Clinical *Burkholderia pseudomallei* isolates from Queensland carry diverse
*bimA*<sub>Bm</sub> genes that are associated with central nervous system disease and
are phylogenomically distinct from other Australian strains

Delaney Burnard<sup>1</sup>, Michelle J Bauer<sup>1</sup>, Caitlin Falconer<sup>1</sup>, Ian Gassiep<sup>1,2</sup>, Robert E Norton<sup>3</sup>,
David L Paterson<sup>1</sup>, Patrick N A Harris<sup>4</sup>

<sup>1</sup>University of Queensland Centre for Clinical Research, Royal Brisbane and Woman's Hospital,
Herston, Queensland, 4029, Australia.

<sup>2</sup>Department of Infectious Diseases, Mater Hospital Brisbane, South Brisbane, Queensland, 4101,
Australia.

<sup>3</sup>Pathology Queensland, Townsville University Hospital, Townsville, Queensland, 4814, Australia.

<sup>4</sup>Pathology Queensland, Royal Brisbane & Women's Hospital, Herston, Queensland, 4029, Australia.

*Corresponding author*: Dr Patrick N A Harris

University of Queensland Centre for Clinical Research, Building 71/918 Royal Brisbane &
Women's Hospital Campus, Herston, QLD, 4029

Email: p.harris@uq.edu.au; Tel: +61 (0) 7 3346 6081

Keywords: *Burkholderia pseudomallei*, melioidosis, central nervous system disease,
virulence factor

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Abstract

*Burkholderia pseudomallei* is an environmental gram-negative bacterium that causes the disease melioidosis and is endemic in many countries of the Asia-Pacific region. In Australia, the mortality rate remains high at approximately 10%. The bacterium is almost exclusively found in the endemic region, which spans the Northern Territory, Western Australia and North Queensland. However, Queensland infections remain understudied compared to those of the Northern Territory. This pilot study aimed to assess the prevalence of central nervous system disease associated variant *bimA*\textsubscript{Bm}, identify circulating antimicrobial resistance mutations and genetically distinct strains from Queensland, via comparative genomics. From 76 clinical isolates, we identified the *bimA*\textsubscript{Bm} variant in 20 (26.3%) isolates and in 9 (45%) of the isolates with documented central nervous system infection (n=18). Explorative analysis suggests a significant association between isolates carrying the *bimA*\textsubscript{Bm} variant and central nervous system disease (OR 2.8, 95% CI 1.3-6.0, P=0.009) compared with isolates carrying the wildtype *bimA*\textsubscript{Bp} with the possibility of additional driving factors. Furthermore, 50% of isolates were identified as novel multi-locus sequence types (STs), while the *bimA*\textsubscript{Bm} variant was more commonly identified in isolates with novel sequence STs, compared to those of previously described STs. Additionally, mutations associated with acquired antimicrobial resistance were only identified in 14.5% of all genomes. The findings of this research have provided clinically relevant genomic data of *B. pseudomallei* in Queensland and suggest that the *bimA*\textsubscript{Bm} variant may enable risk stratification for the development CNS complications and be a potential therapeutic target.
Introduction

*Burkholderia pseudomallei* is an environmental gram-negative pathogen present in Asia, South America, Africa and the Pacific. Infections cause the disease melioidosis, with an extensive range of disease manifestations including, but not limited to, pneumonia, sepsis, and skin and soft tissue abscesses[1]. Successful treatment can be compromised by the diagnosis of central nervous system disease (e.g. encephalomyelitis, abscess, meningitis, cranial nerve impairment)[2]. As these infections are often sub-clinical, exhibit varying times to clinical presentation, have diverse clinical presentations and affect multiple other organs; they can be challenging cases to manage[2,3].

Central nervous system infections have been associated with the virulence factor bim$_{A_B p}$ mutant, bim$_{A_B m}$. These bimA genes encode as intracellular motility factor A, type V effector proteins that utilise host cellular actin. This motility results in the ability to evade the host immune system and invade the central nervous system via inter-cellular host cell migration[4]. Patients with infections possessing the bim$_{A_B m}$ gene are suggested to be 14 times more likely to present with neurological disease, while the wild-type variant bim$_{A_B p}$ is linked to pneumonia[5,6].

