Whole-Genome Sequences of 80 Environmental and Clinical Isolates of *Burkholderia pseudomallei*


Bioenergy and Biome Sciences Group, Bioscience Division, Los Alamos National Laboratory, Los Alamos, New Mexico, USA; College of Public Health, Medical and Veterinary, James Cook University, Townsville, Queensland, Australia; Global and Tropical Health Division, Menzies School of Health Research, Casuarina, Northern Territory, Australia; Institute for Glycomics, Griffith University, Queensland, Australia; Division of Microbiology and Infectious Diseases, PathWest Laboratory Medicine Western Australia, Nedlands, Western Australia, Australia; School of Pathology and Laboratory Medicine, The University of Western Australia, Nedlands, Western Australia, Australia; National Biodefense Analysis and Countermeasures Center, Frederick, Maryland, USA

Here, we present the draft genome sequences of 80 isolates of *Burkholderia pseudomallei*. The isolates represent clinical cases of melioidosis and environmental isolates from regions in Australia and Papua New Guinea where *B. pseudomallei* is endemic. The genomes provide further context for the diversity of the pathogen.

*B. pseudomallei* is the causative agent of melioidosis and is endemic in parts of the tropical world, including northern Australia, Papua New Guinea, and Southeast Asia (1–3). Studies of pathogen phylogeny or diversity using whole-genome sequencing have been dominated by Asian strains, for which more genome sequences were available (4, 5). We report here the whole-genome sequences of 80 *B. pseudomallei* isolates from both Australian clinical cases and environmental sampling of geographically diverse regions in northern Australia and Papua New Guinea. The genomes will contribute to our understanding of the global diversity of *B. pseudomallei*.

High-quality, high-molecular-weight genomic DNA was sequenced using a combination of Illumina, 454, and PacBio technologies, depending on the isolate. For those with only Illumina short-insert data (100-bp reads, noted as “I” in Table 1) assemblies were generated with IDBA version 1.1.1 (6). For those that also included Roche 454 data (noted as “R”) or Illumina long-insert data (insert sizes 8 to 10 kb, noted as “L”), the libraries were assembled together in Newbler version 2.6 (Roche) and the consensus sequences computationally shredded into 2-kb overlapping fake reads (shreds). The raw reads were also assembled in Velvet and those consensus sequences computationally shredded into 1.5-kbp overlapping shreds (7). Draft data from all platforms were assembled together with AllPaths (8), and if Pacific Biosciences data was available (noted in Table 1 as “P”) and at 100X coverage or greater, assembled using HGAP (9). Consensus sequences from all assemblers were computationally shredded and assembled with a subset of read pairs from the long-insert library using Phrap (10, 11). The resulting assemblies were manually and computationally improved using Consed (12) and in-house scripts.

For strains MSH62 and MSHR3997, a 10-kbp insert library was sequenced on the Pacific Biosciences platform. The assembly was generated with Celera Assembler version 8.0 (13) by previously described methods (14). The longest 25X of corrected sequences were assembled, and contigs composed of fewer than 10 sequences were omitted. Contigs were manually merged based on identified end overlaps to obtain the final assembly. The MSHR62 10-kbp insert assembly was used to assist in gap closure and correction of the short-read assembly.

For all genomes, annotations were completed at the Los Alamos National Laboratory (LANL) using the Ergatis workflow manager (15) and in-house scripts. Of the 80 *B. pseudomallei* genomes assembled, nine are at finished quality (<1 error per 100,000 bp [16]), 49 are either noncontiguous finished or improved high-quality draft (IHQD) and available as scaffolded draft assemblies, and 22 assemblies are unscaffolded drafts.

**Nucleotide sequence accession numbers.** Genome accession numbers for the assemblies deposited in DDBJ/ENA/GenBank are listed in Table 1.

**ACKNOWLEDGMENTS**

We thank Richard Robison and Annette Bunnell for extracting genomic DNA from the isolates.

This project was funded by the DHS Science and Technology Directorate through the Agreement between the Governments of the United States of America and Australia on Cooperation in Science and Technology for Homeland/Domestic Security Matters, signed 21 December 2005. The contributions of S.K. and M.J.R. were funded under Agreement HSHQDC-07-C-00020 awarded by the Department of Homeland Security Science and Technology Directorate (DHS/S&T) for the management and operation of the National Biodefense Analysis and Countermeasures Center (NBACC), a Federally Funded Research and Development Center.

