Whole-Genome Sequencing Confirms that *Burkholderia pseudomallei* Multilocus Sequence Types Common to Both Cambodia and Australia Are Due to Homoplasy

Birgit De Smet,[a,f] Derek S. Sarovich,[b] Erin P. Price,[b] Mark Mayo,[b] Vanessa Theobald,[b] Chun Kham,[d] Seiha Heng,[e] Phe Thong,[d] Matthew T. G. Holden,[g,h] Julian Parkhill,[g] Sharon J. Peacock,[g,i] Brian G. Spratt,[c] Jan A. Jacobs,[a,j] Peter Vandamme,[f] Bart J. Currie[b]

Department of Clinical Sciences, Institute of Tropical Medicine, Antwerp, Belgium[a]; Global and Tropical Health Division, Menzies School of Health Research, Darwin, Australia[b]; Imperial College, London, United Kingdom[c]; Sihanouk Hospital Centre of HOPE, Phnom Penh, Cambodia[d]; Institut Pasteur du Cambodge, Phnom Penh, Cambodia[e]; Laboratory of Microbiology, Faculty of Sciences, Ghent University, Ghent, Belgium[f]; Wellcome Trust Sanger Institute, Cambridge, United Kingdom[g]; University of St. Andrews, St. Andrews, United Kingdom[h]; University of Cambridge, Cambridge, United Kingdom[i]; Department of Microbiology and Immunology, University of Leuven, Leuven, Belgium[j]

***Burkholderia pseudomallei* isolates with shared multilocus sequence types (STs) have not been isolated from different continents. We identified two STs shared between Australia and Cambodia. Whole-genome analysis revealed substantial diversity within STs, correctly identified the Asian or Australian origin, and confirmed that these shared STs were due to homoplasy.**

*B*urkholderia pseudomallei*, the bacterium that causes melioidosis, is well-recognized in the regions of northern Australia and Southeast Asia where melioidosis is prevalent. The prevalence of *B. pseudomallei* is increasingly being recognized in other tropical regions including parts of Africa, the Americas, and other Asian regions such as India (1). Melioidosis is a potentially fatal disease with mortality rates ranging from 10 to 50% of infected individuals (2). The majority of *B. pseudomallei* infections are acquired from the environment following percutaneous inoculation, inhalation, or ingestion of contaminated soil or surface water; human-to-human transmission is exceedingly rare (2). The nature of melioidosis acquisition, coupled with restricted *B. pseudomallei* environmental dissemination patterns, has contributed to the evolution of localized genetic populations with finite geographic distribution (3–5). Phylogeographic studies using multilocus sequence typing (MLST) (6) and whole-genome sequencing (4) have identified two distinct populations of *B. pseudomallei* corresponding to Asia and Australia (3, 4, 7, 8). Knowledge of this population structure has facilitated source attribution for unusual melioidosis cases, particularly those occurring in regions where melioidosis is not endemic, such as returning travelers (9, 10). Despite the success of previous studies in identifying robust phylogeographic patterns within *B. pseudomallei* populations, the inherently high recombination rate of this bacterium and greater sampling efforts were predicted to inevitably reveal shared sequence types (STs) between these distinct geographic locations (11). In the present study, we identify for the first time two such instances of *B. pseudomallei* STs being shared between Cambodian and Australian isolates.

B. pseudomallei sequence type 105 (ST105) and ST849 isolates were analyzed from both Australia and Cambodia; a total of four isolates were analyzed. MLST was performed as previously described (6). MSHR282, the Australian ST105 isolate, was obtained in 1994 from an Australian patient enrolled in the Darwin Prospective Melioidosis Study (12), and the Cambodian ST105 isolate, CAM41, was isolated from a Cambodian melioidosis patient in 2008. To date, no other ST105 isolates have been submitted to the MLST database (http://bpseudomallei.mlst.net/).

MSHR4004, the Australian ST849 isolate, was isolated from an Australian soil sample in 2010, and the Cambodian ST849 isolate, SHCH2430, was isolated from a Cambodian melioidosis patient in the same year. Clinical isolates were from patients with strong epidemiological data to support local acquisition of their infections, including no documented travel history to other regions where melioidosis is endemic. Ethics were approved by the Human Research Ethics Committee of the Northern Territory Department of Health and Families, the Menzies School of Health Research, and the Cambodian National Ethical Committee.

