Improved multilocus sequence typing of *Burkholderia pseudomallei* and closely related species

Erin P. Price, Barbara MacHunter, Brian G. Spratt, David M. Wagner, Bart J. Currie and Derek S. Sarovich

1Global and Tropical Health Division, Menzies School of Health Research, Darwin, Northern Territory, Australia
2Faculty of Medicine, Imperial College London, London, UK
3Center for Microbial Genetics and Genomics, Northern Arizona University, Flagstaff, AZ, USA
4Department of Infectious Diseases and Northern Territory Medical Program, Royal Darwin Hospital, Darwin, Northern Territory, Australia

The *Burkholderia pseudomallei* multilocus sequence typing (MLST) database (http://pubmlst.org/bpseudomallei/) contains the largest global sequence repository for *B. pseudomallei* and its closest genetic relatives. Using conventional MLST and in silico MLST data derived from publicly available whole-genome sequences, we first defined the phylogenetic relatedness of *B. pseudomallei* and its nearest neighbours. Based on this analysis, we propose that the recently described *B. pseudomallei* complex (Bpc) should be expanded to encompass *B. pseudomallei*, *Burkholderia humptydooensis* (proposed), *B. mallei*, *B. oklahomensis*, *B. thailandensis* and three unassigned *Burkholderia* Clades A, B and C (represented by type strains BDU 5, BDU 8 and MSMB0265, respectively). Of note, the MLST narK locus is present in all Bpc species but is missing in all other *Burkholderia* spp., including all *Burkholderia cepacia* complex species, with the exception of most *Burkholderia ubonensis* strains, which contain narK but encode genetically distinct sequences. The presence of narK is thus indicative of a Bpc strain. Next, we revisited in silico the performance of the existing MLST primers, which prompted redesign of primers targeting the gmhD, lepA, lipA, narK and ndh loci to encompass genetic diversity among Bpc strains and to address amplification/sequencing issues. We show in silico and in vitro that the redesigned primers yield good-quality amplification and sequencing results for the gmhD, lepA, lipA, narK and ndh loci in Bpc species. These primers provide an alternative for amplification and sequencing of MLST loci in Bpc species in cases when poor-quality amplification or sequencing data are obtained using the original MLST primers.

INTRODUCTION

*Burkholderia pseudomallei* is a Gram-negative soil- and water-dwelling bacterium that causes melioidosis. Until recently, melioidosis endemicity was thought to be largely confined to northern Australia and parts of Southeast Asia; however, *B. pseudomallei* presence is now being unmasked in many tropical, subtropical and temperate regions worldwide (Currie, 2015). The global burden of melioidosis in 2015 alone has recently been predicted to be ca. 165 000 cases, with ca. 89 000 deaths, rates that are comparable to measles (Limmathurotsakul et al., 2016). Melioidosis symptoms are non-specific and can affect almost any organ, with presentation occurring in either chronic or acute form (Wiersinga et al., 2012). In Thailand, up to 40% of melioidosis patients still die despite treatment, and in Australia, 10% of treated cases result in mortality (Limmathurotsakul et al., 2010; Currie, 2015).

*Burkholderia mallei*, an equine-adapted clone of *B. pseudomallei* that causes the zoonotic disease glanders, has been successfully eradicated from most countries, although endemic foci remain in parts of Africa, Asia, Central and

**Abbreviations**: Bcc, *Burkholderia cepacia* complex; Bpc, *Burkholderia pseudomallei* complex; MLST, multilocus sequence typing; ST, sequence type.

One supplementary table is available with the online Supplementary Material.
South America and the Middle East (Lehavi et al., 2002; Naureen et al., 2007; Dance, 2009). Compared with B. pseudomallei, the B. mallei genome has undergone substantial reduction and, as a consequence, does not survive well outside the mammalian host (Nierman et al., 2004). Both B. pseudomallei and B. mallei are listed as Tier 1 select agents in the USA owing to concerns about their bioweaponization potential. These concerns are reflected in the deliberate use of B. mallei as a biological warfare agent during the First and Second World Wars (Lehavi et al., 2002; Dance, 2009).