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of
<table>
<thead>
<tr>
<th>Strain name</th>
<th>Isolation source</th>
<th>GenBank accession no.</th>
<th>Sequence data type(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSHR44</td>
<td>Clinical, Australia</td>
<td>JQIM00000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>MSHR62</td>
<td>Clinical, Australia</td>
<td>CP009235, CP009234</td>
<td>I, R, P</td>
</tr>
<tr>
<td>MSHR303</td>
<td>Clinical, Australia</td>
<td>JQD0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>MSHR332</td>
<td>Clinical, Australia</td>
<td>JQFM0000000000</td>
<td>I, R</td>
</tr>
<tr>
<td>MSHR435</td>
<td>Clinical, Australia</td>
<td>JRF0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>MSHR449</td>
<td>Clinical, Australia</td>
<td>JQFO0000000000</td>
<td>I, R</td>
</tr>
<tr>
<td>MSHR456</td>
<td>Clinical, Australia</td>
<td>JQFN0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>MSHR465</td>
<td>Clinical, Australia</td>
<td>JPZ0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>MSHR543</td>
<td>Clinical, Australia</td>
<td>JPZ0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>MSHR640</td>
<td>Clinical, Australia</td>
<td>JQFP0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>MSHR684</td>
<td>Clinical, Australia</td>
<td>JQDC0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>MSHR733</td>
<td>Clinical, Australia</td>
<td>JQEE0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>MSHR3016</td>
<td>Clinical, Australia</td>
<td>JQEH0000000000</td>
<td>I, R</td>
</tr>
<tr>
<td>MSHR3335</td>
<td>Clinical, Australia</td>
<td>JQFO0000000000</td>
<td>I, R</td>
</tr>
<tr>
<td>MSHR3709</td>
<td>Clinical, Australia</td>
<td>JQEH0000000000</td>
<td>I, R</td>
</tr>
<tr>
<td>MSHR1153</td>
<td>Clinical, Australia</td>
<td>CP009271, CP009272</td>
<td>I, R, P</td>
</tr>
<tr>
<td>MSHR1357</td>
<td>Clinical, Australia</td>
<td>JQDA0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>MSHR2138</td>
<td>Clinical, Australia</td>
<td>JQIJ0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>MSHR2243</td>
<td>Clinical, Australia</td>
<td>CP009270, CP009269</td>
<td>I, R, P</td>
</tr>
<tr>
<td>MSHR2451</td>
<td>Clinical, Australia</td>
<td>JQEG0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>MSHR2990</td>
<td>Clinical, Australia</td>
<td>JQHV0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>MSHR3016</td>
<td>Clinical, Australia</td>
<td>JQEE0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>MSHR3335</td>
<td>Clinical, Australia</td>
<td>JQEH0000000000</td>
<td>I, R</td>
</tr>
<tr>
<td>MSHR3458</td>
<td>Clinical, Australia</td>
<td>JQOE0000000000</td>
<td>I, R</td>
</tr>
<tr>
<td>MSHR3709</td>
<td>Clinical, Australia</td>
<td>JQF0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>ABCPW 1</td>
<td>Clinical, Australia</td>
<td>JQHI0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>ABCPW 30</td>
<td>Clinical, Australia</td>
<td>JPVF0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>ABCPW 91</td>
<td>Clinical, Australia</td>
<td>JPUY0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>ABCPW 107</td>
<td>Clinical, Australia</td>
<td>JQDN0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>ABPCW 111</td>
<td>Clinical, Australia</td>
<td>JQWT0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>A79A</td>
<td>Clinical, Australia</td>
<td>CP009165, CP009164</td>
<td>I, R, P</td>
</tr>
<tr>
<td>A79C</td>
<td>Clinical, Australia</td>
