sen testas på *B. anthracis*-sporer på ett skyddsnivå-3-laboratorium. De har också använts som kontroller i diagnostik av jord- och omgivningsprover från områden som haft mjältbrandsutbrott.

7.2.2 Mjukvara för att jämföra bakteriella genom

För att dra slutsatser om en bakteries genom så hjälper det om man kan jämföra genomet med andra bakteriers genom. Att jämföra hela genom mot varandra är svårt när antalet genom överstiger cirka 20 stycken, mycket p.g.a. av att normala datorer inte klarar av mängden beräkningar som krävs.

Mjukvaran Gegenees har därför utvecklats för att se konservering av genetiskt material mellan ett stort antal genom. Programmet delar upp DNA-sekvenserna i små bitar och jämför alla mot alla istället för att jämföra hela genomen mot varandra direkt. Detta gör att en normal arbetsdator kan jämföra flera hundra genom samtidigt, beroende på genom-storlek. Gegenees lämpar sig väl för att utveckla ny diagnostik då unika områden i en bakteries DNA snabbt kan identifieras med programmet och specifika diagnostiska markörer placeras där. Gegenees kan också hjälpa användaren att hitta orsaker till extrem patogenicitet eller antibiotikaresistens i en bakterie.

7.2.3 Optimera diagnostikmarkörer

Eftersom jordprover nästan alltid innehåller andra *Bacillus*-arter så kan det skapa falskt positiva analysresultat om någon av dem innehåller DNA-sekvenser som tidigare antagits vara unika för *B. anthracis*. För att säkerställa att så specifika markörer som möjligt används så utfördes en jämförelsestudie av alla publicerade DNA-markörer för detektion av *B. anthracis*.

Programmet Gegenees jämförde alla markörsekvenser mot alla publikt tillgängliga *Bacillus*-genom. Sekvenserna sorterades sedan för specificitet vilket betyder att sekvenser som får en perfekt träff i alla *B. anthracis*-genom men inte en enda perfekt träff i andra *Bacillus* sorteras fram först. Tre markörsekvenser var enligt Gegenees 100 % specifika så de valdes ut tillsammans med några med lägre specificitet och markörer som WHO rekommenderade. Dessa markörer analyserades sedan i en ringstudie mellan fem europeiska laboratorier för att jämföra Gegenees-resultaten med verkliga analysresultat. Resultaten överensstämde mellan laboratorierna och med vad Gegenees förutspått. De tre teoretiskt bästa sekvenserna var även bäst i verkligheten. Denna metodik visade sig vara effektiv på att jämföra diagnostiska DNA-markörsekvenser mot varandra och kan säkerställa att bästa möjliga sekvenser används.

7.2.4 Antibiotikaresistens

Bacillus anthracis är normalt känslig för penicillin men efter penicillinbehandling av djuren under utbrottet 2011 så utvecklades resistens och resistenta bakterier isolerades från döda djur. Dessa isolat helgenomsekvenserades och jämfördes med isolat som fortfarande var känsliga

för att se vad som hänt i genomet hos de resistenta isolaten. Det visade vara en så kallad anti-sigmafaktor, RsiP, som hade muterats på olika sätt i olika isolat men med samma effekt – att denna negativa sigmafaktor inte längre fungerade. Denna RsiP och dess positiva sigmafaktor SigP styr uttrycket av betalaktamasgener som behövs för att motverka penicillinet. Att den negativa faktorn nu var söndermuterad ger att den positiva faktorn går på högvarv och betalaktamaser uttrycks hela tiden och bakterien kan inte styra detta.

Det visade sig också att i bakterieisolat som sparats i frys så fanns det en subpopulation av bakterier som också hade muterade, och därmed defekta, positiva sigmafaktorer. Dessa bakterier var som väntat inte längre resistenta mot penicillin. Att det bara krävs en enda slumpmässig mutation i *rsiP*-genen för att *B. anthracis* ska bli resistent betyder att om det finns tillräckligt många bakterier så kommer någon av dem ha ”rätt” mutation för att vara resistent. Detta ger bakterien ingen vinst, snarare tvärtom, men när penicillin sätts in som behandling blir den ensam kvar och kan växa och döda djuret. Att det kan finnas bakterier som sen har råkat stänga av sin resistens för gott kan ha effekter på resistensbedömningar på laboratoriet.

7.2.5 Smittspåra med hjälp av helgenomsekvensering

Vid smittspårningar av bakterier så undersöks vissa områden i bakteriens genom som har stor variabilitet. Jämförelser av dessa kan då användas för att skilja mellan olika stammar. Stammar av *B. anthracis* från hela världen har väldigt låg variation och ser väldigt lika ut, genomiskt sett. Det gör att väldigt högupplösta metoder måste användas för att kunna skilja mellan isolat från olika ursprung. Helgenomsekvensering representerar den högsta nivån av upplösning eftersom alla baser undersöks efter mutationer som är diskriminerande. Under ett utbrott av *B. anthracis* så är det oftast en enda stam som infekterar en grupp djur så bakteriernas DNA är i stort sett identiskt men ett fåtal mutationer borde ändå uppkomma under en passage genom ett djur.

För att kvantifiera hur många mutationer som uppkommer så helgenomsekvenserades flera isolat från flera olika döda djur från utbrottet i Kvismaren. Det visade sig att varje isolat hade åtminstone ett till två baspar muterade jämfört med de andra isolaten. Detta tolkades som att om identiska bakterier infekterar olika djur så kommer de skilja sig på en till två positioner efter att djuren dött. Men efter att bakterier från den troliga källan till utbrottet sekvenserats så var även de heterogena. Källan i detta fall misstänks vara en mjältbrandsgrav från 1943. Detta betyder att vi inte kan veta om mutationerna vi såg 2011 uppkom redan 1943 och senare ärvdes in i djuren 2011 eller om de faktiskt uppkom 2011.

Sekvenseringar av blodprover från olika delar av djuren visade att det finns rumsliga variationer i djuren. Det betyder att bakterier med olika mutationer kommer ha olika sannolikheter för att isoleras på olika provställen på djuret.

Helgenomsekvensering för att karaktärisera ett mjältbrandsisolat visade sig vara effektivt men fler isolat från samma djur bör sekvenseras för att kunna dra slutsatser om hur populationen ser ut. Kunskap om 2011-isolaten användes när mjältbrand återigen påvisades 2013 på en gård inte långt från utbrottsområdet 2011. En jämförelse av det nya isolatet mot de från 2011 gav snabbt svaret att det var samma stam och en ny källa till 2013-års utbrott kunde uteslutas.

