DETECTION AND PHYLOGENOMICS OF THE ENVIRONMENTAL OPPORTUNISTIC PATHOGEN, *BURKHOLDERIA PSEUDOMALLEI*, AND ITS CLOSE PHLYOGENETIC RELATIVES IN THE *B. PSEUDOMALLEI* COMPLEX

By Carina Marie Hall

A Dissertation

Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

in Biology

Northern Arizona University

December 2021

Approved:

Dave Wagner, Ph.D., Chair

Paul Keim, Ph.D.

Jason Sahl, Ph.D.

David Engelthaler, Ph.D.

ABSTRACT

DETECTION AND PHYLOGENOMICS OF THE ENVIRONMENTAL OPPORTUNISTIC PATHOGEN, *BURKHOLDERIA PSEUDOMALLEI*, AND ITS CLOSE PHLYOGENETIC RELATIVES IN THE *B. PSEUDOMALLEI* COMPLEX

CARINA MARIE HALL

Burkholderia pseudomallei, a Gram-negative environmental bacterium, is the causative agent of the disease melioidosis. Most melioidosis cases are reported from tropical regions in northern Australia and Southeast Asia where *B. pseudomallei* is endemic in the environment but the true global distribution appears to be much larger. The *B. pseudomallei* complex (Bpc) is a phylogenetic group of *Burkholderia* species that consist of *B. pseudomallei* and its closest phylogenetic relatives. Of the six described species in the Bpc, *B. pseudomallei* has adapted to opportunistic pathogenicity, *B. mallei* to obligate pathogenicity, and *B. thailandensis*, *B. oklahomensis*, *B. humptydooensis*, and *B. singularis* to environmental saprophytism with occasional pathogenicity (except *B. humptydooensis*).

Determining where *B. pseudomallei* is present in the environment is crucial for understanding the potential risk to humans of acquiring melioidosis since almost all infections with *B. pseudomallei* are independently acquired from the environment. In Chapter 2, our goal was to gain a better understanding of the prevalence and geographic distribution of *B. pseudomallei* and other *Burkholderia* spp. in the environment in Puerto Rico. The results from Chapter 2 suggest that *B. pseudomallei* is widely dispersed but rare in the environment in Puerto Rico. In Chapter 3, we isolated a member of the Bpc, *B. thailandensis*, from the Western Hemisphere for the first time. *B. thailandensis* is of interest from a public health standpoint due to its potential to cause infection as a fortuitous pathogen and to the melioidosis community

since *B. thailandensis* is often used as a surrogate in the laboratory instead of *B. pseudomallei* due to the reduced restrictions required. In Chapter 4, we proposed two novel Bpc species, *B. mayonis* sp. nov. and *B. savannae* sp. nov., isolated from soils from multiple locations in northern Australia. The addition of *B. mayonis* sp. nov. and *B. savannae* sp. nov. results in a total of eight species within this significant complex of bacteria that are available for future studies. Understanding the genomics and virulence of *B. pseudomallei* and its closest relatives in Bpc is important for identifying robust diagnostic targets specific to *B. pseudomallei* and understanding evolution of virulence in *B. pseudomallei*.

ACKNOWLEGMENTS

- Dave Wagner. There are not enough words to express my gratitude towards my mentor, Dave. I have learned so much from him over the years and I would not be where I am today if not for him. Dave has been my supervisor for the past decade and has always challenged me to be a better scientist while giving me enough scientific freedom to pave my path. Over the years he has given me many opportunities to grow professionally whether that was presenting to DTRA in D.C. or traveling to the Caribbean to lead field studies. He always pushed me to be my best self while providing me with the support to achieve my goals. Dave also is a great field biologist and I learned how to persevere while digging holes alongside him in a jungle maze in Maunabo, Puerto Rico. I can only hope that one day I'll be half as good of a scientist as he is.
- To each of my committee members. Jason Sahl for all his assistance with the bioinformatic analysis and always making time (and having the patience) for my often-trivial questions. Paul Keim for providing all his incredible insight in a way that only he can offer. Dave Engelthaler for supplying an outside perspective on my chapters and all his knowledge of pathogen dispersal that allowed me to ponder aspects on pathogen dispersal that I had not yet considered.
- To the Wagner team, Joseph Busch, Dawn Birdsell, Nate Stone, and Roxanne Nottingham. Each one of you has helped and contributed to the success of these studies and it is truly such a pleasure to enjoy the people who I get to work with. I'm so grateful to know and work with such wonderful people!
- To my life partner and husband, Paul Phillips. Thank you for always supporting me and providing perspective throughout this journey. I love that you're always willing to talk

through the science with me. I am so grateful for you and to be able to nerd out on anything and everything with you!

- To my Mama, Barbara Ann Hall, in every way imaginable. As a child, you pushed me to always do my best. As a teenager, you showed me how incredibly strong and independent a woman can be on their own. As an adult you supported me in whatever I wanted to pursue and you were a great travel buddy to Germany and New Zealand. Thank you, Mama!!
- To my Dad, George Hall. Thank you for always encouraging my scientific curiosity and telling me at a young age that being a biologist can be a profession.
- To my hetero-life mate, Chandler Roe. Watching your path has always inspired me. Thank you for paving the way. Seeing you challenge yourself made me want to challenge myself as well. You bring out the better person in me.
- To my Grandma Hall. For always believing in me academically and encouraging me to finish college before getting married. I love and miss you so much.
- To my brother, Derek, sister-in-law, Kindra, and the Phillips, Jane and Sam. Thank you for all the encouragement, support, and fun times over the years.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

DEDICATION

To my little pepita, Emberly Ann Phillips. You have brought more joy to my life in your first year on this planet than I've experienced in my whole life (and it's been a very happy life so far!). I hope you know that with love, hard work, and perseverance that you can do anything you put your mind to.

PREFACE

The format of this dissertation is in the journal format meaning that Chapters 2-4 are a series of papers that have been submitted or are being prepared for submission to professional journals and that Chapter 1 is an introductory and concluding chapter. The pronoun "we" is used throughout this dissertation to represent both my contributions and various other co-authors that contributed to these studies. Chapter 2, "*Burkholderia pseudomallei*, the causative agent of melioidosis, is rare but ecologically established and widely dispersed in the environment in Puerto Rico," was published in the journal "PLOS Neglected Tropical Diseases" in September 2019. Chapter 3, "For the first time *Burkholderia thailandensis* is isolated from the environment in the Western Hemisphere including the United States," has been prepared as a dispatch for submission to the journal "Emerging Infection Diseases." Chapter 4, "Expanding the *Burkholderia pseudomallei* complex with the addition of two novel species: *Burkholderia mayonis* sp. nov. and *Burkholderia savannae* sp. nov.," was accepted in ASM's Journal "Applied and Environmental Microbiology" in October 2021.

CHAPTER 1 Introduction

Burkholderia pseudomallei is a Gram-negative soil bacterium that is the causative agent of the disease melioidosis (1). *B. pseudomallei* is one of the more pathogenic species of the genus *Burkholderia* which was recently divided into *Burkholderia sensu stricto*, *Paraburkholderia*, *Caballeronia*, *Robbsia*, and *Pararobbsia*. Together, these taxonomic groups comprise over 100 described species (http://www.bacterio.net/) that can have pathogenic, mutualistic, and/or commensal relationships with plants, animals, and/or humans (2-4). This division resulted in *Burkholderia sensu stricto* containing most of the opportunistic pathogens belonging to one of two major phylogenetic groups: the *B. pseudomallei* complex (Bpc) or the *B. cepacia* complex (Bcc). New species are regularly described in *Burkholderia sensu stricto* (5-10), and the majority of these species are naturally found in the environment, primarily in soil and water (11). The *Burkholderia sensu stricto* is a group of diverse, primarily soil saprophytic, Gram-negative bacteria that have many strategies to survive and persist in soil. Species in the Bpc can also cause disease in humans and animals, with *B. pseudomallei* and *B. mallei* resulting in more severe disease; they are causative agents of the diseases melioidosis and glanders, respectively (1). Interestingly, *B. mallei* is a clone of *B. pseudomallei* that became the only obligate intracellular species in the *Burkholderia* genus following significant genome reduction in the process of becoming host-adapted to equines (12). In addition to *B. pseudomallei* and *B. mallei*, there are currently four other *Burkholderia* species assigned to the Bpc: *B. thailandensis* (13), *B. oklahomensis* (14), *B. humptydooensis* (6), and *B. singularis* (7). The taxonomy of this genus remains incomplete and new species are continued to be described (8, 15, 16).

Due to the pathogenicity of *B. pseudomallei*, its potential to be aerosolized, and the often-vague symptoms of melioidosis, *B. pseudomallei* was classified as a Tier 1 Select Agent by the

US Centers for Disease Control and Prevention (CDC) (17, 18). Melioidosis can be contracted via cutaneous inoculation, inhalation, or ingestion, and can present with extremely varied symptoms (19); these vague symptoms and diverse clinical presentations, along with culture-based diagnostic anomalies, make it difficult to properly diagnose in clinical settings (18, 20). Because melioidosis can be difficult to diagnose, it is possible that *B. pseudomallei* is also present in the environment in other regions of the world and causing human disease in these areas but going undetected (21). Most melioidosis cases are reported from tropical regions in northern Australia and Southeast Asia where *B. pseudomallei* is endemic in the environment but the true global distribution appears to be much larger (21). As the awareness of *B. pseudomallei* grows, more melioidosis cases have been reported from areas outside of the commonly reported regions (22-25). In fact, a retrospective study identified 12 melioidosis cases in Panama during a ten period (26). Despite being commonly thought of as a tropical pathogen, *B. pseudomallei* causes disease and has been isolated from the environment in non-tropical regions around the world (22, 27-29). Interestingly, extreme weather events have been shown to lead to increased melioidosis cases in both endemic and non-endemic regions of the world (22, 29, 30).

Currently, there are no vaccines against *B. pseudomallei* available, making rapid detection and specific antibiotic treatment crucial for favorable outcomes in infected humans. Successful antibiotic treatment typically includes a strict and long regimen of intravenous antibiotics, such as ceftazidime or meropenem, for at least two weeks, followed by oral antibiotics, such as co-trimoxazole, for up to six months (17, 18). However, treatment can be complicated by the fact that *B. pseudomallei* is intrinsically resistant to several clinically relevant antibiotics (31). Importantly, other *Burkholderia* species that co-exist with *B. pseudomallei* in the environment are known to have intrinsic resistance to other clinically relevant antibiotics, such as

meropenem resistance in *B. ubonensis* (32), and thereby represent a possible source of similar resistance in *B. pseudomallei*.

Burkholderia thailandensis is often used as a surrogate in the laboratory for *B. pseudomallei* (33-35) due to the similarities between the closely related bacteria and that *B. thailandensis* has reduced restrictions as a non-Select Agent, BSL-2 organism. *B. thailandensis*, is also an environmental Gram-negative bacterium and is of interest from a public health standpoint due to its potential to cause infection as an opportunistic pathogen. *B. thailandensis* is the closest phylogenetic relative to *B. pseudomallei*. Similar to *B. pseudomallei*, *B. thailandensis* has primarily been described from the environment in Southeast Asia and northern Australia (36-38). Recently *B. thailandensis* was identified in the environment in Africa, specifically in Sierra Leone (39), but overall the global distribution of *B. thailandensis* is largely unknown. *B. thailandensis* is difficult to distinguish from *B. pseudomallei* in the clinic due to the same biochemical phenotypes, except for the ability to assimilate L-arabinose (36, 40, 41). *B. thailandensis* was first described after researchers observed reduced virulence in what they thought was *B. pseudomallei* but the 16S rRNA gene phylogeny revealed a novel *Burkholderia* species that was later named *B. thailandensis* after the origin of the type strain, E264^{T}, from Thailand (36). Human infections of *B. thailandensis* are not common but do occur (42). There have been three reported clinical cases of *B. thailandensis* in the Western Hemisphere, all from the southern U.S. in Louisiana in 1997, Texas in 2003 (40), and Arkansas in 2017 (42). Even with environmental sampling related to these clinical cases, *B. thailandensis* was never recovered from the environment. To date, the natural occurrence of *B. thailandensis* in the Western Hemisphere is poorly understood. It is important to note that until this study *B. thailandensis* had never been isolated from the environment in the Western Hemisphere.

Determining where the *B. pseudomallei* complex is present in the environment is crucial for understanding the potential risk to humans of acquiring melioidosis and other disease caused by Bpc. This is because almost all infections with Bpc (except for *B. mallei*) are independently acquired from the environment (27); human to human transmission of melioidosis is extremely rare (43). Understanding the genetic diversity within the Bpc is also of the utmost importance for *B. pseudomallei* molecular diagnostics. This is because as new species and strains are discovered the *B. pseudomallei* core genome continues to reduce and the accessory genome expand (5).

The genome of *B. pseudomallei* is larger and more complex compared to most bacterial genomes. While the average bacterial genome is around 5 Mb (44), *B. pseudomallei* has a genome size of 7.2 Mb that comprises of two chromosomes. This large genome has about twice as much accessory genome as the core genome (45). The *B. pseudomallei* genome also has high amounts of homologous recombination it has about twice as much homologous recombination as *Streptococcus pneumonia*, an organism known for its high recombination (46). The largest chromosome consists of many functional genes, while the smaller chromosome has mostly accessory genes (47). Known areas of high recombination are termed genomic islands (48). Areas of high recombination are not randomly found throughout the genome but are more common in regions associated with virulence and survival. The recombination hotspots are found on both chromosomes but interestingly they are identified more often on chromosome 2. To further complicate this already complex genome, *B. pseudomallei* also has distinct restriction-modification systems between different phylogenetic clades, allowing for gene exchange within clades but not across distinct clades (45). All the above items are important to take into consideration when building a phylogeny to ensure that the correct evolutionary model is being used and thus the correct phylogenetic conclusions are reached.

We used a systematic approach for the detection and isolation of *B. pseudomallei*, *B. thailandensis*, and two novel *Burkholderia* spp. belonging to the Bpc from soil and water samples collected in Texas, Puerto Rico, and Australia. The information gained from this study is vitally important to help understand where and how the Bpc persists in the environment and to better characterize the genetic diversity in the environment to assist with traceback studies of clinical strains to determine a potential geographic origin and source of exposure.

References

1. Currie BJ. 2015. Melioidosis: Evolving Concepts in Epidemiology, Pathogenesis, and Treatment. Seminars in Respiratory and Critical Care Medicine 36:111-125.

2. Eberl L, Vandamme P. 2016. Members of the genus Burkholderia: good and bad guys. F1000Res 5.

3. Beukes CW, Palmer M, Manyaka P, Chan WY, Avontuur JR, van Zyl E, Huntemann M, Clum A, Pillay M, Palaniappan K, Varghese N, Mikhailova N, Stamatis D, Reddy TBK, Daum C, Shapiro N, Markowitz V, Ivanova N, Kyrpides N, Woyke T, Blom J, Whitman WB, Venter SN, Steenkamp ET. 2017. Genome Data Provides High Support for Generic Boundaries in Burkholderia Sensu Lato. Front Microbiol 8:1154.

4. Lin QH, Lv YY, Gao ZH, Qiu LH. 2020. Pararobbsia silviterrae gen. nov., sp. nov., isolated from forest soil and reclassification of Burkholderia alpina as Pararobbsia alpina comb. nov. International Journal of Systematic and Evolutionary Microbiology 70:1412-1420.

