Elsevier required licence: © <2019>. This manuscript version is made available under the CC-BY-NC-ND 4.0 license http://creativecommons.org/licenses/by-nc-nd/4.0/
The definitive publisher version is available online at
[https://www.sciencedirect.com/science/article/pii/S0379073818306480?via%3Dihub]

Manuscript Details

Manuscript number FSI_2018_160

Title Bacillus species at the Canberra Airport: A comparison of real-time polymerase chain reaction and massively parallel sequencing for identification

Article type Research Paper

Abstract

Anthrax, caused by the Gram-positive, spore forming bacterium Bacillus anthracis, is a disease with naturally occurring outbreaks in many parts of the world, primarily in domestic and wild herbivores. Due to the movement of people and stock, B. anthracis could, however, be at transportation hubs including airports. The continuous threat to national and international security from a biological agent release, or hoax attack, is a very real concern. Sensitive, robust and rapid methods to identify biological agents, including B. anthracis, and distinguish pathogenic from non-pathogenic species, is an essential cornerstone to national security. The aim of this project was to determine the presence of Bacillus species at the Canberra Airport using two massively parallel sequencing (MPS) approaches and compare with previous results using real-time polymerase chain reaction (qPCR). Samples were collected daily for seven days each month from August 2011 – July 2012 targeting movement of people, luggage and freight into and out of the Canberra Airport. Extracted DNA was analysed using qPCR specific for B. anthracis. A subset of samples was analysed using two MPS approaches. Approach one, using the Ion PGM™ (Thermo Fisher Scientific; TFS) and an in-house assay, targeted the two B. anthracis virulence plasmids (cya and capB genes) and a single conserved region of the 16S rRNA gene. Approach two, using the Ion S5™ (TFS) and the commercial Ion 16S™ Metagenomics Kit (TFS), targeted multiple regions within the bacterial 16S rRNA gene. Overall there was consistency between the two MPS approaches and between MPS and qPCR, however, MPS was more sensitive, particularly for plasmid detection. Whilst the broad-range 16S genomic target(s) used in both MPS approaches in this study was able to generate a metagenomic fingerprint of the bacterial community at the Canberra Airport, it could not resolve Bacillus species beyond the level of the B. cereus group. The inclusion of B. anthracis virulence plasmid targets in the in-house assay did allow for the potential presumptive identifications of pathogenic species. No plasmid targets were in the Ion 16S™ Metagenomics Kit. This study shows the choice of target(s) is key in MPS assay development and should be carefully considered to ensure the assay is fit for purpose, whether as an initial screening (presumptive) or a more specific (but not entirely confirmatory) test. Identification approaches may also benefit from a combination of MPS and qPCR as each has benefits and limitations.

Keywords Bacillus; massively parallel sequencing, real-time polymerase chain reaction; Canberra Airport

Taxonomy Forensic Genetics, Genetics, Non-Human DNA Typing, Security, Terrorism, Biological Evidence

Corresponding Author Dennis McNevin

Corresponding Author's Institution University of Technology Sydney

Order of Authors Michelle Gahan, Sorelle Bowman, Rory Chevalier, Rebecca Rossi, Michelle Nelson, Paul Roffey, Bo Xu, Daniel Power, Dennis McNevin

Suggested reviewers Bruce Budowle

Submission Files Included in this PDF

File Name [File Type]

Highlights formatted for submission.docx [Highlights]

Title page.docx [Title Page (with Author Details)]

Manuscript.docx [Manuscript (without Author Details)]

To view all the submission files, including those not included in the PDF, click on the manuscript title on your EVISE Homepage, then click 'Download zip file'.

Highlights

- Consistency between two MPS approaches and between MPS and qPCR
- Greater detection of *Bacillus* at the Canberra Airport with MPS compared to qPCR
- Selected MPS target(s) generated metagenomic fingerprint of bacterial community
- Selected MPS target(s) could not resolve *Bacillus* beyond the *B. cereus* group
- Choice of target(s) is key to differentiate pathogenic from non-pathogenic species

Bacillus species at the Canberra Airport: A comparison of real-time polymerase chain reaction and massively parallel sequencing for identification

M.E. Gahan[1], S. Bowman[1], R. Chevalier[1], R. Rossi[1], M. Nelson[1], P. Roffey[1,2], B. Xu[3], D. Power[3], D. McNevin[1,4]

[1] National Centre for Forensic Studies, University of Canberra, Bruce ACT, 2617.
michelle.gahan@canberra.edu.au
[2] Specialist Operations-Forensics, Australian Federal Police, PO Box 401 Canberra ACT, 2601
[3] Thermo Fisher Scientific, Victoria, Australia
[4] Centre for Forensic Science, School of Mathematical & Physical Sciences (MaPS), Faculty of Science, University of Technology Sydney, Broadway NSW, 2007. dennis.mcnevin@uts.edu.au

Corresponding author: D. McNevin

Key words: _Bacillus_; massively parallel sequencing, real-time polymerase chain reaction; Canberra Airport.

Conflict of interest: This study was partially funded by Thermo Fisher Scientific. D. Power and B. Xu are employed by Thermo Fisher Scientific. All Thermo Fisher Scientific products are for research use only and not for use in diagnostic procedures.

Bacillus species at the Canberra Airport: A comparison of real-time polymerase chain reaction and massively parallel sequencing for identification

Abstract

Anthrax, caused by the Gram-positive, spore forming bacterium *Bacillus anthracis*, is a disease with naturally occurring outbreaks in many parts of the world, primarily in domestic and wild herbivores. Due to the movement of people and stock, *B. anthracis* could, however, be at transportation hubs including airports. The continuous threat to national and international security from a biological agent release, or hoax attack, is a very real concern. Sensitive, robust and rapid methods to identify biological agents, including *B. anthracis*, and distinguish pathogenic from non-pathogenic species, is an essential cornerstone to national security. The aim of this project was to determine the presence of *Bacillus* species at the Canberra Airport using two massively parallel sequencing (MPS) approaches and compare with previous results using real-time polymerase chain reaction (qPCR).

Samples were collected daily for seven days each month from August 2011 – July 2012 targeting movement of people, luggage and freight into and out of the Canberra Airport. Extracted DNA was analysed using qPCR specific for *B. anthracis*. A subset of samples was analysed using two MPS approaches. Approach one, using the Ion PGM™ (Thermo Fisher Scientific; TFS) and an in-house assay, targeted the two *B. anthracis* virulence plasmids (*cya* and *capB* genes) and a single conserved region of the 16S rRNA gene. Approach two, using the Ion S5™ (TFS) and the commercial Ion 16S™ Metagenomics Kit (TFS), targeted multiple regions within the bacterial 16S rRNA gene.

Overall there was consistency between the two MPS approaches and between MPS and qPCR, however, MPS was more sensitive, particularly for plasmid detection. Whilst the broad-range 16S genomic target(s) used in both MPS approaches in this study was able to generate a metagenomic fingerprint of the bacterial community at the Canberra Airport, it could not resolve *Bacillus* species beyond the level of the *B. cereus* group. The inclusion of *B. anthracis* virulence plasmid targets in the in-house assay did allow for the potential presumptive identifications of pathogenic species. No plasmid targets were in the Ion 16S™ Metagenomics Kit.

This study shows the choice of target(s) is key in MPS assay development and should be carefully considered to ensure the assay is fit for purpose, whether as an initial screening (presumptive) or a more specific (but not entirely confirmatory) test. Identification approaches may also benefit from a combination of MPS and qPCR as each has benefits and limitations.

Key words: *Bacillus*; massively parallel sequencing, real-time polymerase chain reaction; Canberra Airport.

1.0 Introduction
Anthrax, caused by the Gram-positive, spore forming bacterium *Bacillus anthracis*, is a disease with naturally occurring outbreaks in many parts of the world, primarily in domestic and wild herbivores such as sheep, cattle, goats and horses. Outbreaks are due to reservoirs of dormant spores which persist in the environment for decades and germinate into disease causing bacteria upon favourable conditions such as a live host. Within Australia, outbreaks are mainly limited to livestock in grazing regions of Victoria and New South Wales, in a region known as the 'anthrax belt' (Durrheim *et al.*, 2009), however, due to the movement of people and stock, *B. anthracis* could be present outside this region, including at transport hubs like airports.

Natural outbreaks in the human population are uncommon and are generally related to occupational exposure to infected animals or their products. However, in addition to natural outbreaks, *B. anthracis* has gained renewed attention since 2001 as a bioterrorism agent. This was a result of the highly publicised United States (US) letter attacks in which *B. anthracis* (Ames strain) spores were sent through the mail causing 22 cases of anthrax, five deaths and the widespread contamination of the US postal system (Jernigan *et al.*, 2002). Furthermore, there has been a significant increase in reports of suspected hoaxes, ranging from 'everyday' white powders (such as flour and baking powder) to biological hoax agents (such as non-pathogenic species) (Leask *et al.*, 2003). Biological hoax agents are difficult to detect if they share similar phenotypic and genetic characteristics with hazardous biological agents. This is especially the case for the *Bacillus* genus, specifically *B. thuringiensis*, a member of the *B. cereus* group, which is a well-established biological hoax agent, comparable genetically and microscopically to *B. anthracis* (Leask *et al.*, 2003). Bioinsecticides containing *B. thuringiensis* spores such as Dipel (Nature's Way Caterpillar Killer, Yates) are readily available for public purchase.