This gene has been identified in both clinical and environmental isolates from the Northern Territory; yet isolates from North Queensland have not been as extensively characterised, despite both states being endemic regions[5,6]. Clinical presentations of neurological disease have been documented in a retrospective Queensland isolate collection[7]; however, the diversity of bim$_{A_B m}$ and the association with central nervous system disease in Queensland remains unknown.
This study aimed to use comparative genomics to screen for the presence of the virulence factor \( bimA_bm \), which has been previously linked to neurological disease in \( B. pseudomallei \) infections in patients of the Northern Territory. Furthermore, genomic diversity and antimicrobial resistance profiles are described.
Methods

Isolates

Clinical B. pseudomallei isolates have been collected prospectively over the last 22 years (1996-2018) from participating hospital pathology facilities (Cairns, Townsville and Central laboratories of Pathology Queensland). A total of 76 isolates were selected between postcodes 4895-4500 to include isolates from all of Queensland, from a collection of 400 clinical isolates. All isolates with documented central nervous system presentation were included (n=18) and the remaining 57 isolates were selected to include a variety of other common B. pseudomallei clinical presentations (Table 1). An additional 24 publicly available reference genomes were included in this study, of these, 13 were derived from the Northern Territory and one representative each from Papua New Guinea, China and Thailand (Table S1).

DNA extraction and WGS

The 76 isolates were recovered from -80°C storage and subcultured twice to ensure purity. DNA extraction was performed with the QIAGEN DNAeasy ultra-pure DNA extraction kit according to manufacturer’s instructions. Sequencing libraries were generated using the Nextera Flex DNA library preparation kit and sequenced on the MiniSeq™ System (Illumina Inc., San Diego, CA, USA) on a high output 300 cycle cartridge according to the manufacturer’s instructions. Five strains with unique BimA_m sequence and/or genomic diversity were prepped for long read sequencing using the MinION (Oxford Nanopore Technologies, Oxford, UK), where sequencing libraries were generated using the Rapid Barcoding Sequencing Kit (SQK-RBK004) and run on a flow cell R9.4.1 for 72 hrs.
Genomic Analysis

Illumina reads were trimmed with Trimmmomatic v0.36[8] and quality assessed with multiQC[9], genomes were assembled with SPAdes v3.14.0[10] and annotated with Prokka v1.13[11]. Long reads were filtered, quality checked, assembled, and polished with Illumina reads where applicable using the MicroPIPE pipeline[12]. Genomes derived from this study are publicly available at NCBI BioProject ID: PRJNA717363.

All read mapping was performed with BWA-MEM[13]. MLST sequence types were determined with MLST[14] (https://github.com/tseemann/mlst) and read mapped to alleles retrieved from pubMLST, where SNPs were suspected[14]. Genotypic antimicrobial resistance was determined with ArDaP[15]. Whole genome alignment (4,950,632bp) and phylogenomic analysis of the 24 reference genomes and 76 genomes derived from this study (n=100) were achieved using parSNP[16] JModelTest[17] (TVM+F+I+G4) and IQtree[18] (1000 bootstrap (BS) replicates). Reference genomes were downloaded from NCBI, under the “complete” genome filter, where Australian reference strains were included if they were annotated with significant metadata (Table S1). The reference strains for virulence factors bimA<sub>Bm</sub> and bimA<sub>Bp</sub> were MSHR0668 ( locus BH02_RS34290) NZ_CP009546.1 and K96243 ( locus BPSS1492) NC_006351.1, respectively.

Statistical analysis

Using Stata and the csi command for unstratified cumulative incidence data[19], and isolates with bimA<sub>Bm</sub> or CNS presentation (n=29), Odds Ratios (ORs) and Chi-Square were calculated, with a P value <0.05 considered significant.

Ethics
This study received ethical approval from the Royal Brisbane & Women's Hospital Ethics Committee as a low or negligible-risk study (LNR/2020/QRBW/65573), with site specific authority obtained from the Townsville Hospital and Health Service and approval under the Queensland Public Health Act.

**Results**

In total, we identified the \( \text{bimA}_{\text{Bm}} \) virulence factor in 20/76 Queensland isolates (26.3%), geographically limited to the northern half of Queensland (Figure 1). Of the 18 isolates with known CNS disease, nine were found to carry the \( \text{bimA}_{\text{Bm}} \) virulence factor and nine carried the wild type \( \text{bimA}_{\text{Bp}} \). A total of 20 isolates were confirmed to carry the \( \text{bimA}_{\text{Bm}} \) virulence factor, of which eleven had no clinical record of CNS disease. The Stata csi analysis supported \( \text{bimA}_{\text{Bm}} \) being associated with CNS infection (Odds Ratio 2.8, 95% CI 1.3-6.1, \( P=0.009 \)).