5. Sahl JW, Vazquez AJ, Hall CM, Busch JD, Tuanyok A, Mayo M, Schupp JM, Lummis M, Pearson T, Shippy K, Colman RE, Allender CJ, Theobald V, Sarovich DS, Price EP, Hutcheson A, Korlach J, LiPuma JJ, Ladner J, Lovett S, Koroleva G, Palacios G, Limmathurotsakul D, Wuthiekanun V, Wongsuwan G, Currie BJ, Keim P, Wagner DM. 2016. The Effects of Signal Erosion and Core Genome Reduction on the Identification of Diagnostic Markers. MBio 7.

6. Tuanyok A, Mayo M, Scholz H, Hall CM, Allender CJ, Kaestli M, Ginther J, Spring-Pearson S, Bollig MC, Stone JK, Settles EW, Busch JD, Sidak-Loftis L, Sahl JW, Thomas A, Kreutzer L, Georgi E, Gee JE, Bowen RA, Ladner JT, Lovett S, Koroleva G, Palacios G, Wagner DM, Currie BJ, Keim P. 2017. Burkholderia humptydooensis sp nov., a New Species Related to

Burkholderia thailandensis and the Fifth Member of the Burkholderia pseudomallei Complex. Applied and Environmental Microbiology 83.

7. Vandamme P, Peeters C, De Smet B, Price EP, Sarovich DS, Henry DA, Hird TJ, Zlosnik JEA, Mayo M, Warner J, Baker A, Currie BJ, Carlier A. 2017. Comparative Genomics of Burkholderia singularis sp nov., a Low G plus C Content, Free-Living Bacterium That Defies Taxonomic Dissection of the Genus Burkholderia. Frontiers in Microbiology 8.

8. Martina P, Leguizamon M, Prieto CI, Sousa SA, Montanaro P, Draghi WO, Stammler M, Bettiol M, de Carvalho C, Palau J, Figoli C, Alvarez F, Benetti S, Lejona S, Vescina C, Ferreras J, Lasch P, Lagares A, Zorreguieta A, Leitao JH, Yantorno OM, Bosch A. 2018. Burkholderia puraquae sp. nov., a novel species of the Burkholderia cepacia complex isolated from hospital settings and agricultural soils. Int J Syst Evol Microbiol 68:14-20.

9. Ong KS, Aw YK, Lee LH, Yule CM, Cheow YL, Lee SM. 2016. Burkholderia paludis sp. nov., an Antibiotic-Siderophore Producing Novel Burkholderia cepacia Complex Species, Isolated from Malaysian Tropical Peat Swamp Soil. Front Microbiol 7:2046.

10. Takeshita K, Tamaki H, Ohbayashi T, Meng XY, Sone T, Mitani Y, Peeters C, Kikuchi Y, Vandamme P. 2018. Burkholderia insecticola sp. nov., a gut symbiotic bacterium of the bean bug Riptortus pedestris. Int J Syst Evol Microbiol 68:2370-2374.

11. Compant S, Nowak J, Coenye T, Clement C, Barka EA. 2008. Diversity and occurrence of Burkholderia spp. in the natural environment. Fems Microbiology Reviews 32:607-626.

12. 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 41:2068-79.

13. Brett PJ, Deshazer D, Woods DE. 1997. Characterization of Burkholderia pseudomallei and Burkholderia pseudomallei-like strains. Epidemiology and Infection 118:137-148.

14. Glass MB, Steigerwalt AG, Jordan JG, Wilkins PP, Gee JE. 2006. Burkholderia oklahomensis sp nov., a Burkholderia pseudomallei-like species formerly known as the Oklahoma strain of Pseudomonas pseudomallei. International Journal of Systematic and Evolutionary Microbiology 56:2171-2176.

15. Tuanyok A, Mayo M, Scholz H, Hall CM, Allender CJ, Kaestli M, Ginther J, Spring-Pearson S, Bollig MC, Stone JK, Settles EW, Busch JD, Sidak-Loftis L, Sahl JW, Thomas A, Kreutzer L, Georgi E, Gee JE, Bowen RA, Ladner JT, Lovett S, Koroleva G, Palacios G, Wagner DM, Currie BJ, Keim P. 2017. Burkholderia humptydooensis sp. nov., a New Species Related to Burkholderia thailandensis and the Fifth Member of the Burkholderia pseudomallei Complex. Appl Environ Microbiol 83.

16. Vandamme P, Peeters C, De Smet B, Price EP, Sarovich DS, Henry DA, Hird TJ, Zlosnik JEA, Mayo M, Warner J, Baker A, Currie BJ, Carlier A. 2017. Comparative Genomics of Burkholderia singularis sp. nov., a Low G+C Content, Free-Living Bacterium That Defies Taxonomic Dissection of the Genus Burkholderia. Front Microbiol 8:1679.

17. Wiersinga WJ, Virk HS, Torres AG, Currie BJ, Peacock SJ, Dance DAB, Limmathurotsakul D. 2018. Melioidosis. Nature Reviews Disease Primers 4.

18. Currie BJ. 2015. Melioidosis: evolving concepts in epidemiology, pathogenesis, and treatment. Semin Respir Crit Care Med 36:111-25.

19. Yee KC, Lee MK, Chua CT, Puthucheary SD. 1988. Melioidosis, the great mimicker: a report of 10 cases from Malaysia. J Trop Med Hyg 91:249-54.

20. Doker TJ, Quinn CL, Salehi ED, Sherwood JJ, Benoit TJ, Glass Elrod M, Gee JE, Shadomy SV, Bower WA, Hoffmaster AR, Walke HT, Blaney DD, DiOrio MS. 2014. Fatal Burkholderia pseudomallei infection initially reported as a Bacillus species, Ohio, 2013. Am J Trop Med Hyg 91:743-6.

21. Limmathurotsakul D, Golding N, Dance DA, Messina JP, Pigott DM, Moyes CL, Rolim DB, Bertherat E, Day NP, Peacock SJ, Hay SI. 2016. Predicted global distribution of Burkholderia pseudomallei and burden of melioidosis. Nat Microbiol 1.

22. Alvarez-Hernandez G, Cruz-Loustaunau D, Ibarra JA, Rascon-Alcantar A, Contreras-Soto J, Meza-Radilla G, Torres AG, Estrada-de los Santos P. 2021. Description of two fatal cases of melioidosis in Mexican children with acute pneumonia: case report. Bmc Infectious Diseases 21.

23. Gee JE, Gulvik CA, Castelo-Branco D, Sidrim JJC, Rocha MFG, Cordeiro RA, Brilhante RSN, Bandeira T, Patricio I, Alencar LP, da Costa Ribeiro AK, Sheth M, Deka MA, Hoffmaster AR, Rolim D. 2021. Genomic Diversity of Burkholderia pseudomallei in Ceara, Brazil. mSphere 6.

24. Rodriguez JY, Morales-Lopez SE, Rodriguez GJ, Alvarez-Moreno CA, Esquea K, Pinzon H, Ramirez LR, Moreno L, Ocampo W, Cepeda ML. 2019. Case Series Study of Melioidosis, Colombia. Emerging Infectious Diseases 25:1531-1534.

25. Stone NE, Hall CM, Browne AS, Sahl JW, Hutton SM, Santana-Propper E, Celona KR, Guendel I, Harrison CJ, Gee JE, Elrod MG, Busch JD, Hoffmaster AR, Ellis EM, Wagner DM. 2020. Burkholderia pseudomallei in Soil, US Virgin Islands, 2019. Emerging Infectious Diseases 26:2773-2775.

26. Arauz AB, Castillo K, Santiago E, Quintero Y, Adames E, Castillo B, Rodriguez-French A, Henostroza G. 2020. Geographic Distribution and Incidence of Melioidosis, Panama. Emerging Infectious Diseases 26:118-121.

27. Cossaboom CM, Marinova-Petkova A, Strysko J, Rodriguez G, Maness T, Ocampo J, Gee JE, Elrod MG, Gulvik CA, Liu LD, Bower WA, Hoffmaster AR, Blaney DD, Salzer JS, Yoder JS, Mattioli MC, Sidwa TJ, Ringsdorf L, Morrow G, Ledezma E, Kieffer A. 2020. Melioidosis in a Resident of Texas with No Recent Travel History, United States. Emerging Infectious Diseases 26:1295-1299.

28. Webb JR, Buller N, Rachlin A, Golledge C, Sarovich DS, Price EP, Mayo M, Currie BJ. 2020. A Persisting Nontropical Focus of Burkholderia pseudomallei with Limited Genome Evolution over Five Decades. Msystems 5.

29. Rolim DB, Vilar DCFL, Sousa AQ, Miralles IS, de Oliveira DCA, Harnett G, O'Reilly L, Howard K, Sampson I, Inglis TJJ. 2005. Melioidosis, northeastern Brazil. Emerging Infectious Diseases 11:1458-1460.

30. Chai LYA, Fisher D. 2018. Earth, wind, rain, and melioidosis Comment. Lancet Planetary Health 2:E329-E330.

31. Schweizer HP. 2012. Mechanisms of antibiotic resistance in Burkholderia pseudomallei: implications for treatment of melioidosis. Future Microbiol 7:1389-99.

32. Price EP, Sarovich DS, Webb JR, Hall CM, Jaramillo SA, Sahl JW, Kaestli M, Mayo M, Harrington G, Baker AL, Sidak-Loftis LC, Settles EW, Lummis M, Schupp JM, Gillece JD, Tuanyok A, Warner J, Busch JD, Keim P, Currie BJ, Wagner DM. 2017. Phylogeographic, genomic, and meropenem susceptibility analysis of Burkholderia ubonensis. PLoS Negl Trop Dis 11:e0005928.

33. Gallagher LA, Ramage E, Patrapuvich R, Weiss E, Brittnacher M, Manoil C. 2013. Sequence-Defined Transposon Mutant Library of Burkholderia thailandensis. MBio 4.

34. Morici LA, Heang J, Tate T, Didier PJ, Roy CJ. 2010. Differential susceptibility of inbred mouse strains to Burkholderia thailandensis aerosol infection. Microb Pathog 48:9-17.

35. West TE, Frevert CW, Liggitt HD, Skerrett SJ. 2008. Inhalation of Burkholderia thailandensis results in lethal necrotizing pneumonia in mice: a surrogate model for pneumonic melioidosis. Trans R Soc Trop Med Hyg 102 Suppl 1:S119-26.

36. Brett PJ, DeShazer D, Woods DE. 1998. Burkholderia thailandensis sp. nov., a Burkholderia pseudomallei-like species. Int J Syst Bacteriol 48:317-320.

37. Ginther JL, Mayo M, Warrington SD, Kaestli M, Mullins T, Wagner DM, Currie BJ, Tuanyok A, Keim P. 2015. Identification of Burkholderia pseudomallei Near-Neighbor Species in the Northern Territory of Australia. PLoS Negl Trop Dis 9.

38. Levy A, Merritt AJ, Aravena-Roman M, Hodge MM, Inglis TJJ. 2008. Expanded range of Burkholderia species in Australia. American Journal of Tropical Medicine and Hygiene 78:599-604.

39. Birnie E, van't Hof S, Bijnsdorp A, Mansaray Y, Huizenga E, van der Ende A, Hugenholtz F, Grobusch MP, Wiersinga WJ. 2019. Identification of Burkholderia thailandensis with novel genotypes in the soil of central Sierra Leone. PLoS Negl Trop Dis 13.

40. Glass MB, Gee JE, Steigerwalt AG, Cavuoti D, Barton T, Hardy RD, Godoy D, Spratt BG, Clark TA, Wilkins PP. 2006. Pneumonia and septicemia caused by Burkholderia thailandensis in the United States. J Clin Microbiol 44:4601-4.

41. Smith MD, Angus BJ, Wuthiekanun V, White NJ. 1997. Arabinose assimilation defines a nonvirulent biotype of Burkholderia pseudomallei. Infect Immun 65:4319-4321.

42. Gee JE, Elrod MG, Gulvik CA, Haselow DT, Waters C, Liu L, Hoffmaster AR. 2018. Burkholderia thailandensis Isolated from Infected Wound, Arkansas, USA. Emerg Infect Dis 24:2091-2094.

43. Ralph A, McBride J, Currie BJ. 2004. Transmission of Burkholderia pseudomallei via breast milk in northern Australia. Pediatr Infect Dis J 23:1169-71.

44. Land M, Hauser L, Jun SR, Nookaew I, Leuze MR, Ahn TH, Karpinets T, Lund O, Kora G, Wassenaar T, Poudel S, Ussery DW. 2015. Insights from 20 years of bacterial genome sequencing. Functional & Integrative Genomics 15:141-161.

45. Nandi T, Holden MTG, Didelot X, Mehershahi K, Boddey JA, Beacham I, Peak I, Harting J, Baybayan P, Guo Y, Wang S, How LC, Sim B, Essex-Lopresti A, Sarkar-Tyson M, Nelson M, Smither S, Ong C, Aw LT, Hoon CH, Michell S, Studholme DJ, Titball R, Chen SL, Parkhill J, Tan P. 2015. Burkholderia pseudomallei sequencing identifies genomic clades with distinct recombination, accessory, and epigenetic profiles (vol 25, pg 129, 2015). Genome Research 25:608-608.

46. Price EP, Currie BJ, Sarovich DS. 2017. Genomic Insights Into the Melioidosis Pathogen, Burkholderia pseudomallei. Current Tropical Medicine Reports 4:95-102.

47. 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 101:14240-5.

48. Spring-Pearson SM, Stone JK, Doyle A, Allender CJ, Okinaka RT, Mayo M, Broomall SM, Hill JM, Karavis MA, Hubbard KS, Insalaco JM, Mcnew LA, Rosenzweig CN, Gibbons HS, Currie BJ, Wagner DM, Keim P, Tuanyok A. 2015. Pangenome Analysis of Burkholderia pseudomallei: Genome Evolution Preserves Gene Order despite High Recombination Rates. Plos One 10.

CHAPTER 2 *Burkholderia pseudomallei*, the causative agent of melioidosis, is rare but ecologically established and widely dispersed in the environment in Puerto Rico

Abstract

Background

Burkholderia pseudomallei is a soil-dwelling bacterium and the causative agent of melioidosis. The global burden and distribution of melioidosis is poorly understood, including in the Caribbean. *B. pseudomallei* was previously isolated from humans and soil in eastern Puerto Rico but the abundance and distribution of *B. pseudomallei* in Puerto Rico as a whole has not been thoroughly investigated.

Methodology/Principal Findings

We collected 600 environmental samples (500 soil and 100 water) from 60 sites around Puerto Rico. We identified *B. pseudomallei* by isolating it via culturing and/or using PCR to detect its DNA within complex DNA extracts. Only three adjacent soil samples from one site were positive for *B. pseudomallei* with PCR; we obtained 55 isolates from two of these samples. The 55 *B. pseudomallei* isolates exhibited fine-scale variation in the core genome and contained four novel genomic islands. Phylogenetic analyses grouped Puerto Rico *B. pseudomallei* isolates into a monophyletic clade containing other Caribbean isolates, which was nested inside a larger clade containing all isolates from Central/South America. Other *Burkholderia* species were commonly observed in Puerto Rico; we cultured 129 isolates from multiple soil and water samples collected at numerous sites around Puerto Rico, including representatives of *B. anthina*, *B. cenocepacia*, *B. cepacia*, *B. contaminans*, *B. glumae*, *B. seminalis*, *B. stagnalis*, *B. ubonensis*, and several unidentified novel *Burkholderia* spp.

