Evaluating the use of whole-genome sequencing for outbreak investigations in the lack of closely related reference genome

Mohamed M.H. Abdelbary[a,*], Laurence Senn[a], Estelle Moulin[a], Guy Prod'hom[b], Antony Croxatto[b], Gilbert Greub[b], Dominique S. Blanc[a,b]

[a] *Service of Hospital Preventive Medicine, Lausanne University Hospital, Lausanne, Switzerland*
[b] *Institute of Microbiology, Lausanne University Hospital, Lausanne, Switzerland*

ARTICLE INFO

ABSTRACT

Whole-genome sequencing (WGS) has emerged as a powerful molecular typing method for outbreak analysis enabling the rapid discrimination between outbreak and non-outbreak isolates. However, such analysis can be challenging in the absence of closely related reference genomes. In this study, we assessed the use of WGS in investigating an outbreak of a relatively understudied bacterial pathogen with no publicly available closely related reference genome. Eleven *Burkholderia cepacia* complex (Bcc) isolates (seven from patients and four from disposable dermal gloves packages) that were collected during an outbreak were sequenced using the Illumina MiSeq platform. Our results showed that mapping the 11 sequenced Bcc outbreak isolates against a genetically distant reference genome yield loses coverage (31.6–48.3%) and a high number of detected false single-nucleotide polymorphisms (SNPs) (1123–2139). Therefore, a reference genome consensus from an outbreak clinical isolate was generated by combining both *de novo* assembly and mapping approaches. Based on this approach, we were able to demonstrate that the Bcc outbreak isolates were closely related and were phylogenetically distinct from the 11 publically available Bcc genomes. In addition, the pairwise SNP distance analysis detected only 1 to 6 SNPs differences among the outbreak isolates, confirming that contaminated disposable dermal gloves were the cause of the outbreak.

1. Introduction

Microbial whole-genome sequencing (WGS) has become a significant tool for molecular typing and outbreak investigations. The availability of rapid benchtop sequencers and the advent of genomics becoming affordable have significantly increased the use of WGS for routine microbiological analysis. Furthermore, WGS provides higher discriminatory power compared to other molecular typing methods (*e.g.* pulsed-field gel electrophoresis), which enables the detection of genetic relatedness between different bacterial isolates belonging to the same clone.

In recent years, numerous studies have demonstrated the effective use of WGS in tracing transmission events during outbreak investigations of various pathogens within a short period of time (Leekitcharoenphon et al., 2014; Zhou et al., 2016). For example, a phylogenetic reconstruction based on WGS data can be used to differentiate between outbreak and non-outbreak isolates, and to determine associations between environmental and clinical isolates in case of outbreak. Such phylogenetic reconstruction can be achieved by

reference-based or reference-free approaches. In contrast to the reference-based method, reference-free approaches may exhibit higher number of false positive SNPs due to errors that rise from *de novo* genome assemblies' processes (Pettengill et al., 2014; Pightling et al., 2014). However, the reference-based method can be challenging by the lack of reference genome or when available genomic data remain limited such as in the case of *Burkholderia cepacia* complex (Bcc).

Members of Bcc are known as important opportunistic pathogens associated with various kinds of human infections such as respiratory tract infections and catheter-related infections in hospitalized and immunocompromised patients (mainly in cystic fibrosis patients) (Saffra & Moriarty, 2014). Bcc was found to be highly resistant to a wide range of antimicrobial agents, which limits treatment options (Rhodes & Schweizer, 2016).

Previous reports have described outbreaks of Bcc infections that are linked to contaminated medical devices and pharmaceutical products used for patient care in the hospital settings (*e.g.* mouthwash solution, chlorhexidine solutions, mannitol solution, and antiseptic solution) (Singhal et al., 2015; Martin et al., 2012; Ko et al., 2015). Currently, the

Centers for Disease Control and Prevention (CDC) and the Food and Drug Administration (FDA) are investigating a multistate outbreak of Bcc infections associated with a contaminated liquid docusate product leading to a cluster of 60 confirmed cases from eight different states in the United States (https://www.cdc.gov/hai/outbreaks/b-cepacia/index.html).

In July 2016, a warning was launched by Swissmedic (https://www.swissmedic.ch/medizinprodukte/02635/02639/03480/index.html?lang=en) stating that pre-moisturized disposable dermal gloves from the brand Sinaqua (Welcare, Italy), which were delivered to Switzerland, were contaminated with Bcc and might be associated with serious infections in various hospital settings. Following this warning, an epidemiological investigation was conducted in the University Hospital of Lausanne, Switzerland. A case was considered as a patient with Bcc recovered in clinical samples and who did not suffer from cystic fibrosis and was not previously colonized with Bcc. While 0 to 2 cases per year were identified between 2012 and 2015, 7 cases were reported between May and July 2016.

Little is known about the genetic diversity between Bcc isolates due to a lack of WGS-based studies assessing population structure, hospital outbreaks and transmission of Bcc. Hence, we aimed in this study to demonstrate the potential challenges and to evaluate the use of WGS in investigating an outbreak of relatively understudied bacterial pathogen such as Bcc.

2. Materials and methods

2.1. Bacterial isolates and DNA sequencing

For this study, a total of 11 Bcc isolates, collected between May and August 2016 at the university hospital of Lausanne, were sequenced to determine their phylogenetic relatedness. All isolates were part of the suspected outbreak. Seven were collected from patients, while the remaining four isolates were collected from different disposable washing gloves packages (Table 1).

