

RESEARCH ARTICLE

Tracing the environmental footprint of the *Burkholderia pseudomallei* lipopolysaccharide genotypes in the tropical “Top End” of the Northern Territory, Australia

Jessica R. Webb^{1*}, Audrey Rachlin¹, Vanessa Rigas¹, Derek S. Sarovich^{1,2}, Erin P. Price^{1,2}, Mirjam Kaestli^{1,3}, Linda M. Ward¹, Mark Mayo¹, Bart J. Currie^{1,4}

1 Global and Tropical Health Division, Menzies School of Health Research, Darwin, Northern Territory, Australia, **2** GeneCology Research Centre, University of the Sunshine Coast, Sippy Downs, Queensland, Australia, **3** Research Institute for the Environment and Livelihoods, Charles Darwin University, Darwin, Northern Territory, Australia, **4** Department of Infectious Diseases and Northern Territory Medical Program, Royal Darwin Hospital, Darwin, Northern Territory, Australia

* Jessica.Webb@menzies.edu.au



OPEN ACCESS

Citation: Webb JR, Rachlin A, Rigas V, Sarovich DS, Price EP, Kaestli M, et al. (2019) Tracing the environmental footprint of the *Burkholderia pseudomallei* lipopolysaccharide genotypes in the tropical “Top End” of the Northern Territory, Australia. *PLoS Negl Trop Dis* 13(7): e0007369. <https://doi.org/10.1371/journal.pntd.0007369>

Editor: Alfredo G. Torres, University of Texas Medical Branch, UNITED STATES

Received: April 4, 2019

Accepted: July 4, 2019

Published: July 26, 2019

Copyright: © 2019 Webb et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All of the 175 genomes used for comparative analysis are available on NCBI and the corresponding accession numbers are provided in [S1 Table](#).

Funding: The research was funded under the Australian National Health and Medical Research Council grant numbers 1046812, 1098337, and 1131932 (The HOT NORTH initiative). DSS and EPP are supported by Advance Queensland Fellowships (AQR13016-17RD2 and

Abstract

The Tier 1 select agent *Burkholderia pseudomallei* is an environmental bacterium that causes melioidosis, a high mortality disease. Variably present genetic markers used to elucidate strain origin, relatedness and virulence in *B. pseudomallei* include the *Burkholderia* intracellular motility factor A (*bimA*) and filamentous hemagglutinin 3 (*fhaB3*) gene variants. Three lipopolysaccharide (LPS) O-antigen types in *B. pseudomallei* have been described, which vary in proportion between Australian and Asian isolates. However, it remains unknown if these LPS types can be used as genetic markers for geospatial analysis within a contiguous melioidosis-endemic region. Using a combination of whole-genome sequencing (WGS), statistical analysis and geographical mapping, we examined if the LPS types can be used as geographical markers in the Northern Territory, Australia. The clinical isolates revealed that LPS A prevalence was highest in the Darwin and surrounds (n = 660; 96% being LPS A and 4% LPS B) and LPS B in the Katherine and Katherine remote and East Arnhem regions (n = 79; 60% being LPS A and 40% LPS B). Bivariate logistics regression of 999 clinical *B. pseudomallei* isolates revealed that the odds of getting a clinical isolate with LPS B was highest in East Arnhem in comparison to Darwin and surrounds (OR 19.5, 95% CI 9.1–42.0; $p < 0.001$). This geospatial correlation was subsequently confirmed by geographically mapping the LPS type from 340 environmental Top End strains. We also found that in the Top End, the minority *bimA* genotype *bimA_{Bm}* has a similar remote region geographical footprint to that of LPS B. In addition, correlation of LPS type with multi-locus sequence typing (MLST) was strong, and where multiple LPS types were identified within a single sequence type, WGS confirmed homoplasmy of the MLST loci. The clinical, sero-diagnostic and vaccine implications of geographically-based *B. pseudomallei* LPS types, and their relationships to regional and global dispersal of melioidosis, require global collaborations with further analysis of larger clinically and geospatially-linked datasets.

AQIRF0362018, respectively). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Author summary

Burkholderia pseudomallei is a pathogenic soil bacterium that causes the disease melioidosis, which occurs in many tropical regions globally and in recent years has emerged in non-tropical regions. Melioidosis has been predicted to affect 165,000 people every year resulting in an estimated 89,000 deaths. Person to person transmission is rare with most cases linked to exposure to the bacterium from the environment. The genetic background of *B. pseudomallei* has been well studied and variably present genes have been linked to distinct melioidosis disease states and geographic regions, however we still need a stronger understanding of the association of genes with geography. Three lipopolysaccharide types exist in *B. pseudomallei* and the prevalence of the lipopolysaccharide genes vary between melioidosis endemic regions, but it is unknown if the lipopolysaccharide genes can be used as geographical markers in a single melioidosis-endemic region. In this study, we used a combination of whole-genome sequencing, statistics and geographical mapping to elucidate if the three lipopolysaccharide genes can be used as geographical markers within the Northern Territory, Australia. We show that the three LPS types have distinct but overlapping geographical footprints within a single melioidosis region and can be used as geographic markers alongside a number of other important variably present *B. pseudomallei* genes.

Introduction

Burkholderia pseudomallei is the causative agent of melioidosis, an often fatal disease that is endemic in tropical regions globally, including the “Top End” of the Northern Territory, Australia [1]. Clinical presentations of melioidosis are highly varied [2–5] with pneumonia being the most common presentation [6]. Melioidosis mortality rates range from <10% in the NT to >40% in regions of Asia and Southeast Asia where prompt diagnosis and accessibility to drugs and hospital care can be limited [2, 6, 7]. Global melioidosis mortality has been predicted to be considerably higher than mortality from dengue and leptospirosis combined and similar to that from measles [8]. In addition, the bio-threat status of *B. pseudomallei* and the lack of a current vaccine makes this organism of high importance to public health in many regions globally.

Variable genetic markers in *B. pseudomallei* with known geographical associations include the mutually exclusive *Burkholderia* intracellular motility factor A (BimA) variants *bimA*_{Bm} and *bimA*_{Bp}, the filamentous hemagglutinin 3 (*fhaB3*) gene and the *Burkholderia thailandensis*-like flagellum and chemotaxis (BTFC/YLF) gene clusters [9, 10]. These genetic markers have been used to elucidate strain relatedness, origin and virulence potential, with *bimA*_{Bm} being strongly linked to neurological melioidosis in Australian studies [9]. The lipopolysaccharide (LPS) of *B. pseudomallei*, a type II O-polysaccharide [11], consists of A, B and B2 variants [12]. As with *bimA*_{Bm}/*bimA*_{Bp}, YLF/BTFC, and *fhaB3*, the distribution of LPS genotypes varies between Australia and Southeast Asia [9, 12] and it is unknown what LPS types predominant in other endemic regions. In Thailand and Australia, LPS type A is the most abundant, followed by LPS B, whereas LPS B2 has not been detected in Thailand but has been detected in Australia [12]. Despite regional differences in LPS genotype prevalence, the distribution of these types across different geographical regions within a single melioidosis-endemic region has not been thoroughly investigated. Elucidating the geospatial patterns of LPS genotypes is potentially important for *B. pseudomallei* virulence studies, the development of LPS-based sero-diagnostics, and for consideration of vaccines that have an LPS component.