Conclusions/Significance

B. pseudomallei was only detected in three soil samples collected at one site in north central Puerto Rico with only two of those samples yielding isolates. All previous human and environmental *B. pseudomallei* isolates were obtained from eastern Puerto Rico. These findings suggest *B. pseudomallei* is ecologically established and widely dispersed in the environment in Puerto Rico but rare. Phylogeographic patterns suggest the source of *B. pseudomallei* populations in Puerto Rico and elsewhere in the Caribbean may have been Central or South America.

Author Summary

The objective of this study was to examine the distribution and abundance of *Burkholderia pseudomallei* in the environment in Puerto Rico. *B. pseudomallei* is a microbe that lives in soil and causes the disease melioidosis. We conducted sampling around Puerto Rico to survey for the presence of *B. pseudomallei* in the environment. Of the 600 environmental samples collected, we isolated live *B. pseudomallei* from just two soil samples collected from the same site, which was in a region of the island where *B. pseudomallei* had never been previously reported. These results suggest *B. pseudomallei* is widely dispersed but rare in the environment in Puerto Rico. *B. pseudomallei* isolates from Puerto Rico are most closely related to other strains from the Caribbean. Caribbean strains are inside a larger group that contained all analyzed isolates from Central/South America, suggesting that *B. pseudomallei* populations in the Caribbean may have been introduced from Central or South America.

Introduction

The *Burkholderia* genus is a group of diverse, primarily soil-dwelling, Gram-negative bacteria that have many strategies to survive and persist in soil, including acid tolerance (1) and

intrinsic antibiotic resistance (2). These species employ a wide variety of ecological strategies, including degradation of common pollutants, mutualistic relationships with plants, and also pathogenic relationships plants, humans, and/or animals (3-10). The taxonomy of this genus remains incomplete and new species are regularly described (3-5). The genus is commonly separated into two major phylogenetic groups: the *B. pseudomallei* complex (Bpc), consisting of *B. pseudomallei* and its most closely related phylogenetic relatives, and the *B. cepacia* complex (Bcc) (6). The Bcc includes a number of species that can be opportunistic pathogens of immunocompromised individuals, especially cystic fibrosis (CF) patients (7, 8). Some other *Burkholderia* species are not assigned to either of these complexes, including the important plant pathogens *B. glumae* and *B. gladioli*; *B. glumae* causes bacterial panicle blight, a devastating disease in rice plants (9).

B. pseudomallei is the causative agent of the disease melioidosis and considered a Tier 1 Select Agent by the US Centers for Disease Control and Prevention (CDC) (2, 10). Melioidosis can be contracted via cutaneous inoculation, inhalation, or ingestion, and can present with extremely varied symptoms (11); these vague symptoms and diverse clinical presentations, along with culture-based diagnostic anomalies, make it difficult to properly diagnose in clinical settings (2, 12). No vaccines against *B. pseudomallei* are currently available, making rapid detection and specific antibiotic treatment crucial for favorable outcomes in infected humans. Successful antibiotic treatment typically includes a strict and long regimen of intravenous antibiotics, such as ceftazidime or meropenem, for at least two weeks, followed by oral antibiotics, such as co-trimoxazole, for up to six months (2, 10). However, treatment can be complicated by the fact that *B. pseudomallei* is intrinsically resistant to several clinically relevant antibiotics (13). Importantly, other *Burkholderia* species that co-exist with *B. pseudomallei* in the

environment are known to have an intrinsic resistance to other clinically relevant antibiotics, such as meropenem resistance in *B. ubonensis* (14), and thereby represent a possible source of similar resistance in *B. pseudomallei*.

B. pseudomallei has an "open genome" that can readily incorporate new genomic content via lateral gene transfer (15). As a result, it has a relatively large accessory genome (i.e. the genomic features variable present among different *B. pseudomallei* strains) and a relatively small core genome (i.e. the genomic features present in all *B. pseudomallei* strains). The core genome is currently estimated at ~1,600 genes but will likely continue to decrease due to a process known as core genome decay, just as the accessory genome will continue to increase (6). This is because, as additional *B. pseudomallei* genomes are generated from novel isolates, components previously identified as part of the core genome will be missing in some of the new genomes and completely novel components will also be identified, both of which increase the size of the accessory genome (6). Genomic islands, often associated with tRNA sequences (16), contribute much of the genomic diversity observed in the *B. pseudomallei* accessory genome and some are hypothesized to contain virulence components (16, 17). The adaptive potential of the large accessory genome in *B. pseudomallei* may be substantial, but remains poorly understood.

Determining where *B. pseudomallei* is present in the environment is crucial for understanding the potential risk to humans of acquiring melioidosis. This is because almost all infections with *B. pseudomallei* are independently acquired from the environment (27); human to human transmission of melioidosis is extremely rare (18). *B. pseudomallei* has long been known to be endemic in tropical regions in northern Australia and Southeast Asia but the true global distribution appears to be much larger. Because melioidosis can be difficult to diagnose, it is

possible that *B. pseudomallei* is also present in the environment in other regions of the world and causing human disease in these areas but going undetected (19).

The majority of Puerto Rico experience a tropical rainforest climate (Af) based on the Köppen climate classification which is commonly associated with the presence of *B. pseudomallei* and human melioidosis cases have been previously reported from the island. Since 1982, there have been a total of seven reported human cases from Puerto Rico, all from the more populated eastern portion of the island (Fig 1) (20, 21). A recent human melioidosis case from Puerto Rico, in 2012, occurred in the southeast municipality of Maunabo. Subsequent soil sampling in this region in 2013 resulted, for the first time, in the isolation of *B. pseudomallei* from the environment in Puerto Rico (20). These previous human melioidosis cases (all but one with no travel history) and the isolation of *B. pseudomallei* from soil indicated that *B. pseudomallei* was present in the environment in Puerto Rico. The primary goal of this study was to gain a better understanding of the prevalence and geographic distribution of *B. pseudomallei* and other *Burkholderia* spp. in the environment in Puerto Rico. To achieve this goal, we conducted widespread soil and water sampling around the island and analyzed the samples using PCR and culture-based approaches to identify the presence of *B. pseudomallei* and other *Burkholderia* species.

Materials and Methods

Environmental sampling in Puerto Rico

Methods for environmental sampling were based upon international consensus guidelines for sampling for *B. pseudomallei* in the environment (22), with additional modifications developed by the Menzies School of Health Research in Darwin, Australia (23).

Site selection. In April 2017, we surveyed for *B. pseudomallei* in Puerto Rico by collecting and analyzing 600 environmental samples (500 soil and 100 water) from 50 soil sites and 10 water sites around the island. During the month sampling prior to sample collection, Puerto Rico received 5.48 inches of rain in March 2017 and during the month of sample collection Puerto Rico received 4.68 inches of rain in April 2017. This was 192% above the normal rainfall expected in March and 138% above the normal rainfall expected in April (24, 25). At the start of this study, limited information was available regarding the presence of *B. pseudomallei* in the environment in Puerto Rico; a single positive soil sample was identified from a site associated with a previous melioidosis cases (20). Because of this, we did not perform systematic sampling across the entire island but, rather, focused our sampling efforts on locations that we suspected would be most likely to harbor *B. pseudomallei*. Within many endemic regions, such as the tropical Top End of Australia (26) and Laos (27, 28), *B. pseudomallei* is found more often at lower elevations. For this reason, we focused our sampling efforts in lower elevation areas near the outer perimeter of Puerto Rico (Fig 1). In addition, in some endemic regions a greater abundance of *B. pseudomallei* has been documented from soils in agriculture lands as compared to non-agriculture lands (29, 30). We primarily targeted agricultural lands even though other studies have also shown non-agriculture land can also contain a high prevalence of *B. pseudomallei* (31, 32). As a result, we targeted agricultural lands with farm animals, farming, and/or irrigation present. The 60 sites where we collected environmental samples included 53 sites located on agriculture lands, three sites located in natural reserves with little human impact, and four sites on public lands when nearby agriculture sites were not suitable for environmental sampling or were inaccessible. Permission was received from landowners to collect soil and/or water samples on their property and, when necessary, permits were obtained to collect soil and/or

water samples from reserve lands. Because all previous reports of *B. pseudomallei* from humans and the environment in Puerto Rico originated in the eastern portion of the island, suggesting that it may be limited to that geographic region (33), we sampled more intensively in that region (Fig 1).

Soil sampling. At each soil collection site, separate samples were collected at a depth of 30cm below the surface from ten different holes spaced 2.5 meters apart along a single linear transect. Approximately 30g of soil from each sample was placed into a clean 50mL conical tube and stored in a covered insulated container in the shade to prevent direct UV exposure. Digging tools and other equipment were cleaned and decontaminated between samples to prevent cross contamination between samples and sites: equipment was first scrubbed with water to remove any large soil particles and then sprayed with 70% isopropyl alcohol to decontaminate. Soil pH for each sample was measured using a calibrated handheld pH meter: approximately 10g of soil was placed into a 50mL conical tube containing 40mL of DI water and the soil/water mixture was shaken by hand until the mixture was well homogenized and, following a 1-minute incubation at ambient temperature, the pH was measured using an Oakton EcoTestr® pH 2 Waterproof pH Tester, with the pH reading recorded once the value stabilized.

Water sampling. At each water collection site, ten samples were collected along a single linear transect with samples collected approximately 2.5 meters apart, when possible. For each sample, 1L of water was collected into a Whirl-Pak® bag, utilizing an extendable sampling pole when needed. At sites with flowing water, samples were collected near the water edge where there was less disturbance from the current. Water pH was measured from the first and last water samples at each site using a calibrated Oakton EcoTestr® pH 2 Waterproof pH Tester and the pH readings were recorded once the pH values stabilized. Water samples were stored out of direct sunlight at

ambient temperature until ready for filtering. The water samples were filtered in Puerto Rico using a Sartorius water filtration device that consisted of a Combisart® 3 branch manifold with 250mL sterile funnels containing Microsart® filters (cellulose nitrate, 47mm diameter, 0.2µM pore size) and a Microsart® EJet pump. Sterile Minisart® syringe filters (25mm, 0.2µm PTFE) were attached to each branch on the apparatus for sterile venting. Each water sample was split into three parts so that each water sample was filtered through three different filters. Once a water sample was completely filtered, all three filters were collected using sterile forceps and placed into a single sterile 50mL conical tube. All soil samples and filters from water samples were shipped at ambient temperatures to Northern Arizona University (NAU) and stored in the dark at ambient temperature until processed.

Culturing *Burkholderia* species. All culturing activities occurred at NAU and were conducted within containment using a biosafety cabinet in a BSL-2 laboratory, or in a Select Agent BSL-3 facility when *B. pseudomallei* was identified; all requisite entities were notified after detection of *B. pseudomallei*. Culturing followed international consensus guidelines (22) with specific modifications as previously described (34). In short, 20g of soil was aseptically placed into a sterile 50mL bio-reaction tube with a hydrophobic membrane cap for venting (CellTreat®, Pepperell, MA) that contained 20mL of sterile water. The soil and water mixture was vortexed until homogenized while using Parafilm® to cover the tube to prevent the filter cap from becoming saturated. The samples were then incubated for 48 hours at 37°C while shaking at a speed to achieve aeration. After 48 hours, the shaking was stopped and the samples were allowed to settle for one hour before handling. A glycerol stock was created by adding 1mL of the top layer of the soil/water solution into a 2mL cryovial containing 500µL of concentrated Luria-Bertani (LB) broth and glycerol, resulting in a final concentration of 1 x LB broth with 20%

glycerol. This glycerol stock was stored indefinitely at -80°C to serve as a backup culturing reserve. Then, 10µL from the top layer of the water/soil solution was plated on a small portion of half of an Ashdown's agar plate (containing 4mg/mL of gentamycin) and streaked for isolation using the rest of the same half of the plate. 100µL of the top layer of the water/soil solution was plated on the other half of the Ashdown's agar plate.

An enrichment culture in Ashdown's broth was then initiated to favor growth of *B. pseudomallei* and other *Burkholderia* spp. 10mL of the soil/water solution was transferred into a new 50mL filter cap conical tube containing 30mL of Ashdown's broth (containing 50mg/L of colistin). For water samples, all three filters from one water sample were added to a 50mL filter cap conical tube containing 30mL of Ashdown's broth (containing 50mg/L of colistin). Both soil and water samples in Ashdown's broth were incubated at 37°C for seven days while shaking at 130rpm. During incubation, we sampled repeatedly from these enriched broth cultures to test for the presence of *B. pseudomallei*. First, 3mL of the Ashdown's broth incubated 2-5 days was placed into a new tube and pelleted at 3,750 x g for 10 minutes. The supernatant was removed and the pellet was stored at -20°C until ready for DNA extraction (described below). Second, we removed 10µL and 100µL from the top layer of the Ashdown's broth to culture onto a new Ashdown's agar plate after both two and seven days of incubation. Each of the above Ashdown's plates was examined after 48 hours of incubation at 37°C for any colonies of interest. A sub-culture was performed onto a new Ashdown's agar plate if a colony had an appearance similar to *B. pseudomallei*: lavender to purple colonies, dry, slightly textured, with a raised dome or fried-egg morphology, and dimpled/wrinkled centers (35-37). All sub-culture plates were incubated at 37°C for 48 hours until DNA extraction.

Meropenem susceptibility. Meropenem susceptibility was determined for 11 *B. ubonensis* strains and two *B. pseudomallei* strains (Bp9039 and Bp9110) using Etests (bioMérieux, Durham, NC). The meropenem minimal inhibitory concentration (MIC) for six of the 11 *B. ubonensis* strains was previously described (14) and all meropenem Etests were conducted using the same methods described in that previous work. Briefly, an isolate was grown with the Etest on Mueller Hinton agar for 24 hours at 37°C and then zones of inhibition were recorded.

Detection of *B. pseudomallei* and *Burkholderia* spp. PCR was used for the detection of *B. pseudomallei* and *Burkholderia* spp. in DNA extracts taken from enrichment cultures of Ashdown's broth. The microbial community in the Ashdown's broth was screened for the presence of *B. pseudomallei* using a real-time PCR assay that targets *orf2* in the type three secretion system 1 (TTS1) cluster of *B. pseudomallei* (38). DNA was extracted from the stored pelleted broth using QIAamp® Fast DNA Stool Mini Kit (QIAGEN, Germantown, MD) following the manufacturer's instructions, after first re-suspending the broth pellet with 1mL of the InhibitEX Buffer. All Ashdown's broth DNA extractions were diluted to 1/30 using molecular grade water. As a quality control step, the DNA extractions were first screened with a real-time SYBR PCR assay using published conditions (39) of universal 16S rRNA primers (40). If the 16S PCR confirmed that the DNA extractions were successful, then the DNA dilutions were screened in duplicate with a TaqMan assay targeting TTS1 to detect the presence of *B. pseudomallei* DNA (38). Any broth extractions that were initially positive for *B. pseudomallei* with the TTS1 assay were then screened again in triplicate using the same assay; all real-time TaqMan assays were run on ThermoFisher 7900 instruments. Any sample in which the Ashdown's broth community had a signal for *B. pseudomallei* resulted in even more intensive culturing efforts for that particular soil sample, all within NAU's Select Agent BSL-3 facility.

These additional culturing efforts included plating stored glycerol stocks that were created from the soil/water solution (before Ashdown's broth inoculation) onto fresh Ashdown's agar plates. All Ashdown's agar plates were heavily sub-cultured for colonies with morphologies of interest (see above). In some cases, the culturing process was also repeated with a new aliquot of 20g of raw soil.