The variation observed in the 20 \( \text{bimA}_{\text{Bm}} \) protein sequences was significant amongst isolates, with 18 proteins possessing unique sequences, while none of the sequences produced in this study were identical to that of the reference strain MSHR668. The variation was observed almost only in the proline-rich region of the protein (90-160aa) (Figure 2). However, five isolates carried an amino acid variant \( \Delta N213S \) (TSV1, TSV152, TSV141, TSV164, CAM60), of which four isolates had documented CNS disease (one individual died before diagnosis, CAM60), and four cases resulted in death (one individual made an unexpected recovery while in intensive care, TSV 152). These isolates were collected from as early as 1996, with the latest in 2012 and were significantly geographically dispersed (350-850 km apart). Additionally, four of these same isolates (TSV1, TSV152, TSV164, CAM60) and isolate TSV294 also possessed an upstream mutation \( \Delta R11H \). Additionally, all isolates
carrying \textit{bimA}_{Bm} also possess a truncated \textit{bimC} gene which plays a role in intracellular spread and lies upstream of \textit{bimA}_{Bm}.

From the 76 isolates we were able to identify 27 known multi-locus sequence types (STs) from 38 isolates (50%) (Table 3). From these, at least three sequence types (109, 151, 1667) have also been identified in the Northern Territory. Of the 38 novel sequence types, 26 isolates were comprised of previously described alleles in a novel combination, and 12 were due to a SNP present in one of the seven alleles (Figure 3). Novel allele variants were identified across six of the alleles, with ace the only allele without any variants (n=12). In each novel case, sequence types were confirmed via read mapping to all seven alleles. In total 34 of the 38 isolates carried unique novel STs, as two of the STs occurred twice and one ST occurred three times in the sample set. Among the 20 isolates carrying the \textit{bimA}_{Bm} variant, 13 (65%) carried novel STs, over previously described STs (n=7, 35%; Figure 3).

Phylogenomic analysis revealed the majority of Queensland isolates with novel STs branched significantly closer to the Queensland \textit{bimA}_{Bm} reference genome MSHR668, than to the Northern Territory and Asian genomes. This was also true for the \textit{bimA}_{Bm} variant. Two separate lineages can be identified in the phylogeny (point of divergence is approximately in the middle of the phylogeny), with the three clades at the bottom half of the phylogeny carrying 33 of the 38 novel STs identified (86.8%) and 18 of the 20 \textit{bimA}_{Bm} variants present in the sample set (90%). MLST diversity did not appear to be influenced by or correlated with the date the isolate was collected (Figure 3). No Queensland isolate was identical to that of a Northern Territory isolate, with limited clustering with Northern Territory or other Australian isolates. Additionally, five of the Queensland isolates included in this study clustered within the Australian/Asian clade. In this clade the reference genome K96243 from
Thailand is the most basal, with BPC006 from China branching within the Australian isolates CAM2, 112, 189, TSV38 and 95. All of the isolates within this clade were of known ST and did not carry the \( bimA_{Bm} \) variant.

Approximately one third of the genomes did not contain any AMR related mutations at all (35.5%). The pre-cursor mutation to imipenem resistance (\( \text{penA} \Delta T153A \)) was identified in 57.9% of all isolates. Only mutations associated with ceftazidime and meropenem resistance were predicted within the sample set. Ceftazidime resistance (loss of the PBP3 homolog BPSS1219) was only predicted from one isolate (1.3%), while meropenem resistance was predicted to be caused by the loss of function of the \( amrR \) efflux pump (BPSL1805) in 11 of the 76 isolates (14.5%). Meropenem resistance was encoded in 15% of \( bimA_{Bm} \) isolates and 16% of \( bimA_{Bp} \) isolates, respectively (Table 2).
Discussion

Variation in the \textit{bimA}_{Bm} gene was present and extensive. Twenty isolates carried the virulence factor and only two of the sequences produced from this study were identical (TSV39, TSV25). No sequence generated from this study was identical to any of the previously described reference sequences, suggesting depth of variation in the \textit{bimA}_{Bm} gene is extensive and yet to be seen (Figure 2). Since the 2000’s sequence variation of \textit{BimA}_{Bm} in \textit{B. pseudomallei} has not been discussed in great detail, despite being studied recently\cite{20}. For example, five isolates possessed a ΔN213S substitution, positioned between the proline rich region and the YadA-like head domain, a highly conserved region\cite{21}. Interestingly, the isolates with this mutation had a mortality rate of 80% and documented CNS disease at a rate of 80% (one patient died before CNS disease could be diagnosed). Unfortunately, similar to previous studies the effect these variants have upon the function of the protein and resulting virulence remain unclear\cite{21}. Reference sequences from MSHR33, 491 and 172 all possess the mutation as well, suggesting this is not a recent event\cite{21}. However, such high mortality rates and disease severity suggests that isolates with this type of mutation may be more virulent than those without.