In this study, we used a combination of WGS, bivariate logistic regression (with the cluster feature) and geographical mapping to examine the geographical distribution of LPS types in the Top End region. We first investigated LPS geographic correlations using a large number of clinical isolates ($n = 1005$) from our 28-year Darwin Prospective Melioidosis study (DPMS). We then investigated LPS type distributions for 340 environmental strains collected from known Top End locations to determine whether LPS genotypes for clinical isolates, which have usually only presumptive infection locale data, matched that of the environmental isolates, where location is definite. Lastly, we determined *bimA*_{Bm}/*bimA*_{Bp} diversity in the 340 environmental strains to determine whether the *bimA* variants and LPS types share similar geographical footprints in the NT.

Methods

Ethics statement

This study was approved by the Human Research Ethics Committee of the NT Department of Health and the Menzies School of Health Research (HREC 02/38).

B. pseudomallei isolates used in this study

A total of 1,345 *B. pseudomallei* isolates from the Top End region were used in this study for bivariate analysis and geographical mapping. 1,005 clinical isolates were included in this study and 999 (excluding the 6 B2 strains) were used for bivariate statistical analysis, and environmental isolates ($n = 340$) were used for geographical mapping. The 1,005 clinical isolates were primary *B. pseudomallei* isolates corresponding to 1,005 patients enrolled in the 28-year DPMS. The clinical isolates were assigned to one of four Top End geographic regions based on the presumptive location of infection, which is assigned based on the individual patient's history of illness and activities at their residential location or elsewhere if considered relevant. These regions have been segregated and assigned as outlined by the Australian Bureau of Statistics (<https://www.abs.gov.au/>; statistical area level 3): Darwin and surrounds; Darwin remote (includes West Arnhem), Katherine and Katherine remote (representing one region, Fig 1), and East Arnhem (Fig 1). Melioidosis is very uncommon in the southern half of the NT and this region was not included in our study.

The 340 Top End environmental isolates were processed and subsequently *B. pseudomallei* was confirmed as previously described [13–17]. The location for each of the 340 environmental isolates was determined using Global Positioning System coordinates, with mapping of each isolate coordinate using ArcGIS (<https://www.arcgis.com/index.html>) and maps were exported as JPEG and manipulated using INKSCAPE 0.92 (<https://inkscape.org/release/inkscape-0.92/>). For comparison, we included an additional 61 isolates of Queensland ($n = 38$), Western Australian ($n = 12$) and Papua New Guinean ($n = 11$) origin.

LPS and *bimA* genotypes

The assignment of LPS types for the 1,005 clinical genomes has previously been determined [18] using the Basic Local Alignment Search Tool (BLAST). The clinical genomes were not assigned a *bimA* variant as this has previously been performed on our clinical strains [10]. LPS and *bimA* genotypes were determined for each of the 340 Top End environmental genomes, with an LPS type being assigned to those additional 61 isolates from Western Australia, Papua New Guinea and Queensland, Australia. In brief LPS A (*wbil* to *apaH* in K96243 [GenBank ref: NC_006350]), LPS B (*BUC_3392* to *apaH* in *B. pseudomallei* 579 [GenBank ref: NZ_ACCE01000003]), LPS B2 (*BURP840_LPSb01* to *BURP840_LPSb21* in *B. pseudomallei*

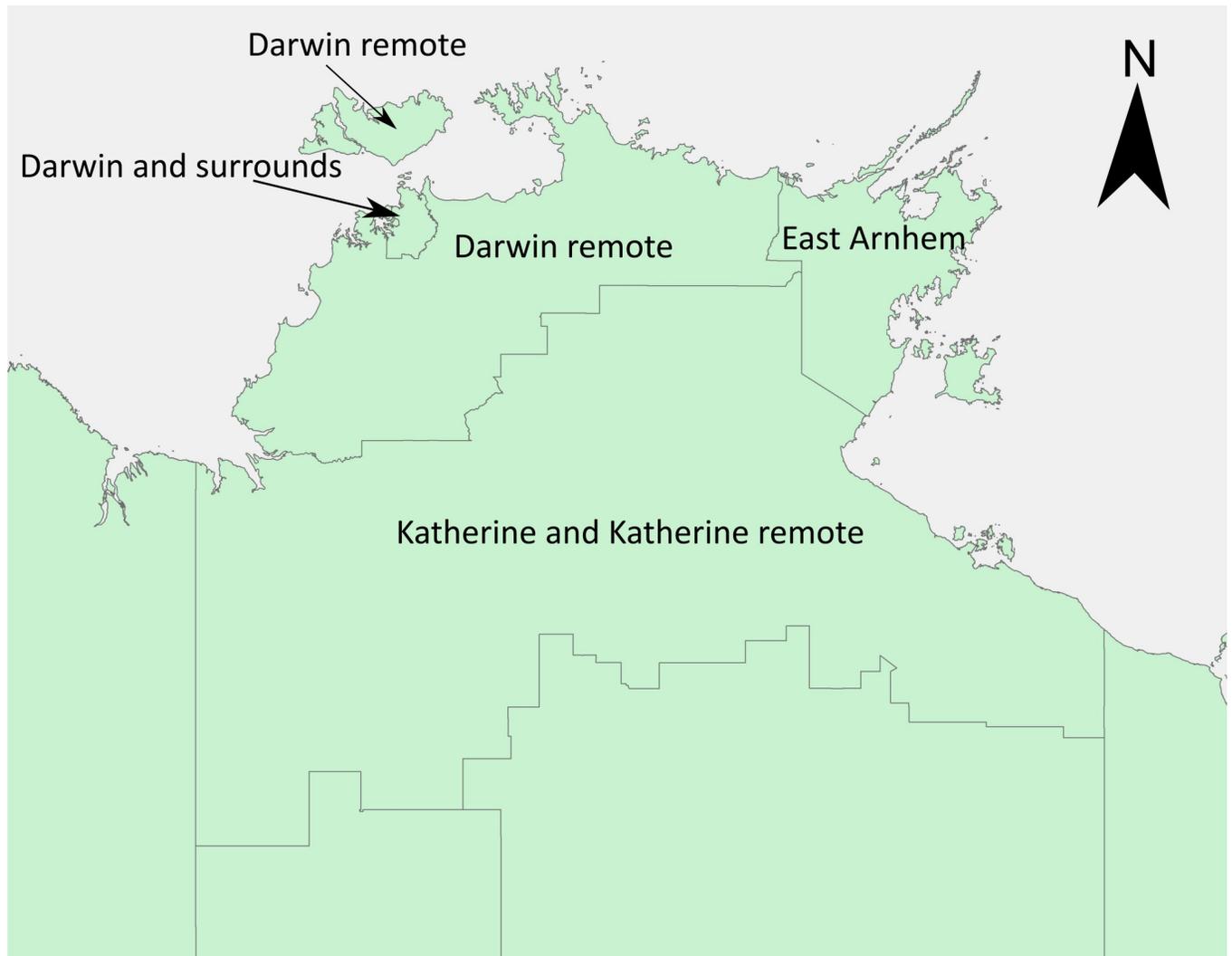


Fig 1. Geographical origin of the *Burkholderia pseudomallei* clinical and environmental isolates within the Top End region of the Northern Territory, Australia, that were used in this study. The map was created using ArcGIS (<https://www.arcgis.com/index.html>).