Single sample *B. pseudomallei* isolates. To investigate the diversity of *B. pseudomallei* within single soil samples (23-07 and 23-09), multiple suspected *B. pseudomallei* colonies were selected from the Ashdown's plates created from the soil glycerol stock. All confirmed *B. pseudomallei* isolates were whole genome sequenced (see below).

Other *Burkholderia* spp. A crude DNA extraction followed by an assay that is largely specific to *Burkholderia* was performed to determine if a colony morphology of interest was a *Burkholderia* spp. DNA was extracted from sub-cultured colonies from the Ashdown's agar plates using a 5% Chelex®-100 heat soak method (41, 42). Using standard PCR, these DNA extracts were screened with *Burkholderia* specific primers BUR3 (43) and BUR5 (35), which target a 365bp region of the *recA* gene; PCR conditions were as previously described (34). The PCR product was run on an agarose gel and, if a band was present at the target size (365bp), Sanger sequencing was conducted using methods for both procedures as previously described (34). Resulting amplicon sequences were searched using NCBI BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) to identify isolates to genus. These approaches also identified other soil-dwelling genera from our samples, such as *Cupravidus*, *Delftia*, *Pseudomonas*, and *Ralstonia*; however, only *Burkholderia* spp. are reported herein.

All molecularly-confirmed *B. pseudomallei* and other *Burkholderia* spp. were processed for long-term storage and whole genome sequencing. A total of three isolation streaks were

performed using Ashdown's agar plates. A single colony was then selected to produce a lawn on a Luria-Bertani (LB) agar plate that was used to create glycerol stocks that are stored indefinitely at -80°C for future use. High quality genomic DNA for whole genome sequencing was extracted from purified isolates using a DNeasy® Blood & Tissue Kit (QIAGEN, Germantown, MD), following the manufacturer's instructions. Prior to sequencing, DNA extractions were screened again with the TTS1 or *recA* PCR assays described above to confirm species identification. Controls were used for all real-time and standard PCR reactions. These included DNA from a reference *B. pseudomallei* strain (K96243) as a positive control, and water for no-template controls (NTCs).

Whole genome sequencing. DNA library construction for whole-genome sequencing (WGS) was performed using KAPA Hyper Prep Kits (Roche, Pleasanton, CA) for Illumina NGS platforms per manufacturer's protocol, with double-sided size-selection performed after sonication. Dual indexing was used (44) with adapters and 8bp index oligos from IDT® (Integrated DNA Technologies, San Diego, CA) used in place of those supplied in the KAPA kit. The final libraries were quantified on an Applied Biosystems™ QuantStudio™ 7 Flex Real-Time PCR System (Invitrogen, ThermoFisher) using the KAPA SYBR® FAST ROX Low qPCR Master Mix (Roche, Pleasanton, CA) for Illumina platforms. The libraries were then pooled together at equimolar concentrations and quality was assessed with a Bioanalyzer DNA 1000 chip (Agilent Technologies, Santa Clara, CA). Final quantitation by qPCR preceded sequencing of the final library. Final pools were sequenced on the Illumina MiSeq platform (Illumina, San Diego, CA) with the 600-cycle v3 kit for 250 cycles.

Genome assembly. Genomes were assembled with SPAdes v3.11.0 (45). Contigs that showed an anomalously low depth of coverage or aligned to known contaminants based on BLASTN (46) alignments against the GenBank (47) nt database were manually removed.

MLST. Multi-locus sequence typing (MLST) was performed *in silico* on the *B. pseudomallei* and other *Burkholderia* spp. genomes, respectively, using information from the existing MLST typing scheme for *B. pseudomallei* (48) and the existing MLST typing scheme for the *B. cepacia* complex (49). All MSLT data, including novel allele sequences and sequence types (STs), were submitted to PubMLST databases, either the *B. pseudomallei* MLST database (https://pubmlst.org/ bpseudomallei/) or *B. cepacia* complex MLST database (https://pubmlst.org/bcc/) (50). Novel STs found in this study are presented in S1 Table.

SNP calling and phylogenetics. To construct a *Burkholderia* spp. phylogeny (not including *B. pseudomallei*), genome assemblies from *Burkholderia* spp. collected in this study (S1 Table), along with a set of reference *Burkholderia* spp. genomes (S2 Table), were aligned against the reference *B. pseudomallei* genome K96243 (51) using NUCmer (52); single nucleotide polymorphisms (SNPs) were then identified using NASP (53). To construct a *B. pseudomallei* only phylogeny, raw reads were aligned against K96243 with BWA-MEM (54) and SNPs were called with the UnifiedGenotyper method in GATK (55). For both methods, SNPs that fell within duplicated regions, based on NUCmer reference self-alignment, were filtered from downstream analyses. Maximum likelihood phylogenies were inferred from concatenated SNP alignments using IQ-TREE v1.6.1 (56) and 1,000 bootstrap replicates. Additionally, to estimate Bayesian time to most recent common ancestor (TMRCA) for each separate *B. pseudomallei* chromosome, individual SNP matrices and phylogenies were generated that included genome assemblies from eight *B. pseudomallei* isolates from Puerto Rico (n=7) and Trinidad (n=1) with

three *B. pseudomallei* isolates from Martinique (S2 Table) to serve as an outgroup. Although both chromosomes 1 and 2 were analyzed separately, no molecular clock signal was detected for chromosome 2. As such, only chromosome 1 was used for all subsequent molecular clock analyses.

Comparative genomics. A pan-genome analysis was performed on all new *B. pseudomallei* genomes generated in this study (S3 Table) using the Large scale BLAST score ratio (LS-BSR) pipeline with a 0.95 BSR threshold (57) and the blat (58) alignment option. Coding regions that were variably conserved were extracted from the matrix and visualized with the Interactive Tree of Life (59). Similar approaches were used to determine, within the set of reference genomes (S2 Table), the presence/absence in the reference genomes of coding regions that were variably conserved in the new *B. pseudomallei* genomes from Puerto Rico (S3 Table).

Genomic island identification. From the BSR results, we identified regions that were variably conserved in the genomes of *B. pseudomallei* isolates obtained from site 23 but were absent from other, geographically diverse *B. pseudomallei* genomes. From the NCBI PGAP annotation, we identified coding and flanking regions that were associated with identified genomic islands. To screen for the presence of these genomic islands in other *Burkholderia* spp., we screened all coding regions associated with genomic islands against all *Burkholderia* spp. genomes with LS-BSR, using a BSR threshold of >0.8 for presence; a lower threshold was used as diverse species were being screened. The structures of the genomic islands were visualized with the genoPlotR R package (60) using the PGAP annotation.

Root-to-tip regression analysis. Using the SNP alignment of chromosome 1 from the Puerto Rico sample subset, the program Gubbins (61) was used to test for and remove recombination, as this can confound divergence-dating analyses; all subsequent timing analyses were performed

with the resulting data. A temporal signal was assessed in the program TempEst version 1.5.1 (62) using regression analysis implementing root-to-tip genetic distance as a function of the sample year. A measure of clocklike behavior was assessed using the determination coefficient R^2, with the best-fitting root selected to maximize R^2. To evaluate the significance of the regression analysis, we performed 10,000 random permutations of the sampling dates over the sequences (63).

Estimations of divergence times. A Bayesian relaxed molecular clock using tip dating was applied using the BEAST version 1.8.4 software package (64) to estimate the TMRCA for chromosome 1 of eight Puerto Rico isolates using three *B. pseudomallei* isolates from Martinique as an outgroup. The best nucleotide substitution model was inferred using the Bayesian information criterion and MEGA7 software (65). BEAST analysis was run with a correction for invariant sites by specifying a Constant Patterns model in the BEAST xml file. A "path and stepping stone" sampling marginal-likelihood estimator was used to determine the best-fitting clock and demographic model combinations (66). The log marginal likelihood was used to assess the statistical fits of 10 clock and demographic model combinations. Four independent chains of 750 million iterations each were run for the best clock and demographic model combination. Convergence among the four chains was confirmed in the program Tracer (version 1.6.0) (67). Molecular clock and demographic combination testing was performed using strict and relaxed molecular clocks in combination with five demographic priors. Model combinations that failed to converge were discarded.

YLF and ITS typing. The presence/absence of the *Yersinia*-like fimbrial (YLF) gene (68) in all new *B. pseudomallei* strains collected in this study was determined by LS-BSR (accession sequence YP_110141.1). In addition, all new isolates were typed for length polymorphisms in

the 16S-23S internal transcribed spacer (ITS) (69) by LS-BSR [accession sequences type C (FJ981718.1), type G (FJ981723.1), and type E (FJ981706.1)].

Data accession. All short reads and genome assemblies were submitted to GenBank under BioProject accession PRJNA451205. Accession numbers for individual genomes are shown in S1-2 Tables.

Results

B. pseudomallei was found in the environment at one new location in Puerto Rico

Our broad environmental survey in April 2017 resulted in the identification of *B. pseudomallei* from soil samples collected at only one new location, in the northern municipality of Arecibo (Fig 1). Just three of the DNA extracts obtained from the 600 enriched Ashdown's broth samples contained *B. pseudomallei* DNA; *B. pseudomallei* was not detected in any of the 100 water samples. The positive samples originated from three adjacent soil samples collected from a single sampling site (site 23; Fig 1). Site 23 was located on a farm in the municipality of Arecibo where swine, goats, chickens, and cattle were present. DNA extractions from three Ashdown's broth samples (23-07, 23-08, and 23-09) tested positive with the *B. pseudomallei*-specific TTS1 PCR assay (run in triplicate). It is important to note that the presence of *B. pseudomallei* DNA in these complex DNA samples did not definitively indicate that live *B. pseudomallei* was present in the enriched broth samples or the original soil samples. However, this information allowed us to refocus our culturing efforts on these samples to attempt to isolate *B. pseudomallei*. Two of the three soil samples did yield *B. pseudomallei* cultures: *B. pseudomallei* isolate Bp9039 was obtained from soil sample 23-07 on the first culturing attempt, and a second round of culturing from another 20g of soil from sample 23-09 yielded *B.*

pseudomallei isolate Bp9110. Additional rounds of culturing yielded other isolates from 23-07 and 23-09 (S3 Table) but no *B. pseudomallei* isolates were ever obtained from soil sample 23-08 despite a positive *B. pseudomallei* DNA signal from the Ashdown's broth extraction and multiple attempts at culturing; this is not an uncommon occurrence when surveying for *B. pseudomallei* in the environment (22). *B. pseudomallei* isolates Bp9039 and Bp9110 were both susceptible to meropenem; other collected *B. pseudomallei* isolates were not tested.

pH of environmental samples

The pH of the soil samples collected around the island varied greatly from highly acidic to highly alkaline, whereas the water samples varied from a neutral pH to a highly alkaline pH (S4 Table). All soil samples across all sites had an average pH of 7.3 with a range of 3.2-11 and all water samples from all sites had an average pH of 7.8 with a range of 6.8-10.2. The three *B. pseudomallei*-positive soil samples from site 23 (07, 08, and 09) yielded pH values of 4.9, 4.9, and 5.1, respectively. The pH of soil samples 01-06 from site 23 were 7.6, 7.2, 7.0, 5.1, 6.2, and 6.8, respectively; the pH of soil sample 10 from site 23 was 4.9. No association between soil pH and the occurrence of *B. pseudomallei* was detected in this study, which was not unexpected given the very small number of *B. pseudomallei*-positive soil samples (n=3).

***B. pseudomallei* isolates from this study are similar to previous isolates from Puerto Rico**

Within a core genome phylogeny of 414 globally diverse *B. pseudomallei* isolates (Fig 2, S2 Table), two *B. pseudomallei* isolates from the municipality of Arecibo (Bp9039 and Bp9110) are highly similar to each other (both isolates were assigned to MLST ST297) and are nested within a large monophyletic group that contains all included *B. pseudomallei* isolates from other locations in the Caribbean, Central and South America, Mexico, and Africa (Fig 2, panel B). Within this larger group, the new isolates from Puerto Rico and the previous *B. pseudomallei*

isolates obtained from Puerto Rico form a distinct subgroup together with one isolate from Trinidad. Within that subgroup are two distinct lineages: one including the new environmental isolates from Arecibo with some previous clinical isolates from Puerto Rico, and a second lineage including the previous environmental isolates collected near the 2012 clinical isolate from Maunabo, as well as the one 2012 clinical isolate from Trinidad.

The root-to-tip regression analysis identified weak clocklike behavior among the Puerto Rico and Martinique sample set with an R^2 value of 0.1201. However, the positive regression slope indicates molecular clock analysis is still reliable for mutation rate estimation (70). The best-fitting nucleotide substitution model implemented based on MEGA7 model testing was GTR. The 10,000-date randomization permutation testing produced a p-value of 0.184, suggesting that the R^2 value produced in the root-to-tip regression analysis was not statistically different than random chance. Stepping-stone and path-sampling analyses did not show marked differences; a relaxed clock and extended Bayesian skyline plot was selected as the model combination for this analysis. The BEAST timing analysis had a mean estimate of the year 1950 (95% HPD, 1923 to 1975; S1 Fig) for the TMRCA of chromosome 1 for the eight *B. pseudomallei* isolates from Puerto Rico and Trinidad. The evolutionary rate was estimated at 5.01E-6 (95% HPD, 2.81E-6 to 8.28E-6) for all eight samples and the Martinique outgroup. This is in contrast to another study that found an evolutionary rate of 1.80E-6 (95% HPD, 1.36E-6 to 2.66E-6) for chromosome 1 for multiple *B. pseudomallei* isolates from the Americas (71).

B. pseudomallei genomic diversity observed from a single sampling site

We observed fine-scale genomic diversity among multiple *B. pseudomallei* isolates obtained from a single sampling site and even from a single soil sample. A total of 55 *B. pseudomallei* isolates were isolated from two soil samples at site 23, with 50 isolates from soil

sample 23-07 and five isolates from soil sample 23-09 (S3 Table). It is important to note that these isolates were obtained from enriched culture medium so it is possible that less than 55 individual *B. pseudomallei* cells were present in the original samples. All 55 isolates were similar in regards to being assigned to the same ST (297) and to ITS type G; they all also contained the YLF gene cassette. However, variation was still observed among these strains in the core genome phylogeny (48 unique SNP genotypes were identified among the 55 isolates). There are three distinct clades (A-C) observed in the core genome phylogeny for these isolates, with isolates from soil sample 07 (*n*=50) assigning to all three clades, while all isolates from soil sample 09 (*n*=5) assigned to just clade A along with two of the isolates from soil sample 07 (S2 Fig). Interestingly, there were two *B. pseudomallei* isolates from different soil samples (Bp9046-sample07 and Bp9110-sample09) that were very similar: these two strains exhibit no SNP differences in the core genome (S2 Fig).