7.3 Slutsatser

De i avhandlingen beskrivna studierna visar att man med hjälp av de senaste sekvenseringsteknikerna och genomikstudier avsevärt kunnat förbättra förutsättningarna för snabb och effektiv hantering av ett mjältbrandsutbrott. Diagnostiska markörer har utvärderats och modellsporer har utvecklats för att lättare utveckla metoder och underlätta övningar. Slutligen har genomiska karaktäriseringar av utbrottsisolat blivit lättare och mer kraftfulla att utföra under själva utbrottet. Detta kan svara på frågan om huruvida bakterierna kommer från ett laboratorium, har några speciella attribut eller om de är helt normala bakterier som förväntas kunna hittas i utbrottsområdet.

Viktiga forsknings- och utvecklingsområden för framtiden borde vara användarvänligare mjukvara som kan hantera den exponentiella utvecklingen på själva sekvenseringsmaskinerna och upprättandet av en databas med genomiska data från *B. anthracis*-stammar. Om alla länder, i åtminstone Europa, samarbetade så skulle ett nytt utbrottsisolat kunna jämföras med all tillgänglig kunskap om tidigare stammar. Detta skulle vara väldigt viktigt även vid ett utbrott med antagonistiskt ursprung.

8 References

AFS.2005:1 *Microbiological Work Environment Risks – Infection, Toxigenic Effect, Hypersensitivity*: The Swedish Work Environment Authority.

Agaisse, H., Gominet, M., Okstad, O.A., Kolsto, A.B. & Lereclus, D. (1999). PlcR is a pleiotropic regulator of extracellular virulence factor gene expression in *Bacillus thuringiensis*. *Mol Microbiol,* 32(5), pp. 1043-53.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990). Basic local alignment search tool. *J Mol Biol,* 215(3), pp. 403-10.

Andrews, S. *FastQC: a quality control tool for high throughput sequence data*. http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Aronson, A.I., Beckman, W. & Dunn, P. (1986). *Bacillus thuringiensis* and related insect pathogens. *Microbiol Rev,* 50(1), pp. 1-24.

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(2), pp. 424-8.

Baize, S., Pannetier, D., Oestereich, L., Rieger, T., Koivogui, L., Magassouba, N., Soropogui, B., Sow, M.S., Keita, S., De Clerck, H., Tiffany, A., Dominguez, G., Loua, M., Traore, A., Kolie, M., Malano, E.R., Heleze, E., Bocquin, A., Mely, S., Raoul, H., Caro, V., Cadar, D., Gabriel, M., Pahlmann, M., Tappe, D., Schmidt-Chanasit, J., Impouma, B., Diallo, A.K., Formenty, P., Van Herp, M. & Gunther, S. (2014). Emergence of Zaire Ebola Virus Disease in Guinea - Preliminary Report. *N Engl J Med*.

Barnes, J.M. (1947). Penicillin and *B. anthracis*. *The Journal of pathology and bacteriology,* 59(1-2), pp. 113-25.

Bavykin, S.G., Lysov, Y.P., Zakhariev, V., Kelly, J.J., Jackman, J., Stahl, D.A. & Cherni, A. (2004). Use of 16S rRNA, 23S rRNA, and gyrB gene sequence analysis to determine phylogenetic relationships of *Bacillus cereus* group microorganisms. *J Clin Microbiol,* 42(8), pp. 3711-30.

Bellan, S.E., Turnbull, P.C., Beyer, W. & Getz, W.M. (2013). Effects of experimental exclusion of scavengers from carcasses of anthrax-infected herbivores on *Bacillus anthracis* sporulation, survival, and distribution. *Appl Environ Microbiol,* 79(12), pp. 3756-61.

Beres, S.B., Carroll, R.K., Shea, P.R., Sitkiewicz, I., Martinez-Gutierrez, J.C., Low, D.E., McGeer, A., Willey, B.M., Green, K., Tyrrell, G.J., Goldman,

T.D., Feldgarden, M., Birren, B.W., Fofanov, Y., Boos, J., Wheaton, W.D., Honisch, C. & Musser, J.M. (2010). Molecular complexity of successive bacterial epidemics deconvoluted by comparative pathogenomics. *Proc Natl Acad Sci U S A,* 107(9), pp. 4371-6.

Bergman, N.H. (ed.) (2011). *Bacillus anthracis and Anthrax*. New Jersey: John Wiley & Sons, Inc.

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(3), p. e1534.

Blackburn, J.K. & Goodin, D.G. (2013). Differentiation of springtime vegetation indices associated with summer anthrax epizootics in west Texas, USA, deer. *J Wildl Dis,* 49(3), pp. 699-703.

Braack, L.E. & De Vos, V. (1990). Feeding habits and flight range of blow-flies (Chrysomyia spp.) in relation to anthrax transmission in the Kruger National Park, South Africa. *Onderstepoort J Vet Res,* 57(2), pp. 141-2.

Brannigan, J.A., Tirodimos, I.A., Zhang, Q.Y., Dowson, C.G. & Spratt, B.G. (1990). Insertion of an extra amino acid is the main cause of the low affinity of penicillin-binding protein 2 in penicillin-resistant strains of *Neisseria gonorrhoeae*. *Mol Microbiol,* 4(6), pp. 913-9.

Buckley, P., Rivers, B., Katoski, S., Kim, M.H., Kragl, F.J., Broomall, S., Krepps, M., Skowronski, E.W., Rosenzweig, C.N., Paikoff, S., Emanuel, P. & Gibbons, H.S. (2012). Genetic barcodes for improved environmental tracking of an anthrax simulant. *Appl Environ Microbiol,* 78(23), pp. 8272-80.

Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L., Vandesompele, J. & Wittwer, C.T. (2009). The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem,* 55(4), pp. 611-22.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. & Madden, T.L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics,* 10, p. 421.

Canchaya, C., Fournous, G. & Brussow, H. (2004). The impact of prophages on bacterial chromosomes. *Mol Microbiol,* 53(1), pp. 9-18.

CDC (2002). Suspected cutaneous anthrax in a laboratory worker--Texas, 2002. *MMWR Morb Mortal Wkly Rep,* 51(13), pp. 279-81.

CDC *Centers for Disease Control and Prevention*. http://www.bt.cdc.gov/agent/agentlist-category.asp [July 3].

Chambers, H.F. & Sachdeva, M. (1990). Binding of beta-lactam antibiotics to penicillin-binding proteins in methicillin-resistant *Staphylococcus aureus*. *J Infect Dis,* 161(6), pp. 1170-6.

Chen, Y., Succi, J., Tenover, F.C. & Koehler, T.M. (2003). Beta-lactamase genes of the penicillin-susceptible *Bacillus anthracis* Sterne strain. *Journal of bacteriology,* 185(3), pp. 823-30.