Genomic DNA was extracted using the Wizard Genomic DNA Purification kit (Promega, Madison, WI, USA) according to the manufacturer's instructions. Subsequently, the extracted genomic DNA was quantified using the Qubit double-stranded DNA high-sensitivity (HS) assay kit (Life Technologies, Waltham, MA, USA). Sequencing library preparation was performed using Nextera XT DNA Library Preparation Kit (Illumina, San Diego, CA, USA) following the manufacturer's guidelines, followed by sequencing using version 2 chemistry protocol on the Illumina MiSeq platform (Illumina, San Diego, CA, USA) generating 2×150 bp paired-end reads.

Moreover, additional sequencing data from 11 Bcc publicly available genomes retrieved from different Bioprojects were included in this study (Table 1). The idealized genome coverage was calculated from sequence data using the formula coverage = (read count × read length) / total genome size. All Sequence reads were pre-processed, trimmed for adaptors and then *de novo* assembled using A5-Miseq pipeline (Coil et al., 2015). Subsequently, the draft *de novo* assembled genomes contigs were reordered against the Bcc published reference genome ATCC 25416 (accession numbers NZ_CP012981-NZ_CP012983) using MAUVE version 2.4.0 (Darling et al., 2004; Rissman et al., 2009). For each of the 22 Bcc genomes, the reordered contigs were concatenated into one contig, which was then annotated using the Prokka annotation pipeline (Seemann, 2014).

Multi-locus sequence typing (MLST) analysis and identification of sequence types (STs) were determined from the sequence reads and *de novo* assembled genomes using SRST2 (Inouye et al., 2014) and MLST-check pipelines (AJPaBTaJA, 2016), respectively.

2.2. Generating the reference genome assembly

A reference genome assembly was generated from the clinical Bcc isolate H32570. Sequence reads were pre-processed and assembled using MGAP pipeline (https://github.com/dsarov/MGAP—Microbial-Genome-Assembler-Pipeline), which implements several tools and assembly programs, such as Trimmomatic (Bolger et al., 2014), Velvet (Zerbino & Birney, 2008), VelvetOptimiser (https://github.com/tseemann/VelvetOptimiser), GapFiller (Boetzer & Pirovano, 2012), ABACAS (Assefa et al., 2009), IMAGE (Tsai et al., 2010) and SSPACE (Boetzer et al., 2011). Subsequently, the assembled H32570 genome was corrected for SNPs, insertions and deletions that were detected by mapping H32570 sequence reads to the H32570 assembled genome

Table 1
Burkholderia cepacia complex genomes analyzed in the present study.

Isolate number	Isolation date	Source	Country of origin	Institution	BioSample accession	Idealized coverage	Sequence type (ST)
H32570	28.05.2016	Patient 1	Switzerland	Lausanne University Hospital	Present study	62.4×	ST1095
H32571	17.06.2016	Patient 2	Switzerland	Lausanne University Hospital	Present study	64.7×	ST1095
H32572	27.06.2016	Patient 3	Switzerland	Lausanne University Hospital	Present study	64.0×	ST1095
H32573	09.07.2016	Patient 4	Switzerland	Lausanne University Hospital	Present study	84.3×	ST1095
H32574	18.07.2016	Patient 5	Switzerland	Lausanne University Hospital	Present study	69.1×	ST1095
H32575	14.07.2016	Patient 6	Switzerland	Lausanne University Hospital	Present study	69.3×	ST1095
H32576	27.07.2016	Patient 7	Switzerland	Lausanne University Hospital	Present study	79.6×	ST1095
H32577	02.08.2016	Gloves 1	Switzerland	Lausanne University Hospital	Present study	82.1×	ST1095
H32578	02.08.2016	Gloves 2	Switzerland	Lausanne University Hospital	Present study	61.6×	ST1095
H32579	02.08.2016	Gloves 3	Switzerland	Lausanne University Hospital	Present study	72.3×	ST1095
H32580	02.08.2016	Gloves 4	Switzerland	Lausanne University Hospital	Present study	78.0×	ST1095
SRR1956207	2011	Soil	Australia	Northern Arizona University	SAMN03449556	86.6×	Not assigned
SRR1956269	2010	Soil	Australia	Northern Arizona University	SAMN03449618	73.4×	Not assigned
SRR1956258	2010	Soil	Australia	Northern Arizona University	SAMN03449607	93.5×	Not assigned
SRR1956222	2011	Soil	Australia	Northern Arizona University	SAMN03449571	79.2×	ST698
SRR1956221	2011	Soil	Australia	Northern Arizona University	SAMN03449570	71.6×	ST698
SRR1956505	2010	Soil	Australia	Northern Arizona University	SAMN03449282	33.6×	Not assigned
SRR1618489	2005	Aerosol	Australia	Los Alamos National Laboratory	SAMN02797514	305×	ST807
SRR1655356	Unknown	Medical/Surgical	USA, WA	The University of Washington Medical Center	SAMN03197124	24.8×	ST10
SRR1655367	Unknown	Medical/Surgical	USA, WA	The university of Washington Medical Center	SAMN03197135	21.0×	ST10
SRR1956381	2007	Soil	Thailand: Ubon, Trakam	Northern Arizona University	SAMN03449130	83.4×	Not assigned
SRR2846099	1989	Sputum	The United Kingdom: Edinburgh	Los Alamos National Laboratory	SAMN03144973	702.3	ST28