Previous studies have suggested that \textit{bimA}_{Bm} carrying isolates are 14 times more likely to develop CNS disease than the wild-type \textit{bimA}_{Bp} isolates\cite{5}. Explorative analysis of this data suggested a significant association between the \textit{bimA}_{Bm} variant and the development of CNS disease, compared to the wild type. However, as this dataset was selected for particular factors (CNS disease and postcode) these numbers should be viewed with caution. Perhaps what is more representative of the Queensland population is the 11 \textit{bimA}_{Bm} variants identified from the 58 non-CNS disease isolates, suggesting 19% of \textit{B. pseudomallei}
infections will carry the virulence factor. Indeed, larger sample sizes collected at random will be needed to confirm these numbers and this should be considered for future studies. Sarovich et al., also assessed the likelihood of clinical co-variants that may be associated with CNS disease, but did not find other associations[5]. Due to the lack of complete clinically relevant data surrounding the isolates in this study, we were unable to calculate if this was applicable to our results.

In this study, CNS disease isolates were evenly split between \(\text{bim}_{A_{bp}}\) and \(\text{bim}_{A_{bm}}\) genes and more isolates presented without CNS disease and \(\text{Bim}_{A_{bm}}\) (n=11), than with both of these variables (n=9). Therefore, it is likely this gene is not the only factor driving the development of CNS disease. This was also evident in the Northern Territory, with 85.5% of isolates carrying the variant, not diagnosed with CNS disease. Additionally, 1% presented with CNS disease, but no variant[5]. In both studies this is a significant proportion of \(\text{Bim}_{A_{bm}}\) isolates that have not developed CNS disease. An environmental study[6] of both \(\text{Bim}_{A_{bm}}\) and another virulence factor lipopolysaccharide (LPS) suggested the \(\text{Bim}_{A_{bm}}\) virulence factor was more likely to occur with LPS genotype B, the more prevalent genotype in Australia. Genotype A, is more prevalent in Thailand and Southeast Asia, where CNS disease is less frequently reported[5,6]. Furthermore, the gene upstream of \(\text{bim}_{A_{bm}}, \text{bim}_{C}\) was also assessed in this study. As both \(\text{bim}_{A}\) and \(\text{bim}_{C}\) genes have been shown to play a role in the intracellular spread of \(B.\ pseudomallei\)[22]. As all \(\text{bim}_{A_{bm}}\) isolates in this study possessed a truncated \(\text{bim}_{C}\) gene, the affect this truncation has upon virulence and the role of \(\text{bim}_{C}\) in intracellular spread remains unknown. However, it suggests the variant is significant to virulence in these strains and should be further investigated. There is no way to show a definitive correlation between these two virulence factors and \(\text{bim}_{A_{bm}}\) in Queensland without including both clinical and environmental isolates in future studies. The authors
suggest a pan-genomic analysis coupled with extensive clinical metadata from as many isolates as possible representing both states evenly, to be the best approach at identifying all genes associated with CNS disease. It is possible that there may be some combination of genes derived only from the unique genetic diversity of Australian isolates responsible for the increased incidences of CNS disease observed here in Australia compared to those overseas.