<td>JQHE0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>A79D</td>
<td>Clinical, Australia</td>
<td>JQHE0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>BDU 2</td>
<td>Clinical, Australia</td>
<td>JPVG0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>K42</td>
<td>Clinical, Australia</td>
<td>CP009151, CP009150</td>
<td>I, R, P</td>
</tr>
<tr>
<td>MSHR3951</td>
<td>Clinical, Australia</td>
<td>JQVA0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>MSHR3960</td>
<td>Clinical, Australia</td>
<td>JPVJ0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>MSHR3964</td>
<td>Clinical, Australia</td>
<td>JPVJ0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>MSHR3965</td>
<td>Clinical, Australia</td>
<td>JPVJ0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>MSHR3997</td>
<td>Clinical, Australia</td>
<td>JQH0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>MSHR4000</td>
<td>Clinical, Australia</td>
<td>JPVJ0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>MSHR4003</td>
<td>Clinical, Australia</td>
<td>JPVJ0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>MSHR4009</td>
<td>Clinical, Australia</td>
<td>JPVJ0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>MSHR4012</td>
<td>Clinical, Australia</td>
<td>JPVJ0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>MSHR4018</td>
<td>Clinical, Australia</td>
<td>JQIK0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>MSHR4032</td>
<td>Clinical, Australia</td>
<td>JQIK0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>MSHR4299</td>
<td>Clinical, Australia</td>
<td>JQIK0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>MSHR4300</td>
<td>Clinical, Australia</td>
<td>JPVU0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>MSHR4303</td>
<td>Clinical, Australia</td>
<td>JPVU0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>MSHR4304</td>
<td>Clinical, Australia</td>
<td>JPVU0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>MSHR4308</td>
<td>Clinical, Australia</td>
<td>JPVU0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>MSHR4372</td>
<td>Clinical, Australia</td>
<td>JPVU0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>MSHR4375</td>
<td>Clinical, Australia</td>
<td>JPVU0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>MSHR4377</td>
<td>Clinical, Australia</td>
<td>JPVU0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>MSHR4378</td>
<td>Clinical, Australia</td>
<td>JPVU0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>MSHR4462</td>
<td>Clinical, Australia</td>
<td>JPVU0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>MSHR4503</td>
<td>Clinical, Australia</td>
<td>JPVU0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>MSHR4868</td>
<td>Clinical, Australia</td>
<td>JQDP0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>MSHR5492</td>
<td>Clinical, Australia</td>
<td>JQDP0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>MSHR3569</td>
<td>Clinical, Australia</td>
<td>JQDL0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>MSHR5596</td>
<td>Clinical, Australia</td>
<td>JQDE0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>MSHR5608</td>
<td>Clinical, Australia</td>
<td>JPWB0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>MSHR5609</td>
<td>Clinical, Australia</td>
<td>JPWB0000000000</td>
<td>I, R, P</td>
</tr>
<tr>
<td>MSHR5613</td>
<td>Clinical, Australia</td>
<td>JPWB0000000000</td>
<td>I, R, P</td>
</tr>
</tbody>
</table>