<https://doi.org/10.1371/journal.pntd.0007369.g001>

MSHR840 [GenBank ref: GU574442]), *bimA_{Bm}* (BURPS668_A2118 in *B. pseudomallei* MSHR668 [GenBank ref: NZ_CP009545]) and *bimA_{Bp}* (BPSS1492 in *B. pseudomallei* K96243) were BLAST-searched against an in-house *B. pseudomallei* database containing all genomes in this study using the nucleotide BLAST (BLASTn) parameter.

Statistical analysis of LPS type, locale, and multilocus sequence type (MLST) using the clinical *B. pseudomallei* dataset (n = 999)

For each clinical strain, the presumptive geographical location of infection and its sequence type (ST) (<https://pubmlst.org/bpseudomallei/>) were analysed using Stata version 14.2 (Stata-Corp LP, College Station, TX, USA). Bivariate logistics regression using the cluster feature was performed for LPS type with locale using the Darwin and surrounding region as the reference level, and bivariate Pearson's χ^2 (frequency >5 in all cells) was performed for LPS type with ST or BimA variants: A $P < 0.05$ was considered significant.

Phylogenomic analysis of 175 *B. pseudomallei* genomes

We performed comparative analyses of 175 *B. pseudomallei* genomes that represent both the global diversity of this bacterium and within-Northern Territory population diversity (S1 Table). Of these 175 genomes, 144 were publicly available and 31 were sequenced as part of this study. Genomic DNA for 31 isolates lacking WGS data was extracted as previously described [19]. These isolates were sequenced at Macrogen, Inc. (Gasan-dong, Seoul, Republic of Korea) or Australian Genome Research Facility Ltd. (Melbourne, Australia) using the Illumina HiSeq2000 and Illumina HiSeq2500 platforms (Illumina, Inc., San Diego, CA) and these data are available on NCBI (S1 Table)

Comparative analysis of the 175 genomes was carried out to investigate geographical clustering of the LPS types, to determine the genetic relatedness of suspected ST homoplasmy cases [20, 21], and to ascertain if suspected ST homoplasmy cases were intercontinental or intracontinental. Orthologous biallelic single-nucleotide polymorphisms (SNPs) were identified from the WGS data using the default settings of SPANDx v3.2 [22]. The closed Australian *B. pseudomallei* genome MSHR1153 [23] was used as the reference for read mapping. A maximum-parsimony (MP) phylogenetic tree was reconstructed based on 219,075 SNPs identified among the 175 genomes using PAUP* 4.0.b5 [24]. Recombinogenic SNPs were identified using Gubbins v2.2.0 (default parameters) [25]. STs were determined using the BIGSdb tool [26].

Results

LPS B is linked to remote Top End regions of the Northern Territory

BLAST analysis demonstrated that overall in the Top End, LPS A was dominant in both environmental *B. pseudomallei* strains (A = 89%; LPS B = 10% and LPS B2 = 1%) and in clinical strains (LPS A = 87%; LPS B = 12% and LPS B2 = 1%) (Table 1). Using the clinical *B. pseudomallei* dataset, we performed bivariate logistics regression (with the cluster feature) of LPS type A and B correlations and presumptive infecting locale, using Darwin urban as the reference level. Low LPS B2 numbers (n = 6) precluded statistical analysis of this genotype, leaving 999 DPMS isolates for the analysis. Of these six B2 isolates, five were from presumptive infections occurring in Katherine and Katherine remote communities, with the remaining LPS B2 strain linked to infection in the Darwin and surrounding region. Bivariate logistics regression adjusted with the cluster feature demonstrated that the odds of getting a clinical isolate with LPS B in the Darwin remote, Katherine and Katherine remote and East Arnhem regions, respectively is 4.7 (95% CI 2.4–9.6; $p < 0.001$), 14.6 (95% CI 6.1–34.7; $p < 0.001$) and 19.5 (95% CI 9.1–42.0; $p < 0.001$) times higher compared to the odds in the Darwin and surrounding region. LPS B prevalence was highest in patients infected in the Katherine and Katherine

Table 1. The geographic locale of LPS genotype A and B in *Burkholderia pseudomallei* isolated from melioidosis patients and from the environment in the Top End region, Northern Territory, Australia.

Geographic region	N	Clinical isolates		N	Environmental isolates	
		LPS A (observed %)	LPS B (observed %)		LPS A (observed %)	LPS B (observed %)
Darwin and surrounds	686	660 (96.2)	26 (3.8%)	236	234 (99.2%)	2 (0.8%)
Darwin remote	114	96 (84.2)	18 (15.8)	27	23 (85.2%)	2 (14.8%)
Katherine and Katherine remote	107	68 (63.6)	39 (36.4)	72	42 (58.3%)	30 (41.7%)
East Arnhem	92	52 (56.5)	40 (43.5)	4	2 (50%)	2 (50%)
Isolate total	999	876	123	339	303	36

Abbreviations: LPS, lipopolysaccharide

<https://doi.org/10.1371/journal.pntd.0007369.t001>

remote and East Arnhem regions (n = 79; 64% of LPS B isolates), whereas LPS A prevalence was highest in patients infected in the Darwin and surrounds and Darwin remote regions (n = 756; 86% of LPS A isolates).

Geographic clustering of LPS types was also observed in NT environmental strains

Geographical mapping of LPS types for the environmental isolates also supported the link between LPS and geography. LPS type A was predominant in Darwin and surrounds (Table 1 and Fig 2B), whilst LPS B was found in all four geographical regions but with numbers highest in the Katherine and Katherine remote region (Fig 2A and 2B, number of LPS B strains at each site is indicated). LPS B2 from environmental strains was only observed in the Katherine remote region (n = 1). In non-Top End comparator strains, LPS A was also the dominant genotype (range = 63–91%), followed by B (range = 9–24%) and B2 (range = 0–18%). Numbers of the LPS type A, B, and B2, respectively, were as follows: Western Australia (10, 2, 0), Queensland (24, 9, 5) and Papua New Guinea (8, 1, 2).