The MLST profiles of all Cambodian and Australian *B. pseudomallei* strains in the MLST database (as of 13 June14) were first analyzed with eBURST V3 (13) (Fig. 1). eBURST showed some evidence of ST clustering according to geographic source; however, these groupings contained multiple cases of Asian isolates grouping with isolates of Australian origin (Fig. 1). Overall, eBURST was unreliable for inferring the geographic origin of STs, most likely due to very high rates of recombination in *B. pseu-*
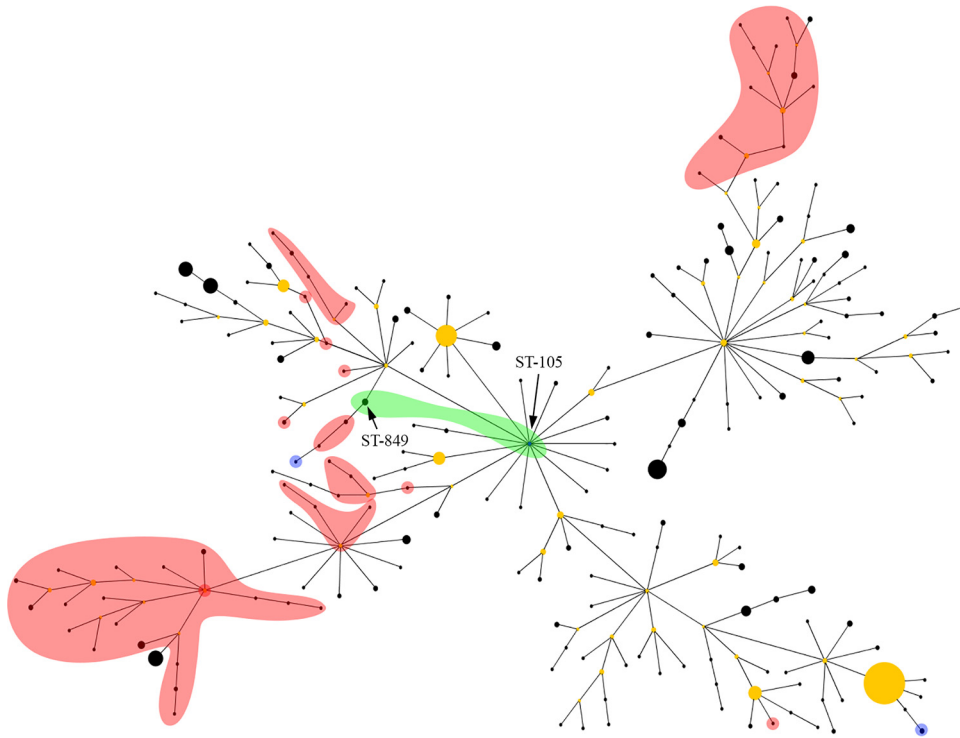
FIG 1 eBURST analysis of 954 *Burkholderia pseudomallei* isolates from Australia and Cambodia, comprising 245 sequence types (STs). Both ST105 and ST849 (shaded green) contain *B. pseudomallei* isolates found in Cambodia and Australia. All other STs have been found solely in Australasia or Asia. STs shaded in red indicate an Asian origin and include isolates found in Cambodia, Thailand, China, and Vietnam. STs shaded in blue represent islands in Australasia. All unshaded STs are of Australian origin.

domallei, which has previously been shown to confound accurate prediction of recent ancestry using this tool (11).