Daffonchio, D., Cherif, A. & Borin, S. (2000). Homoduplex and heteroduplex polymorphisms of the amplified ribosomal 16S-23S internal transcribed

spacers describe genetic relationships in the "*Bacillus cereus* group". *Applied and environmental microbiology*, 66(12), pp. 5460-8.

Darling, A.E., Mau, B. & Perna, N.T. (2010). progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One*, 5(6), p. e11147.

Dauphin, L.A., Moser, B.D. & Bowen, M.D. (2009). Evaluation of five commercial nucleic acid extraction kits for their ability to inactivate *Bacillus anthracis* spores and comparison of DNA yields from spores and spiked environmental samples. *Journal of microbiological methods*, 76(1), pp. 30-7.

Derzelle, S. & Thierry, S. (2013). Genetic diversity of *Bacillus anthracis* in Europe: genotyping methods in forensic and epidemiologic investigations. *Biosecur Bioterror*, 11 Suppl 1, pp. S166-76.

Dixon, T.C., Meselson, M., Guillemin, J. & Hanna, P.C. (1999). Anthrax. *N Engl J Med*, 341(11), pp. 815-26.

Dragon, D.C., Bader, D.E., Mitchell, J. & Woollen, N. (2005). Natural dissemination of *Bacillus anthracis* spores in northern Canada. *Appl Environ Microbiol*, 71(3), pp. 1610-5.

Dragon, D.C. & Rennie, R.P. (2001). Evaluation of spore extraction and purification methods for selective recovery of viable *Bacillus anthracis* spores. *Lett Appl Microbiol*, 33(2), pp. 100-5.

Drysdale, M., Heninger, S., Hutt, J., Chen, Y., Lyons, C.R. & Koehler, T.M. (2005). Capsule synthesis by *Bacillus anthracis* is required for dissemination in murine inhalation anthrax. *EMBO J*, 24(1), pp. 221-7.

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. *Eur J Clin Microbiol Infect Dis*, 31(10), pp. 2783-90.

Edmonds, J.M., Collett, P.J., Valdes, E.R., Skowronski, E.W., Pellar, G.J. & Emanuel, P.A. (2009). Surface sampling of spores in dry-deposition aerosols. *Appl Environ Microbiol*, 75(1), pp. 39-44.

Ehling-Schulz, M. & Messelhausser, U. (2013). *Bacillus* "next generation" diagnostics: moving from detection toward subtyping and risk-related strain profiling. *Front Microbiol*, 4, p. 32.

Ehling-Schulz, M., Svensson, B., Guinebretiere, M.H., Lindback, T., Andersson, M., Schulz, A., Fricker, M., Christiansson, A., Granum, P.E., Martlbauer, E., Nguyen-The, C., Salkinoja-Salonen, M. & Scherer, S. (2005). Emetic toxin formation of *Bacillus cereus* is restricted to a single evolutionary lineage of closely related strains. *Microbiology*, 151(Pt 1), pp. 183-97.

Elena, S.F. & Lenski, R.E. (2003). Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nat Rev Genet*, 4(6), pp. 457-69.

Ellis, R.J., Morgan, P., Weightman, A.J. & Fry, J.C. (2003). Cultivation-dependent and -independent approaches for determining bacterial diversity in heavy-metal-contaminated soil. *Appl Environ Microbiol*, 69(6), pp. 3223-30.

Elvander, M. *Källforskning i syfte att geografiskt kartlägga mjältbrandsgårdar i Sverige 1910 – 1957.* http://www.sva.se/upload/Redesign2011/Pdf/Djurhalsa/Epizootier/Slutrapport_mjaltbrand120508.pdf.

Fasanella, A., Di Taranto, P., Garofolo, G., Colao, V., Marino, L., Buonavoglia, D., Pedarra, C., Adone, R. & Hugh-Jones, M. (2013a). Ground Anthrax Bacillus Refined Isolation (GABRI) method for analyzing environmental samples with low levels of *Bacillus anthracis* contamination. *BMC Microbiol,* 13, p. 167.

Fasanella, A., Garofolo, G., Galella, M., Troiano, P., De Stefano, C., Pace, L., Aceti, A., Serrecchia, L. & Adone, R. (2013b). Suspect vector transmission of human cutaneous anthrax during an animal outbreak in Southern Italy. *Vector Borne Zoonotic Dis,* 13(10), pp. 769-71.

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. *J Clin Microbiol,* 43(7), pp. 3398-401.

Fisher, N. & Hanna, P. (2005). Characterization of *Bacillus anthracis* germinant receptors in vitro. *Journal of bacteriology,* 187(23), pp. 8055-62.

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.F., Dougherty, B.A., Merrick, J.M. & et al. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science,* 269(5223), pp. 496-512.

Fouet, A., Smith, K.L., Keys, C., Vaissaire, J., Le Doujet, C., Levy, M., Mock, M. & Keim, P. (2002). Diversity among French *Bacillus anthracis* isolates. *J Clin Microbiol,* 40(12), pp. 4732-4.

Garofolo, G., Ciammaruconi, A., Fasanella, A., Scasciamacchia, S., Adone, R., Pittiglio, V. & Lista, F. (2010). SNR analysis: molecular investigation of an anthrax epidemic. *BMC Vet Res,* 6, p. 11.

Garofolo, G., Serrecchia, L., Corro, M. & Fasanella, A. (2011). Anthrax phylogenetic structure in Northern Italy. *BMC Res Notes,* 4, p. 273.

Gibbons, H.S., Broomall, S.M., McNew, L.A., Daligault, H., Chapman, C., Bruce, D., Karavis, M., Krepps, M., McGregor, P.A., Hong, C., Park, K.H., Akmal, A., Feldman, A., Lin, J.S., Chang, W.E., Higgs, B.W., Demirev, P., Lindquist, J., Liem, A., Fochler, E., Read, T.D., Tapia, R., Johnson, S., Bishop-Lilly, K.A., Detter, C., Han, C., Sozhamannan, S., Rosenzweig, C.N. & Skowronski, E.W. (2011). Genomic signatures of strain selection and enhancement in *Bacillus atrophaeus* var. globigii, a historical biowarfare simulant. *PLoS One,* 6(3), p. e17836.

Ginetti, F., Perego, M., Albertini, A.M. & Galizzi, A. (1996). *Bacillus subtilis* mutS mutL operon: identification, nucleotide sequence and mutagenesis. *Microbiology,* 142 (Pt 8), pp. 2021-9.

Gohar, M., Okstad, O.A., Gilois, N., Sanchis, V., Kolsto, A.B. & Lereclus, D. (2002). Two-dimensional electrophoresis analysis of the extracellular proteome of *Bacillus cereus* reveals the importance of the PlcR regulon. *Proteomics,* 2(6), pp. 784-91.