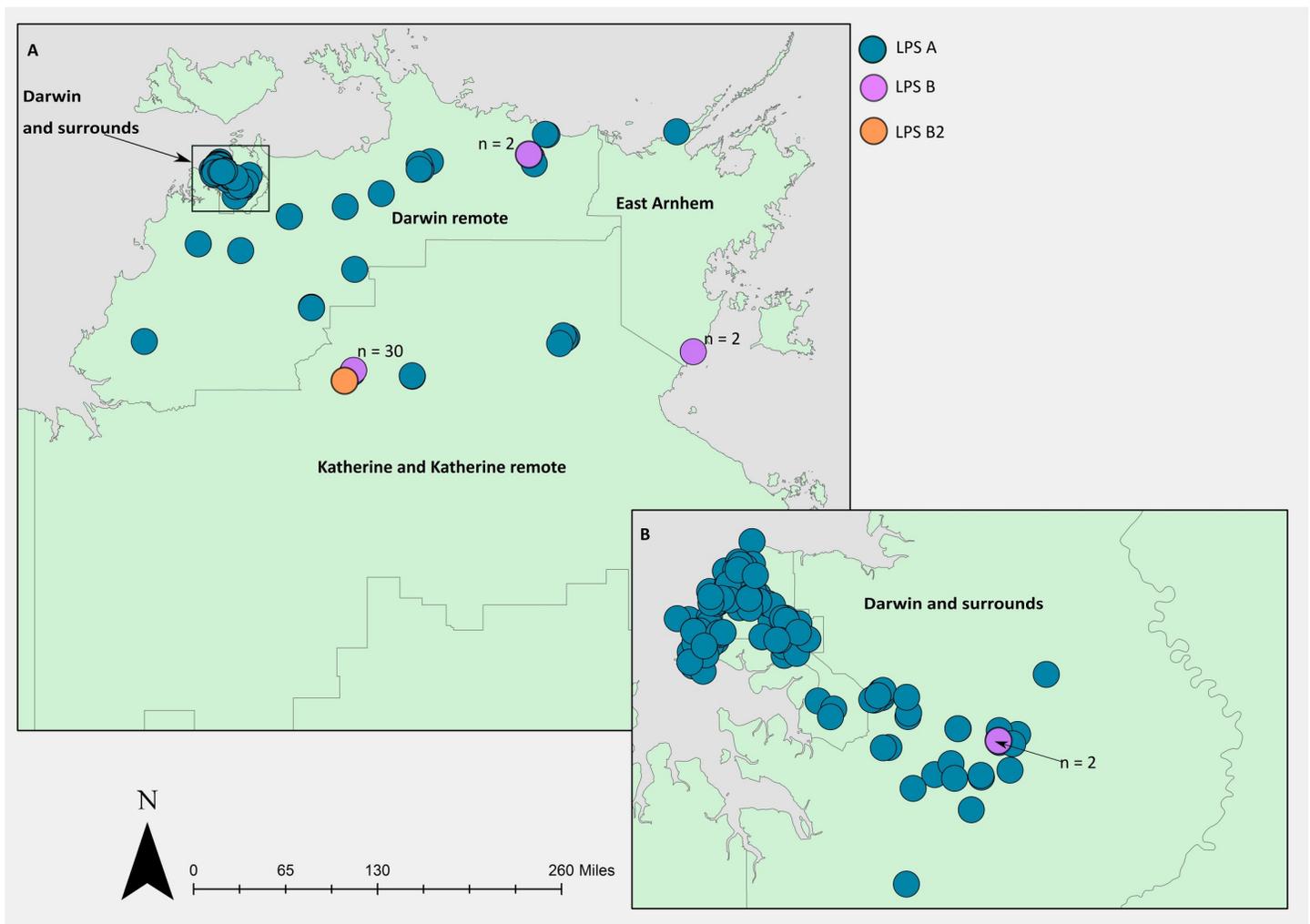


Fig 2. Geographical distribution of LPS genotypes in 340 environmental *Burkholderia pseudomallei* isolates from (A) the Top End region, Northern Territory, Australia and (B) Darwin and surrounds. The number of LPS B strains at each of the four LPS B regions is indicated. The map was created using ArcGIS (<https://www.arcgis.com/index.html>).

<https://doi.org/10.1371/journal.pntd.0007369.g002>

Bivariate analysis revealed a strong correlation between LPS A and STs found only in the Darwin and surrounds and Darwin remote regions

Bivariate analysis of clinical isolates encoding LPS types A and B revealed statistically significant associations between LPS types and STs. LPS A was significantly associated with STs only found in the Darwin and surrounding region, including ST-109 (n = 126; 14% of LPS A's; P<0.001), ST-36 (n = 81; 9% of LPS A's; P<0.001) and ST-132 (n = 67; 8% of LPS A's; P<0.001), whilst LPS B was associated with 88 rare STs that are uncommon in the Darwin regions (S2 Table). These associations therefore reflected the LPS geographical associations observed in clinical strains. Although LPS B2 was not included in the statistical analysis it was found in strains belonging to rare STs (ST-331 n = 1; ST-737 n = 2; ST-770 n = 2; ST-118 n = 1) with four of the STs unique to the Katherine and Katherine remote region. Of the STs identified in the 1,005 DPMS clinical strains, four contained multiple LPS types amongst 18 strains. Strains belonging to ST-468, ST-807 and ST-734 contained LPS A or LPS B, whilst strains belonging to ST-118 contained LPS A or LPS B2. The LPS and ST discrepancy was also observed in the Top End environmental strain cohort, with three strains belonging to ST-1485 containing LPS A or B.

The geographical footprint of *bimA_{Bm}* in environmental isolates is similar to LPS B

BLAST analysis demonstrated that *bimA_{Bp}* was the dominant BimA variant in Top End environmental samples (94.8%), similar to the *bimA_{Bp}* prevalence in clinical isolates from a previous study, whereby 95.7% of clinical isolates from Darwin and surrounds possessed *bimA_{Bp}* [10]. Bivariate statistical analysis demonstrated that the proportion of LPS B strains carrying *bimA_{Bm}* (n = 8, 21%) is higher compared to LPS A (n = 8, 2.6%) and this was significant (P<0.001). We next determined if *bimA_{Bp}* or *bimA_{Bm}* has a similar geographical footprint to LPS A or LPS B in the Top End. [10]. Geographical mapping of the *bimA* variants in the environmental strains demonstrated that *bimA_{Bp}* was linked with Darwin and surrounds (Fig 3B), having a similar geographical footprint as LPS A. In contrast, *bimA_{Bm}* was associated with the East Arnhem and the Katherine and Katherine remote region (13 of 17 (76%) environmental *bimA* strains), similar to LPS B. We also noted that the majority of the LPS B strains in the Katherine and East Arnhem region also carried *bimA_{Bm}* (n = 8/13; 62%), and these strains were concentrated in two geographical regions (Fig 3, star [n = 6/9] and triangle [n = 2/2]).

Phylogenomics revealed that LPS B2 is geographically restricted within the Top End

A phylogenomic analysis was performed to investigate the distribution of LPS types and strain relatedness on the whole-genome level (Fig 4). Overall, good clade structure based on geographical origin was observed, with strains belonging to the same ST clustering together and within the same LPS type, but with LPS A and B being dispersed throughout the phylogeny (Fig 4). However, this analysis revealed that 5/6 LPS B2 genomes of Top End origin clustered together, suggesting that LPS B2 is geographically restricted in the Northern Territory (Fig 4). Similarly, the LPS B2 strains with a Queensland, Australian origin grouped closely together.