B. anthracis is distinguished by two virulence factors: one is responsible for the synthesis of the anthrax toxin and the other produces a poly-$_{D}$-glutamic acid capsule. These factors are encoded onto two virulence plasmids, pXO1 and pXO2, respectively (for a review see Koehler, 2009). Assays to detect pathogenic *B. anthracis* and distinguish it from non-pathogenic strains and other members of the *B. cereus* group, need to target both a conserved genomic marker as well as markers for the two virulence plasmids. Two molecular-based methods for the detection of *Bacillus* species are real-time polymerase chain reaction (qPCR) and massively parallel sequencing (MPS).

We have previously reported background levels of *B. anthracis* at the Canberra Airport over the period August 2011 – July 2012 using single-plex qPCR with a *B. anthracis* genomic (*PL3* gene) and two plasmid (*cya* and *cap*B genes) targets (Gahan *et al.*, 2015). A total of 575 samples were collected from areas targeting the movement of people, luggage and freight in and out of the airport. Of these, fourteen samples were *PL3* positive, 24 were positive for *cya* (pXO1) and five for *capB* (pXO2). No samples were positive for all three targets. Limitations of the qPCR assay are that it only allowed detection of bacteria for which the primers were specifically targeted, and it was a single-plex assay so separate reactions (one for each target) had to be run. Furthermore, it may not have been sensitive enough to pick up trace concentrations of bacteria.

MPS has allowed high-throughput genetic fingerprinting of microorganisms responsible for outbreaks and played an important role in biosurveillance and biosecurity (Chin *et al.*, 2011; Grad *et al.*, 2012; Leekitcharoenphon *et al.*, 2016). MPS enables the detection and characterisation of multiple microorganisms in mixtures, whether abundant or trace, intact or degraded. Until recently a major drawback of MPS has been the long run times. However, the development of bench-top platforms with easy to use, cost-efficient and streamlined workflows, have reduced identification times (Salipante *et al.*, 2014; Whiteley *et al.*, 2012). Targeted sequencing of species-specific and even strain-specific markers on bench-top platforms can enable the differentiation of pathogenic from non-pathogenic strains.

The aim of this project was to determine the presence of *Bacillus* species, in a sub-set of samples collected from the Canberra Airport, using two MPS approaches and compare findings to the previous qPCR results. Approach one used the Ion PGM™ (Thermo Fisher Scientific (TFS)) and an in-

house assay targeting the two *B. anthracis* virulence plasmids (*cya* and *capB* genes) and a conserved region of the 16S rRNA gene. Approach two used the Ion S5™ (TFS) and the commercial Ion 16S™ Metagenomics Kit (TFS) which targeted multiple hypervariable regions of the bacterial 16S rRNA gene.

2.0 Materials and Methods

2.1 Samples

B. anthracis Sterne strain (pXO1^{+} pXO2^{-}) and *B. thuringiensis* subsp. *Kurstaki* were provided by the Australian Federal Police. Bacterial strains were initiated from frozen glycerol stocks containing a 50:50 suspension of nutrient broth media and sterile glycerol stored at -80 °C and cultured in either nutrient broth (Oxoid) or on nutrient agar (Oxoid). Media was sterilised by autoclaving for 20 minutes at 121 °C. Broth cultures were incubated for 18 hours at 37 °C with shaking at 210 – 240 revolutions per minute and agar plates were incubated aerobically at 37 °C for 18 – 20 hours.

Environmental samples were collected and DNA extracted as previously detailed by Gahan *et al.* (2015). Briefly, 575 samples were collected from six different sites within and around the Canberra Airport (Canberra, Australian Capital Territory, Australia) over a 12-month period (August 2011 – July 2012). Collection sites corresponded to the movement of people, luggage and freight (Australian Air Express) arriving and departing the Canberra Airport. A low foot traffic control was also included. DNA was extracted using the QIAamp DNA extraction mini kit (QIAGEN) with an additional bead-beating disruption step using 0.1 mm glass beads (Gahan *et al*., 2015).

Twenty DNA samples that were positive for at least one of the *PL3*, *cya* and *capB* targets reported by Gahan *et al*. (2015) were selected for sequencing using the in-house sequencing assay (MPS 1) and of these, eight were sequenced using the commercial Ion 16S™ Metagenomics Kit (MPS 2) (Table 1).

Table 1: Canberra Airport samples selected for sequencing

Sample name	Description (Date and sample location)	In-house sequencing assay (MPS 1)	Ion 16S™ Metagenomics Kit (MPS 2)
CA9POV20.2	20/02/2012, People out	Yes	Yes
CA9LC20.2	20/02/2012, Low traffic control	Yes	No
CA9PO21.2	21/02/2012, People out	Yes	No
CA9LO21.2	21/02/2012, Luggage out	Yes	No
CA11LO23.4	23/04/2012, Luggage out	Yes	Yes
CA11PO29.4	29/04/2012, People out	Yes	No
CA11LI29.4	29/04/2012, Luggage in	Yes	No
CA9LO20.2	20/02/2012, Luggage out	Yes	Yes
CA9LC24.2	24/02/2012, Low traffic control	Yes	No
CA9LC25.2	25/02/2012, Low traffic control	Yes	Yes
CA9POV21.2	21/02/2012, People out	Yes	Yes
CA13LO21.6	21/06/2012, Luggage out	Yes	No
CA7FI19.12	19/12/2011, Freight in	Yes	Yes
CA13LO20.6	20/06/2012, Luggage out	Yes	Yes
CA9LI20.6	20/06/2012, Luggage in	Yes	No
CA9POV22.2	22/02/2012, People out	Yes	No
CA9LO24.2	24/02/2012, Luggage out	Yes	No
CA9LO25.2	25/02/2012, Luggage out	Yes	Yes
CA11POV29.4	29/04/2012, People out	Yes	No
S56FO20.8	20/08/2011, Freight out	Yes	No

2.2 In-house assay (MPS 1)

2.2.1 PCR enrichment and pooling

Oligonucleotide primers (Table 2) were purchased from GeneWorks. Three regions were targeted, the 16S rRNA conserved genetic region (Antolinos *et al.*, 2012), the *cya* gene on pXO1 and the *capB* gene on pXO2 (Wielinga *et al.*, 2011). Primers were verified using BLAST and Primer-Blast NCBI to ensure specificity to *B. anthracis* for *cya* and *capB* and broad range for the 16S rRNA target. The latter was to ensure detection of a range of *Bacillus* species and *Escherichia coli*.

Table 2: Oligonucleotide primers for polymerase chain reaction enrichment

Primer	Primer sequence (5' → 3')	Gene target	Amplicon length (bp)
cya (F)	AGGTAGATTTATAGAAAAAAACATTACGGG	Oedema factor (pXO1)	150
cya (R)	GCTGACGTAGGGATGGTATT		
capB (F)	AGCAAATGTTGGAGTGATTGTAAATG	Capsule synthesis component B (pXO2)	148
capB (R)	AAAGTAATCCAAGTATTCACTTTCAATAG		
16S rRNA (F)	AATAAGCGTCGTCAGGAACG	16S rRNA	184
16S rRNA (R)	AGCCAGGATCAAACTCTCCA		

Forward (F) and reverse (R) primer sequences with approximate amplicon length in base pairs (bp).

Enrichment PCRs contained 5U/µL MyTaq DNA polymerase and MyTaq Reaction Buffer (QIAGEN), 0.4 µM (16S rRNA and capB) or 0.2 µM (cya) of each forward and reverse primer, 400 ng of bovine serum albumin, up to 1 ng DNA and DNAse free water to a total volume of 50 µL. Positive controls containing a known concentration of *B. anthracis* Sterne strain DNA (85 ng) and *B. thuringiensis* subsp. *Kurstaki* DNA (100 ng) and two negative controls (no-template control (NTC) and TE (10 mM Tris-HCl, 0.1 mM EDTA, pH 8.0)), were included.

PCR enrichment was carried out in a GeneAmp® PCR system 9700 (Applied Biosystems). Reaction conditions comprised of an initial denaturation at 95 °C for 10 minutes followed by 35 cycles of 95 °C for 15 seconds, 55 °C (capB) or 56 °C (cya) or 58 °C (16S rRNA) for 30 seconds and 72 °C for 30 seconds. Amplified products were subjected to a final 72 °C extension for three minutes.

PCR products were visualised using 1 % agarose gels in 1x Tris-acetate-EDTA buffer to confirm correct size and a single product. The DNA yield of each PCR amplification was quantified using PicoGreen® intercalating dye fluorescence (Qubit® 2.0 fluorometer; Invitrogen), according to the manufacturer's protocol.