Other Burkholderia spp. closely related to B. pseudomallei on the genetic level include the environmental species Burkholderia huntydooensis (proposed) (Ginther et al., 2015), Burkholderia oklahomensis and Burkholderia thailandensis (Glass et al., 2006; Depoorter et al., 2016). Although rare, B. thailandensis and B. oklahomensis have been known to cause opportunistic infections in humans (Glass et al., 2006). B. huntydooensis (proposed) is a recently described B. thailandensis-like taxon that was first identified in 1995 from Humpy Doo, Northern Territory, Australia (Gee et al., 2008), but has not yet been associated with disease.

Because of the bioweaponization potential of B. pseudomallei and B. mallei, there has been an interest in understanding and characterizing the genetic relatedness of species most closely related to these pathogens (Ginther et al., 2015). However, there is currently no clear definition of species belonging to the B. pseudomallei complex (Bpc). In comparison, the Burkholderia cepacia complex (Bcc) is a well-established group of species that encompasses the closest genetic relatives of B. cepacia; the Bcc group currently contains 20 validly named Burkholderia spp., with several more informally described species (Peeters et al., 2016). Previous work has defined the Bpc as B. pseudomallei, B. mallei and B. thailandensis (Lowe et al., 2014); however, B. huntydooensis (proposed) and B. oklahomensis also have high sequence homology to B. pseudomallei (Depoorter et al., 2016). There is, therefore, a need to more formally investigate the relatedness of the closest B. pseudomallei neighbours as it is likely that there are additional Bpc species that have not yet been recognized in the literature.

The multilocus sequence typing (MLST) scheme for B. pseudomallei, which was first developed in 2003 (Godoy et al., 2003) and partially revised in 2006 (Vesaratatchev et al., 2006), targets seven housekeeping loci: ace (acetocacetyl coenzyme A reductase), gltB (glutamate synthase), gmhD (ADP-L-glycerol-3-phosphate heptose 6-epimerase), lepA (GTP-binding elongation factor), lipA (lipoic acid synthetase), narK (nitrite extrusion protein) and ndh (NADH dehydrogenase). The B. pseudomallei MLST database (previously housed at http://bpseudomallei.mlst.net; now at http://pubmlst.org/bpseudomallei/) provides the largest repository of B. pseudomallei genotypes in the world. This database currently contains more than 4500 isolates comprising nearly 1500 sequence types (STs). A PubMed search conducted in February 2016 using the terms, ‘Burkholderia pseudomallei’ and ‘multilocus sequence typing’ or ‘MLST’, retrieved 53 peer-reviewed articles. In recent years, whole-genome sequencing (WGS) has provided a rapid alternative for extracting MLST profiles; however, the PCR-based MLST method is still used by many researchers across the globe.

Our group and others have encountered issues with some of the existing MLST primers. In particular, certain loci do not always amplify well or co-amplify non-specific products, leading to poor sequence trace data that can be difficult to interpret. These issues prompted us to redesign some of the MLST primers for improved amplification and sequencing of these housekeeping loci across the Bpc species (Table 1). The first B. pseudomallei genome, K96243, was not published until 2004 (Holden et al., 2004), a year after the initial MLST scheme by Godoy et al. (2003) was developed. As of June 2016, there are 265 Bpc genomes available in public databases, of which 190 are B. pseudomallei. This wealth of genomic information provides a timely opportunity to design better primers that avoid known genetic variation among strains, that are more specific and that amplify efficiently in closely related species. Thus, the purpose of our study was twofold: first, to use MLST data to define the Bpc species by phylogenetic analysis of B. pseudomallei and its closest genetic relatives and, second, to use WGS data to redesign the B. pseudomallei MLST primers to provide alternative primers for amplification and sequencing across all Bpc species and to improve the quality of the resulting sequence data.