The *B. pseudomallei* LPS discrepancy in five MLST STs can be explained by ST homoplasy (isolates with the same ST that are genetically distinct)

The strong link between LPS type, geography and ST suggested that homoplasy may be present in the 18 clinical strains and three environmental strains that belonged to the five STs with more than one LPS type (Table 2). Based on comparative analysis of 175 genomes, isolates

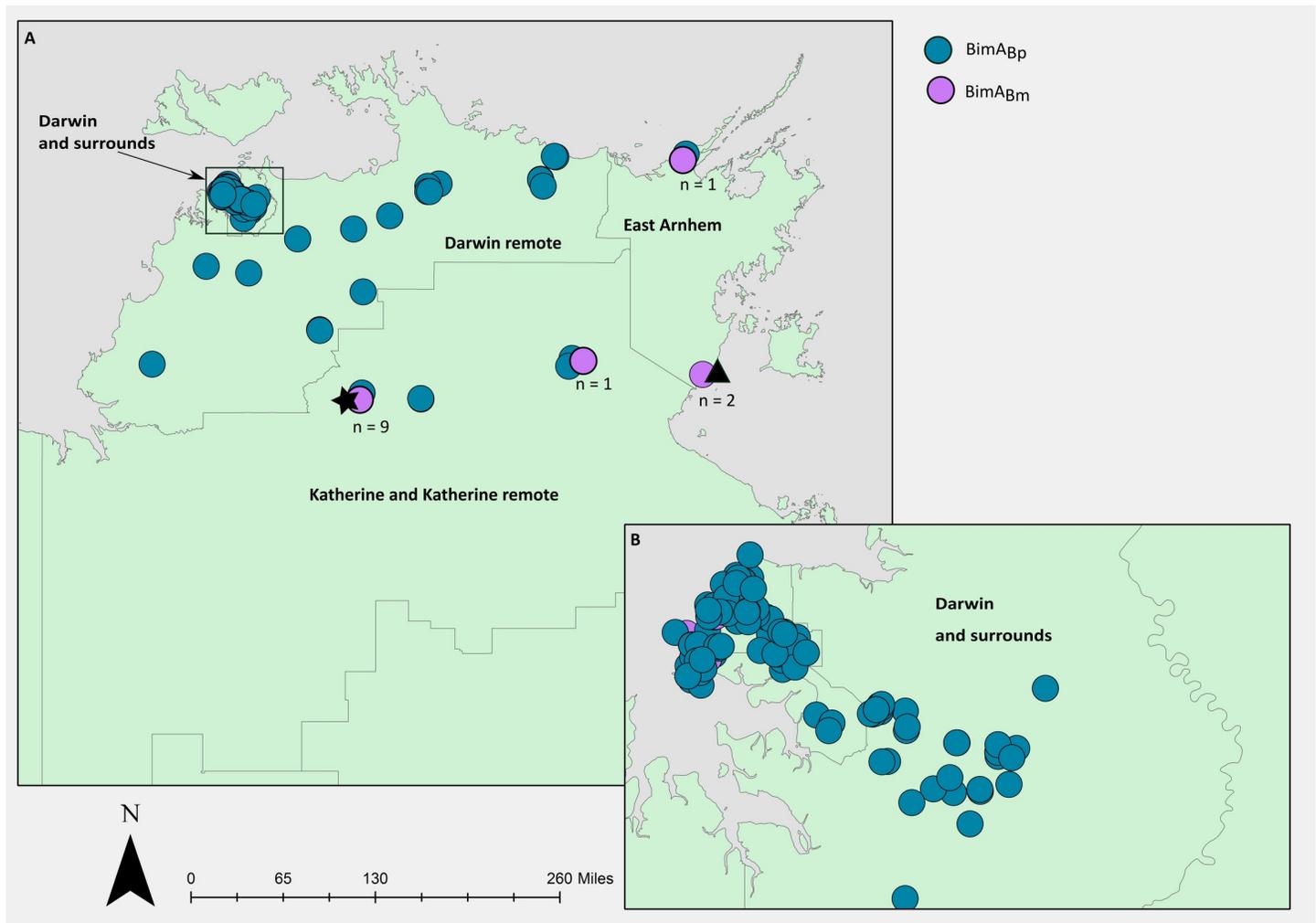


Fig 3. Geographical distribution of *bimA* variants in 340 environmental *Burkholderia pseudomallei* isolates from (A) the Top End region, Northern Territory, Australia and (B) the Darwin and surrounding region. Strains located in two Top End regions carried *bimA*_{Bm} and LPS B, as denoted by a black star (6/9 strains had *bimA*_{Bm} and LPS B) and a black triangle (2/2 strains had *bimA*_{Bm} and LPS B). The map was created using ArcGIS (<https://www.arcgis.com/index.html>).

<https://doi.org/10.1371/journal.pntd.0007369.g003>

with the same ST but mixed LPS type did not group together (Fig 4: “ST homoplasmy cases”) and were separated by a large number of SNPs. This is comparable to previously observed MLST homoplasmy in *B. pseudomallei* [20, 21]. To assess the effect of recombination on phylogenetic inference, SNP density filtering was applied to remove recombinogenic SNPs with Gubbins (v.2.3.1), which did not considerably alter the topology of the phylogenetic tree. We also noted that the isolates belonging to ST-807 and ST-468 were from three Top End geographical regions and varied by a large number of SNPs with the largest SNP difference being ~36,000 and ~29,000 respectively (Table 2). This phenomenon represents further examples of intracontinental homoplasmy but includes homoplasmy occurring in more closely related geographical regions than previously described [20].

Discussion

Zoonotic and person-to-person transmission of *B. pseudomallei* is extremely rare [28], with almost every case of melioidosis therefore reflecting a single infecting exposure event from the

Table 2. *Burkholderia pseudomallei* intracontinental multilocus sequence type (ST) homoplasmy cases.

ST	Strain	Year	Sample type	Location	LPS ⁺ type	Largest SNP difference	Homoplasmy present
468	MSHR0548	1998	Clinical	Darwin and surrounds	B	~29,000	Yes
	MSHR4299	2010	Environmental [#]	Darwin remote	B		
	MSHR1987	2005	Clinical	Katherine and Katherine remote	A		
	MSHR1922	2005	Clinical	Darwin remote	A		
	MSHR4805	2011	Clinical	Darwin remote	A		
807	MSHR3628	2010	Clinical	Katherine and Katherine remote	B	~36,000	Yes
	MSHR11424	1998	Clinical	Katherine and Katherine remote	A		
	MSHR7676	2012	Clinical	East Arnhem	A		
734	MSHR0859	1999	Clinical	Katherine and Katherine remote	B	~29,000	Yes
	MSHR3326	2008	Clinical	Katherine and Katherine remote	B		
	MSHR3457	2009	Clinical	Katherine and Katherine remote	B		
	MSHR6734	2012	Clinical	Katherine and Katherine remote	B		
	MSHR9613	2016	Environmental [#]	Katherine and Katherine remote	B		
	MSHR9932	2017	Environmental [#]	Katherine and Katherine remote	B		
	MSHR11560	2018	Clinical	Katherine and Katherine remote	B		
	MSHR7901	2013	Clinical	East Arnhem	A		
	MSHR8614	2015	Clinical	East Arnhem	A		
	MSHR9237	2016	Clinical	East Arnhem	A		
118	MSHR3325	2008	Clinical	East Arnhem	B2	~38,000	Yes
	MSHR0896	1999	Clinical	Darwin remote	A		
	MSHR2348	2006	Clinical	Katherine and Katherine remote	A		
1485	MSHR2619	2001	Environmental	Darwin and surrounds	A	~33,000	Yes
	MSHR2618	2001	Environmental	Darwin and surrounds	A		
	MSHR8609	2014	Environmental	Katherine and Katherine remote	B		