Green, B.D., Battisti, L., Koehler, T.M., Thorne, C.B. & Ivins, B.E. (1985). Demonstration of a capsule plasmid in *Bacillus anthracis*. *Infect Immun,* 49(2), pp. 291-7.

Greenberg, D.L., Busch, J.D., Keim, P. & Wagner, D.M. (2010). Identifying experimental surrogates for *Bacillus anthracis* spores: a review. *Investig Genet,* 1(1), p. 4.

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., Cai, W., Zimmermann, R. & Schaade, L. (2013). Anthrax among heroin users in Europe possibly caused by same *Bacillus anthracis* strain since 2000. *Euro Surveill,* 18(13).

Guidi-Rontani, C., Weber-Levy, M., Labruyere, E. & Mock, M. (1999). Germination of *Bacillus anthracis* spores within alveolar macrophages. *Mol Microbiol,* 31(1), pp. 9-17.

Guidi, V., Patocchi, N., Luthy, P. & Tonolla, M. (2011). Distribution of *Bacillus thuringiensis* subsp. israelensis in Soil of a Swiss Wetland reserve after 22 years of mosquito control. *Appl Environ Microbiol,* 77(11), pp. 3663-8.

Gulledge, J.S., Luna, V.A., Luna, A.J., Zartman, R. & Cannons, A.C. (2010). Detection of low numbers of *Bacillus anthracis* spores in three soils using five commercial DNA extraction methods with and without an enrichment step. *J Appl Microbiol,* 109(5), pp. 1509-20.

Harris, S.R., Cartwright, E.J.P., Török, M.E., Holden, M.T.G., Brown, N.M., Ogilvy-Stuart, A.L., Ellington, M.J., Quail, M.A., Bentley, S.D., Parkhill, J. & Peacock, S.J. (2013). Whole-genome sequencing for analysis of an outbreak of meticillin-resistant *Staphylococcus aureus*: a descriptive study. *The Lancet Infectious Diseases,* 13(2), pp. 130-136.

Heilbron, K., Toll-Riera, M., Kojadinovic, M. & MacLean, R. (2014). Fitness Is Strongly Influenced by Rare Mutations of Large Effect in a Microbial Mutation Accumulation Experiment. *Genetics,* 114.

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(6), pp. 2627-30.

Hendriksen, R.S., Price, L.B., Schupp, J.M., Gillece, J.D., Kaas, R.S., Engelthaler, D.M., Bortolaia, V., Pearson, T., Waters, A.E., Upadhyay, B.P., Shrestha, S.D., Adhikari, S., Shakya, G., Keim, P.S. & Aarestrup, F.M. (2011). Population genetics of *Vibrio cholerae* from Nepal in 2010: evidence on the origin of the Haitian outbreak. *MBio,* 2(4), pp. e00157-11.

Hoffmann, B., Scheuch, M., Hoper, D., Jungblut, R., Holsteg, M., Schirrmeier, H., Eschbaumer, M., Goller, K.V., Wernike, K., Fischer, M., Breithaupt, A., Mettenleiter, T.C. & Beer, M. (2012). Novel orthobunyavirus in Cattle, Europe, 2011. *Emerg Infect Dis,* 18(3), pp. 469-72.

Hoffmaster, A.R., Fitzgerald, C.C., Ribot, E., Mayer, L.W. & Popovic, T. (2002). Molecular subtyping of *Bacillus anthracis* and the 2001 bioterrorism-associated anthrax outbreak, United States. *Emerg Infect Dis,* 8(10), pp. 1111-6.

Hoffmaster, A.R., Ravel, J., Rasko, D.A., Chapman, G.D., Chute, M.D., Marston, C.K., De, B.K., Sacchi, C.T., Fitzgerald, C., Mayer, L.W., Maiden, M.C., Priest, F.G., Barker, M., Jiang, L., Cer, R.Z., Rilstone, J., Peterson, S.N., Weyant, R.S., Galloway, D.R., Read, T.D., Popovic, T. & Fraser, C.M. (2004). Identification of anthrax toxin genes in a *Bacillus cereus* associated with an illness resembling inhalation anthrax. *Proc Natl Acad Sci U S A*, 101(22), pp. 8449-54.

Holty, J.E., Bravata, D.M., Liu, H., Olshen, R.A., McDonald, K.M. & Owens, D.K. (2006). Systematic review: a century of inhalational anthrax cases from 1900 to 2005. *Ann Intern Med*, 144(4), pp. 270-80.

Hugh-Jones, M. & Blackburn, J. (2009). The ecology of *Bacillus anthracis*. *Mol Aspects Med*, 30(6), pp. 356-67.

Inglesby, T.V., O'Toole, T., Henderson, D.A., Bartlett, J.G., Ascher, M.S., Eitzen, E., Friedlander, A.M., Gerberding, J., Hauer, J., Hughes, J., McDade, J., Osterholm, M.T., Parker, G., Perl, T.M., Russell, P.K. & Tonat, K. (2002). Anthrax as a biological weapon, 2002: updated recommendations for management. *JAMA : the journal of the American Medical Association*, 287(17), pp. 2236-52.

Jernigan, D.B., Raghunathan, P.L., Bell, B.P., Brechner, R., Bresnitz, E.A., Butler, J.C., Cetron, M., Cohen, M., Doyle, T., Fischer, M., Greene, C., Griffith, K.S., Guarner, J., Hadler, J.L., Hayslett, J.A., Meyer, R., Petersen, L.R., Phillips, M., Pinner, R., Popovic, T., Quinn, C.P., Reefhuis, J., Reissman, D., Rosenstein, N., Schuchat, A., Shieh, W.J., Siegal, L., Swerdlow, D.L., Tenover, F.C., Traeger, M., Ward, J.W., Weisfuse, I., Wiersma, S., Yeskey, K., Zaki, S., Ashford, D.A., Perkins, B.A., Ostroff, S., Hughes, J., Fleming, D., Koplan, J.P. & Gerberding, J.L. (2002). Investigation of bioterrorism-related anthrax, United States, 2001: epidemiologic findings. *Emerg Infect Dis*, 8(10), pp. 1019-28.

Kao, R.R. (2002). The role of mathematical modelling in the control of the 2001 FMD epidemic in the UK. *Trends Microbiol*, 10(6), pp. 279-86.

Keim, P., Price, L.B., Klevytska, A.M., Smith, K.L., Schupp, 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*. *Journal of bacteriology*, 182(10), pp. 2928-36.

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. *Infect Genet Evol*, 4(3), pp. 205-13.

Kim, H.S., Sherman, D., Johnson, F. & Aronson, A.I. (2004). Characterization of a major *Bacillus anthracis* spore coat protein and its role in spore inactivation. *Journal of bacteriology*, 186(8), pp. 2413-7.