### METHODS

#### Assignment of STs for all publicly available Bpc genomes.

The first step of this study was to determine MLST profiles for all publicly available genome-sequenced Burkholderia strains not previously submitted to the MLST database for which unambiguous and full-length sequences for the seven loci could be assigned. We carried out this task using the BLGSdb tool (Jolley et al., 2010), which is integrated into the B. pseudomallei MLST website (http://pubmlst.org/bpseudomallei/). MLST profiles for B. ubonensis were derived in silico from reference-assisted

---

**Table 1. Redesigned MLST amplification and sequencing primers for Bpc species**

<table>
<thead>
<tr>
<th>Locus</th>
<th>Forward primer (5’–3’)</th>
<th>Reverse primer (5’–3’)</th>
<th>Amplicon size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>gmhD</td>
<td>gmhD_F: GCYTCGGTTGAAGCAGCTGTTA</td>
<td>gmhD_R: CGAGCCGGCTTTCGCGAG</td>
<td>526</td>
</tr>
<tr>
<td>lepA</td>
<td>lepA_F: CCGCACRTAGTTGTCGAACCA</td>
<td>lepA_R: CACTGAAATGGATCATATTCGCAA</td>
<td>621</td>
</tr>
<tr>
<td>lipA</td>
<td>lipA_F: ACCCGGACGTTCATGTATCA</td>
<td>lipA_R: CCATCAGGCGCCGATTTCC</td>
<td>457</td>
</tr>
<tr>
<td>narK</td>
<td>narK_F: TCAAGGAGTGGCTATGTC</td>
<td>narK_R: ATGAACGGCACCCACACC</td>
<td>665</td>
</tr>
<tr>
<td>ndh</td>
<td>ndh_F: CCACTGGCTCCAGGAG</td>
<td>ndh_R: GCTCACCATTGCCAGGT</td>
<td>535</td>
</tr>
</tbody>
</table>

http://jmm.microbiologyresearch.org

Downloaded from www.microbiologyresearch.org by IP: 54.70.40.11
On: Mon, 11 Feb 2019 20:29:13
Phylogenetic analysis of STs in the B. pseudomallei MLST database. Following the MLST database update, concatenated sequences for all 1478 STs in the B. pseudomallei MLST database (as of 17 June 2016) were obtained using the ‘sequence/profile definitions’ link by performing a search using the parameter ‘ST=0’, followed by export of the concatenated sequences in multi-Fasta format. Sequences were imported into ClustalW V2.1 (Larkin et al., 2007), and a neighbour-joining phylogenetic tree was constructed from these concatenated data. From this initial analysis, four STs (ST-1039, ST-1326, ST-1365 and ST-1388) were excluded owing to data integrity concerns. The final phylogenetic analysis of 1474 STs was used to determine the Bpc species boundary. Tree visualization was carried out using FigTree v1.4.0 (http://tree.bio.ed.ac.uk/software/figtree/).

Redesign of MLST loci to incorporate all Bpc species. Using the extensive publicly available genomic data for Bpc species, conserved regions were identified in the sequences flanking the MLST loci. Primers were designed using Primer Express v3.0 (Applied Biosystems) based on an optimum melting temperature of 60 °C. All paired primers were examined for secondary structure using NetPrimer (http://www.premierbiosoft.com/netprimer/). The redesigned primers were tested in silico using Microbial Genomes BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) across 190 B. pseudomallei, 35 B. mallei, 17 B. thailandensis, 6 B. huminumdoensis (proposed), 2 B. oklahomensis, 2 Burkholderia sp. Clade A, 2 Burkholderia sp. Clade B and 11 Burkholderia sp. Clade C genomes (Table S1; Data S1) to confirm specificity and conservation of the primer-binding sites. New primers for gltB and ace were not designed as we and others have found the Godoy et al. (2003) primers consistently work well for amplification and sequencing of these loci in Bpc species.