Equimolar pools of the three amplicons (capB, cya and 16S rRNA) for each sample, in a total volume of 40 µL, were transferred into 1.5 µL LoBind tubes (Eppendorf), homogenised and purified using an AMPure XP bead (Beckman Coulter) clean-up step with MagnaRack™ (TFS). Briefly, 36 µL (0.9 x pooled PCR volume) of AMPure XP beads were homogenised and incubated with the pooled DNA targets at room temperature (RT) for 5 minutes then placed on the MagnaRack™ for 5 minutes. To remove larger DNA fragments, 70 µL of supernatant was transferred to a new 1.5 mL LoBind tube and the beads were discarded. To bind target DNA an additional 64 µL (1.6 x pooled PCR volume) AMPure XP beads were homogenised and incubated with the pooled targets at RT for 5 minutes then placed on the MagnaRack™ for 5 minutes. The supernatant was removed, and the beads were washed twice with 200 µL of 80 % ethanol then dried for 15 minutes. To elute the target DNA, 28 µL of nuclease-free water was added. A total of 25 µL of supernatant was stored at -4 °C.

DNA quantity and quality post-purification was evaluated on the 2100 Bioanalyzer instrument (Agilent Technologies) with the DNA 1000 Bioanalyzer kit and chips (Agilent Technologies) according to the manufacturer's protocol.

2.2.2 Library preparation

Library construction for targeted Ion PGM™ sequencing involved the following QIAGEN workflow: end-repair of PCR products (GeneRead DNA library L core kit), adapter ligation (GeneRead DNA Adapter L set 12 plex kit), DNA purification post-adapter ligation (Appendix B: Library Construction Using the GeneRead Library Prep Kits for Ion PGM™ Sequencer/Proton, from the GeneRead DNAseq Targeted Panels V2 Handbook (2015)), amplification of library (GeneRead DNA L Amp kit) and a final purification step.

For end-repair, 20.5 µL (10 – 200 ng) of purified target DNA is recommended. However, since the DNA concentrations of the purified amplicons were low, these were not diluted prior to end-repair and the maximum volume (25 µL) was used. For adapter-ligation, 12 barcode (Bc) adapters from QIAGEN were utilised to label amplicons and corresponded to sample 1 (Bc 1) – sample 12 (Bc 12) on each of two Ion 314™ semiconductor chips, giving a total of 24 samples.

The 2100 Bioanalyzer with a DNA 1000 kit was employed for quality control after the final purification step. DNA molarity (nM) was measured by integrating the area under the amplicon size range using the 2100 Bioanalyzer smear analysis software according to the manufacturers' protocol. The DNA controls of *B. anthracis* Sterne strain and *B. thuringiensis* subsp. *Kurstaki* were diluted to 8 pM prior to library pooling. Canberra Airport samples were used undiluted because they were all <8 pM. Aliquots of 2.5 µL of each of the 12 purified samples (per chip) were pooled and homogenised for a total of 2 × 30 µL. For template preparation, subsequent aliquots of 2 × 25 µL from the previous pooling step were utilised.

2.2.3 Template preparation

Template preparation was performed on an Ion OneTouch™ 2 (OT2) (TFS) and an Ion OneTouch™ 2 ES (OT-ES; TFS) using the Ion PGM™ Hi-Q OT2 200 kit (TFS) according to the Quick Reference Ion PGM™ Hi-Q Template guide (TFS) from step 2. Enriched Ion Sphere Particles (ISPs) were briefly stored at -4 °C prior to sequencing.

2.2.4 Sequencing

Sequencing was performed on an Ion PGM™ sequencer using two Ion 314™ v2 Chips (TFS) in combination with the Ion PGM™ Hi-Q Sequencing kit chemistry (TFS) according to table 3 and the Quick Reference Ion PGM™ Hi-Q Sequencing guide (TFS). The three target BED files for 16S rRNA, *cya* and *capB* gene regions are located in the supplementary material.

Table 3: Ion PGM™ sequencing conditions

Reference genomes	16S rRNA	BA_genome(Bacillus anthracis str. 'Ames ancestor' complete genome)
	cya gene	BA_pXO1(Bacillus anthracis str. 'Ames ancestor' plasmid pXO1 complete genome)
	capB gene	BA_pXO2(Bacillus anthracis str. 'Ames ancestor' plasmid pXO2 complete genome)
Target regions	16S rRNA	ba_genome(27_4).bed
	cya gene	ba_pxo1(27_4).bed
	capB gene	ba_pxo2(27_4).bed
Target technique		TargetSeq DNA
Template kit		Ion PGM™ Hi-Q OT2 kit – 200
Sequencing kit		Ion PGM™ Hi-Q Sequencing kit
Number of flows		500
Chip type		Ion 314™ Chip v2
Barcode set		GeneRead DNA Adapter L set 12 plex kit
Plugins		Coverage analysis and FileExporter

2.2.5 Data analysis

Sequencing output files were exported in FASTQ format by the FileExporter plugin in Torrent™ Suite Software (TFS).

FASTQC (Andrews, 2010) was utilised to extract read quality control data from the FASTQ files. Reads that passed default parameters were mapped to the genome of *B. anthracis* Ames ancestor (NCBI entries: chromosome NC_007530.2; pXO1 NC_007322.2; pXO2 NC_007323.3) using *Bowtie2* (Langmead *et al.*, 2012). *SAMtools* (Li *et al.*, 2009) was used to convert from SAM to BAM format and to generate variant call files (VCF). Sorted and indexed BAM files were converted to FASTA format by *Picard* tools (Picard, 2014) and *Seqtk* (Li, GitHub, year 2012). The FASTA files were then imported into MATLAB® using the *Bioinformatics* toolbox to assist in graphically visualising the target clusters with stacked bar graphs.

2.3 Ion 16S™ Metagenomics Kit (MPS 2)

Amplicons were prepared using the Ion 16S™ Metagenomics Kit (TFS) according to the manufacturer's user guide (Revision C.0) using 2 µL sample and with PCR cycles increased from 25 to 30. Positive (*E. coli*) and negative controls, included with the kit, were also prepared and sequenced. The kit includes two primer sets (V2-4-8 and V3-6, 7-9) that amplify the corresponding hypervariable regions of the 16S region in bacteria. Libraries were prepared using the Ion Plus Fragment Library Kit (TFS) and quantified using the Ion Universal Library Quantitation Kit (TFS) both according to the manufacturer's protocol. Libraries were diluted to 30 pM and templating and sequencing preformed using an Ion 520™ & 530™Kit-Chef (TFS). Sequencing was done using a single Ion 520™ chip using 850 nucleotide flows. Basecalling and alignment was done using Torrent Suite Software v5.4.

Data was analysed using the 16S Metagenomics workflow within the Ion Reporter Software (TFS). Briefly, the software reports the sequencing results separately for each hypervariable 16S region targeted with the Ion 16S™ Metagenomics Kit. Sequencing identifications are reported from the phylum down to the species level, along with scores for aspects including % identification, % total reads, % mapped reads and total counts. It is important to note the 16S™ Metagenomics Kit can only report identification down to the *B. cereus* group level which is given as *Bacillus anthracis/Bacillus cereus/Bacillus mycoides/Bacillus thuringiensis/Bacillus weihenstephanensis*.

3.0 Results

3.1 In-house assay (MPS 1)

Using agarose gel electrophoresis, the optimised 16S rRNA assay was found to detect PCR products from *B. anthracis* Sterne strain, *B. thuringiensis* subsp. *Kurstaki*, *B. thuringiensis* subsp. *Israelensis*, *B. cereus*, *B. subtilis* and *E. coli* (data not shown). The cya assay was specific for the intended plasmid target for *B. anthracis* Sterne strain, showing no cross-reactivity amongst closely related *Bacillus* species and *E. coli* (data not shown). As pathogenic strains of *B. anthracis* (pXO1$^+$ pXO2$^+$) are highly regulated Tier one agents, this study used a non-pathogenic model strain of Sterne (pXO1$^+$ pXO2$^-$). However, the capB assay targets the *capB* gene, which cannot be detected in pXO2-negative strains. Therefore, whilst the capB assay was negative for the bacterial species tested (data not shown), a limitation was there was no positive control for capB.

The PCR enriched targets were quantitated using fluorometry (Qubit® 2.0 fluorometer) and ranged in concentration from 2.39 – 13.6 ng/µL for the Canberra Airport samples. No product was detected for the NTC or TE controls.

Prior to the library preparation steps, pooled DNA quantity and quality post-purification was evaluated by microfluidic capillary electrophoresis on the Bioanalyzer. Electropherograms of the *B. thuringiensis* and *B. anthracis* Sterne strain (pXO1$^+$ pXO2$^-$) controls produced peaks at 184 base pairs (bp) and 150 bp indicating the presence of the 16S rRNA and *cya* (pXO1) gene targets, respectively (Figure 1).