[#]included in study to increase ST numbers. Abbreviations: ⁺LPS, Lipopolysaccharide

<https://doi.org/10.1371/journal.pntd.0007369.t002>

The LPS O-antigen moiety of *B. pseudomallei* is highly varied with three serotypes described [12]. Nevertheless we recently found that previously published *in vitro* differences in virulence between LPS A and LPS B did not translate to differences in mortality in DPMS patients, supporting the dominant role in human melioidosis of host risk factors in determining disease severity and outcomes [18]. Regional differences in LPS serotype prevalence have been noted in Southeast Asia and Australia with LPS A being the most prevalent, followed by LPS B and with LPS B2 being the rarest in Australian strains. As yet LPS B2 has only been detected in *B. pseudomallei* strains from Australia and Papua New Guinea [12]. Despite the reporting of these regional differences in LPS prevalence, no studies have investigated the geographical footprint of the LPS types in a contiguous melioidosis-endemic region. We used a combination of statistical analysis, WGS and geographical mapping to examine the geographical footprint of the LPS genotypes in the Top End of the Northern Territory.

We first examined the prevalence of the LPS types across the environmental and clinical samples, and our data confirm that the prevalence of the LPS types is consistent between the Northern Territory environmental (LPS A = 89%; LPS B = 10% and LPS B2 = 1%) and clinical isolate collection (LPS A = 87%; LPS B = 12% and LPS B2 = 1%) [18], with LPS A being the most prevalent. High prevalence of LPS A is reflected in it also being present in many near-neighbour species [31], with the possibility of horizontal gene transfer of LPS A between these related species. LPS A was also distributed throughout the *B. pseudomallei* phylogeny, also supporting horizontal transfer of the entire LPS A loci. Based on our large clinical dataset

(n = 999) we found strong geographical clustering of LPS types in the Northern Territory, Australia, with the highest odds of having a clinical isolate with LPS B being in the East Arnhem region in comparison to the Darwin region (OR 19.5, 95% CI 9.1–42.0; $p < 0.001$). A similar trend was observed for strains isolated from the Northern Territory environment (n = 340), and in the environmental strains LPS B was not detected in urban Darwin. The successful use of LPS genotypes as geographical markers of *B. pseudomallei* within the Northern Territory can now be replicated in other melioidosis regions globally to ascertain the global LPS footprint alongside other variable markers such as *bimA* and *fhb3*.

LPS provides protective immunity to melioidosis in *in vitro* models of disease making it a potential vaccine candidate. It has also been demonstrated that antibodies from the two dominant LPS types, LPS A and B are not cross-reactive due to structural differences of the O-antigen, with consequent implications for an LPS based vaccine [12]. The sole presence of LPS A in the Darwin city region where the majority of melioidosis cases occur in the Northern Territory [6], suggests that an LPS A based vaccine strategy would be appropriate in this region. However, in regions where LPS A and LPS B strains are both present such as the Darwin remote region, an LPS A alone vaccine strategy is unlikely to provide adequate protection. For other melioidosis endemic regions globally, determination of the LPS genotype footprints will be critical if LPS vaccines are to be used.

Despite the strong link noted between ST and LPS type, we detected an LPS discrepancy in 21 isolates belonging to five rare STs: ST-118 (n = 3), ST-734 (n = 10), ST-468 (n = 5), ST-807 (n = 3) and ST-1485 (n = 3). Phylogenomic analysis of 175 genomes confirmed that the isolates belonging to the five STs represented five new ST homoplasy occurrences, explaining the LPS discrepancy noted in the five STs. Isolates of each ST were separated by a large number of SNPs (~29,000–~39,000), which is characteristic of Australian isolates that belong to different STs [20, 21]. Despite the large number of SNPs separating isolates within each ST, they all still resided within the Australian clade, representing five novel intracontinental ST homoplasy cases. These findings further highlight the low resolution of MLST and the need for high resolution techniques such as WGS to resolve unexpected genotyping results, particularly for *B. pseudomallei* which is a highly recombinogenic pathogen [32].

In addition to the phylogeographical role of the different *B. pseudomallei* LPS genotypes, the potential for differential virulence amongst LPS genotypes and potential interactions with other variable genotypes requires consideration. *In vitro* models of LPS types have been associated with differential virulence [12, 33, 34]. However, as noted above this did not translate to clinical significance in our large clinical cohort (>1000 melioidosis patients) where mortality, rates of bacteraemia and septic shock were the same for patients with LPS A and LPS B [18]. Nevertheless, it is known that virulence mechanisms work in concert for *B. pseudomallei* to invade and cause disease and further genomic studies are required to assess whether there are regional combinations of virulence genes including LPS variants that will have predictive capability for variable disease presentations and severity. For example, an association between *bimA*_{Bm} and neurological melioidosis has been previously described [10], and a new finding from this study was that eight environmental strains from the Katherine and Katherine remote and East Arnhem region had a combination of *bimA*_{Bm} and LPS B. Whether this specific *bimA*_{Bm}–LPS B combination has implications for neurotropism of *B. pseudomallei* now requires elucidation.

The present study encompassed a number of limitations. One limitation of this study is that the assigned presumptive location of infection for cases of melioidosis is dependent on good epidemiological history obtained from patients and will sometimes not be the true location of infection. Nevertheless, the close correlation of geographical LPS footprints between the clinical case isolates and the environmental isolates (which are 100% location accurate) is

reassuring that the prospective nature of the Darwin melioidosis study is providing accurate epidemiological data. A second limitation is that the four regions (Darwin and surrounding [persons, 148,564; geographic size, 316,391hectares], Darwin remote [persons, 17,902; geographic size, 11,229,485hectares], Katherine and Katherine remote [persons, 20,839; geographic size, 32,625,027hectares] and East Arnhem [persons, 14,519; geographic size, 3,360,659hectares] <https://www.abs.gov.au/>) vary substantially in both geographic size and population density, which results in a greater number of isolates obtained from the Darwin and surrounding region and ultimately clustered data. We accounted for the clustered nature of our data by using bivariate logistics regression with the cluster feature.