Klee, S.R., Brzuszkiewicz, E.B., Nattermann, H., Bruggemann, H., Dupke, S., Wollherr, A., Franz, T., Pauli, G., Appel, B., Liebl, W., Couacy-Hymann, E., Boesch, C., Meyer, F.D., Leendertz, F.H., Ellerbrok, H., Gottschalk, G., Grunow, R. & Liesegang, H. (2010). The genome of a *Bacillus* isolate causing anthrax in chimpanzees combines chromosomal properties of *B. cereus* with *B. anthracis* virulence plasmids. *PLoS One*, 5(7), p. e10986.

Klee, S.R., Ozel, M., Appel, B., Boesch, C., Ellerbrok, H., Jacob, D., Holland, G., Leendertz, F.H., Pauli, G., Grunow, R. & Nattermann, H. (2006). Characterization of *Bacillus anthracis*-like bacteria isolated from wild great apes from Cote d'Ivoire and Cameroon. *Journal of bacteriology*, 188(15), pp. 5333-44.

Knisely, R.F. (1966). Selective medium for *Bacillus anthracis*. *Journal of bacteriology*, 92(3), pp. 784-6.

Knutsson, R., Båverud, V., Elvander, M., Olsson Engvall, E., Eliasson, K. & Sternberg Lewerin, S. (2012). Managing and learning from an anthrax outbreak in a Swedish beef cattle herd. In: Hoorfar, J. (ed. *Case studies in food safety and authenticity. Lessons from real-life situations*. Cornwall, UK: Woodhead Publishing Limited, pp. 151-160.

Koch, R. (1876). Die Ätiologie der Milzbrand-Krankheit, begründet auf die Entwicklungsgeschichte des *Bacillus Anthracis*. *Beitr. Biol. Pflanz.*, 2, pp. 277–310.

Kolsto, A.B., Tourasse, N.J. & Okstad, O.A. (2009). What sets *Bacillus anthracis* apart from other *Bacillus* species? *Annual review of microbiology*, 63, pp. 451-76.

Koren, S., Harhay, G.P., Smith, T.P., Bono, J.L., Harhay, D.M., McVey, S.D., Radune, D., Bergman, N.H. & Phillippy, A.M. (2013). Reducing assembly complexity of microbial genomes with single-molecule sequencing. *Genome Biol*, 14(9), p. R101.

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. *J Clin Microbiol*, 48(8), pp. 2821-9.

Kuske, C.R., Barns, S.M., Grow, C.C., Merrill, L. & Dunbar, J. (2006). Environmental survey for four pathogenic bacteria and closely related species using phylogenetic and functional genes. *J Forensic Sci*, 51(3), pp. 548-58.

Köser, C.U., Holden, M.T., Ellington, M.J., Cartwright, E.J., Brown, N.M., Ogilvy-Stuart, A.L., Hsu, L.Y., Chewapreecha, C., Croucher, N.J., Harris, S.R., Sanders, M., Enright, M.C., Dougan, G., Bentley, S.D., Parkhill, J., Fraser, L.J., Betley, J.R., Schulz-Trieglaff, O.B., Smith, G.P. & Peacock, S.J. (2012). Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. *The New England journal of medicine*, 366(24), pp. 2267-75.

Leishman, O.N., Labuza, T.P. & Diez-Gonzalez, F. (2010). Hydrophobic properties and extraction of *Bacillus anthracis* spores from liquid foods. *Food Microbiol*, 27(5), pp. 661-6.

Lewerin, S.S., Elvander, M., Westermark, T., Hartzell, L.N., Norstrom, A.K., Ehrs, 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, p. 7.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L. & Law, M. (2012). Comparison of next-generation sequencing systems. *J Biomed Biotechnol*, 2012, p. 251364.

Luna, V.A., Gulledge, J., Cannons, A.C. & Amuso, P.T. (2009). Improvement of a selective media for the isolation of *B. anthracis* from soils. *Journal of microbiological methods*, 79(3), pp. 301-6.

Luo, C., Tsementzi, D., Kyrpides, N., Read, T. & Konstantinidis, K.T. (2012). Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample. *PLoS One*, 7(2), p. e30087.

Macia, M.D., Blanquer, D., Togores, B., Sauleda, J., Perez, J.L. & Oliver, A. (2005). Hypermutation is a key factor in development of multiple-antimicrobial resistance in *Pseudomonas aeruginosa* strains causing chronic lung infections. *Antimicrobial agents and chemotherapy*, 49(8), pp. 3382-6.

Makino, S., Uchida, I., Terakado, N., Sasakawa, C. & Yoshikawa, M. (1989). Molecular characterization and protein analysis of the cap region, which is essential for encapsulation in *Bacillus anthracis*. *Journal of bacteriology*, 171(2), pp. 722-30.

Manchee, R.J., Broster, M.G., Stagg, A.J. & Hibbs, S.E. (1994). Formaldehyde Solution Effectively Inactivates Spores of *Bacillus anthracis* on the Scottish Island of Gruinard. *Appl Environ Microbiol*, 60(11), pp. 4167-71.

Marston, C.K., Beesley, C., Helsel, L. & Hoffmaster, A.R. (2008). Evaluation of two selective media for the isolation of *Bacillus anthracis*. *Lett Appl Microbiol*, 47(1), pp. 25-30.

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 Microbiol*, 6, p. 22.

Mellmann, A., Harmsen, D., Cummings, C.A., Zentz, E.B., Leopold, S.R., Rico, A., Prior, K., Szczepanowski, R., Ji, Y., Zhang, W., McLaughlin, S.F., Henkhaus, J.K., Leopold, B., Bielaszewska, M., Prager, R., Brzoska, P.M., Moore, R.L., Guenther, S., Rothberg, J.M. & Karch, H. (2011). Prospective genomic characterization of the German enterohemorrhagic *Escherichia coli* O104:H4 outbreak by rapid next generation sequencing technology. *PLoS One*, 6(7), p. e22751.

Meselson, M., Guillemin, J., Hugh-Jones, M., Langmuir, A., Popova, I., Shelokov, A. & Yampolskaya, O. (1994). The Sverdlovsk anthrax outbreak of 1979. *Science*, 266(5188), pp. 1202-8.

Meynell, E. & Meynell, G.G. (1964). The Roles of Serum and Carbon Dioxide in Capsule Formation by *Bacillus Anthracis*. *J Gen Microbiol*, 34, pp. 153-64.

Miller, J.M., Hair, J.G., Hebert, M., Hebert, L., Roberts, F.J., Jr. & Weyant, R.S. (1997). Fulminating bacteremia and pneumonia due to *Bacillus cereus*. *J Clin Microbiol*, 35(2), pp. 504-7.