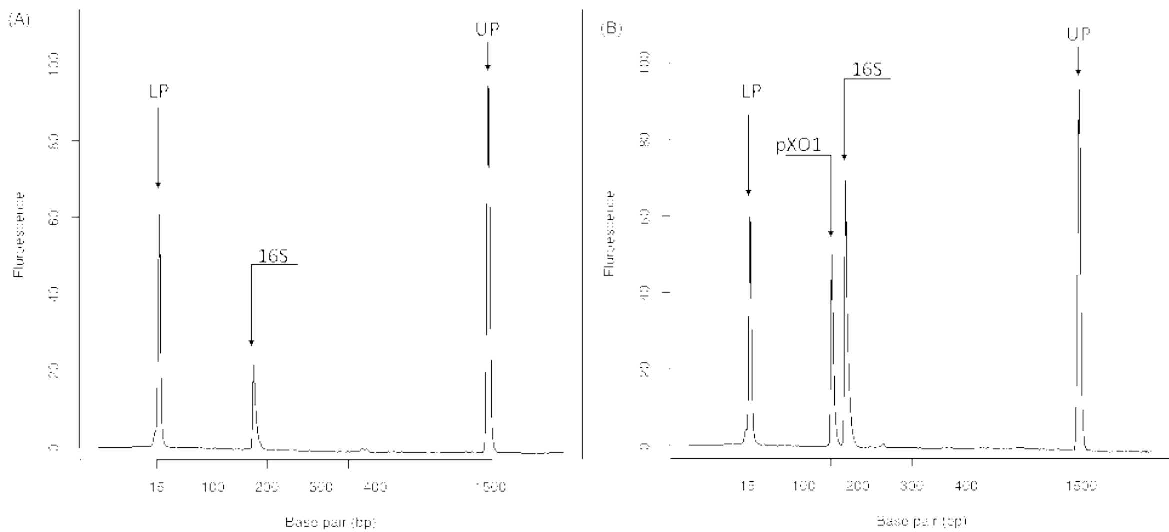

Figure 1: Electropherograms of *B. thuringiensis* and *B. anthracis* controls post-target pooling
Plots display (A) *B. thuringiensis* subsp. *Kurstaki* and (B) *B. anthracis* Sterne strain ($pXO1^{+}$ $pXO2^{-}$) controls post-target pooling. LP = lower ladder peak; UP = upper ladder peak; 16S = peak indicating 16S rRNA gene; pXO1 = peak indicating *cya* gene harboured on pXO1.

Electropherograms of the Canberra Airport samples were found to contain low DNA quantities and hence low fluorescent intensity (data not shown). For this reason, the electropherograms of both *B. thuringiensis* and *B. anthracis* Sterne strain controls were evaluated for the successful ligation of the 60 bp adapter. Electropherograms performed post-library preparation resulted in controls with peaks at 245 bp and 210 bp, indicating a presence of the 16S rRNA and *cya* genes with the additional adapter, respectively (Figure 2). Furthermore, there was no dominant peak below 100 bp suggesting no adapter dimer. DNA concentration was measured by smear analysis from the expected range of 100 – 300 bp and the *B. anthracis* and *B. thuringiensis* controls had concentrations of 2.50 ng/µL and 0.44 ng/µL, respectively. The DNA concentrations of the Canberra Airport samples were all less than 0.28 ng/µL.

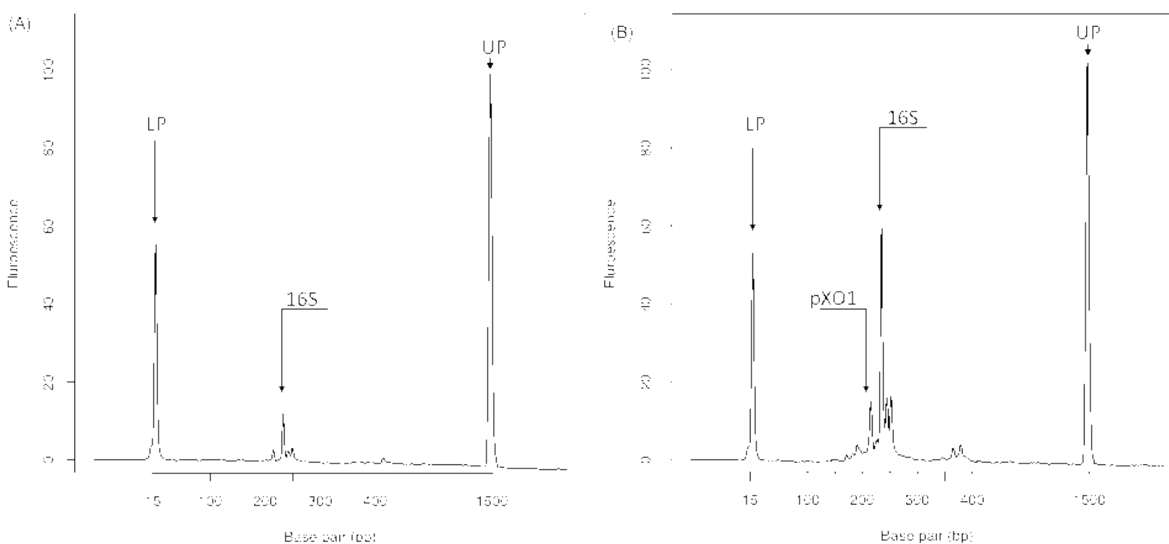

Figure 2: Electropherograms of *B. thuringiensis* and *B. anthracis* controls post-library preparation
Plots display (A) *B. thuringiensis* subsp. *Kurstaki* and (B) *B. anthracis* Sterne strain ($pXO1^{+}$ $pXO2^{-}$) controls post-library construction. LP = lower ladder peak; UP = upper ladder peak; 16S = peak indicating 16S rRNA gene; pXO1 = peak indicating *cya* gene harboured on pXO1.

Chip 1 and chip 2 were loaded to 73 % and 74 % capacity, respectively. For chip 1, 912,207 library ISPs were enriched, of which 16.4 % (or 150,040) produced usable reads. For chip 2, a total of 924,698 ISP loading was reported and from these 4.9 % (or 45,345) produced reads that could be utilised for downstream applications. The usable read percentage was calculated by the Torrent Suite™ after the ISPs were filtered for being of low quality (chip 1 = 20.8 %; chip 2 = 11.3 %), polyclonal (chip 1 = 54 %; chip 2 = 80.2 %) or exhibiting adapter dimer (chip 1 = 8.2 %; chip 2 = 1.6 %). For both chips, read lengths ranged from 148-184 bp which was consistent with the expected read lengths of 184 bp (16S rRNA), 150 bp (*cya*) and 148 bp (*capB*). A large number of reads were below 50 bp, indicating adapter dimer.

Individual sample FASTQ files were imported into the Mac OS X Terminal and underwent quality control measures prior to *Bowtie2* alignment. An alignment of reads from the positive controls to the *B. anthracis* Ames ancestor genome reference (NC_007530.2) resulted in a high abundance of 16S rRNA reads for both *B. thuringiensis* subsp. *Kurstaki* and *B. anthracis* Sterne strain (Figure 3). The latter also had reads that mapped to the *cya* gene harboured on pXO1. Conversely, the negative controls (TE buffer and NTC) returned no read alignment. A total of 16 of the 20 (80 %) Canberra Airport samples were positive for the *B. cereus* group target of 16S rRNA across chip 1 (Figure 3A) and chip 2 (Figure 3B). These samples were collected in December 2011 and February, April and June 2012. There was a trend towards a higher abundance of 16S rRNA reads at the Canberra Airport in the luggage out (LO) samples and the Australian Air Express freight departing (FO). Nine of the 20 (40 %) Canberra Airport samples aligned to the *B. anthracis* Ames ancestor pXO1 reference (NC_007322.2) and all were collected in February 2012. Of the samples positive for *cya*, there was a higher abundance and occurrence in two different sites for people leaving (PO and POV) the airport. A total of 15 of the 20 (70 %) samples aligned to the *B. anthracis* Ames ancestor pXO2 reference (NC_007323.3) with two of these collected from the Australian Air Express site (freight departing (FO) and arriving (FI)) in August and December 2011. The sampling sites with the highest incidence of *cya* alignment across all months were people (PO and POV) and luggage (LO) departing the Canberra Airport. Five samples were positive for all three targets (16S rRNA, *cya* and *capB*). These were collected in February 2012 and comprised of two samples from luggage departing and one sample from each of the people departing and low traffic control (LC) sites. As a consequence of mapping a moderate set of reads to a small set of targets, a high proportion of the data remained unmapped, particularly for the controls of NTC and TE buffer (100 %) and CA9LI20.6 (99 %).