To gain further insight into the genetic relatedness and geographic origin of these four isolates, we performed whole-genome sequencing (WGS) using the Illumina HiSeq2000 platform (Illumina Inc., San Diego, CA). WGS data have been deposited into the NCBI SRA database with the following accession numbers: CAM41 (ERR539773, ERR539807, and ERR539841), MSHR282 (ERR298339), SHCH240 (ERR298360), and MSHR4004 (ERR298343). Single-nucleotide polymorphisms (SNPs) in the core genome were identified with SPANDx v2.3 (14) using *B. pseudomallei* K96243 as a reference sequence (15). Additional reference *B. pseudomallei* genomes were incorporated into the analysis by inclusion of simulated Illumina data using ART v2.1.8 and a quality shift of 10 (16). Using the default settings of SPANDx, 84,839 core genome SNPs were identified. Maximum likelihood phylogenetic analysis of these SNPs using RAxML (17) grouped the Australian isolates (MSHR282 and MSHR4004) with other Australian isolates; likewise, the Cambodian isolates (CAM41 and SHCH2430) grouped most closely with isolates of Asian origin (Fig. 2). To assess the effect of recombination on phylogenetic inference, recombinogenic regions were removed using altered SNP-filtering parameters with GATK (18) based on a SNP density of more than two SNPs within 300 bp or with gubbins based on default parameters (19). GATK filtering or gubbins analysis removed 37,213 (44%) or 24,216 (13.5%) SNPs, respectively, used in the phylogenetic reconstruction but did not alter geographic attribution of strains or tree topology (see Fig. S1 and S2 in the supplemental material). Overall, these findings suggest that both
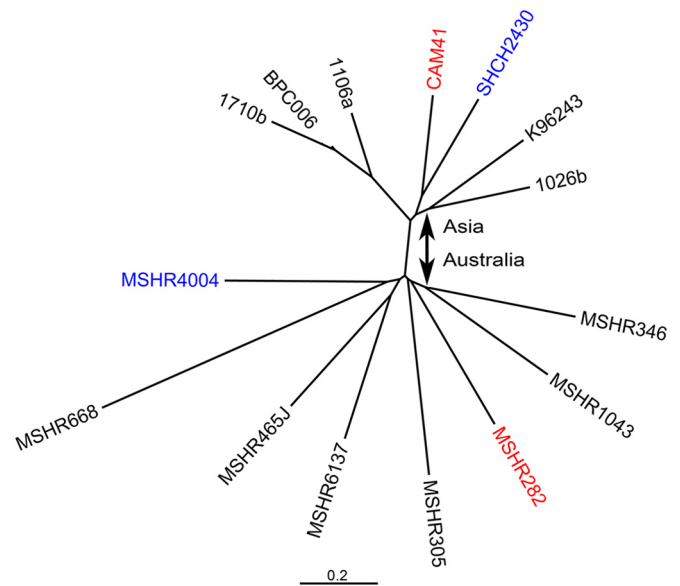

FIG 2 Maximum likelihood phylogenetic analysis of two Cambodian (CAM41 and SHCH2430) and two Australian (MSHR282 and MSHR4004) *Burkholderia pseudomallei* genomes in comparison to *B. pseudomallei* reference genomes. A total of 84,839 core genome single-nucleotide polymorphisms were used to construct the phylogeny. Based on multilocus sequence typing (MLST), both MSHR282 and CAM41 isolates are ST105 (red), and isolates MSHR4004 and SHCH2430 are ST849 (blue). However, whole-genome phylogenetic analysis clearly groups these strains based upon geographic origin, i.e., MSHR282 and MSHR4004 group with other Australian isolates, and CAM41 and SHCH2430 group with other Asian strains. Thus, MLST of *B. pseudomallei* can, in rare cases, be confounded by ST homoplasy. The scale bar represents the average number of nucleotide substitutions per site.

ST105 and ST849 convergence was a consequence of both mutation and multiple recombination events over considerable evolutionary time rather than from recent recombination involving the MLST loci (Fig. S1).

To complement the whole-genome SNP phylogeny findings, several variable genetic markers with known geographic associations were interrogated *in silico*; specifically, the virulence factors encoded by *Burkholderia pseudomallei* or *Burkholderia mallei bimA* ($bimA_{Bp/Bm}$) [20] and *fhaB3* [21], and the BTFC (*Burkholderia thailandensis*-like fimbrial cluster) and YLF (*Yersinia*-like fimbrial cluster) loci, the latter two of which are mutually exclusive. All four isolates possessed *fhaB3* and the $bimA_{Bp}$ subtype. Only the MSHR4004 isolate was positive for BTFC; all other isolates possessed the YLF allele. The *fhaB3*, $bimA_{Bp}$, and YLF markers are more common in *B. pseudomallei* isolates of Asian origin, and $bimA_{Bm}$ has yet to be observed in any Southeast Asian isolate but is present in 12% of Australian isolates [22, 23]. These genotypes correlate with their expected prevalence in Australian and Southeast Asian *B. pseudomallei*. The $bimA_{Bm}$ and BTFC loci were not observed in the two Cambodian isolates. The discordant YLF/BTFC profiles in the two ST849 isolates is highly unusual. To the best of our knowledge, this has not been reported previously in isolates with identical STs from the same geographic region. Taken together, these results support the expected geographic origin of these isolates and further consolidate the convergent nature of these STs.