We identified four distinct genomic islands (GI1-GI4) among the 55 *B. pseudomallei* isolates obtained from site 23 (Fig 3), and these contain a subset of genes not found in any other *B. pseudomallei* genomes. The insertion of these genomic islands appears to be associated with tRNA gene loci (data not shown), which is similar to previous patterns described from *B. pseudomallei* (16). GI1 (comprised of 61 genes; S5 Table) is conserved across all 55 of the *B. pseudomallei* isolates from site 23, whereas GI2 (comprised of 15 genes; S5 Table), GI3 (comprised of 29 genes; S5 Table), and GI4 (comprised of 5 genes; S5 Table) are variably present among the 55 isolates from site 23 (Fig 3B). The accessory genome of the 55 *B. pseudomallei* isolates from site 23 is comprised of 58 genes: 49 (1-49; S2 Fig, S5 Table) are contained in GI2, GI3, and GI4 and nine others occur at different genomic locations (50-58; S2 Fig, S5 Table). GI2 is conserved among clade A isolates but not found in the other two clades,

GI3 is conserved among clade B isolates but not found in the other two clades, and GI4 is conserved among clade B isolates and variably present in clade A and clade C isolates (S2 Fig). None of the four genomic islands were found in a complete form in 412 other globally diverse *B. pseudomallei* genomes that were examined, including the genomes of *B. pseudomallei* isolates obtained from other locations in Puerto Rico (Fig 3B, S5 Table), nor in the genomes of 781 other *Burkholderia* spp. isolates (S5 Table). A majority (*n*=44) of the 61 genes within GI1 were found in at least one of the genomes of the 1,193 other *B. pseudomallei* and/or *Burkholderia* spp. isolates that were examined, but the other 17 genes in GI1 were only found in the site 23 isolates (S5 Table); none of the genes in GI2 were found in these other genomes (S5 Table). A majority of the genes within GI3 (25/29) and GI4 (4/5) were not found in any of the other 1,193 genomes, but all of the nine accessory genes that occurred outside of the genomic islands were found in other *B. pseudomallei* genomes and some were also found in the genomes of other *Burkholderia* spp. (S5 Table). Interestingly, it was more common for genes from GI1 and GI3 to be found in the genomes of other *Burkholderia* spp. than in the genomes of other, global *B. pseudomallei* isolates (S5 Table).

Widespread environmental dispersal of many other *Burkholderia* species

A number of other *Burkholderia* species are widespread and common in both soil and water throughout Puerto Rico (S4 Table). A total of 686 sub-cultures were selected from Ashdown's agar plates from 301 soil samples (collected from all 50 soil collection sites) and 77 water samples (collected from all 10 water collection sites). Of these sub-cultures, 129 were identified as members of the *Burkholderia* genus according to the sequence of a *recA* gene fragment. Most of the 129 *Burkholderia* isolates (*n*=104) were isolated from 61 different soil samples (originating from 20 of the 50 soil sampling sites), with only 25 isolated from 22

different water samples (but originating from seven of the 10 water sampling sites). *Burkholderia* spp. were cultured from the environment in 21 of the 41 sampled municipalities within Puerto Rico. It is important to note that this does not indicate that there were not *Burkholderia* spp. present in the environmental samples collected at the other 20 municipalities, only that we did not successfully culture any *Burkholderia* spp. from environmental samples collected from those locations using our methods.

The 129 *Burkholderia* spp. isolates were identified in a whole genome phylogeny (Fig 4) as follows: *B. anthina* (*n*=2), *B. cenocepacia* (*n*=29), *B. cepacia* (*n*=15), *B. contaminans* (*n*=5), *B. glumae* (*n*=1), *B. seminalis* (*n*=2), *B. stagnalis* (*n*=36), *B. ubonensis* (*n*=11), and other unidentified novel *Burkholderia* spp. (*n*=28) (S1 Table). A total of 332 novel MLST alleles were identified from the 129 isolates, resulting in 102 novel STs (S1 Table). All *Burkholderia* isolates cultured from this study belong to the *B. cepacia* complex with the exception of *B. glumae,* which is genetically distinct from both the Bpc and Bcc. The single *B. glumae* isolate was identified from a water sample collected in Patillas, Puerto Rico. *B. ubonensis* appears widespread throughout Puerto Rico, with 11 isolates obtained from five municipalities spread across the island [Barceloneta (*n*=1), Cabo Rojo (*n* =4), Ceiba (n=4), Juncos (*n* =1), and Maunabo (*n*=1); S4 Table]. All 11 *B. ubonensis* isolates were resistant to meropenem (>32μg/mL) (S1 Table).

Discussion

B. pseudomallei is ecologically established and widely dispersed in the environment in Puerto Rico but rare. It has now been isolated from soil samples from two regions of the island separated by >100 kilometers (Fig 1). Despite this widespread geographic distribution, it is also

quite rare in the environment in Puerto Rico. Even with extensive sampling at 60 different soil and water sites located around the island (Fig 1), *B. pseudomallei* was detected at just one of the 50 soil sampling sites and was not detected at any of the 10 water sampling sites. The new location where 55 *B. pseudomallei* soil isolates were obtained is in the central northern municipality of Arecibo; to date, there have been no reports of human melioidosis nor collection of environmental *B. pseudomallei* isolates from this region. The only previously reported human melioidosis cases from Puerto Rico were reported from municipalities located on the eastern region of the island, and the same is true for the two environmental soil isolates (obtained from the same soil sample) previously reported from this region (20) (Fig 1). Of note, in this study we did not detect *B. pseudomallei* from environmental samples collected in the eastern portion of the island even though we sampled more extensively in this region because the previous human and environmental isolates were obtained there. This result is also suggestive of the overall rarity of *B. pseudomallei* in the environment in Puerto Rico.

B. pseudomallei also appears to be rare locally in the environment of Puerto Rico at sites where it is present. When *B. pseudomallei* was previously detected in southeastern Puerto Rico it was only isolated from one of 20 soil samples collected from a single neighborhood (20). Similarly, at the one site where *B. pseudomallei* was isolated from soil in this current study it was only detected in three of the 10 soil samples collected at that site, and these three *B. pseudomallei*-positive soil samples were adjacent to one another (thereby separated by <5 m). These findings suggest a locally clumped distribution for *B. pseudomallei* in the environment in Puerto Rico, a pattern that has also been reported from highly-endemic regions, such as northeast Thailand (72). In addition, it is possible that *B. pseudomallei* is also rare at the level of individual soil samples in Puerto Rico because we made multiple attempts to successfully isolate it from one of the soil

samples (23-09) and we were never able to isolate it from another soil sample (23-08) despite multiple attempts and a PCR result that indicated that *B. pseudomallei* DNA was present in DNA extracted from that same soil sample. This is in contrast to patterns from highly endemic regions, such as Thailand and northern Australia, where hundreds of *B. pseudomallei* isolates can often be obtained from a single soil sample (31, 73, 74).

The overall rarity of *B. pseudomallei* in the environment in Puerto Rico may be due to unsuitable environmental conditions. A recent study (19) estimated global environmental suitability for *B. pseudomallei* based upon predicted models developed using location data from >22,000 documented human and animal cases, which were primarily from highly endemic settings in southeast Asia and northern Australia. In the western hemisphere, this analysis predicted high environmental suitability for *B. pseudomallei* in large areas of northern South America, portions of Central America and Mexico, and several small areas in the southern United States. In contrast, it predicted low environmental suitability for *B. pseudomallei* for most locations in the Caribbean with a few exceptions. One of those exceptions was along the northwest coast of Puerto Rico, including the location in the municipality of Arecibo where we isolated *B. pseudomallei* in this study. That said, it is important to note that this same model predicts low environmental suitability for *B. pseudomallei* for the rest of Puerto Rico, including the eastern portion of the island where the previous *B. pseudomallei*-positive soil sample was collected and all previous known human melioidosis cases occurred (Fig 1).

It seems likely that *B. pseudomallei* was introduced to Puerto Rico relatively recently, possibly from other locations in the Caribbean. A recent introduction of *B. pseudomallei* would provide another explanation for why *B. pseudomallei* is rare in the environment in Puerto Rico. Several previous phylogenomic studies of *B. pseudomallei* have all noted a consistent pattern in

which isolates from the Americas cluster together on a single branch that emerges from the larger clade African clade. Thus, the leading hypothesis for the introduction of *B. pseudomallei* to the Americas from Africa suggest that it occurred via the transatlantic human slave trade in the 16^{th}-19^{th} centuries (17, 71, 75, 76); molecular clock estimates in one of these studies support this proposed timeline (71). Our phylogenetic results are consistent with this pattern of strains from the Americas forming a monophyletic clade that is nested within a larger clade containing all known *B. pseudomallei* isolates from Africa (Fig 2) but also provide further insights because we included additional strains from the Caribbean and other locations from the Americas. Within the monophyletic clade from the Americas, we found that all isolates from the Caribbean, with the exception of one isolate from Aruba, group together in a smaller monophyletic clade, and all of the seven known isolates from Puerto Rico grouped together in a smaller clade with one isolate from Trinidad. The molecular clock estimates from this study support a recent introduction of *B. pseudomallei* to Puerto Rico, within the last 70 years (S1 Fig, S2 Table). Together, these findings suggest that *B. pseudomallei* may have been first introduced to other regions of the Americas from Africa and then, more recently, was introduced to the Caribbean from these other regions of the Americas. Because isolates from Puerto Rico group together within a larger clade containing all but one of the other isolates from Caribbean, it is tempting to suggest that *B. pseudomallei* was introduced to Puerto Rico from other locations in the Caribbean. However, it is important to note that all of these ideas are based upon analysis of currently available *B. pseudomallei* isolates. As there are numerous countries where *B. pseudomallei* is thought to occur but has not yet been detected, especially in the western hemisphere (19), the global phylogeographic patterns of *B. pseudomallei* will almost certainly change as additional isolates are obtained from new

locations and sequenced. In particular, more environmental sampling is necessary to better understand the occurrence and spread of *B. pseudomallei* in the Caribbean.

All included isolates from Puerto Rico, together with a single isolate from Trinidad, share a recent common ancestor in the global phylogeny (Fig 2), which is suggestive of a single introduction to Puerto Rico. In addition, the new *B. pseudomallei* soil isolates from Arecibo, the two previous soil isolates from Maunabo, and the previous human isolates from Maunabo have all been assigned to ST297, as has the human isolate from Trinidad. This pattern is not unexpected as ST297 in *B. pseudomallei* is typically associated with isolates from the Western Hemisphere (75). This overall lack of diversity is in contrast to patterns observed in highly-endemic settings, such as northeast Thailand and northern Australia, where high levels of genetic diversity are observed at multiple spatial scales, including among multiple isolates obtained from single soil samples, using multiple genotyping approaches, including MLST, mutli-locus variable number tandem repeat analysis, and pulse field gel electrophoresis (31, 77). That said, there are also two distinct lineages among the Puerto Rican isolates and one of these lineages contains two human isolates from Puerto Rico that were assigned to ST92 and ST95 (20) (Fig 2). Thus, it is also plausible that *B. pseudomallei* has been introduced to Puerto Rico multiple times. Again, additional environmental sampling in the Caribbean, including Puerto Rico, is needed to better understand these patterns.

As an alternative to the hypothesis of human-mediated dispersal, *B. pseudomallei* may have been introduced to Puerto Rico via hurricanes or other extreme weather events. In Australia and Asia, *B. pseudomallei* can become aerosolized during extreme weather events like cyclones (i.e., hurricanes), leading to subsequent increases in human disease events (78-80). The predominate path of Atlantic hurricanes is generally from the east-southeast to the west-

northwest, essentially directly through the Caribbean (81), which would be consistent with hurricanes dispersing *B. pseudomallei* to Puerto Rico from other locations to the southeast of it in South America or other regions of the Caribbean. Long distance dispersal of *B. pseudomallei* during extreme weather events also offers potential explanations for several other patterns observed in the phylogeny (Fig 2). For example, the single isolate from Trinidad, from a 2012 clinical case, clusters together with a 2012 clinical isolate from Puerto Rico (and two soil isolates collected in 2013 near the residence of the 2012 Puerto Rico case) rather than with isolates from more nearby locations in South America. In addition, a 2012 clinical isolate from Aruba clusters together most closely with isolates from Mexico and Central and South America, rather than with other isolates from the Caribbean (Fig 2). Of note, the 2012 Atlantic hurricane season was more active than normal, including 10 different hurricanes (82).

The four unique genomic islands identified from the genomes of *B. pseudomallei* isolates from site 23 may be indicative of adaptation to local ecological conditions. Identifying novel genomic components from new *B. pseudomallei* genomes is not at all unexpected as the accessory genome of this species is quite large and continues to grow as more isolates are sequenced (6), likely because this species has an "open genome" that can readily acquire new genomic content via lateral gene transfer (15). However, what is striking is that the four genomic islands, as well as a majority of the genes within them, are only found in *B. pseudomallei* isolates from site 23 and not in other *B. pseudomallei* isolates or isolates from other *Burkholderia* spp. (S2 Table). Given the almost complete absence of these genomic islands in globally diverse isolates of *B. pseudomallei*, the genes contained within these genomic islands may have been obtained locally from other soil dwelling species as a means of adapting to fine-scale environmental conditions. In some cases, other *Burkholderia* spp. also shared a portion of these

genomic islands, providing a potential source of these accessory genes. However, over half of the accessory genes found in the *B. pseudomallei* isolates from site 23 were completely absent from all of the other global *B. pseudomallei* and *Burkholderia* spp. genomes that were examined (S2 Table), suggesting that the source of many of these accessory genes may be species outside the *Burkholderia* genus that co-occur in the soil. Additional studies of the accessory genome of multiple *B. pseudomallei* isolates obtained from soil samples collected across small spatial scales will be important for yielding new insights into the possibility of the acquisition of new accessory genes representing a mechanism for *B. pseudomallei* to adapt to local ecological conditions.

Other diverse *Burkholderia* spp. are widespread in the environment in Puerto Rico and may be the source of some of the unique accessory genes in the *B. pseudomallei* isolates from site 23 in the municipality of Arecibo. Indeed, a number of the novel *B. pseudomallei* genes present in the four genomic islands described here were also identified from other *Burkholderia* spp. (S5 Table). Although there was no evidence of other members of the Bpc, such as *B. thailandensis*, *B. oklahomensis*, or *B. humptydooensis*, being present in Puerto Rico, many different *Burkholderia* spp. from the Bcc were isolated from both soil and water in Puerto Rico, including some potentially novel species (S4 Table; Fig 4). Due to the large number of novel Bcc MLST alleles and STs identified among these Bcc isolates, it appears that Puerto Rico harbors many unique and diverse Bcc strains that have yet to be classified. And it is important to note that our survey almost certainly provides a limited understanding of the true diversity of *Burkholderia* spp. present in the environment in Puerto Rico as we only examined *Burkholderia* species that were capable of growing on Ashdown's selective medium. Horizontal gene transfer from these other diverse and widespread *Burkholderia* spp. may facilitate adaptation of *B.*

pseudomallei to local environmental conditions in Puerto Rico and elsewhere. One such concern is the transfer of intrinsic antibiotic resistance to clinically relevant antibiotics, such as the potential transfer of the meropenem resistance observed in *B. ubonensis*, to *B. pseudomallei*.

Overall, *B. ubonensis* was quite widespread in Puerto Rico: we isolated it from soil samples collected from five locations in the southwest, southeast, eastern, and north central portions of the island; we did not isolate it from any water samples (S4 Table). *B. ubonensis* has been previously described from the environment only from countries where *B. pseudomallei* is highly endemic, including Australia, Malaysia, Thailand, and Papua New Guinea (6). To our knowledge, the *B. ubonensis* strains collected from Puerto Rico in this study are the first instance of this species being isolated from the environment in the Caribbean and the western hemisphere (14). However, our results from Puerto Rico suggest that it may be widespread in the Caribbean and elsewhere in the western hemisphere. Previous studies have found that *B. ubonensis* is the most common Bcc species to be co-isolated with *B. pseudomallei* (36). The first isolates from Puerto Rico had a distant phylogenetic relationship to isolates from Australia (14). Interestingly, we did not find any evidence of *B. ubonensis* from the municipality of Arecibo where *B. pseudomallei* was isolated. However, we did find *B. ubonensis* from the municipality of Maunabo, where *B. pseudomallei* was previously found in the soil in 2013. We found that all of the *B. ubonensis* strains that we collected from Puerto Rico (11 of 11) were resistant to meropenem (Table S1), an antibiotic used to treat patients with advanced melioidosis, such as sepsis (2). A previous study investigating the meropenem resistance of *B. ubonensis* found that 21% of tested strains from Australia and 67% of tested strains from Thailand were meropenem resistant (14).