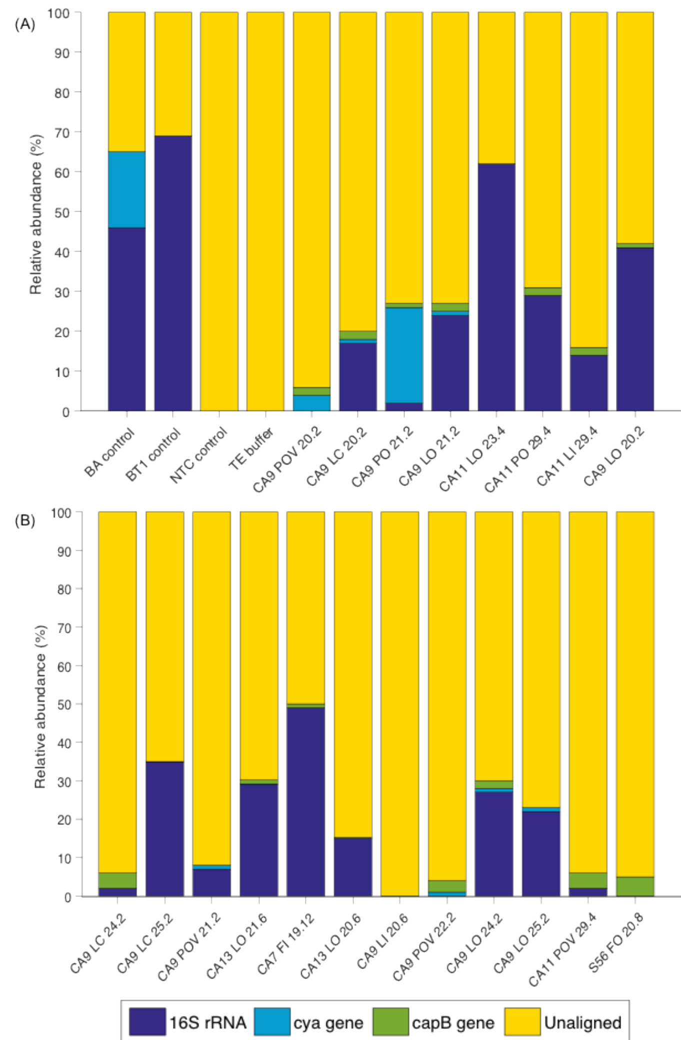

Figure 3: Relative abundance and bacterial target composition of the Canberra Airport samples and controls

Plots display relative abundance of sample sequence reads in chip 1 (A) and chip 2 (B) that aligned to specific targets as well as unaligned reads. The x-axis represents the individual samples and the y-axis the percentage of reads reported. BA = *B. anthracis* Sterne strain; BT1 = *B. thuringiensis* subsp. *Kurstaki*; NTC = no-template control; TE = Tris-EDTA buffer solution. Colours are based on the three targets, navy blue = 16S rRNA gene; light blue = *cya* gene on pXO1; green = *capB* gene on pXO2; yellow = unaligned reads.

3.2 Ion 16S™ Metagenomics Kit (MPS 2)

For MPS 2, the Ion 520 chip was loaded to 93 % capacity with 100 % enrichment. Of the 11,603,369 total reads 4,618,388 (40 %) were considered usable. The usable read percentage was calculated by the Torrent Suite™ after the ISPs were filtered for being of low quality (28.3 %), polyclonal (31.8 %) or exhibiting adapter dimer (0.1 %). The mean, median and mode read lengths were 241 bp, 261 bp and 289 bp, respectively.

For the positive (*E. coli*) control, a total of 91 identifications were reported (across all primer targets) with all being members of the *Enterobacteriaceae* family. For the negative control a total of 168 sequence identifications were reported (across all primer targets) with these predominantly comprising of members of the *Proteobacteria* and *Firmicutes* phylum with smaller numbers of members of *Cyanobacteria*, *Bacteroidetes* and *Actinobacteria*. The top three identifications for the negative control at the species level, in terms of % of mapped reads, were *Corynebacterium tuberculostearicum* (27.94 %), *Propionibacterium acnes* (7.79 %) and *Cupriavidus/Ralstonia sp.* (6.76 %). At the species level for the negative control there was only one report for *Bacillus* and this was with primer V2 for *Bacillus mycoides/Bacillus weihenstephanensis* at 0.25 % of mapped reads.

For the Canberra Airport samples the number of sequencing identifications ranged from 862 for CA7FI19.12 up to 2643 for CA9LC25.2 (Table 4). The average number of phyla identified was 15 with a minimum of 11 (CA0POV21.2) and a maximum of 21 (CA11LO23.4). At the genus/family level there was a variation in the predominant bacteria in each sample, however, across all eight samples, *Staphylococcus*, *Nostocaceae* and *Propionibacterium* were the most predominant bacteria appearing in the top three identifications for five, four and four of the eight Canberra Airport samples, respectively (Table 4).

Table 4: Summary of sequencing results for Ion 16S™ Metagenomics Kit

Sample	Total number of ID	Number of phyla	Top 3 ID (% mapped reads)
CA9LO20.2	2091	15	*Nostocaceae* (F, 2.55) *Staphylococcus* (G, 2.08) *Propionibacterium* (G, 1.41)
CA9LC25.2	2643	17	*Propionibacterium* (G, 3.2) *Nostocaceae* (F, 2.06) *Propionibacterium* (G, 1.64)
CA9POV20.2	1237	12	*Methylobacterium* (G, 8.84) *Methylobacterium* (G, 8.84) *Methylobacterium* (G, 6.75)
CA9LO25.2	1364	13	*Nostocaceae* (F, 3.2) *Staphylococcus* (G, 2.69) *Rickettsiella* (G, 2.61)
CA9POV21.2	1419	11	*Staphylococcus* (G, 3.41) *Acinetobacter* (G, 2.44) *Nostocaceae* (F, 2.06)
CA7FI19.12	862	14	*Pseudomonas* (G, 2.96) *Acinetobacter/Arsenophonus* (G, 2.52) *Enterococcus* (G, 1.56)
CA11LO23.4	2582	21	*Staphylococcus* (G, 3.36) *Rickettsiella* (G, 1.68) *Propionibacterium* (G, 1.58)
CA13LO20.6	1845	19	*Staphylococcus* (G, 5.13) *Propionibacterium* (G, 2.76) *Corynebacterium* (G, 1.35)

Results are reported across all primer targets down to the genus (G) or the family (F) where genus level resolution was not reported. ID = identifications.

To enable a comparison with MPS 1, the data from MPS 2 was mined for all *Bacillus* species. At the genus level *Bacillus* was identified in all eight Canberra Airport samples ranging from 0.73 % (CA9LO25.2) to 2.1 % (CA9LO20.2) of all identifications (Table 5). The *B. cereus* group (*B. anthracis/B. cereus/B. mycoides/B. thuringiensis/B. weihenstephanensis*) was also identified in all eight samples and, with the exception of CA9POV20.2 and CA9POV21.2, was within the top three *Bacillus* identifications. For all eight samples *B. decisifrondis* was identified in the top three *Bacillus* identifications (Table 5).

Table 5: Summary of *Bacillus* sequencing results for Ion 16S™ Metagenomics Kit

Sample	ID to *Bacillus* genus	*Bacillus* ID as a percentage of all ID	Top 3 *Bacillus* species (% mapped reads)
CA9LO20.2	44	2.1	1. *B. anthracis* / *B. cereus* / *B. mycoides* / *B. thuringiensis* / *B. weihenstephanensis* (0.23) 2. *B. decisifrondis* (0.13) 3. *B. aquimaris* / *B. carboniphilus* / *B. mojavensis* / *B. pseudalcaliphilus* / *Bacillus sp. DSM* 8802 / *B. subtilis* / *B. tequilensis* (0.10)
CA9LC25.2	25	0.95	1. *B. anthracis* / *B. cereus* / *B. mycoides* / *B. thuringiensis* / *B. weihenstephanensis* (0.15) 2. *B. aquimaris* / *B. carboniphilus* / *B. mojavensis* / *B. pseudalcaliphilus* / *Bacillus sp. DSM* 8802 / *B. subtilis* / *B. tequilensis* (0.12) 3. *B. decisifrondis* (0.10)
CA9POV20.2	10	0.80	1. *B. decisifrondis* (0.17) 2. *B. mojavensis* / *Bacillus sp. DSM* 8802 (0.07) 3. *B. aquimaris* / *B. carboniphilus* / *B. mojavensis* / *B. pseudalcaliphilus* / *Bacillus sp. DSM* 8802 / *B. subtilis* / *B. tequilensis* (0.07)
CA9LO25.2	10	0.73	1. *B. decisifrondis* (0.22) 2. *B. anthracis* / *B. cereus* / *B. mycoides* / *B. thuringiensis* / *B. weihenstephanensis* (0.11) 3. *B. anthracis* / *B. cereus* / *B. mycoides* / *B. thuringiensis* / *B. weihenstephanensis* (0.04)
CA9POV21.2	13	0.92	1. *B. decisifrondis* (0.23) 2. *B. clausii* / *B. rhizosphaerae* (0.05) 3. *B. rhizosphaerae* (0.03)
CA7FI19.12	10	1.16	1. *B. anthracis* / *B. cereus* / *B. mycoides* / *B. thuringiensis* / *B. weihenstephanensis* (0.47) 2. *B. cereus* / *B. thuringiensis* (0.27) 3. *B. anthracis* / *B. cereus* / *B. mycoides* / *B. thuringiensis* / *B. weihenstephanensis* (0.22)
CA11LO23.4	30	1.16	1. *B. decisifrondis* (0.22) 2. *B. anthracis* / *B. cereus* / *B. mycoides* / *B. thuringiensis* / *B. weihenstephanensis* (0.15) 3. *B. cereus* / *B. thuringiensis* (0.06)
CA13LO20.6	20	1.08	1. *B. decisifrondis* (0.35) 2. *B. anthracis* / *B. cereus* / *B. mycoides* / *B. thuringiensis* / *B. weihenstephanensis* (0.06) 3. *B. flexus* / *B. megaterium* (0.05)

ID = identifications

3.3 Comparison of sequencing and real-time PCR for identification of *Bacillus* species

With the exception of CA9POV20.2, which was detected by MPS 2 only, albeit not within the top three *Bacillus* identifications, both MPS approaches consistently identified the *B. cereus* group in the eight common samples (Table 6). For the chromosomal target there was 70 % (14/20 samples) and 63 % (5/8 samples) consistency between MPS 1 and qPCR and between MPS 2 and qPCR, respectively. Three samples (CA9POV21.2, CA7FI19.12 and CA13LO20.6) were MPS positive (both approaches) and qPCR negative. Two samples (CA9LI20.6 and CA9POV22.2) were positive by qPCR but negative by MPS 1 and not tested using MPS 2 (Table 6).