Mock, M. & Fouet, A. (2001). Anthrax. *Annual review of microbiology*, 55, pp. 647-71.

Nicholson, W.L., Munakata, N., Horneck, G., Melosh, H.J. & Setlow, P. (2000). Resistance of *Bacillus* endospores to extreme terrestrial and extraterrestrial environments. *Microbiol Mol Biol Rev,* 64(3), pp. 548-72.

Nishi, J.S., Ellsworth, T.R., Lee, N., Dewar, D., Elkin, B.T. & Dragon, D.C. (2007). Northwest Territories. An outbreak of anthrax (*Bacillus anthracis*) in free-roaming bison in the Northwest Territories, June-July 2006. *Can Vet J,* 48(1), pp. 37-8.

OIE (2013). *Terrestrial Code*: World Organisation for Animal Health (OIE).

Okinaka, R.T., Cloud, K., Hampton, O., Hoffmaster, A.R., Hill, K.K., Keim, P., Koehler, T.M., Lamke, G., Kumano, S., Mahillon, J., Manter, D., Martinez, Y., Ricke, D., Svensson, R. & Jackson, P.J. (1999). Sequence and organization of pXO1, the large *Bacillus anthracis* plasmid harboring the anthrax toxin genes. *Journal of bacteriology,* 181(20), pp. 6509-15.

Pearson, T. (2004). Phylogenetic discovery bias in *Bacillus anthracis* using single-nucleotide polymorphisms from whole-genome sequencing. *Proceedings of the National Academy of Sciences,* 101(37), pp. 13536-13541.

Pike, R.M., Sulkin, S.E. & Schulze, M.L. (1965). Continuing Importance of Laboratory-Acquired Infections. *Am J Public Health Nations Health,* 55, pp. 190-9.

Pile, J.C., Malone, J.D., Eitzen, E.M. & Friedlander, A.M. (1998). Anthrax as a potential biological warfare agent. *Arch Intern Med,* 158(5), pp. 429-34.

Pinho, M.G., de Lencastre, H. & Tomasz, A. (2001). An acquired and a native penicillin-binding protein cooperate in building the cell wall of drug-resistant staphylococci. *Proc Natl Acad Sci U S A,* 98(19), pp. 10886-91.

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(8), pp. 1307-13.

Priest, F.G., Barker, M., Baillie, L.W., Holmes, E.C. & Maiden, M.C. (2004). Population structure and evolution of the *Bacillus cereus* group. *Journal of bacteriology,* 186(23), pp. 7959-70.

Quail, M.A., Smith, M., Coupland, P., Otto, T.D., Harris, S.R., Connor, T.R., Bertoni, A., Swerdlow, H.P. & Gu, Y. (2012). A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics,* 13, p. 341.

Ramsay, C.N., Stirling, A., Smith, J., Hawkins, G., Brooks, T., Hood, J., Penrice, G., Browning, L.M. & Ahmed, S. (2010). An outbreak of infection with *Bacillus anthracis* in injecting drug users in Scotland. *Euro Surveill,* 15(2).

Rasko, D.A., Webster, D.R., Sahl, J.W., Bashir, A., Boisen, N., Scheutz, F., Paxinos, E.E., Sebra, R., Chin, C.S., Iliopoulos, D., Klammer, A., Peluso, P., Lee, L., Kislyuk, A.O., Bullard, J., Kasarskis, A., Wang, S., Eid, J., Rank, D., Redman, J.C., Steyert, S.R., Frimodt-Moller, J., Struve, C., Petersen, A.M., Krogfelt, K.A., Nataro, J.P., Schadt, E.E. & Waldor, M.K.

(2011). Origins of the *E. coli* strain causing an outbreak of hemolytic-uremic syndrome in Germany. *N Engl J Med,* 365(8), pp. 709-17.

Read, T.D., Peterson, 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., Okstad, 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., Kolsto, A.B. & Fraser, C.M. (2003). The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria. *Nature,* 423(6935), pp. 81-6.

Redmond, C., Pearce, M.J., Manchee, R.J. & Berdal, B.P. (1998). Deadly relic of the Great War. *Nature,* 393(6687), pp. 747-8.

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(9241), pp. 1574-5.

Rokicki, J., Knox, D., Dowell, R.D. & Copley, S.D. (2014). CodaChrome: a tool for the visualization of proteome conservation across all fully sequenced bacterial genomes. *BMC Genomics,* 15, p. 65.

Rosenberg, M., Gutnick, D. & Rosenberg, E. (1980). Adherence of bacteria to hydrocarbons: A simple method for measuring cell-surface hydrophobicity. *FEMS Microbiology Letters,* 9, pp. 29-33.

Ross, C.L., Thomason, K.S. & Koehler, T.M. (2009). An extracytoplasmic function sigma factor controls beta-lactamase gene expression in *Bacillus anthracis* and other *Bacillus cereus* group species. *Journal of bacteriology,* 191(21), pp. 6683-93.

Roth, N.G., Livery, D.H. & Hodge, H.M. (1955). Influence of oxygen uptake and age of culture on sporulation of *Bacillus anthracis* and *Bacillus globigii*. *Journal of bacteriology,* 69(4), pp. 455-9.

Rutqvist L, S.O. (1957). Epizootologiska och bakteriologiska undersökningar
vid mjältbrandsepizootien i Sverige 1956-1957. *Nord Vet Med,* 9, pp. 641-663.

Sanger, F., Nicklen, S. & Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A,* 74(12), pp. 5463-7.

Sastalla, I., Maltese, L.M., Pomerantseva, O.M., Pomerantsev, A.P., Keane-Myers, A. & Leppla, S.H. (2010). Activation of the latent PlcR regulon in *Bacillus anthracis*. *Microbiology,* 156(Pt 10), pp. 2982-93.

Segerman, B. (2012). The genetic integrity of bacterial species: the core genome and the accessory genome, two different stories. *Front Cell Infect Microbiol,* 2, p. 116.

Segerman, B., De Medici, D., Ehling Schulz, M., Fach, P., Fenicia, L., Fricker, M., Wielinga, P., Van Rotterdam, B. & Knutsson, R. (2011). Bioinformatic tools for using whole genome sequencing as a rapid high resolution diagnostic typing tool when tracing bioterror organisms in the food and feed chain. *Int J Food Microbiol,* 145 Suppl 1, pp. S167-76.