Until this study, *B. glumae*, a USDA-APHIS regulated plant pathogen, had not been described from Puerto Rico since 2004, when it was identified from an onion plant (83). We detected *B. glumae* in a water sample from the municipality of Patillas (S4 Table). This plant pathogen causes bacterial panicle blight and can be devastating to various types of crops, including rice (9). As a result, the appropriate regulating agencies within Puerto Rico and the United States were notified of the presence of *B. glumae* at this location. The knowledge of the presence of this plant pathogen at this location can serve as vital information when investigating potential crop infestations.

Conclusions

Widespread environmental surveys for *B. pseudomallei* in the environment in Puerto Rico identified the pathogen from soil samples collected in a region of Puerto Rico from which it had never been previously detected in the environment or in humans. This study demonstrates that although *B. pseudomallei* is present in the environment in several widespread locations in Puerto Rico, it is also rare. Given how rare it is in the environment, *B. pseudomallei* does not appear to pose a large public health risk in Puerto Rico. There have been no known human melioidosis cases reported from the specific location in Arecibo where we detected it in the environment, or even that general region of the island. However, *B. pseudomallei* is clearly ecologically established in Puerto Rico and, as previously suggested (20), both the public and clinicians in Puerto Rico should be made more aware of it. Of note, all but one of the previous human melioidosis cases in Puerto Rico occurred in immunocompromised individuals (20). The International Diabetes Federation stated there were over 400,600 cases of diabetes in Puerto Rico in 2017 with a total prevalence of diabetes in adults of 15.4% (84). As diabetes is an important

risk factor for melioidosis, clinicians should particularly be aware of the possibility of melioidosis in these individuals.

Acknowledgments

We would like to thank all the Animal Health Technicians from USDA-APHIS-VS in Puerto Rico that assisted with sampling locations in the field. We would also like to thank Austin Dikeman, Kim Celona, Daryn Erickson, Ryan Lancione, Ryelan McDonough, Amalee Nunnally, Chase Ridenour, and Bryce Schmidt for their assistance in the laboratory.

Tables and Figures

Chapter 2 Tables

Supporting Tables

S1 Table. Genome accession numbers, MLST, and other data for 129 *Burkholderia* spp. isolates obtained in this study. *Species identification was based upon whole genome sequence comparisons. [a] PS-4 previously described (6). [b] *B. ubonensis* strain resistant to meropenem (>32µg/mL). Yellow highlighted cells indicate novel MLST alleles and STs identified in this study. This supporting table was too large to include in this document. The excel spreadsheet containing this table is available to all at this link (https://doi.org/10.1371/journal.pntd.0007727.s003).

S2 Table. Genome accession numbers for additional *Burkholderia* genomes used to construct the phylogenies presented in Figs. 2 - 4. n/a = not applicable

This supporting table has 1,265 rows and was too large to include in this document. The excel spreadsheet containing this table is available to all at this link (https://doi.org/10.1371/journal.pntd.0007727.s004).

S3 Table. Genome accession numbers and epidemiological data for 55 *Burkholderia pseudomallei* isolates obtained in this study. All of these *B. pseudomallei* isolates were isolated from site 23 in the municipality of Arecibo. These 55 isolates are also presented in Figure 3 and S2 Figure.

ID	Read Accession Number	Assembly Accession Number	Puerto Rico Municipality	Hole ID	Source
Bp9039	SRR7457394	QTQT00000000	Arecibo	07	soil
Bp9041	SRR7457397	n/a	Arecibo	07	soil
Bp9043	SRR7457396	n/a	Arecibo	07	soil
Bp9044	SRR7457399	n/a	Arecibo	07	soil
Bp9045	SRR7457398	n/a	Arecibo	07	soil
Bp9046	SRR7457401	n/a	Arecibo	07	soil
Bp9047	SRR7457400	n/a	Arecibo	07	soil
Bp9048	SRR7457403	n/a	Arecibo	07	soil
Bp9050	SRR7457402	RPDD00000000	Arecibo	07	soil
Bp9051	SRR7457469	n/a	Arecibo	07	soil
Bp9052	SRR7457470	n/a	Arecibo	07	soil
Bp9053	SRR7457471	n/a	Arecibo	07	soil
Bp9054	SRR7457472	n/a	Arecibo	07	soil
Bp9055	SRR7457473	n/a	Arecibo	07	soil
Bp9056	SRR7457476	n/a	Arecibo	07	soil
Bp9057	SRR7457372	n/a	Arecibo	07	soil
Bp9058	SRR7457391	n/a	Arecibo	07	soil
Bp9059	SRR7457474	n/a	Arecibo	07	soil
Bp9060	SRR7457475	n/a	Arecibo	07	soil
Bp9061	SRR7457458	n/a	Arecibo	07	soil
Bp9062	SRR7457457	n/a	Arecibo	07	soil
Bp9063	SRR7457456	n/a	Arecibo	07	soil
Bp9064	SRR7457455	n/a	Arecibo	07	soil
Bp9065	SRR7457454	n/a	Arecibo	07	soil
Bp9066	SRR7457453	n/a	Arecibo	07	soil
Bp9067	SRR7457452	n/a	Arecibo	07	soil
Bp9068	SRR7457451	n/a	Arecibo	07	soil
Bp9069	SRR7457450	n/a	Arecibo	07	soil
Bp9070	SRR7457449	n/a	Arecibo	07	soil
Bp9071	SRR7457485	n/a	Arecibo	07	soil
Bp9072	SRR7457486	n/a	Arecibo	07	soil

Bp9073	SRR7457483	n/a	Arecibo	07	soil
Bp9074	SRR7457484	n/a	Arecibo	07	soil
Bp9075	SRR7457481	n/a	Arecibo	07	soil
Bp9076	SRR7457482	n/a	Arecibo	07	soil
Bp9077	SRR7457479	n/a	Arecibo	07	soil
Bp9078	SRR7457480	n/a	Arecibo	07	soil
Bp9079	SRR7457477	n/a	Arecibo	07	soil
Bp9080	SRR7457478	n/a	Arecibo	07	soil
Bp9081	SRR7457319	n/a	Arecibo	07	soil
Bp9082	SRR7457318	n/a	Arecibo	07	soil
Bp9083	SRR7457321	n/a	Arecibo	07	soil
Bp9084	SRR7457320	n/a	Arecibo	07	soil
Bp9085	SRR7457315	n/a	Arecibo	07	soil
Bp9086	SRR7457314	n/a	Arecibo	07	soil
Bp9107	SRR7457420	RSFB00000000	Arecibo	07	soil
Bp9108	SRR7457419	n/a	Arecibo	07	soil
Bp9109	SRR7457418	n/a	Arecibo	07	soil
Bp9110	SRR7457423	QTQS00000000	Arecibo	09	soil
Bp9111	SRR7457422	n/a	Arecibo	09	soil
Bp9112	SRR7457375	RPDC00000000	Arecibo	09	soil
Bp9113	SRR7457376	n/a	Arecibo	07	soil
Bp9114	SRR7457373	n/a	Arecibo	07	soil
Bp9115	SRR7457374	n/a	Arecibo	09	soil
Bp9116	SRR7457379	n/a	Arecibo	09	soil

S4 Table. Information on 60 sampling sites where environmental samples were collected in Puerto Rico. Soil was collected at sites 1-50 and water was collected at sites 51-60. Bolded site ID indicates the site where *B. pseudomallei* was isolated. * Potentially novel species; ** Soil profile downloaded for each site from USDA NRCS Web Soil Survey; n/a = not applicable.
This supporting table was too large to include in this document. The excel spreadsheet containing this table is available to all at this link (https://doi.org/10.1371/journal.pntd.0007727.s006).

S5 Table. Protein accession numbers and descriptions for the 119 *B. pseudomallei* accessory genes detected among 55 *Burkholderia pseudomallei* isolates obtained in this study. These genes were detected using BSR (see text).
This supporting table was too large to include in this document. The excel spreadsheet containing this table is available to all at this link (https://doi.org/10.1371/journal.pntd.0007727.s007).

Chapter 2 Figures

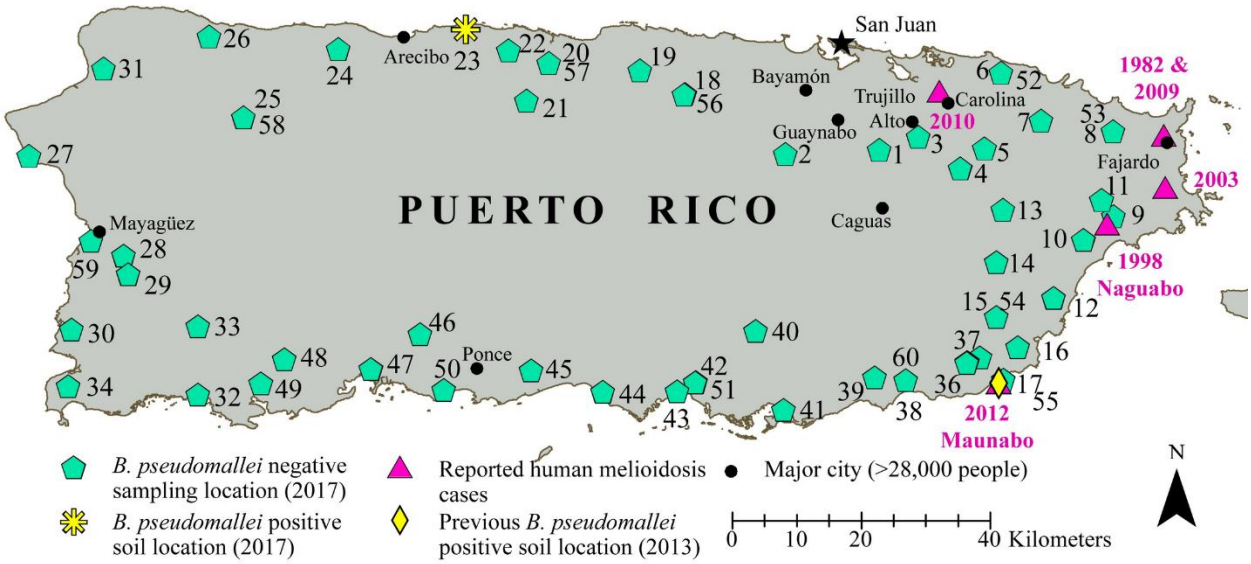


Figure 1. Locations in Puerto Rico where 1) environmental samples were collected in this study, 2) previous human melioidosis cases occurred, and 3) one previous *B. pseudomallei*-positive soil sample was collected. Site numbers are located next to the shapes indicating the 60 sampling locations from this study; soil samples were collected at sites 1-50 and water samples were collected at sites 51-60. The locations of past human melioidosis cases and the one previous *B. pseudomallei* positive soil location (20) are indicated. This map was created using ArcGIS software by Esri.

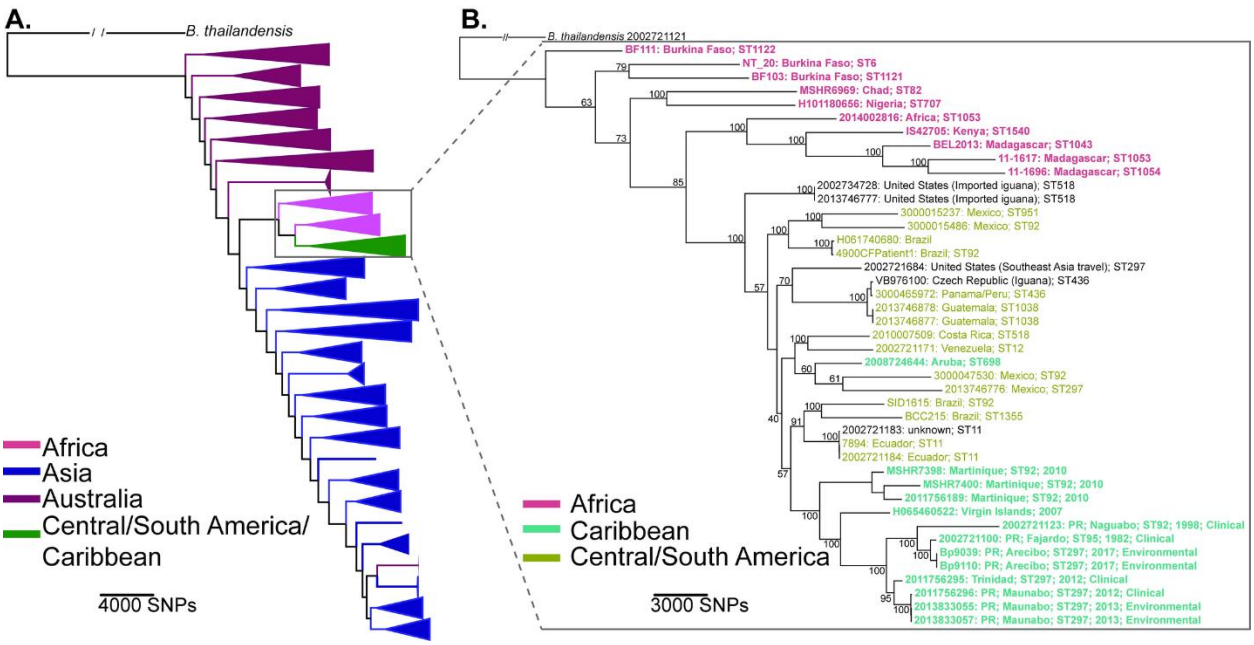


Figure 2. *Burkholderia pseudomallei* global whole genome phylogeny. Maximum-likelihood phylogeny of 414 globally-diverse *B. pseudomallei* isolates rooted with a *B. thailandensis* isolate; bootstrap values are reported on nodes. Two of the 55 *B. pseudomallei* isolates obtained from site 23 in this study are included (Bp9039 and Bp9110). (A) All 414 *B. pseudomallei* genomes with nodes collapsed. (B) Expanded nodes within the monophyletic group containing all included isolates (*n*=44) from Africa, Central and South America, Mexico, and the Caribbean.