Overall, MPS 1 was able to detect a higher incidence of plasmid markers in the Canberra Airport samples, compared with qPCR (Table 6). For cya, MPS 1 detected the target in an additional four samples, although for two samples, CA9LC24.2 and CA9LC25.2, MPS did not detect cya which contradicts the qPCR. For capB, MPS detected the target in an additional 12 samples compared to qPCR but in contrast did not detect the target in CA9POV21.1 (Table 6). It should be noted the additional MPS 1 plasmid detections were at low abundance, suggesting they were below qPCR quantifiable levels.

Table 6: Comparison of real-time polymerase chain reaction (qPCR) and massively parallel sequencing (MPS) identifications

	pXO1			pXO2			Chromosome		
	cya			*capB*			*PL3*	16S rRNA (*B. cereus* group)	16S rRNA (*B. cereus* group)
	qPCR	MPS 1	MPS 2	qPCR	MPS 1	MPS 2	qPCR	MPS 1	MPS 2
CA9POV20.2	+	+	nt	-	+	nt	+	-	+
CA9LC20.2	+	+	nt	-	+	nt	+	+	nt
CA9PO21.2	+	+	nt	-	+	nt	+	+	nt
CA9LO21.2	+	+	nt	-	+	nt	+	+	nt
CA11LO23.4	-	-	nt	-	-	nt	+	+	+
CA11PO29.4	-	-	nt	-	+	nt	+	+	nt
CA11LI29.4	-	-	nt	-	+	nt	+	+	nt
CA9LO20.2	-	-	nt	-	+	nt	+	+	+
CA9LC24.2	+	-	nt	-	+	nt	+	+	nt
CA9LC25.2	+	-	nt	-	-	nt	+	+	+
CA9POV21.2	+	+	nt	+	-	nt	-	+	+
CA13LO21.6	-	+	nt	-	+	nt	+	+	nt
CA7FI19.12	-	-	nt	+	+	nt	-	+	+
CA13LO20.6	-	-	nt	+	+	nt	-	+	+
CA9LI20.6	-	-	nt	-	-	nt	+	-	nt
CA9POV22.2	-	+	nt	-	+	nt	+	-	nt
CA9LO24.2	-	+	nt	-	+	nt	+	+	nt
CA9LO25.2	-	+	nt	-	-	nt	+	+	+
CA11POV29.4	-	-	nt	-	+	nt	+	+	nt
S56FO20.8	-	-	nt	+	+	nt	-	-	nt

qPCR = target detected through qPCR as reported by Gahan *et al.* (2015); MPS 1 = in-house assay; MPS 2 = Ion 16S™ Metagenomics Kit; nt = not tested; - = not detected; + = detected.

4.0 Discussion

The continuous threat to national and international security from a biological agent release, or a hoax attack, is a very real concern. Sensitive, robust and rapid methods to identify a broad range of biological agents, including *B. anthracis*, and distinguish pathogenic from non-pathogenic species, is an essential cornerstone to national security. *B. anthracis* occurs naturally in soil. In Australia although infections primarily occur within anthrax belt, cases of natural outbreaks have been reported outside of this region, possibly due to populace, freight and vehicle travel. As such, it is of high importance to determine the background frequency of *B. anthracis* at transport locations, including airports. Furthermore, with a range of detection assays available, it is important to assess the applicability of each and determine the most appropriate for the situation. We have previously reported background levels of *B. anthracis* at the Canberra Airport over the period August 2011 – July 2012 using single-plex qPCR with a *B. anthracis* genomic (*PL3* gene) and two plasmid (*cya* and

*cap*B genes) targets (Gahan *et al*., 2015). The aim of this study was to use two MPS approaches, on a sub-set of the Canberra Airport samples, and compare findings to the previous qPCR results.

The qPCR results showed a peak of positive *B. anthracis* markers in February 2012, with secondary peaks in April and June 2012. Eleven of the 20 samples selected for the in-house assay were from February 2012, four were from April and three were from June 2012. The majority were from the luggage departing sampling area. The five samples identified by Gahan *et al*. (2015) as *B. anthracis* positive for the genomic marker *PL3* and *cya* gene on pXO1 were included in this study. Two virulence plasmids (*cya* and *capB* genes) and a conserved *B. cereus* group (16S rRNA gene) marker were targeted with the in-house assay which allowed for the presumptive identification of pathogenic and non-pathogenic *B. anthracis*, however, it must be noted that, in contrast to the qPCR chromosomal target (*PL3*), the 16S rRNA target was not specific for *B. anthracis* and can only resolve identification to the *B. cereus* group. This allowed us to survey a broader spectrum of bacteria. Eight of these 20 samples were also sequenced using the commercial Ion 16S™ Metagenomics Kit which targeted multiple hypervariable regions of the bacterial 16S region, and, like the in-house assay, could only resolve identification to the *B. cereus* group.

For MPS 1, chips were loaded to 73 % (chip 1) and 74 % (chip 2) capacity with 16.4 % (chip 1) and 4.9 % (chip 2) reported as usable. The greatest loss of potential reads was due to polyclonality (54 % chip 1 and 80.2 % chip 2) implying that more than one unique amplicon attached to each ISP. For MPS 2 there was a higher percentage (40 %) of useable reads but there was still a high polyclonal percentage (31 %). High polyclonality is generally caused by overloading the ISPs and may have been the result of inaccurate (underestimated) library quantitation before pooling.

Both MPS 1 negative controls (NTC and TE) received significant coverage (352 and 626 mapped reads, respectively) and bacterial DNA was also detected in the negative control for MPS 2. The top three identifications for the MPS 2 negative control at the species level were *C. tuberculostearicum*, *P. acnes* and *Cupriavidus/Ralstonia sp.* which are all skin contaminants and not surprising given the high number of skin cells shed by individuals. While none of the reads in negative controls mapped to the *B. anthracis* targets, the presence of bacterial amplicons demonstrates the difficulty in removing all bacterial DNA from water and buffers and hence the importance of including negative controls in MPS assays using broad 16S rRNA targets.

With the exception of CA9POV20.2, which was detected by MPS 2 only, both MPS approaches consistently identified the *B. cereus* group in the eight common samples (Table 6). Although it could be hypothesized that this difference is due to the different sequencers used (Ion PGM for MPS 1 and Ion S5 for MPS 2) a more likely explanation is the increased number of primer targets for MPS 2. In contrast to MPS 1 which targeted only a single chromosomal target, MPS 2 has six chromosomal targets which is likely to increase the sensitivity and broad range of the assay to detect bacteria. Whilst for both assays the chromosomal target(s) could only resolve identification to the *B. cereus* group, a further limitation of MPS 2 is that, in contrast to MPS 1, it does not contain targets for the *B. anthracis* virulence plasmids so cannot provide any information on pathogenicity.

For the chromosomal target there was 70 % (14/20 samples) and 63 % (5/8 samples) consistency between MPS 1 and qPCR and between MPS 2 and qPCR, respectively. Samples which were positive by both MPS and qPCR can be reported as positive for *B. anthracis* however, based on the chromosomal target alone, it is unknown whether this is pathogenic or non-pathogenic. Three samples (CA9POV21.2, CA7FI19.12 and CA13LO20.6) were MPS positive (both approaches) and qPCR negative suggesting these samples contain a member of the *B. cereus* group other than *B. anthracis*.

Overall MPS 1 was able to detect a higher incidence of *B. anthracis* plasmid markers in Canberra Airport samples, particularly *capB*, compared with qPCR, and has detected both plasmid markers in five samples (Table 6), along with a 16S rRNA *B. cereus* target. This is likely due to the sensitivity of MPS which can theoretically identify a single DNA molecule in a noisy background. Whilst this could suggest pathogenic *B. anthracis* may be present, without culture there is no way to confirm that all three targets came from the same live cell and in fact that this is pathogenic *B. anthracis* as there

have also been reported cases of pXO-like plasmids in other *Bacillus* species (Luna *et al*. 2006; Pannucci *et al*. 2002; Wright *et al*. 2011).