(Continued on following page)
Homeland Security. In no event shall the DHS, NBACC, or Battelle Na-
tional Biodefense Institute (BNBI) have any responsibility or liability for
any use, misuse, inability to use, or reliance upon the information con-
tained herein. The Department of Homeland Security does not endorse
any use, misuse, inability to use, or reliance upon the information con-
tained herein. The Department of Homeland Security does not endorse
any products or commercial services mentioned in this publication. This
manuscript is approved by LANL for unlimited release (LA-UR-14-
19286).

REFERENCES
3. Currie BJ, Dance DA, Cheng AC. 2008. The global distribution of Burk-
Hyg 102(Suppl 1):S1–S4. http://dx.doi.org/10.1111/j.1365-
9203(08)70002-6.
Tuanyok A, Price EP, Glass MB, Ledem B, Beckstrom-Sternberg JS,
Allan GJ, Foster JT, Wagner DM, Okinaka RT, Sim SH, Pearson O, Wu
Keim P. 2009. Phylogeographic reconstruction of a bacterial species with high levels of lateral gene transfer. BMC Biol
5. Engelthaler DM, Bowers J, Schupp JA, Pearson T, Ginther J, Hornstra HM,
LB, Contente T, Beckstrom-Sternberg SM, Blaney DD, Wagner DM, Mayo
acquired case of melioidosis in southern AZ. PLoS Neglected Trop Dis
iterative de Bruijn graph de novo assembler. Lect Notes Comput Sci 6044:
dx.doi.org/10.1101/gr.074492.107.
8. Butler J, MacCallum I, Kleber M, Shlyakhter IA, Belmonte MK, Lande
ES, Nusbaum C, Jaffe DB. 2008. ALLPATHS: de novo assembly of whole-
10.1101/gr.7357908.
9. Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C,
Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J.
2013. Nonhybrid, finished microbial genome assemblies from long-read
10.1038/nmeth.2474.
sequencer traces using Phred. I. Accuracy assessment. Genome Res
bioinformatics/bts515.
13. Koren S, Schatz MC, Walenz BP, Martin J, Howard JT, Ganapathy G,
Wang Z, Rasko DA, McCombie WR, Jarvis ED, Phillippy AM. 2012. Hybrid
error correction and de novo assembly of single-molecule sequenc-
nbt.2280.
14. Koren S, Harhay GP, Smith TP, Bono JL, Harhay DM, McVey SD,
Radune D, Bergman NH, Phillippy AM. 2013. Reducing assembly com-
plexity of microbial genomes with single-molecule sequencing. Genome
An Ergatis-based prokaryotic genome annotation Web server. Bioinforma-
16. Chain PSG, Graham DV, Fulton RS, FitzGerald MG, Hostetler J,
S, Field D, Garry GM, Gibbs R, Graves T, Han CS, Harrison SH,
Highlander S, Hugenholtz P, Khouri HM, Kodira CD, Kolker E, Kyrip-
des NC, Lang D, Lapidus A, Malfatti SA, Markowitz V, Matha T, Nelson
KE, Parkhill J, Pitluck S, Qin X, Read TD, Schmutz J, Sozhamannan S,
Sterk P, Straussberg RL, Sutton G, Thomson NR, Tiedje JM, Weinstock
G, Wollam A, Consortium GSCHMPJ,bettler DC. 2009. Genome project
dx.doi.org/10.1126/science.1180614.

TABLE 1 (Continued)

<table>
<thead>
<tr>
<th>Strain name</th>
<th>Isolation source</th>
<th>GenBank accession no.</th>
<th>Sequence data type(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSHR7334</td>
<td>–13.1708260, 130.6744830</td>
<td>JQDF00000000</td>
<td>I</td>
</tr>
<tr>
<td>MSHR7343</td>
<td>–13.1709770, 130.6739790</td>
<td>JQDM00000000</td>
<td>I</td>
</tr>
<tr>
<td>MSHR7498</td>
<td>–14.1288333, 134.4440333</td>
<td>JQDF00000000</td>
<td>I</td>
</tr>
<tr>
<td>MSHR7500</td>
<td>–14.1420167, 134.4274833</td>
<td>JRE00000000</td>
<td>I</td>
</tr>
<tr>
<td>MSHR7504</td>
<td>–14.1103500, 134.4069500</td>
<td>JPWR00000000</td>
<td>I</td>
</tr>
<tr>
<td>MSHR7527</td>
<td>–14.1903333, 134.3715833</td>
<td>JPWS00000000</td>
<td>I</td>
</tr>
<tr>
<td>TSV5</td>
<td>–19.2573333, 146.7928056</td>
<td>JQGY00000000</td>
<td>I</td>
</tr>
<tr>
<td>TSV25</td>
<td>–19.2643611, 146.7998611</td>
<td>JPVK00000000</td>
<td>I, L, P</td>
</tr>
<tr>
<td>TSV28</td>
<td>–19.2630528, 146.7966556</td>
<td>JQHU00000000</td>
<td>I</td>
</tr>
<tr>
<td>TSV31</td>
<td>–19.2601667, 146.7941111</td>
<td>JPVE00000000</td>
<td>I, L, P</td>
</tr>
<tr>
<td>TSV32</td>
<td>–19.2546944, 146.8012222</td>
<td>JQHT00000000</td>
<td>I</td>
</tr>
<tr>
<td>TSV43</td>
<td>–19.2601667, 146.7941111</td>
<td>JPQK00000000</td>
<td>I, L, P</td>
</tr>
<tr>
<td>TSV44</td>
<td>–19.2630528, 146.7966556</td>
<td>JQGX00000000</td>
<td>I</td>
</tr>
<tr>
<td>TSV48</td>
<td>–19.2564694, 146.7898111</td>
<td>CP009161, CP009160</td>
<td>I, L, P</td>
</tr>
<tr>
<td>TSV202</td>
<td>–19.2806167, 147.0308833</td>
<td>CP009157, CP009156, CP009155, CP009154</td>
<td>I, L, P</td>
</tr>
</tbody>
</table>

a Isolation source is reported as clinical or as latitude and longitude for environmental isolates.
b Sequence data types are Illumina short-insert (I), Roche 454 (R), Illumina long-insert (L), and Pacific Biosciences (P).