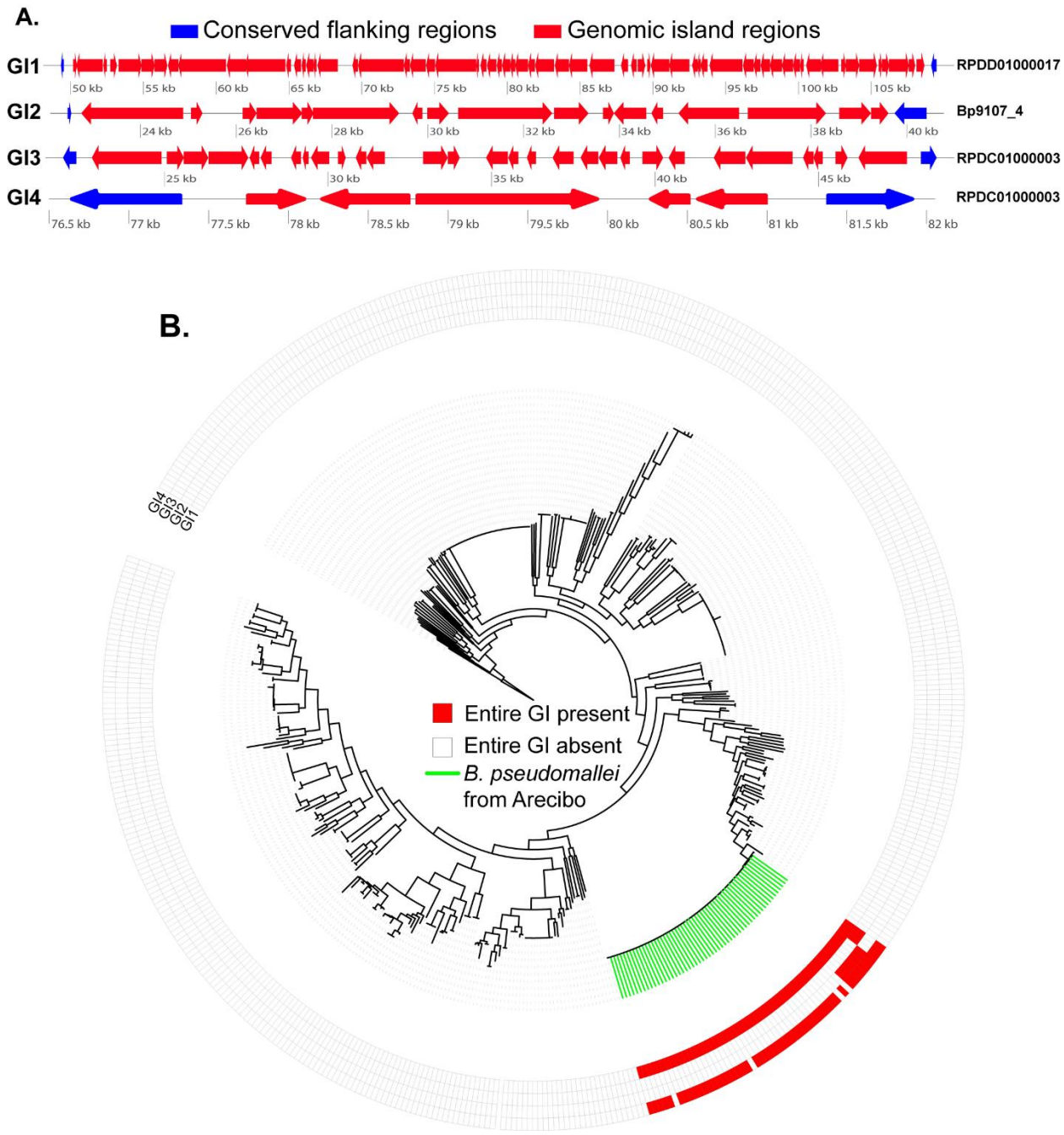


Figure 3. Genomic islands present in *B. pseudomallei* isolates from a single location in Puerto Rico. (A) Reveals the structure of the four novel *B. pseudomallei* genomic islands (GI1-4) that were discovered in this study; contig names are listed on the far right. The red arrows reflect *B. pseudomallei* genes found within the genomic islands and the blue arrows reflect conserved flanking regions commonly found in other *B. pseudomallei* strains. (B) Circular phylogeny with

genomic islands mapped on the outside of the phylogeny. This phylogeny was constructed using the same 414 *B. pseudomallei* isolates used to generate Figure 2 plus the 53 additional *B. pseudomallei* isolates from site 23 that were not included in Figure 2.

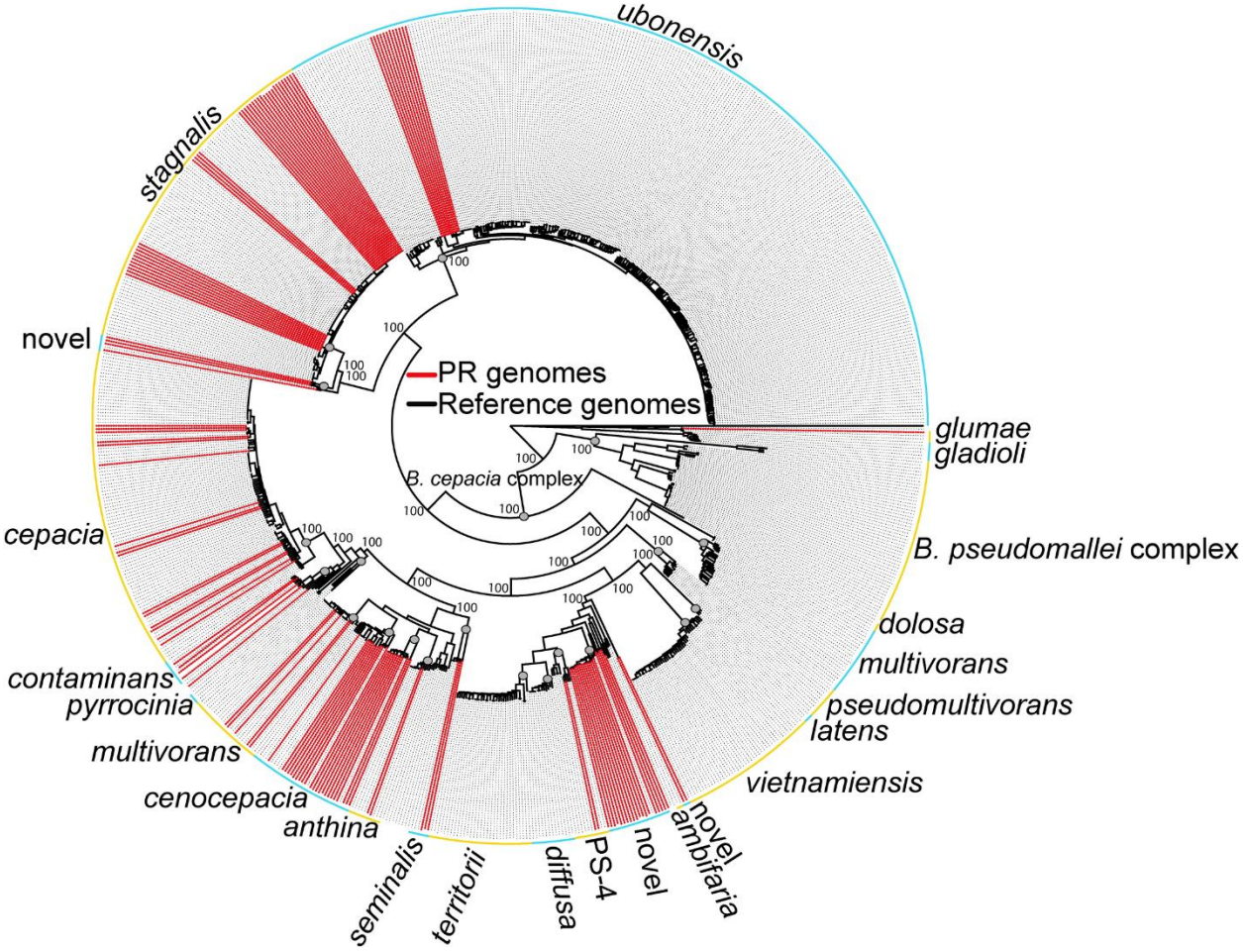
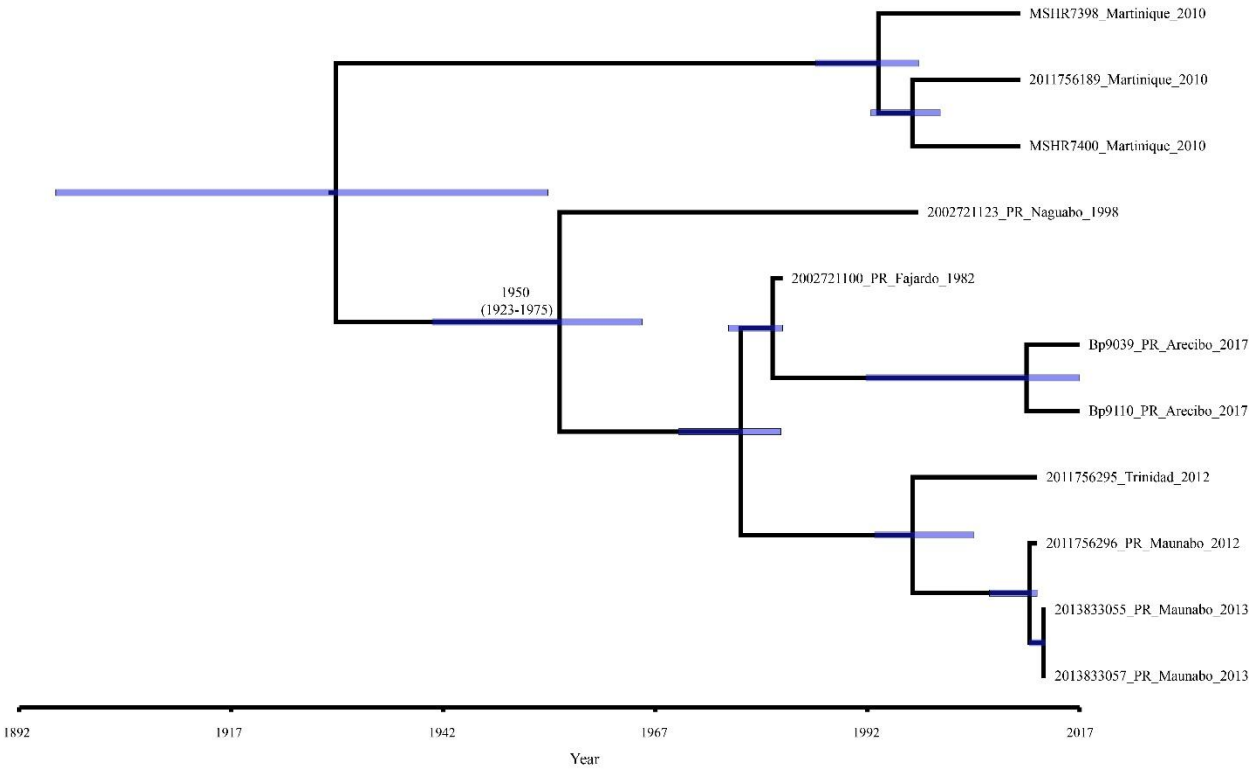
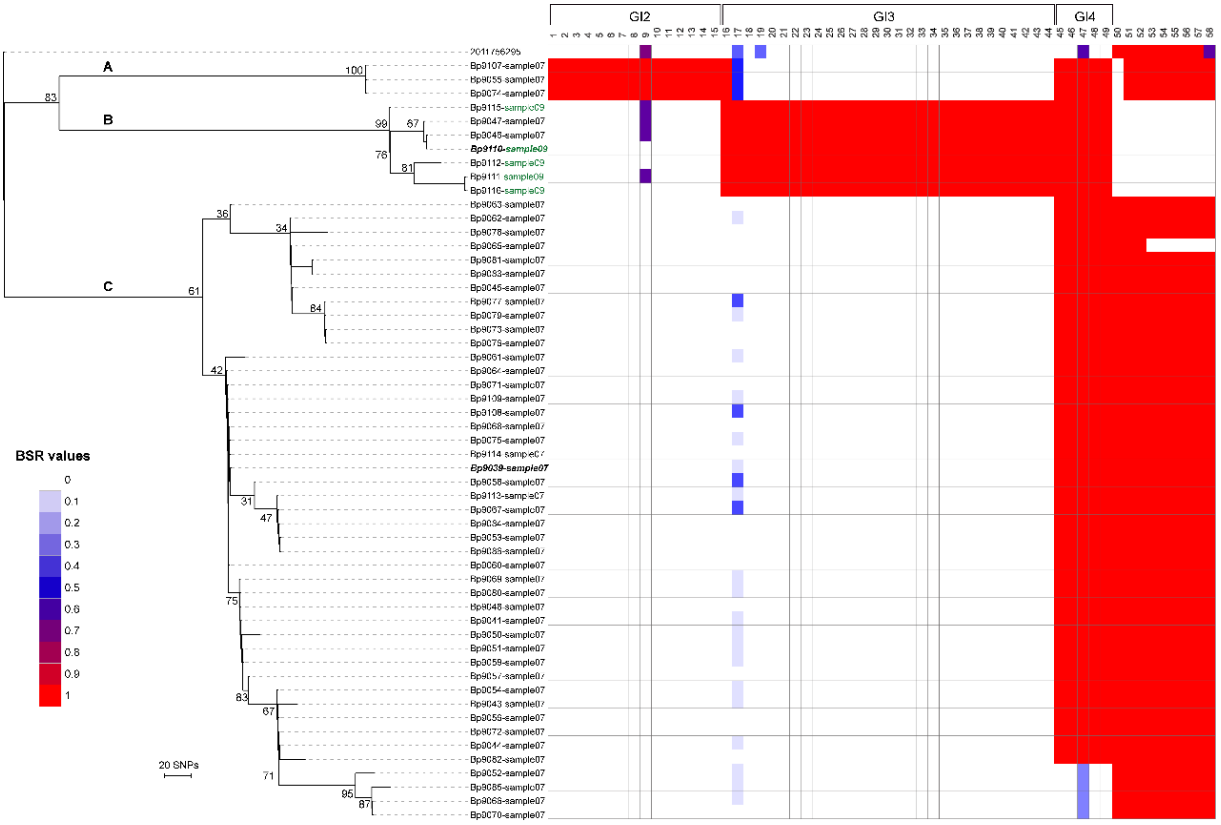


Figure 4. *Burkholderia* spp. whole genome phylogeny. Core genome maximum-likelihood phylogeny of 781 *Burkholderia* spp. isolates rooted with *Cupriavidus*. The 781 isolates include 129 isolates obtained from Puerto Rico in this study (indicated with red lines; S1 Table) and 651 publicly available *Burkholderia* spp. isolates (S5 Table). Bootstrap values are reported on nodes. *B. pseudomallei* is not included in this phylogeny.

Supporting Figures



S1 Figure. BEAST molecular clock estimations (chromosome 1) of eight *B. pseudomallei* isolates from Puerto Rico and Trinidad. BEAST phylogeny with error bars showing 95% highest posterior density (HPD); three *B. pseudomallei* isolates from Martinique were used as an outgroup.



S2 Figure. Comparative genomics of *B. pseudomallei* from a single location (site 23). Left: Core genome phylogeny of 55 *B. pseudomallei* isolates obtained from site 23 containing three major clades (A-C) with a consistency index of 0.99. The phylogeny is rooted with a *B. thailandensis* isolate and bootstrap values are reported on nodes. Right: Presence/absence of the 58 genes that comprise the accessory genome (S5 Table) of this group of isolates; 49 of these accessory genes are in GI1, GI2, and GI3. Each cell provides the BSR values (0 represents no alignment; 1 represents an identical nucleotide alignment) for different accessory genes (columns) in the individual genomes (rows). Green text indicates *B. pseudomallei* isolates from soil sample 09, whereas black text represents *B. pseudomallei* isolates from soil sample 07. Bolded and italicized text indicates the two isolates (Bp9039 and Bp9110) included in Figure 2.

References

1. Stopnisek N, Bodenhausen N, Frey B, Fierer N, Eberl L, Weisskopf L. Genus-wide acid tolerance accounts for the biogeographical distribution of soil Burkholderia populations. Environmental microbiology. 2014;16(6):1503-12. Epub 2013/08/16.

2. Currie BJ. Melioidosis: evolving concepts in epidemiology, pathogenesis, and treatment. Seminars in respiratory and critical care medicine. 2015;36(1):111-25. Epub 2015/02/03.