There were cases where samples previously reported as positive by qPCR were found to be negative by MPS 1. These were CA9LI20.6 and CA9POV22.2 for the chromosomal target, CA9LC24.2 and CA9LC25.2 for *cya* and CA9POV21.2 for *capB*. The DNA used for MPS was stored since 2012 and during the optimisation of MPS target enrichment and subsequent library preparation, underwent multiple freeze-thaw cycles. As a result, there may have been a loss in concentration and as such the targets may have degraded. These samples were not tested with MPS 2.

The MPS 1 and MPS 2 assays target broad 16S rRNA target(s) instead of the *PL3 B. anthracis* genomic target used in the qPCR. Whilst this means the MPS is not specifically targeted to the detection of *B. anthracis* it does provide for a broader range screening approach. This could be seen with the MPS 2 results (Table 4) in which the top three identifications were commonly *Staphylococcus*, *Nostocaceae* and *Propionibacterium* species but also included other species such as *Pseudomonas* and *Rickettsiella*. These findings are not surprising, given many of these are members of the normal flora commonly found in the environment, and a broad range target allows for greater information to be obtained from samples and a metagenomic fingerprint to be created. Although the MPS 1 data was only mined for *Bacillus* there was a high number of unmapped reads (i.e. not specific to *Bacillus*) so further work is recommended to focus on this data and compare the metagenomic fingerprint of both sequencing approaches. This will further contribute to an understanding of the overall bacterial composition of the Canberra Airport and enable additional conclusions to be drawn regarding the sensitivity and detection ability of both MPS assays.

Monthly samples collected by Gahan *et al.* (2015) at several locations from porous and non-porous surfaces at the Canberra Airport and Australian Air Express freight between August 2011 - July 2012, allowed for a detailed analysis of seasonal and locational variation. The key finding was a higher incidence of positive *B. anthracis* targets in February, April and June 2012. This finding was supported by MPS. Of the 15 samples positive for the *B. cereus* group by MPS 1, two samples were also pXO1-positive in February 2012 and seven were also pXO2-positive (December 2011 and June 2012). A total of five samples aligned to all three markers, of which four were collected in February 2012. As discussed in Gahan *et al.* (2015), Queensland and NSW experienced flooding during December 2011 – January 2012, which may have caused a peak in *B. anthracis*. Environmental and climatic factors have a great influence on the ecology of *B. anthracis*, especially rainfall and temperature (Blackburn *et al.*, 2007). Daytime temperatures in February 2012 at the Canberra Airport were above the average at 30 – 31 °C with moderate winds (2 – 24 km/h) and relatively high humidity (approximately 74 %) (Time and Date, 2012). Another likely scenario may be increased bacterial prevalence due to heightened foot traffic in February associated with returning passengers from Christmas and school holidays or passengers celebrating the Lunar New Year period from late January to early February (Freed, 20 March 2015).

Due to the low number of samples sequenced no conclusions can be drawn from MPS regarding whether the positive samples were associated with the movement of luggage, people and/or freight in and/or out of the airport. Interestingly, one of the low traffic control samples which is situated under an escalator, and so away from foot traffic and less likely to be cleaned regularly, was positive for all three markers by MPS. This suggests that once bacteria enter the airport they disperse, possibly due to the use of the heating–ventilation–air-conditioning (HVAC) units in operation throughout the building interior. There have been a number of studies which have demonstrated the dispersal and migration of microorganisms by people, air movement, HVAC and electrostatic forces in the indoor urban environment (Li *et al.*, 2007; Prussin *et al.*, 2015; Tringe *et al.*, 2008). Sextro *et al.* (2002) in particular, employed predictive modelling to examine the dispersion of *B. anthracis* spores in a building from a single point of origin, while Van Cuyk *et al.* (2012) used an aerosolised simulant (*B. thuringiensis*) to model the movement from an outdoor release into a building. These dispersal studies highlight that in future work sampling of the air duct surfaces, in conjunction with air sampling, is recommended.

Like any method for bacterial species identification, there are drawbacks associated with MPS. Limiting factors include the cost, time and current restriction of analyses in a laboratory environment. Having said this, sequencing costs are falling and third generation sequencing, such as single-molecule real-time sequencing (SMRT), eliminates the time-consuming steps in template amplification and provides for a field-portable system that can plug into a standard universal serial bus (USB 3.0) port on a computer (for a review see Schadt *et al*., 2010). SMRT platforms have been integrated into systems developed by Oxford Nanopore Technologies (MinION™, SmidgION™ and PromethION™; for a review see Lu *et al*., 2016) and Pacific Biosciences (Sequel system and PacBio RS II; for a review see Rhoads *et al*., 2015). SMRT platforms have already demonstrated their diagnostic capabilities for rapid outbreak management in the field during the Ebola epidemic in Guinea, West Africa (Quick *et al*., 2016).

5.0 Conclusions

Overall, the sensitivity of MPS technology allowed for greater detection of *Bacillus* species at the Canberra Airport compared to qPCR. Whilst the broad-range 16S genomic target(s) used in both MPS approaches was able to generate a metagenomic fingerprint of the bacterial community at the Canberra Airport, unlike the *PL3* target used in the qPCR, it could not resolve *Bacillus* species beyond the level of the *B. cereus* group. The inclusion of *B. anthracis* virulence plasmid targets in MPS 1 did allow for presumptive identification of pathogenic species. These targets were not present in MPS 2. The choice of target(s) is therefore key in MPS assay development and should be carefully considered to ensure the assay is fit for purpose, whether as an initial screening (presumptive) or a more specific test. Identification approaches may also benefit from a combination of MPS and qPCR. A limitation of molecular techniques, including qPCR and MPS, is that they are unable to provide a true confirmation because even if all three *B. anthracis* targets (chromosomal, pXO1 and pXO2) are identified in the one sample, it does not necessarily mean that a live pathogenic strain is present, rather, just the DNA sequences are present. The source may not be cellular in origin and could be dead cells or spores or there may be multiple bacteria present each with a different target. Nevertheless, the presence of all three would provide justification to send the sample for “gold standard” identification by culturing which is time-consuming and requires a high-level containment laboratory.

There is room for further work on the full collection of the 2011 – 2012 Canberra Airport samples. Sequencing of additional samples, and detailed analysis of the full sequencing data generated, would provide a greater understanding of the bacterial community at the Canberra Airport and enhance our understanding of the relationship of bacterial occurrence with: (1) fluctuations in flight volume and foot traffic during the holiday seasons and (2) the effects of the summer climate which draw heightened conditioned air from the HVAC system. Studies could be extended to other transportation hubs, including those within the anthrax belt, to provide a large-scale database of background levels of microorganisms.

6.0 References

1. Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.
2. Antolinos, V., P. S. Fernandez, M. Ros-Chumillas, P. M. Periago and J. Weiss (2012). Development of a high-resolution melting-based approach for efficient differentiation among *Bacillus cereus* group isolates. Foodborne Pathog Dis 9(9): 777-785.
3. Blackburn, J. K., K. M. McNyset, A. Curtis and M. E. Hugh-Jones (2007). Modeling the geographic distribution of *Bacillus anthracis*, the causative agent of anthrax disease, for the contiguous United States using predictive ecological [corrected] niche modeling. Am J Trop Med Hyg 77(6): 1103-1110.
4. Chin, C. S., J. Sorenson, J. B. Harris, W. P. Robins, R. C. Charles, R. R. Jean-Charles, J. Bullard, D. R. Webster, A. Kasarskis, P. Peluso, E. E. Paxinos, Y. Yamaichi, S. B. Calderwood, J. J. Mekalanos, E. E. Schadt and M. K. Waldor (2011). The origin of the Haitian cholera outbreak strain. N Engl J Med 364(1): 33-42.
5. Durrheim, D. N., P. Freeman, I. Roth and M. Hornitzky (2009). Epidemiologic questions from anthrax outbreak, Hunter Valley, Australia. Emerg Infect Dis 15(5): 840-842.