Conventional MLST analysis using the redesigned primers. We performed PCRs using the newly designed primers for gmkD, lepA, lipA, narK and ndh (Table 1) and the original (Godoy et al., 2003) ace and gltB primers across two B. pseudomallei, two B. huminumdoensis (proposed), two B. thailandensis, one Burkholderia ubonensis and one B. cepacia to confirm the primer and amplicon specificity observed in silico (see Fig. 2). The latter two species belong to the Bcc, which has its own MLST scheme (http://pubmlst.org/bcc/). PCRs were performed using 0.2 μM each primer (Macrogen), 0.2 mM dNTPs, 1.5 mM MgCl₂, 1.2 M Q-Solution, 0.08 U HotStarTag DNA polymerase, 1× HotStarTag PCR buffer (Qiagen) and molecular grade H₂O to a final volume of 20 μL. PCR thermocycling was conducted using an initial denaturation of 95 °C for 15 min, 35 cycles of denaturation at 95 °C for 30 s, annealing at 60 °C for 30 s and extension at 72 °C for 10 min. Samples were sequenced using the standard dyeoxy sequencing method (Macrogen).

RESULTS AND DISCUSSION

Phylogenetic analysis of B. pseudomallei MLST data to define Bpc species

Phylogenetic reconstruction of the 1474 STs in the B. pseudomallei MLST database revealed a clear delineation between Bpc species and B. ubonensis, a member of the Bcc group (Fig. 1). BLAST analysis confirmed that B. ubonensis is the only non-Bpc species that contains narK, with all other non-Bpc Burkholderia spp. lacking this locus. Based on our analysis, we posit that there are at least seven distinct clades representing at least eight species that define the Bpc group. Three of these clades belong to four known Burkholderia spp. (B. pseudomallei/B. mallei, B. thailandensis and B. oklahomensis), with a fourth clade representing B. huminumdoensis (proposed). The remaining three clades (Clades A, B and C) have not been assigned species designations but have been grouped within the Bpc according to MLST (Fig. 1). These three clades are most closely related to B. oklahomensis and are represented solely by environmental isolates obtained from northern Australia (ABCPW 8, ABCPW 14, ABCPW 111, ABCPW BDU 5, BDU 6, BDU 8, BDU 18, BDU 19, BDU 20, MSMB0263, MSMB0265, MSMB0266, MSMB0617, MSMB0852, MSMB1498, MSMB2040, MSMB2041 and MSMB2042) (Table S1). It is worth noting that B. oklahomensis has not yet been isolated outside the USA and that B. huminumdoensis (proposed) has not been found outside northern Australia. Similarly, it is possible that Clades A, B and C are only found in the northern Australian region, although these species may have been largely dismissed previously owing to genetic or morphological dissimilarities to B. pseudomallei.

Our phylogenetic analysis of MLST sequence data has greatly expanded the number of putative species belonging to the Bpc and is the first time an attempt has been made to define this group using extensive molecular data. Given the relatively high specificity of the narK locus for detecting Bpc species, we encourage all researchers identifying non-B. pseudomallei isolates found to be positive for narK by PCR or genome sequencing to fully characterize them by MLST and to submit these data to the B. pseudomallei MLST database.

Reassignment of incorrectly speciated Bpc genomes

We used the phylogeny reconstructed from MLST data to reassign species for a small number of incorrectly speciated Bpc isolates in the NCBI Nucleotide database (Table S1). Species designations for B. thailandensis MSMB43, B. thailandensis MSMB121, Burkholderia sp. MSMB122, Burkholderia sp. MSMB1552, Burkholderia sp. MSMB1588 and Burkholderia sp. MSMB1589 were reassigned as B. huminumdoensis (proposed); Burkholderia sp. TSV 202 was reassigned as B. pseudomallei; B. pseudomallei ABCPW 14 was reassigned as Burkholderia sp. (Clade B); and B. pseudomallei ABCPW 111 was reassigned as Burkholderia sp. (Clade C) (Fig. 1). MLST, thus, provides a valuable tool for correctly determining species designations for Bpc isolates.

In silico performance of the Godoy et al. (2003) and redesigned MLST primers

We performed extensive in silico analysis of both primer sets to determine their performance across all Bpc genome-sequenced strains (Table S1). We confirmed that the Godoy
et al. (2003) primers are sufficiently specific for performing MLST in B. pseudomallei and B. mallei, although the gmhD-dn and narK-up primers are directly adjacent to the allelic start sequence, which diminishes the length of requisite sequence data obtained in those directions. The redesigned primers gmhD_F and nark_F (Table 1) overcome this issue.