Although Queensland has been included in MLST diversity analyses before, novel STs have not been described, which is significant given 50% of isolates in this study were novel STs (Figure 3, Table 3)[23,24]. A large number of isolates (34%) were comprised of a previously undescribed combination of alleles to generate a novel ST. It also appears that isolates with the \textit{bimA\textsubscript{bm}} virulence factor are more likely to be novel STs. This implies that Queensland isolates are exchanging genetic material[24]. The exchange of genetic material was also evidenced by the lack of clonality seen in both previously described and novel STs in Queensland. ST 70 was the most commonly described (6.5%) and was identified in both Brisbane and Cairns isolates suggesting one ST does not dominate certain spatial areas in this study, despite this being reported for the Northern Territory[6,23]. Furthermore, the same studies reported that Queensland and the Northern Territory do not share any common STs. However, in this study we identified at least three STs shared between the states (STs 109, 151, 1667), with the potential of more, as the exact location was not recorded for all STs in pubMLST. The identification of shared STs here, may be due to the previous under-representation, or exclusion of Queensland in \textit{B. pseudomallei} isolates in previous studies[23–26].
Genomic analyses also identified the loss of function of the amrR regulator (amrAB-OprA efflux pump), associated with meropenem resistance in 14.5% of isolates, however this has not been validated phenotypically[15]. A pre-cursor mutation in penA ΔT153A, was common amongst the isolates and in combination with other missense and promoter mutations can cause imipenem and amoxicillin-clavulanic acid resistance[15]. This is not overly worrisome as imipenem is not used to treat B. pseudomallei in Australia, meropenem is the preferred carbapenem, however amoxicillin-clavulanic acid is sometimes used in eradication therapy[3]. Phenotypically, antimicrobial resistance circulating in the Northern Territory comprises of approximately 3% for doxycycline and 0.9% for trimethoprim-sulfamethoxazole[27], while isolates remain susceptible to meropenem and ceftazidime[1,27,28]. This data suggests AMR may be rarely encountered in Queensland isolates and the current selection of antimicrobials for use against B. pseudomallei infections will be effective.

**Conclusion**

This study has revealed a significant amount of genetic diversity in Queensland B. pseudomallei isolates, such as novel MLST sequence types and unique bimA<sub>BM</sub> gene sequences. The virulence factor bimA<sub>BM</sub> is linked to CNS disease, yet it is suspected there are multiple drivers for this type of infection. Further exploration into the virulence factors responsible for CNS disease should focus on maximum sample size, pan-genomics, detailed clinical metadata and include environmental samples. Identification of CNS disease drivers may provide an excellent therapeutic target or act as a screening test to warn clinicians and prompt additional investigations such as a lumbar puncture or MRI.
Funding

This study/project has been funded by a Research Grant 2020 of the European Society of Clinical Microbiology and Infectious Diseases (ESCMID) to DB.

Acknowledgements

The authors would like to thank Valentine Murigneux for the use of, and their assistance in their long-read assembly pipeline MicroPIPE.

Disclosures

DLP has received research support from Shionogi, MSD and Pfizer and honoraria for speaking or advisory boards from MSD, Pfizer, Sumitomo, Biomerieux, Accelerate and Lyosvant. PNAH reports grants from Shionogi, Merck (MSD) and Sandoz and honoraria for speaking or advisory boards from Pfizer, Sandoz and Sumitomo. All other authors declare no conflicts of interest.
Author contribution

DB: conceptualization, methodology, analysis, writing. MJB: methodology, analysis, writing-review and edits. CF: methodology, analysis, writing-review and edits. IG: methodology, writing-review and edits. REN: data curation, writing-review and edits. PNAH: supervisor, writing-review and edits. DLP: supervisor, writing-review and edits.
Table 1. Metadata regarding the *B. pseudomallei* isolates included in this study.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Date range of collection</th>
<th>Geographical range</th>
<th>Gender</th>
<th>Age range</th>
<th>Central Nervous System Disease</th>
<th>Mortality*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1996-2018</td>
<td>4000-4895 Queensland, Australia</td>
<td>31 (40.8%) Female, 45 (59.2%) Male</td>
<td>6-84 years</td>
<td>18 (23.7%)</td>
<td>15 (19.7%)</td>
</tr>
</tbody>
</table>

**Isolation sites:**

<table>
<thead>
<tr>
<th>Isolation sites</th>
<th>Isolates (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blood</td>
<td>42</td>
</tr>
<tr>
<td>Pus</td>
<td>8</td>
</tr>
<tr>
<td>Sputum</td>
<td>7</td>
</tr>
<tr>
<td>Cerebral spinal fluid</td>
<td>4</td>
</tr>
<tr>
<td>Tissue</td>
<td>4</td>
</tr>
<tr>
<td>Brain</td>
<td>2</td>
</tr>
<tr>
<td>Endotracheal aspirate</td>
<td>2</td>
</tr>
<tr>
<td>Bronchoalveolar Lavage</td>
<td>1</td>
</tr>
<tr>
<td>Gastrointestinal Tract</td>
<td>1</td>
</tr>
<tr>
<td>Liver</td>
<td>1</td>
</tr>
<tr>
<td>Lung aspirate</td>
<td>1</td>
</tr>
<tr>
<td>Lymph</td>
<td>1</td>
</tr>
<tr>
<td>Urine</td>
<td>1</td>
</tr>
<tr>
<td>Unknown</td>
<td>1</td>
</tr>
<tr>
<td><strong>Total:</strong></td>
<td>76</td>
</tr>
</tbody>
</table>

*only information for 23/76 isolates was available.