SFS.1999:657 (1999). *Epizootic Act SFS. 1999:657*.
Shafazand, S., Doyle, R., Ruoss, S., Weinacker, A. & Raffin, T.A. (1999). Inhalational anthrax: epidemiology, diagnosis, and management. *Chest,* 116(5), pp. 1369-76.
Shury, T.K., Frandsen, D. & O'Brodovich, L. (2009). Anthrax in free-ranging bison in the Prince Albert National Park area of Saskatchewan in 2008. *Can Vet J,* 50(2), pp. 152-4.
Sjödin, A., Broman, T., Melefors, O., Andersson, G., Rasmusson, B., Knutsson, R. & Forsman, M. (2013). The need for high-quality whole-genome sequence databases in microbial forensics. *Biosecur Bioterror,* 11 Suppl 1, pp. S78-86.
Slamti, L., Perchat, S., Gominet, M., Vilas-Boas, G., Fouet, A., Mock, M., Sanchis, V., Chaufaux, J., Gohar, M. & Lereclus, D. (2004). Distinct mutations in PlcR explain why some strains of the *Bacillus cereus* group are nonhemolytic. *Journal of bacteriology,* 186(11), pp. 3531-8.
Sozhamannan, S., Chute, M.D., McAfee, F.D., Fouts, D.E., Akmal, A., Galloway, D.R., Mateczun, A., Baillie, L.W. & Read, T.D. (2006). The *Bacillus anthracis* chromosome contains four conserved, excision-proficient, putative prophages. *BMC microbiology,* 6, p. 34.
Spencer, R.C. (2003). *Bacillus anthracis*. *J Clin Pathol,* 56(3), pp. 182-7.
Stenfors Arnesen, L.P., Fagerlund, A. & Granum, P.E. (2008). From soil to gut: *Bacillus cereus* and its food poisoning toxins. *FEMS Microbiol Rev,* 32(4), pp. 579-606.
Sternbach, G. (2003). The history of anthrax. *J Emerg Med,* 24(4), pp. 463-7.
Stratis-Cullum, D.N., Griffin, G.D., Mobley, J., Vass, A.A. & Vo-Dinh, T. (2003). A miniature biochip system for detection of aerosolized *Bacillus globigii* spores. *Anal Chem,* 75(2), pp. 275-80.
Taitt, C.R., Malanoski, A.P., Lin, B., Stenger, D.A., Ligler, F.S., Kusterbeck, A.W., Anderson, G.P., Harmon, S.E., Shriver-Lake, L.C., Pollack, S.K., Lennon, D.M., Lobo-Menendez, F., Wang, Z. & Schnur, J.M. (2008). Discrimination between biothreat agents and 'near neighbor' species using a resequencing array. *FEMS Immunol Med Microbiol,* 54(3), pp. 356-64.
Takahashi, H., Keim, P., Kaufmann, A.F., Keys, C., Smith, K.L., Taniguchi, K., Inouye, S. & Kurata, T. (2004). *Bacillus anthracis* incident, Kameido, Tokyo, 1993. *Emerg Infect Dis,* 10(1), pp. 117-20.
Tebbe, C.C. & Vahjen, W. (1993). Interference of humic acids and DNA extracted directly from soil in detection and transformation of recombinant DNA from bacteria and a yeast. *Appl Environ Microbiol,* 59(8), pp. 2657-65.
Thomas, M.C., Shields, M.J., Hahn, K.R., Janzen, T.W., Goji, N. & Amoako, K.K. (2013). Evaluation of DNA extraction methods for *Bacillus anthracis* spores isolated from spiked food samples. *J Appl Microbiol,* 115(1), pp. 156-62.
Turell, M.J. & Knudson, G.B. (1987). Mechanical transmission of *Bacillus anthracis* by stable flies (*Stomoxys calcitrans*) and mosquitoes (*Aedes aegypti* and *Aedes taeniorhynchus*). *Infect Immun,* 55(8), pp. 1859-61.
Turnbull, P.C.B., Sirianni, N.M., LeBron, C.I., Samaan, M.N., Sutton, F.N., Reyes, A.E. & Peruski, L.F. (2004). MICs of Selected Antibiotics for *Bacillus*

anthracis, *Bacillus cereus*, *Bacillus thuringiensis*, and *Bacillus mycoides* from a Range of Clinical and Environmental Sources as Determined by the Etest. *Journal of clinical microbiology*, 42(8), pp. 3626-3634.

U.S. Congress, O.o.T.A. (1993). *Proliferation of Weapons of Mass Destruction: Assessing the Risks*, *OTA-ISC-559*. Washington, DC: U.S. Government Printing Office.

Uchida, I., Sekizaki, T., Hashimoto, K. & Terakado, N. (1985). Association of the encapsulation of *Bacillus anthracis* with a 60 megadalton plasmid. *J Gen Microbiol*, 131(2), pp. 363-7.

US Department of Agriculture, V.S. *(2007) Differentiation of Naturally Occurring From Non-Naturally Occurring Epizootics of Anthrax in Livestock Populations*. http://www.aphis.usda.gov/animal_health/emergingissues/downloads/finalanthraxnaturalterror.pdf [July 1st].

US Department of Agriculture, V.S. *(2014) National Select Agent Registry website*. www.selectagents.gov [July 3].

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(5), p. e461.

Watson, A. & Keir, D. (1994). Information on which to base assessments of risk from environments contaminated with anthrax spores. *Epidemiol Infect*, 113(3), pp. 479-90.

WHO (2008). *Anthrax in humans and animals – 4th ed. I.World Health Organization. II .Food and Agriculture Organization of the United Nations. III .World Organisation for Animal Health.*: WHO Press, World Health Organization, 20 Avenue Appia, 1211 Geneva 27, Switzerland.

Wielinga, P.R., de Heer, L., de Groot, A., Hamidjaja, R.A., Bruggeman, G., Jordan, K. & van Rotterdam, B.J. (2011a). Evaluation of DNA extraction methods for *Bacillus anthracis* spores spiked to food and feed matrices at biosafety level 3 conditions. *Int J Food Microbiol*, 150(2-3), pp. 122-7.

Wielinga, P.R., Hamidjaja, R.A., Agren, J., Knutsson, R., Segerman, B., Fricker, M., Ehling-Schulz, M., de Groot, A., Burton, J., Brooks, T., Janse, I. & van Rotterdam, B. (2011b). A multiplex real-time PCR for identifying and differentiating *B. anthracis* virulent types. *Int J Food Microbiol*, 145 Suppl 1, pp. S137-44.

Vogler, A.J., Busch, J.D., Percy-Fine, S., Tipton-Hunton, C., Smith, K.L. & Keim, P. (2002). Molecular analysis of rifampin resistance in *Bacillus anthracis* and *Bacillus cereus*. *Antimicrobial agents and chemotherapy*, 46(2), pp. 511-3.

Xu, D. & Cote, J.C. (2003). Phylogenetic relationships between *Bacillus* species and related genera inferred from comparison of 3' end 16S rDNA and 5' end 16S-23S ITS nucleotide sequences. *Int J Syst Evol Microbiol*, 53(Pt 3), pp. 695-704.