In conclusion, by combining genomic data with corresponding strain geographical information we have found that in the tropical north of Australia the LPS types have distinct geographical footprints as well as ST associations, adding to the already known variably present genetic markers *phaB3* and the *bimA* variants. A novel and interesting finding was that the geographical footprint of LPS B and *bimA*_{Bm} are similar and in remote Top End locations. The clinical, sero-diagnostic and vaccine implications of geographically-based *B. pseudomallei* LPS and other gene differentials and their relationships to regional and global dispersal of melioidosis require global collaborations with further analysis of larger clinically and geospatially-linked datasets.

Supporting information

S1 Table. Summary of 175 *Burkholderia pseudomallei* genomes used to construct the Phylogenetic tree with LPS type, ST, origin and GenBank accession no.
(XLSX)

S2 Table. The 88 rare STs that are found amongst the LPS B isolates.
(XLSX)

Acknowledgments

We thank our microbiology laboratory colleagues at Royal Darwin Hospital for their support and expertise in *B. pseudomallei* identification. We also thank Glenda Harrington (Menzies School of Health Research) for laboratory assistance. We thank Jeffrey Warner and Anthony Baker for providing *B. pseudomallei* isolates from Queensland.

Author Contributions

Conceptualization: Jessica R. Webb, Derek S. Sarovich, Erin P. Price, Mirjam Kaestli, Linda M. Ward, Bart J. Currie.

Data curation: Jessica R. Webb, Audrey Rachlin, Vanessa Rigas.

Formal analysis: Jessica R. Webb, Linda M. Ward.

Funding acquisition: Mark Mayo, Bart J. Currie.

Investigation: Jessica R. Webb, Bart J. Currie.

Methodology: Jessica R. Webb, Mark Mayo, Bart J. Currie.

Project administration: Jessica R. Webb, Mark Mayo, Bart J. Currie.

Resources: Jessica R. Webb, Mark Mayo, Bart J. Currie.

Software: Jessica R. Webb.

Supervision: Jessica R. Webb, Mark Mayo, Bart J. Currie.

Validation: Jessica R. Webb.

Visualization: Jessica R. Webb, Bart J. Currie.

Writing – original draft: Jessica R. Webb.

Writing – review & editing: Jessica R. Webb, Derek S. Sarovich, Erin P. Price, Mirjam Kaestli, Mark Mayo, Bart J. Currie.

References

1. Wiersinga WJ, Currie BJ, Peacock SJ. Melioidosis. The New England journal of medicine. 2012; 367(11):1035–44. <https://doi.org/10.1056/NEJMra1204699> PMID: 22970946.
2. White NJ. Melioidosis. Lancet. 2003; 361(9370):1715–22. [https://doi.org/10.1016/s0140-6736\(03\)13374-0](https://doi.org/10.1016/s0140-6736(03)13374-0) PMID: 12767750.
3. Dance DA, White NJ, Suputtamongkol Y, Chaowagul W. *Pseudomonas pseudomallei* and sudden unexplained death in Thai construction workers. Lancet. 1990; 336(8725):1257–8. [https://doi.org/10.1016/0140-6736\(90\)92877-k](https://doi.org/10.1016/0140-6736(90)92877-k) PMID: 1978101.
4. Saravu K, Mukhopadhyay C, Eshwara VK, Shastry BA, Ramamoorthy K, Krishna S, et al. Melioidosis presenting with mediastinal lymphadenopathy masquerading as malignancy: a case report. Journal of medical case reports. 2012; 6:28. <https://doi.org/10.1186/1752-1947-6-28> PMID: 22269609; PubMed Central PMCID: PMC3295726.
5. van Schaik EJ, Tom M, Woods DE. *Burkholderia pseudomallei* isocitrate lyase is a persistence factor in pulmonary melioidosis: implications for the development of isocitrate lyase inhibitors as novel antimicrobials. Infection and immunity. 2009; 77(10):4275–83. <https://doi.org/10.1128/IAI.00609-09> PMID: 19620343; PubMed Central PMCID: PMC2747945.
6. Currie BJ, Ward L, Cheng AC. The epidemiology and clinical spectrum of melioidosis: 540 cases from the 20 year Darwin prospective study. PLoS neglected tropical diseases. 2010; 4(11):e900. <https://doi.org/10.1371/journal.pntd.0000900> PMID: 21152057; PubMed Central PMCID: PMC2994918.
7. Lipsitz R, Garges S, Aurigemma R, Baccam P, Blaney DD, Cheng AC, et al. Workshop on treatment of and postexposure prophylaxis for *Burkholderia pseudomallei* and *B. mallei* Infection, 2010. Emerging infectious diseases. 2012; 18(12):e2. <https://doi.org/10.3201/eid1812.120638> PMID: 23171644; PubMed Central PMCID: PMC3557896.
8. Limmathurotsakul D, Golding N, Dance DA, Messina JP, Pigott DM, Moyes CL, et al. Predicted global distribution of *Burkholderia pseudomallei* and burden of melioidosis. Nature microbiology. 2016; 1:15008. <https://doi.org/10.1038/nmicrobiol.2015.8> PMID: 27517154.
9. Tuanyok A, Auerbach RK, Brettin TS, Bruce DC, Munk AC, Detter JC, et al. A horizontal gene transfer event defines two distinct groups within *Burkholderia pseudomallei* that have dissimilar geographic distributions. Journal of bacteriology. 2007; 189(24):9044–9. <https://doi.org/10.1128/JB.01264-07> PMID: 17933898; PubMed Central PMCID: PMC2168593.
10. Sarovich DS, Price EP, Webb JR, Ward LM, Voutsinos MY, Tuanyok A, et al. Variable virulence factors in *Burkholderia pseudomallei* (melioidosis) associated with human disease. PLoS one. 2014; 9(3): e91682. <https://doi.org/10.1371/journal.pone.0091682> PMID: 24618705; PubMed Central PMCID: PMC3950250.
11. Sarkar-Tyson M, Thwaite JE, Harding SV, Smither SJ, Oyston PC, Atkins TP, et al. Polysaccharides and virulence of *Burkholderia pseudomallei*. Journal of medical microbiology. 2007; 56(Pt 8):1005–10. <https://doi.org/10.1099/jmm.0.47043-0> PMID: 17644705.
12. Tuanyok A, Stone JK, Mayo M, Kaestli M, Gruendike J, Georgia S, et al. The genetic and molecular basis of O-antigenic diversity in *Burkholderia pseudomallei* lipopolysaccharide. PLoS neglected tropical diseases. 2012; 6(1):e1453. <https://doi.org/10.1371/journal.pntd.0001453> PMID: 22235357; PubMed Central PMCID: PMC3250505.
13. Currie BJ, Price EP, Mayo M, Kaestli M, Theobald V, Harrington I, et al. Use of Whole-Genome Sequencing to Link *Burkholderia pseudomallei* from Air Sampling to Mediastinal Melioidosis, Australia. Emerging infectious diseases. 2015; 21(11):2052–4. <https://doi.org/10.3201/eid2111.141802> PMID: 26488732; PubMed Central PMCID: PMC4622230.
14. Mayo M, Kaestli M, Harrington G, Cheng AC, Ward L, Karp D, et al. *Burkholderia pseudomallei* in unchlorinated domestic bore water, Tropical Northern Australia. Emerging infectious diseases. 2011; 17(7):1283–5. <https://doi.org/10.3201/eid1707.100614> PMID: 21762588; PubMed Central PMCID: PMC3381386.