A 3′ ultimate mismatch in the Godoy et al. (2003) narK-up primer in B. thailandensis, B. humptydooensis (proposed) and Clade C Burkholderia sp. genomes is likely to affect amplification efficiency of this locus in these species; in comparison, the redesigned narK primers are 100% specific for the Bpc species. Similarly, the lepA_R and ndh_R primers may be preferable to lepA-up and ndh-dn, respectively, owing to more favourable distance of these primers from the allelic sequence start site.

**Dideoxy sequence analysis of the redesigned MLST primers**

In addition to extensive in silico analysis of the MLST primers, we performed conventional MLST analysis of the Godoy et al. (2003) redesigned primer pairs. All tested Bpc isolates gave unambiguous and overlapping forward and reverse sequence trace data and, where known, the expected STs. We also included two Bcc species, B. ubonensis and B. cepacia (Payne et al., 2005), to demonstrate the limits of the B. pseudomallei MLST scheme toward non-Bpc species.

**Fig. 1.** Midpoint-rooted neighbour-joining tree of 1474 concatenated STs in the B. pseudomallei MLST database. The black arrow points to Bpc species: B. pseudomallei, B. mallei, B. thailandensis, B. oklahomensis, B. humptydooensis (proposed) and three unassigned Burkholderia spp. (labelled Clades A, B and C). Most B. ubonensis strains contain all seven MLST loci, although only four of the seven loci amplify using the conventional MLST method; all other Bcc Burkholderia spp. lack narK and were therefore not included in this tree. Burkholderia sp. Clade A contains two STs (ST-1002 and ST-1003) and is represented by type strain BDU 5. Burkholderia sp. Clade B contains two STs (ST-962 and ST-1472) and is represented by type strain BDU 8. Burkholderia sp. Clade C contains nine STs (ST-646, ST-963, ST-964, ST-1015, ST-1052, ST-1440, ST-1473, ST-1474 and ST-1476) and is represented by type strain MSMB0265.
Only four of the seven loci amplified in *B. ubonensis* (Fig. 2). Similarly, only three loci amplified in *B. cepacia* (Fig. 2). Most Bcc species lack *narK*, and we therefore recommend that PCR-based MLST for these or other Bcc species should instead be determined using the primers and alleles listed for the Bcc MLST scheme (Spilker et al., 2009).

It is important to note that *B. ubonensis* STs have been included on the *B. pseudomallei* database owing to its shared ecological niche with *B. pseudomallei* and the inability to distinguish certain *B. ubonensis* strains from *B. pseudomallei* morphologically using the selective Ashdown’s agar medium (Price et al., 2013). Custom primers would be required to perform amplification and sequencing of the three loci that failed to amplify in *B. ubonensis* using our redesigned *B. pseudomallei* MLST primers.

The redesigned MLST primers described in this study have been designed in the post-genomic era to target the Bpc species *B. pseudomallei*, *B. humptydooensis* (proposed), *B. mallei*, *B. oklahomensis*, *B. thailandensis* and *Burkholderia* sp. Clades A, B and C. We designed these new primers to amplify and sequence the same alleles used in the original MLST scheme, allowing trimmed sequences to be submitted to the *B. pseudomallei* MLST website. These redesigned primers can be used for robust amplification and sequencing of Bpc species in cases where amplification or sequencing issues are encountered with the previously published primers.

**ACKNOWLEDGEMENTS**

E. P. P. wrote the paper, performed bioinformatic analyses and designed the PCR assays; B. M. conducted the laboratory work; E. P. P., B. G. S. and D. S. S. contributed to MLST data curation. E. P. P., B. G. S., D. M. W., B. J. C. and D. S. S. acquired funding for the study. All authors read and approved the final manuscript. The authors declare no conflict of interest. We are grateful to Daniel Godoy.
REFERENCES