Young, J.A. & Collier, R.J. (2007). Anthrax toxin: receptor binding, internalization, pore formation, and translocation. *Annu Rev Biochem*, 76, pp. 243-65.

Zoueki, C.W., Tufenkji, N. & Ghoshal, S. (2010). A modified microbial adhesion to hydrocarbons assay to account for the presence of hydrocarbon droplets. *J Colloid Interface Sci*, 344(2), pp. 492-6.

Ågren, J., Finn, M., Bengtsson, B. & Segerman, B. (2014). Microevolution during an Anthrax outbreak leading to clonal heterogeneity and penicillin resistance. *PLoS One*, 9(2), p. e89112.

Ågren, J., Sundström, A., Håfström, T. & Segerman, B. (2012). Gegenees: fragmented alignment of multiple genomes for determining phylogenomic distances and genetic signatures unique for specified target groups. *PLoS One*, 7(6), p. e39107.

9 Acknowledgements

The studies were carried out at the Department of Bacteriology at the National Veterinary Institute, Uppsala and at the Department of Biomedical Sciences and Veterinary Public Health at the University of Agricultural Sciences, Uppsala. The studies were financially supported by the Swedish Civil Contingencies Agency and the EU-projects Biotracer and AniBioThreat (Grant agreement: HOME2009/ISEC/AG/191) with support from the Prevention of and Fight against Crime Programme of the European Union European Commission – Directorate-General Home Affairs. This thesis 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.

Many people have helped me and contributed to this thesis under these four years and I would like to express my gratitude to you all. In particular I wish to thank:

Eva 'Puck' Olsson Engvall or **Chefen**, my main supervisor, for telling me which toes not to step on and for the excellent support in everything regarding my work (and life sometimes). It has been a pleasure working with you and I hope we will continue with that.

Bo Segerman for being my closest supervisor and for teaching me about genomes and about having the correct scientific approach. This thesis owes a lot to your supervision. I especially appreciate the fun road trips to Kvismaren where I think we took the world's first sampling-anthrax-selfie and also perhaps the first anthrax-grave-fika.

Supervisor **Sofia Boqvist** for very helpful information about the scientific world and academia. For much input into this thesis and for being a really inspiring person. The backwards presentation you held was awesome!

Supervisor **Rickard Knutsson** for being the reason this thesis work was at all possible. Thanks to you I have had the great opportunity to work in two large EU-projects during my thesis work. Also, thanks for all the good times we've had at meetings across Europe.

Everyone in the AniBioThreat project for three amazing years and valuable input into this work. I treasure the memories of our annual meetings in places like the only piece of forest in the Netherlands to a sami village in Jämtland to Doctor Albans favorite hotel in Budapest. I really hope we can start an AniBioThreat 2.
Camilla Beijer and **Susanna Westerberg** for always taking care of everything. **Oskar Karlsson** for your energy and caring nature that makes you work more on other people's projects and careers than on your own. The anthrax team **Charlotta Löfström, Trine Hansen, Radi Hamidjaja, Sylviane Derzelle, Simon Thierry, Miriam Koene, Ingmar Janse, Peter Wielinga** and **Bart van Rotterdam** for good teamwork and good times.
Radi for all the great times at dodgy bars and beer festivals, for always asking that pain-in-the-ass last question to every presentation you attend and for being a good friend.

Co-author **Martina Fricker** for the work we did together on model spores and the great time we had in Dublin.
Co-authors **Maria Finn** and **Björn Bengtsson** for the professionalism and enthusiasm you brought to our collaborative study. Antimicrobial resistance is one of the biggest problems in the world and I hope we can work together again in the future.
Co-author **Anders Sundström** for the realization of all our ideas about what Gegenees could be.
Co-author **Therese Håfström** for collaboration on the Gegenees-paper and great company in London.

RUB: Catarina, Olga, Paula, Annelie and **Moa** for lab-support and discussions about anthrax, music and everything in between.

Ulrika Allard Bengtsson for teaching me the value of a good risk assessment and for discussions about experiments and safety.

All the support-functions at SVA with special thanks to **Johnny** and **Micke** at IT, **Ulla Ginman** and coworkers for help with deliveries, **Berka, Kjell** and **Masoud** for fixing stuff that breaks.

Marianne Elvander and coworkers at ESS for letting a PhD-student be a part of your outbreak investigations – I've learnt a lot.

Therese Westermark for your assistance with sampling around Viskan and for your enthusiasm about everything regarding anthrax.

Thank you **René Kaden** for guidance to beerhouses and food in Munich and for introducing me to important people at the conference.

Big thanks to **Karin Troell** for fixing us a MiSeq to play with.

Viveca Båverud, head of dept. of Bacteriology, for the chance to perform my doctoral studies here at SVA and be a part of the bacteriology team.

The staff at the Department of Bacteriology and especially the **BSL-3 staff** for saving me material to do research on. **Marie, Yvonne and Joann** for coffee breaks and AW's. **Helena Ljung** for always making me laugh with your fantastically harsh sense of humor. These seven years at SVA would not have been the same without you. **Stefan Jernstedt** for the many hours in the SVA gym and for discussions about Russian literature, existentialism, women, God and leg shaving. **Lennart Melin** for always being positive and your inabilities in being quiet and leaving work before 22.00, which are qualities I find fascinating.

The Bacteriology R&D-section (FoU) for scientific discussions, for 'julbord hos chefen' and for being a great team. Many thanks to **Boel Harbom** and **Marianne Persson** for our DNA-sequencing work and for making sure us more UFO-ish persons have pipette tips and micro tubes.

Has-been-kids and **current kids** at the dept. of Bacteriology - **Robert, Sevinc, Susanne, David, Sara, Mikaela, Markus, Tina, Therese, Heidur, Frostis, Tomas, Anders** for all the laughs, good times and insane coffee break-discussions. **Stefan Börjesson**, although far too old to be called a kid and working for another department, I have enjoyed your company at BKTs Wednesday meetings and the music quizzes at Pipes of Scotland. Your knowledge of schlager songs is impressive.

Many thanks to my parents **Torie** and **Kenneth** for always helping out whenever you can and for your excellent attempts at listening interestedly about my work when I need to get it off my mind. My brother **Niklas** and his family **Stephanie, Theodor, Oliver** and **Wilma** for fun times and playdates. Also, thanks for fixing my car. My sister **Sofia** for always making me laugh and for your general awesomeness that I think is infectious. **Noah** för att du är min bäste lekkamrat och det optimala botemedlet mot att inte vara i nuet.

My beautiful son **Elvin** for giving me a new, far better, perspective at life. My beloved **Hanna** for actually creating more distraction than focus during my PhD-studies and for that I am forever grateful. Life is too short with you as travel company.