15. Limmathurotsakul D, Dance DA, Wuthiekanun V, Kaestli M, Mayo M, Warner J, et al. Systematic review and consensus guidelines for environmental sampling of *Burkholderia pseudomallei*. PLoS neglected tropical diseases. 2013; 7(3):e2105. <https://doi.org/10.1371/journal.pntd.0002105> PMID: 23556010; PubMed Central PMCID: PMC3605150.
16. de Lamballerie X, Zandotti C, Vignoli C, Bollet C, de Micco P. A one-step microbial DNA extraction method using "Chelex 100" suitable for gene amplification. Research in microbiology. 1992; 143(8):785–90. PMID: 1298031.
17. Novak RT, Glass MB, Gee JE, Gal D, Mayo MJ, Currie BJ, et al. Development and evaluation of a real-time PCR assay targeting the type III secretion system of *Burkholderia pseudomallei*. Journal of clinical microbiology. 2006; 44(1):85–90. <https://doi.org/10.1128/JCM.44.1.85-90.2006> PMID: 16390953; PubMed Central PMCID: PMC1351940.
18. Webb JR, Sarovich DS, Price EP, Ward LM, Mayo M, Currie BJ. *Burkholderia pseudomallei* lipopolysaccharide genotype does not correlate with severity or outcome in melioidosis: host risk factors remain the critical determinant. Open Forum Infect Dis. 2019; ofz091.
19. Currie BJ, Gal D, Mayo M, Ward L, Godoy D, Spratt BG, et al. Using BOX-PCR to exclude a clonal outbreak of melioidosis. BMC infectious diseases. 2007; 7:68. <https://doi.org/10.1186/1471-2334-7-68> PMID: 17603903; PubMed Central PMCID: PMC1925088.
20. Aziz A, Sarovich DS, Harris TM, Kaestli M, McRobb E, Mayo M, et al. Suspected cases of intracontinental *Burkholderia pseudomallei* sequence type homoplasmy resolved using whole-genome sequencing. Microbial genomics. 2017; 3(11). <https://doi.org/10.1099/mgen.0.000139> PMID: 29208140; PubMed Central PMCID: PMC5729916.
21. De Smet B, Sarovich DS, Price EP, Mayo M, Theobald V, Kham C, et al. Whole-genome sequencing confirms that *Burkholderia pseudomallei* multilocus sequence types common to both Cambodia and Australia are due to homoplasmy. Journal of clinical microbiology. 2015; 53(1):323–6. <https://doi.org/10.1128/JCM.02574-14> PMID: 25392354; PubMed Central PMCID: PMC4290968.
22. Sarovich DS, Price EP. SPANDx: a genomics pipeline for comparative analysis of large haploid whole genome re-sequencing datasets. BMC research notes. 2014; 7:618. <https://doi.org/10.1186/1756-0500-7-618> PMID: 25201145; PubMed Central PMCID: PMC4169827.
23. Johnson SL, Baker AL, Chain PS, Currie BJ, Daligault HE, Davenport KW, et al. Whole-Genome Sequences of 80 Environmental and Clinical Isolates of *Burkholderia pseudomallei*. Genome announcements. 2015; 3(1). <https://doi.org/10.1128/genomeA.01282-14> PMID: 25676747; PubMed Central PMCID: PMC4333647.
24. Swofford DL. PAUP*: Phylogenetic Analysis Using Parsimony (*and other methods) 4.0.b5. Sunderland, MA: Sinauer Associates; 2001.
25. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic acids research. 2015; 43(3):e15. <https://doi.org/10.1093/nar/gku1196> PMID: 25414349; PubMed Central PMCID: PMC4330336.
26. Jolley KA, Maiden MC. BIGSdb: Scalable analysis of bacterial genome variation at the population level. BMC bioinformatics. 2010; 11:595. <https://doi.org/10.1186/1471-2105-11-595> PMID: 21143983; PubMed Central PMCID: PMC3004885.
27. Price EP, Sarovich DS, Currie BJ. Genomic Insights Into the Melioidosis Pathogen, *Burkholderia pseudomallei*. Curr Trop Med Rep. 2017; 4(3):95–102.
28. Cheng AC, Currie BJ. Melioidosis: epidemiology, pathophysiology, and management. Clin Microbiol Rev. 2005; 18(2):383–416. Epub 2005/04/16. <https://doi.org/10.1128/CMR.18.2.383-416.2005> PMID: 15831829; PubMed Central PMCID: PMC1082802.
29. Stevens MP, Stevens JM, Jeng RL, Taylor LA, Wood MW, Hawes P, et al. Identification of a bacterial factor required for actin-based motility of *Burkholderia pseudomallei*. Molecular microbiology. 2005; 56(1):40–53. <https://doi.org/10.1111/j.1365-2958.2004.04528.x> PMID: 15773977.
30. Sim SH, Yu Y, Lin CH, Karuturi RK, Wuthiekanun V, Tuanyok A, et al. The core and accessory genomes of *Burkholderia pseudomallei*: implications for human melioidosis. PLoS pathogens. 2008; 4(10):e1000178. <https://doi.org/10.1371/journal.ppat.1000178> PMID: 18927621; PubMed Central PMCID: PMC2564834.
31. Stone JK, Mayo M, Grasso SA, Ginther JL, Warrington SD, Allender CJ, et al. Detection of *Burkholderia pseudomallei* O-antigen serotypes in near-neighbor species. BMC microbiology. 2012; 12:250. <https://doi.org/10.1186/1471-2180-12-250> PMID: 23126230; PubMed Central PMCID: PMC3541218.
32. Pearson T, Giffard P, Beckstrom-Sternberg S, Auerbach R, Hornstra H, Tuanyok A, et al. Phylogeographic reconstruction of a bacterial species with high levels of lateral gene transfer. BMC biology. 2009; 7:78. <https://doi.org/10.1186/1741-7007-7-78> PMID: 19922616; PubMed Central PMCID: PMC2784454.

33. Arjcharoen S, Wikraiphath C, Pudla M, Limposuwan K, Woods DE, Sirisinha S, et al. Fate of a *Burkholderia pseudomallei* lipopolysaccharide mutant in the mouse macrophage cell line RAW 264.7: possible role for the O-antigenic polysaccharide moiety of lipopolysaccharide in internalization and intracellular survival. *Infection and immunity*. 2007; 75(9):4298–304. Epub 2007/06/20. <https://doi.org/10.1128/IAI.00285-07> PMID: 17576760; PubMed Central PMCID: PMC1951188.
34. Norris MH, Schweizer HP, Tuanyok A. Structural diversity of *Burkholderia pseudomallei* lipopolysaccharides affects innate immune signaling. *PLoS neglected tropical diseases*. 2017; 11(4):e0005571. Epub 2017/04/30. <https://doi.org/10.1371/journal.pntd.0005571> PMID: 28453531; PubMed Central PMCID: PMC5425228.