6. Freed, J. (20 March 2015). Sydney Airport traffic grows in February, *The Sydney Morning Herald*, http://www.smh.com.au/business/aviation/sydney-airport-traffic-grows-in-february-20150319-1m3ih7
7. Gahan, M. E., R. Thomas, R. Rossi, M. Nelson, P. Roffey, M. M. Richardson and D. McNevin (2015). Background frequency of *Bacillus* species at the Canberra Airport: A 12 month study. Forensic Sci Int 257: 142-148.
8. Grad, Y. H., M. Lipsitch, M. Feldgarden, H. M. Arachchi, G. C. Cerqueira, M. Fitzgerald, P. Godfrey, B. J. Haas, C. I. Murphy, C. Russ, S. Sykes, B. J. Walker, J. R. Wortman, S. Young, Q. Zeng, A. Abouelleil, J. Bochicchio, S. Chauvin, T. Desmet, S. Gujja, C. McCowan, A. Montmayeur, S. Steelman, J. Frimodt-Moller, A. M. Petersen, C. Struve, K. A. Krogfelt, E. Bingen, F. X. Weill, E. S. Lander, C. Nusbaum, B. W. Birren, D. T. Hung and W. P. Hanage (2012). Genomic epidemiology of the *Escherichia coli* O104:H4 outbreaks in Europe, 2011. Proc Natl Acad Sci U S A 109(8): 3065-3070.
9. Jernigan, D. B., P. L. Raghunathan, B. P. Bell, R. Brechner, E. A. Bresnitz, J. C. Butler, M. Cetron, M. Cohen, T. Doyle, M. Fischer, C. Greene, K. S. Griffith, J. Guarner, J. L. Hadler, J. A. Hayslett, R. Meyer, L. R. Petersen, M. Phillips, R. Pinner, T. Popovic, C. P. Quinn, J. Reefhuis, D. Reissman, N. Rosenstein, A. Schuchat, W. J. Shieh, L. Siegal, D. L. Swerdlow, F. C. Tenover, M. Traeger, J. W. Ward, I. Weisfuse, S. Wiersma, K. Yeskey, S. Zaki, D. A. Ashford, B. A. Perkins, S. Ostroff, J. Hughes, D. Fleming, J. P. Koplan, J. L. Gerberding and National Anthrax Epidemiologic Investigation Team (2002). Investigation of bioterrorism-related anthrax, United States, 2001: epidemiologic findings. Emerg Infect Dis 8(10): 1019-1028.
10. Koehler, T. M. (2009). *Bacillus anthracis* physiology and genetics. Mol Aspects Med 30(6): 386-396.
11. Langmead, B. and S. L. Salzberg (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods 9(4): 357-359.
12. Leask, A., V. Delpech and J. McAnulty (2003). Anthrax and other suspect powders: initial responses to an outbreak of hoaxes and scares. NSW Public Health Bull 14(11-12): 218-221.
13. Leekitcharoenphon, P., R. S. Hendriksen, S. Le Hello, F. X. Weill, D. L. Baggesen, S. R. Jun, D. W. Ussery, O. Lund, D. W. Crook, D. J. Wilson and F. M. Aarestrup (2016). Global Genomic Epidemiology of *Salmonella enterica* Serovar *Typhimurium* DT104. Appl Environ Microbiol 82(8): 2516-2526.
14. Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin and S. Genome Project Data Processing (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25(16): 2078-2079.
15. Li, Y., G. M. Leung, J. W. Tang, X. Yang, C. Y. Chao, J. Z. Lin, J. W. Lu, P. V. Nielsen, J. Niu, H. Qian, A. C. Sleigh, H. J. Su, J. Sundell, T. W. Wong and P. L. Yuen (2007). Role of ventilation in airborne transmission of infectious agents in the built environment - a multidisciplinary systematic review. Indoor Air 17(1): 2-18.
16. Li, H. (GitHub, year 2012). Seqtk: A toolkit for processing sequences in FASTA/Q formats. https://github.com/lh3/seqtk.
17. Lu, H., F. Giordano and Z. Ning (2016). Oxford Nanopore MinION Sequencing and Genome Assembly. Genomics Proteomics Bioinformatics 14(5): 265-279.
18. Luna, V. A., D. S. King, K. K. Peak, F. Reeves, L. Heberlein-Larson, W. Veguilla, L. Heller, K. E. Duncan, A. C. Cannons, P. Amuso and J. Cattani (2006). *Bacillus anthracis* virulent plasmid pX02 genes found in large plasmids of two other *Bacillus* species. J Clin Microbiol 44(7): 2367-2377.
19. Pannucci, J., R. T. Okinaka, E. Williams, R. Sabin, L. O. Ticknor and C. R. Kuske (2002). DNA sequence conservation between the *Bacillus anthracis* pXO2 plasmid and genomic sequence from closely related bacteria. BMC Genomics 3(1): 34.
20. Picard. (2014). Command line tools for manipulating high-throughput sequencing (HTS) data. Broad Institute Cambridge, MA, https://broadinstitute.github.io/picard/.
21. Prussin, A. J., 2nd and L. C. Marr (2015). Sources of airborne microorganisms in the built environment. Microbiome 3: 78.
22. Quick, J., N. J. Loman, S. Duraffour, J. T. Simpson, E. Severi, L. Cowley, J. A. Bore, R. Koundouno, G. Dudas, A. Mikhail, N. Ouedraogo, B. Afrough, A. Bah, J. H. Baum, B. Becker-Ziaja, J. P. Boettcher, M. Cabeza-Cabrerizo, A. Camino-Sanchez, L. L. Carter, J. Doerrbecker, T. Enkirch, I. G. G. Dorival, N. Hetzelt, J. Hinzmann, T. Holm, L. E. Kafetzopoulou, M. Koropogui, A. Kosgey, E. Kuisma, C. H. Logue, A. Mazzarelli, S. Meisel, M. Mertens, J. Michel, D. Ngabo, K. Nitzsche, E.

Pallash, L. V. Patrono, J. Portmann, J. G. Repits, N. Y. Rickett, A. Sachse, K. Singethan, I. Vitoriano, R. L. Yemanaberhan, E. G. Zekeng, R. Trina, A. Bello, A. A. Sall, O. Faye, O. Faye, N. Magassouba, C. V. Williams, V. Amburgey, L. Winona, E. Davis, J. Gerlach, F. Washington, V. Monteil, M. Jourdain, M. Bererd, A. Camara, H. Somlare, A. Camara, M. Gerard, G. Bado, B. Baillet, D. Delaune, K. Y. Nebie, A. Diarra, Y. Savane, R. B. Pallawo, G. J. Gutierrez, N. Milhano, I. Roger, C. J. Williams, F. Yattara, K. Lewandowski, J. Taylor, P. Rachwal, D. Turner, G. Pollakis, J. A. Hiscox, D. A. Matthews, M. K. O'Shea, A. M. Johnston, D. Wilson, E. Hutley, E. Smit, A. Di Caro, R. Woelfel, K. Stoecker, E. Fleischmann, M. Gabriel, S. A. Weller, L. Koivogui, B. Diallo, S. Keita, A. Rambaut, P. Formenty, S. Gunther and M. W. Carroll (2016). Real-time, portable genome sequencing for Ebola surveillance. Nature 530(7589): 228-232.

23. Rhoads, A. and K. F. Au (2015). PacBio Sequencing and Its Applications. Genomics Proteomics Bioinformatics 13(5): 278-289.
24. Salipante, S. J., T. Kawashima, C. Rosenthal, D. R. Hoogestraat, L. A. Cummings, D. J. Sengupta, T. T. Harkins, B. T. Cookson and N. G. Hoffman (2014). Performance comparison of Illumina and ion torrent next-generation sequencing platforms for 16S rRNA-based bacterial community profiling. Appl Environ Microbiol 80(24): 7583-7591.
25. Schadt, E. E., S. Turner and A. Kasarskis (2010). A window into third-generation sequencing. Hum Mol Genet 19(R2): R227-240.
26. Sextro, R. G., D. M. Lorenzetti, M. D. Sohn and T. L. Thatcher (2002). Modeling the spread of anthrax in buildings. In *Proceedings of Indoor Air 2002 Conference, Monterey, CA* (pp. 506-511).
27. Time and Date. (2012). Past weather in Canberra, Australian Capital Territory , Australia – February 2012.
28. Tringe, S. G., T. Zhang, X. Liu, Y. Yu, W. H. Lee, J. Yap, F. Yao, S. T. Suan, S. K. Ing, M. Haynes, F. Rohwer, C. L. Wei, P. Tan, J. Bristow, E. M. Rubin and Y. Ruan (2008). The airborne metagenome in an indoor urban environment. PLoS One 3(4): e1862.
29. Van Cuyk, S., A. Deshpande, A. Hollander, D. O. Franco, N. P. Teclemariam, J. A. Layshock, L. O. Ticknor, M. J. Brown and K. M. Omberg (2012). Transport of *Bacillus thuringiensis* var. *kurstaki* from an outdoor release into buildings: pathways of infiltration and a rapid method to identify contaminated buildings. Biosecur Bioterror 10(2): 215-227.
30. Whiteley, A. S., S. Jenkins, I. Waite, N. Kresoje, H. Payne, B. Mullan, R. Allcock and A. O'Donnell (2012). Microbial 16S rRNA Ion Tag and community metagenome sequencing using the Ion Torrent (PGM) Platform. J Microbiol Methods 91(1): 80-88.
31. Wielinga, P. R., R. A. Hamidjaja, J. Agren, R. Knutsson, B. Segerman, M. Fricker, M. Ehling-Schulz, A. de Groot, J. Burton, T. Brooks, I. Janse and B. van Rotterdam (2011). A multiplex real-time PCR for identifying and differentiating *B. anthracis* virulent types. Int J Food Microbiol 145 Suppl 1: S137-144.
32. Wright, A. M., S. B. Beres, E. N. Consamus, S. W. Long, A. R. Flores, R. Barrios, G. S. Richter, S. Y. Oh, G. Garufi, H. Maier, A. L. Drews, K. E. Stockbauer, P. Cernoch, O. Schneewind, R. J. Olsen and J. M. Musser (2011). Rapidly progressive, fatal, inhalation anthrax-like infection in a human: case report, pathogen genome sequencing, pathology, and coordinated response. Arch Pathol Lab Med 135(11): 1447-1459.