We report, for the first time, two instances of *B. pseudomallei* isolates with identical STs from two continents. Isolate origins were resolved using whole-genome phylogenetic analysis. Although our study was limited by the availability of Cambodian *B. pseudomallei* isolates for comparative whole-genome phylogenetic analysis, we showed that, in both cases, shared STs between geographic regions were due to ST homoplasy. Our findings rule out recent *B. pseudomallei* transmission between these regions and demonstrate some limitations of MLST for source attribution of highly recombinogenic species.

ACKNOWLEDGMENTS

REFERENCES

1. Currie BJ, Dance DA, Cheng AC. 2008. The global distribution of *Burkholderia pseudomallei* and melioidosis: an update. Trans R Soc Trop Med Hyg 102(Suppl 1):S1–S4. http://dx.doi.org/10.1016/S0035-9203(08)70002-6.
2. Wiersinga WJ, Currie BJ, Peacock SJ. 2012. Melioidosis. N Engl J Med 367:1035–1044. http://dx.doi.org/10.1056/NEJMra1204699.
3. Dale J, Price EP, Hornstra H, Busch JD, Mayo M, Godoy D, Wuthiekanun V, Baker A, Foster JT, Wagner DM, Tuanyok A, Warner J, Spratt BG, Peacock SJ, Currie BJ, Keim P, Pearson T. 2011. Epidemiological tracking and population assignment of the non-clonal bacterium, *Burkholderia pseudomallei*. PLoS Negl Trop Dis 5:e1381. http://dx.doi.org/10.1371/journal.pntd.0001381.
4. Pearson T, Giffard P, Beckstrom-Sternberg S, Auerbach R, Hornstra H, Tuanyok A, Price EP, Glass MB, Leadem B, Beckstrom-Sternberg JS, Allan GJ, Foster JT, Wagner DM, Okinaka RT, Sim SH, Pearson O, Wu Z, Chang J, Kaul R, Hoffmaster AR, Brettin TS, Robison RA, Mayo M, Gee JE, Tan P, Currie BJ, Keim P. 2008. Phylogeographic reconstruction of a bacterial species with high levels of lateral gene transfer. BMC Biol 18:7. http://dx.doi.org/10.1186/1741-7007-7-78.
5. McRobb E, Kaestli M, Price EP, Sarovich DS, Mayo M, Warner J, Spratt BG, Currie BJ. 2014. Distribution of *Burkholderia pseudomallei* in northern Australia, a land of diversity. Appl Environ Microbiol 80:3463–3468. http://dx.doi.org/10.1128/AEM.00128-14.
6. Godoy D, Randle G, Simpson AJ, Aanensen DM, Pitt TL, Kinoshita R, Spratt BG. 2003. Multilocus sequence typing and evolutionary relationships among the causative agents of melioidosis and glanders, *Burkholderia pseudomallei* and *Burkholderia mallei*. J Clin Microbiol 41:2068–2079. http://dx.doi.org/10.1128/JCM.41.5.2068-2079.2003.
7. Vesaratchavest M, Tumapa S, Day NP, Wuthiekanun V, Chierakul W, Holden MT, White NJ, Currie BJ, Spratt BG, Feil EJ, Peacock SJ. 2006. Nonrandom distribution of *Burkholderia pseudomallei* clones in relation to geographical location and virulence. J Clin Microbiol 44:2553–2557. http://dx.doi.org/10.1128/JCM.00629-06.
8. Currie B, Thomas AD, Godoy D, Dance DA, Cheng AC, Ward L, Mayo M, Pitt TL, Spratt BG. 2007. Australian and Thai isolates of *Burkholderia pseudomallei* are distinct by multilocus sequence typing: revision of a case of mistaken identity. J Clin Microbiol 45:3828–3829. http://dx.doi.org/10.1128/JCM.01590-07.
9. Lim RS, Flatman S, Dahm MC. 2013. Sinonasal melioidosis in a returned traveller presenting with nasal cellulitis and sinusitis. Case Rep Otolaryngol 2013:920352. http://dx.doi.org/10.1155/2013/920352.
10. Ezzedine K, Malvy D, Steels E, De Dobbeeler G, Struelens M, Jacobs F, Heenen M. 2007. Imported melioidosis with an isolated cutaneous presentation in a 90-year-old traveller from Bangladesh. Bull Soc Pathol Exot 100:22–25.
11. Turner KM, Hanage WP, Fraser C, Connor TR, Spratt BG. 2007. Assessing the reliability of eBURST using simulated populations with known ancestry. BMC Microbiol 7:30. http://dx.doi.org/10.1186/1471-2180-7-30.
12. Currie BJ, Ward L, Cheng AC. 2010. The epidemiology and clinical spectrum of melioidosis: 540 cases from the 20 year Darwin prospective study. PLoS Negl Trop Dis 4:e900. http://dx.doi.org/10.1371/journal.pntd.0000900.
13. Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG. 2004. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J Bacteriol 186:1518–1530. http://dx.doi.org/10.1128/JB.186.5.1518-1530.2004.
14. Sarovich DS, Price EP. 2014. SPANDx: a genomics pipeline for comparative analysis of large haploid whole genome re-sequencing datasets. BMC Res Notes 7:618. http://dx.doi.org/10.1186/1756-0500-7-618.
15. Holden MT, Titball RW, Peacock SJ, Cerdeno-Tarraga AM, Atkins T, Crossman LC, Pitt T, Churcher C, Mungall K, Bentley SD, Sebaihia M, Thomson NR, Bason N, Beacham IR, Brooks K, Brown KA, Brown NF, Challis GL, Cherevach I, Chillingworth T, Cronin A, Crossett B, Davis P, DeShazer D, Feltwell T, Fraser A, Hance Z, Hauser H, Holroyd S, Jagels K, Keith KE, Maddison M, Moule S, Price C, Quail MA, Rabbinowitsch E, Rutherford K, Sanders M, Simmonds M, Songsivilai S, Stevens K, Tumapa S, Vesaratchavest M, Whitehead S, Yeats C, Barrell BG, Oyston PC, Parkhill J. 2004. Genomic plasticity of the causative agent of melioidosis, Burkholderia pseudomallei. Proc Natl Acad Sci U S A 101:14240–14245. http://dx.doi.org/10.1073/pnas.0403302101.
16. Huang W, Li L, Myers JR, Marth GT. 2012. ART: a next-generation sequencing read simulator. Bioinformatics 28:593–594. http://dx.doi.org/10.1093/bioinformatics/btr708.
17. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313. http://dx.doi.org/10.1093/bioinformatics/btu033.
18. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297–1303. http://dx.doi.org/10.1101/gr.107524.110.
19. Croucher NJ, Harris SR, Fraser C, Quail MA, Burton J, van der Linden M, McGee L, von Gottberg A, Song JH, Ko KS, Pichon B, Baker S, Parry CM, Lambertsen LM, Shahinas D, Pillai DR, Mitchell TJ, Dougan G, Tomasz A, Klugman KP, Parkhill J, Hanage WP, Bentley SD. 2011. Rapid pneumococcal evolution in response to clinical interventions. Science 331:430–434. http://dx.doi.org/10.1126/science.1198545.
20. Stevens MP, Stevens JM, Jeng RL, Taylor LA, Wood MW, Hawes P, Monaghan P, Welch MD, Galyov EE. 2005. Identification of a bacterial factor re-