Table 2. Comparison of clinical presentation, sequence type and antimicrobial resistance of isolates carrying the *bimA<sub>Bm</sub>* or *bimA<sub>Ps</sub>* virulence factor.

<table>
<thead>
<tr>
<th>Variable</th>
<th><em>bimA&lt;sub&gt;Bm&lt;/sub&gt;</em> Isolates (%)</th>
<th><em>bimA&lt;sub&gt;Ps&lt;/sub&gt;</em> Isolates (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNS presentation</td>
<td>45 (%)</td>
<td>16 (%)</td>
</tr>
<tr>
<td>Novel ST</td>
<td>40 (53.6)</td>
<td></td>
</tr>
<tr>
<td>Known ST</td>
<td>60 (46.4)</td>
<td></td>
</tr>
<tr>
<td><em>amrR</em> mutation</td>
<td>15 (16)</td>
<td></td>
</tr>
<tr>
<td>Pre-cursor mutation</td>
<td>45 (64.3)</td>
<td></td>
</tr>
<tr>
<td>No AMR mutation</td>
<td>35 (35.7)</td>
<td></td>
</tr>
<tr>
<td>Mortality*</td>
<td>40 (12.5)</td>
<td></td>
</tr>
</tbody>
</table>
*Information was only available for 23/76 isolates

Table 3. Multilocus sequence types of clinical *Burkholderia pseudomallei* from Queensland, included in this study.

<table>
<thead>
<tr>
<th>Sequence Type</th>
<th>Occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>2</td>
</tr>
<tr>
<td>35</td>
<td>1</td>
</tr>
<tr>
<td>70</td>
<td>5</td>
</tr>
<tr>
<td>109</td>
<td>1</td>
</tr>
<tr>
<td>151</td>
<td>1</td>
</tr>
<tr>
<td>235</td>
<td>1</td>
</tr>
<tr>
<td>252</td>
<td>3</td>
</tr>
<tr>
<td>254</td>
<td>1</td>
</tr>
<tr>
<td>257</td>
<td>2</td>
</tr>
<tr>
<td>283</td>
<td>1</td>
</tr>
<tr>
<td>286</td>
<td>1</td>
</tr>
<tr>
<td>591</td>
<td>2</td>
</tr>
<tr>
<td>593</td>
<td>2</td>
</tr>
<tr>
<td>594</td>
<td>1</td>
</tr>
<tr>
<td>958</td>
<td>1</td>
</tr>
<tr>
<td>1041</td>
<td>1</td>
</tr>
<tr>
<td>1042</td>
<td>1</td>
</tr>
<tr>
<td>1453</td>
<td>1</td>
</tr>
<tr>
<td>-------</td>
<td>---</td>
</tr>
<tr>
<td>1667</td>
<td>2</td>
</tr>
<tr>
<td>1707</td>
<td>1</td>
</tr>
<tr>
<td>1711</td>
<td>1</td>
</tr>
<tr>
<td>1712</td>
<td>1</td>
</tr>
<tr>
<td>1754</td>
<td>1</td>
</tr>
<tr>
<td>1756</td>
<td>1</td>
</tr>
<tr>
<td>1757</td>
<td>1</td>
</tr>
<tr>
<td>1760</td>
<td>1</td>
</tr>
<tr>
<td>1769</td>
<td>1</td>
</tr>
</tbody>
</table>

**Novel**

38

(26 = novel allele combination, 12 = SNP in allele)

**Figure Legends**

**Figure 1.** Distribution of *B. pseudomallei* isolates with the *bimA*<sub>bm</sub> virulence factor in north Queensland.

**Figure 2.** Variation of the *BimA*<sub>bm</sub> proline-rich region identified in Queensland *B. pseudomallei* isolates compared to the Queensland *BimA*<sub>bm</sub> reference genome MSHR668.

**Figure 3.** Maximum likelihood phylogenomic tree of Australian *B. pseudomallei* isolates, rooted to reference MSHR668. Bootstrap support ≥80 is shown at nodes in red and
reference genomes are labelled with strain name stated in NCBI. Tree was built using iQtree

100 genomes and a 4,950,632 bp alignment generated from parSNP.


[19] Stata Statistical Software. College Station, TX: StataCorp LLC; 2019.