quired for actin-based motility of *Burkholderia pseudomallei*. Mol Microbiol **56:**40–53. http://dx.doi.org/10.1111/j.1365-2958.2004.04528.x.

21. **Sim SH, Yu Y, Lin CH, Karuturi RK, Wuthiekanun V, Tuanyok A, Chua HH, Ong C, Paramalingam SS, Tan G, Tang L, Lau G, Ooi EE, Woods D, Feil E, Peacock SJ, Tan P.** 2008. The core and accessory genomes of *Burkholderia pseudomallei*: implications for human melioidosis. PLoS Pathog **4:**e1000178. http://dx.doi.org/10.1371/journal.ppat.1000178.

22. **Sarovich DS, Price EP, Webb JR, Ward LM, Voutsinos MY, Tuanyok A, Mayo M, Kaestli M, Currie BJ.** 2014. Variable virulence factors in *Burkholderia pseudomallei* (melioidosis) associated with human disease. PLoS One **9:**e91682. http://dx.doi.org/10.1371/journal.pone.0091682.

23. **Sitthidet C, Korbsrisate S, Layton AN, Field TR, Stevens MP, Stevens JM.** 2011. Identification of motifs of *Burkholderia pseudomallei* BimA required for intracellular motility, actin binding, and actin polymerization. J Bacteriol **193:**1901–1910. http://dx.doi.org/10.1128/JB.01455-10.