3. Martina P, Leguizamon M, Prieto CI, Sousa SA, Montanaro P, Draghi WO, et al. Burkholderia puraquae sp. nov., a novel species of the Burkholderia cepacia complex isolated from hospital settings and agricultural soils. International journal of systematic and evolutionary microbiology. 2018;68(1):14-20. Epub 2017/11/03.

4. Tuanyok A, Mayo M, Scholz H, Hall CM, Allender CJ, Kaestli M, et al. Burkholderia humptydooensis sp. nov., a New Species Related to Burkholderia thailandensis and the Fifth Member of the Burkholderia pseudomallei Complex. Applied and environmental microbiology. 2017;83(5). Epub 2016/12/18.

5. Vandamme P, Peeters C, De Smet B, Price EP, Sarovich DS, Henry DA, et al. Comparative Genomics of Burkholderia singularis sp. nov., a Low G+C Content, Free-Living Bacterium That Defies Taxonomic Dissection of the Genus Burkholderia. Frontiers in microbiology. 2017;8:1679. Epub 2017/09/22.

6. Sahl JW, Vazquez AJ, Hall CM, Busch JD, Tuanyok A, Mayo M, et al. The Effects of Signal Erosion and Core Genome Reduction on the Identification of Diagnostic Markers. mBio. 2016;7(5). Epub 2016/09/22.

7. LiPuma JJ. The Changing Microbial Epidemiology in Cystic Fibrosis. Clin Microbiol Rev. 2010;23(2):299-+.

8. LiPuma JJ, Spilker T, Gill LH, Campbell PW, Liu LX, Mahenthiralingam E. Disproportionate distribution of Burkholderia cepacia complex species and transmissibility markers in cystic fibrosis. Am J Resp Crit Care. 2001;164(1):92-6.

9. Ham JH, Melanson RA, Rush MC. Burkholderia glumae: next major pathogen of rice? Molecular plant pathology. 2011;12(4):329-39. Epub 2011/04/02.

10. Wiersinga WJ, Virk HS, Torres AG, Currie BJ, Peacock SJ, Dance DAB, et al. Melioidosis. Nature reviews Disease primers. 2018;4:17107. Epub 2018/02/02.

11. Yee KC, Lee MK, Chua CT, Puthucheary SD. Melioidosis, the great mimicker: a report of 10 cases from Malaysia. The Journal of tropical medicine and hygiene. 1988;91(5):249-54. Epub 1988/10/01.

12. Doker TJ, Quinn CL, Salehi ED, Sherwood JJ, Benoit TJ, Glass Elrod M, et al. Fatal Burkholderia pseudomallei infection initially reported as a Bacillus species, Ohio, 2013. The American journal of tropical medicine and hygiene. 2014;91(4):743-6. Epub 2014/08/06.

13. Schweizer HP. Mechanisms of antibiotic resistance in Burkholderia pseudomallei: implications for treatment of melioidosis. Future microbiology. 2012;7(12):1389-99. Epub 2012/12/13.

14. Price EP, Sarovich DS, Webb JR, Hall CM, Jaramillo SA, Sahl JW, et al. Phylogeographic, genomic, and meropenem susceptibility analysis of Burkholderia ubonensis. PLoS neglected tropical diseases. 2017;11(9):e0005928. Epub 2017/09/15.

15. 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 Biol. 2009;7.

16. Tuanyok A, Leadem BR, Auerbach RK, Beckstrom-Sternberg SM, Beckstrom-Sternberg JS, Mayo M, et al. Genomic islands from five strains of Burkholderia pseudomallei. BMC genomics. 2008;9:566. Epub 2008/11/29.

17. Price EP, Currie BJ, Sarovich DS. Genomic Insights Into the Melioidosis Pathogen, Burkholderia pseudomallei. Current Tropical Medicine Reports. 2017;4(3):95-102.

18. Ralph A, McBride J, Currie BJ. Transmission of Burkholderia pseudomallei via breast milk in northern Australia. The Pediatric infectious disease journal. 2004;23(12):1169-71. Epub 2005/01/01.

19. 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(1). Epub 2016/02/16.

20. Doker TJ, Sharp TM, Rivera-Garcia B, Perez-Padilla J, Benoit TJ, Ellis EM, et al. Contact investigation of melioidosis cases reveals regional endemicity in Puerto Rico. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America. 2015;60(2):243-50. Epub 2014/10/02.

21. Sanchez-Villamil JI, Torres AG. Melioidosis in Mexico, Central America, and the Caribbean. Tropical medicine and infectious disease. 2018;3(1). Epub 2018/05/22.

22. 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. Epub 2013/04/05.

23. Kaestli M, Mayo M, Harrington G, Watt F, Hill J, Gal D, et al. Sensitive and specific molecular detection of Burkholderia pseudomallei, the causative agent of melioidosis, in the soil

of tropical northern Australia. Applied and environmental microbiology. 2007;73(21):6891-7. Epub 2007/09/18.

24. NOAA. March 2017 Climate Review for Puerto Rico and U.S. Virgin Islands. weather.gov: National Oceanic and Atmospheric Administration; 2017 [cited 2019 6/28/2019]; Available from: https://www.weather.gov/media/sju/climo/monthly_reports/2017/Mar2017.pdf.

25. NOAA. April 2017 Climate Review for Puerto Rico and U.S. Virgin Islands. weather.gov: National Oceanic and Atmospheric Administration; 2017 [cited 2019 6/28/2019]; Available from: https://www.weather.gov/media/sju/climo/monthly_reports/2017/Apr2017.pdf.

26. Kaestli M, Grist EPM, Ward L, Hill A, Mayo M, Currie BJ. The association of melioidosis with climatic factors in Darwin, Australia: A 23-year time-series analysis. The Journal of infection. 2016;72(6):687-97. Epub 2016/03/08.

27. Rattanavong S, Wuthiekanun V, Langla S, Amornchai P, Sirisouk J, Phetsouvanh R, et al. Randomized soil survey of the distribution of Burkholderia pseudomallei in rice fields in Laos. Applied and environmental microbiology. 2011;77(2):532-6. Epub 2010/11/16.

28. Ribolzi O, Rochelle-Newall E, Dittrich S, Auda Y, Newton PN, Rattanavong S, et al. Land use and soil type determine the presence of the pathogen Burkholderia pseudomallei in tropical rivers. Environmental science and pollution research international. 2016;23(8):7828-39. Epub 2016/01/14.

29. Kaestli M, Mayo M, Harrington G, Ward L, Watt F, Hill JV, et al. Landscape changes influence the occurrence of the melioidosis bacterium Burkholderia pseudomallei in soil in northern Australia. PLoS neglected tropical diseases. 2009;3(1):e364. Epub 2009/01/22.

30. Manivanh L, Pierret A, Rattanavong S, Kounnavongsa O, Buisson Y, Elliott I, et al. Burkholderia pseudomallei in a lowland rice paddy: seasonal changes and influence of soil depth and physico-chemical properties. Scientific reports. 2017;7(1):3031. Epub 2017/06/10.

31. Chantratita N, Wuthiekanun V, Limmathurotsakul D, Vesaratchavest M, Thanwisai A, Amornchai P, et al. Genetic diversity and microevolution of Burkholderia pseudomallei in the environment. PLoS neglected tropical diseases. 2008;2(2):e182. Epub 2008/02/27.

32. Seng R, Saiprom N, Phunpang R, Baltazar CJ, Boontawee S, Thodthasri T, et al. Prevalence and genetic diversity of Burkholderia pseudomallei isolates in the environment near a patient's residence in Northeast Thailand. PLoS neglected tropical diseases. 2019;13(4).

33. Dance DA. Editorial commentary: melioidosis in Puerto Rico: the iceberg slowly emerges. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America. 2015;60(2):251-3. Epub 2014/10/02.

34. Hall CM, Busch JD, Shippy K, Allender CJ, Kaestli M, Mayo M, et al. Diverse Burkholderia Species Isolated from Soils in the Southern United States with No Evidence of B. pseudomallei. PloS one. 2015;10(11):e0143254. Epub 2015/11/26.

35. Ginther JL, Mayo M, Warrington SD, Kaestli M, Mullins T, Wagner DM, et al. Identification of Burkholderia pseudomallei Near-Neighbor Species in the Northern Territory of Australia. PLoS neglected tropical diseases. 2015;9(6):e0003892. Epub 2015/06/30.

36. Price EP, Sarovich DS, Webb JR, Ginther JL, Mayo M, Cook JM, et al. Accurate and rapid identification of the Burkholderia pseudomallei near-neighbour, Burkholderia ubonensis, using real-time PCR. PloS one. 2013;8(8):e71647. Epub 2013/08/24.

37. Trung TT, Hetzer A, Gohler A, Topfstedt E, Wuthiekanun V, Limmathurotsakul D, et al. Highly sensitive direct detection and quantification of Burkholderia pseudomallei bacteria in

environmental soil samples by using real-time PCR. Applied and environmental microbiology. 2011;77(18):6486-94. Epub 2011/08/02.

38. 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. Epub 2006/01/05.

39. Stone NE, Sidak-Loftis LC, Sahl JW, Vazquez AJ, Wiggins KB, Gillece JD, et al. More than 50% of Clostridium difficile Isolates from Pet Dogs in Flagstaff, USA, Carry Toxigenic Genotypes. PloS one. 2016;11(10):e0164504. Epub 2016/10/11.

40. Liu CM, Aziz M, Kachur S, Hsueh PR, Huang YT, Keim P, et al. BactQuant: an enhanced broad-coverage bacterial quantitative real-time PCR assay. BMC microbiology. 2012;12:56. Epub 2012/04/19.

41. Delamballerie X, Zandotti C, Vignoli C, Bollet C, Demicco P. A One-Step Microbial DNA Extraction Method Using Chelex-100 Suitable for Gene Amplification. Research in microbiology. 1992;143(8):785-90.

42. Sarovich DS, Price EP, Von Schulze AT, Cook JM, Mayo M, Watson LM, et al. Characterization of Ceftazidime Resistance Mechanisms in Clinical Isolates of Burkholderia pseudomallei from Australia. PloS one. 2012;7(2).

43. Payne GW, Vandamme P, Morgan SH, Lipuma JJ, Coenye T, Weightman AJ, et al. Development of a recA gene-based identification approach for the entire Burkholderia genus. Applied and environmental microbiology. 2005;71(7):3917-27. Epub 2005/07/08.

44. Kozarewa I, Turner DJ. 96-plex molecular barcoding for the Illumina Genome Analyzer. Methods Mol Biol. 2011;733:279-98. Epub 2011/03/25.

45. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of computational biology : a journal of computational molecular cell biology. 2012;19(5):455-77. Epub 2012/04/18.

46. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. Journal of molecular biology. 1990;215(3):403-10. Epub 1990/10/05.

47. Benson DA, Karsch-Mizrachi I, Clark K, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic acids research. 2012;40(Database issue):D48-53. Epub 2011/12/07.

48. Godoy D, Randle G, Simpson AJ, Aanensen DM, Pitt TL, Kinoshita R, et al. Multilocus sequence typing and evolutionary relationships among the causative agents of melioidosis and glanders, Burkholderia pseudomallei and Burkholderia mallei. Journal of clinical microbiology. 2003;41(5):2068-79. Epub 2003/05/08.

49. Baldwin A, Mahenthiralingam E, Thickett KM, Honeybourne D, Maiden MC, Govan JR, et al. Multilocus sequence typing scheme that provides both species and strain differentiation for the Burkholderia cepacia complex. Journal of clinical microbiology. 2005;43(9):4665-73. Epub 2005/09/08.

50. Jolley KA, Maiden MC. BIGSdb: Scalable analysis of bacterial genome variation at the population level. BMC bioinformatics. 2010;11:595. Epub 2010/12/15.

51. Holden MT, Titball RW, Peacock SJ, Cerdeno-Tarraga AM, Atkins T, Crossman LC, et al. Genomic plasticity of the causative agent of melioidosis, Burkholderia pseudomallei. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(39):14240-5. Epub 2004/09/21.

52. Delcher AL, Phillippy A, Carlton J, Salzberg SL. Fast algorithms for large-scale genome alignment and comparison. Nucleic acids research. 2002;30(11):2478-83. Epub 2002/05/30.

53. Sahl JW, Lemmer D, Travis J, Schupp JM, Gillece JD, Aziz M, et al. NASP: an accurate, rapid method for the identification of SNPs in WGS datasets that supports flexible input and output formats. Microbial genomics. 2016;2(8):e000074. Epub 2017/03/30.

54. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXivorg. 2013;arXiv:1303.3997 [q-bio.GN].

55. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature genetics. 2011;43(5):491-8. Epub 2011/04/12.

56. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular biology and evolution. 2015;32(1):268-74. Epub 2014/11/06.

57. Sahl JW, Caporaso JG, Rasko DA, Keim P. The large-scale blast score ratio (LS-BSR) pipeline: a method to rapidly compare genetic content between bacterial genomes. PeerJ. 2014;2:e332. Epub 2014/04/22.

58. Kent WJ. BLAT--the BLAST-like alignment tool. Genome research. 2002;12(4):656-64. Epub 2002/04/05.

59. Letunic I, Bork P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics. 2007;23(1):127-8. Epub 2006/10/20.

60. Guy L, Kultima JR, Andersson SG. genoPlotR: comparative gene and genome visualization in R. Bioinformatics. 2010;26(18):2334-5. Epub 2010/07/14.

61. 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. Epub 2014/11/22.

62. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus evolution. 2016;2(1):vew007. Epub 2016/10/25.

63. Murray GG, Wang F, Harrison EM, Paterson GK, Mather AE, Harris SR, et al. The effect of genetic structure on molecular dating and tests for temporal signal. Methods in ecology and evolution. 2016;7(1):80-9. Epub 2016/04/26.

64. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular biology and evolution. 2012;29(8):1969-73. Epub 2012/03/01.

65. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Molecular biology and evolution. 2016;33(7):1870-4. Epub 2016/03/24.

66. Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. Molecular biology and evolution. 2012;29(9):2157-67. Epub 2012/03/10.

67. Rambaut A, Suchard M, Drummond A. Tracer version 1.6.0. 2013 [23 Jan 2019]; Available from: http://tree.bio.ed.ac.uk/software/tracer/.

68. 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. Epub 2007/10/16.

69. Liguori AP, Warrington SD, Ginther JL, Pearson T, Bowers J, Glass MB, et al. Diversity of 16S-23S rDNA internal transcribed spacer (ITS) reveals phylogenetic relationships in Burkholderia pseudomallei and its near-neighbors. PloS one. 2011;6(12):e29323. Epub 2011/12/24.

70. Duchene S, Holt KE, Weill FX, Le Hello S, Hawkey J, Edwards DJ, et al. Genome-scale rates of evolutionary change in bacteria. Microbial genomics. 2016;2(11):e000094. Epub 2017/03/30.

71. Chewapreecha C, Holden MT, Vehkala M, Valimaki N, Yang Z, Harris SR, et al. Global and regional dissemination and evolution of Burkholderia pseudomallei. Nature microbiology. 2017;2:16263. Epub 2017/01/24.

72. Limmathurotsakul D, Wuthiekanun V, Chantratita N, Wongsuvan G, Amornchai P, Day NP, et al. Burkholderia pseudomallei is spatially distributed in soil in northeast Thailand. PLoS neglected tropical diseases. 2010;4(6):e694. Epub 2010/06/10.

73. Kaestli M, Harrington G, Mayo M, Chatfield MD, Harrington I, Hill A, et al. What drives the occurrence of the melioidosis bacterium Burkholderia pseudomallei in domestic gardens? PLoS neglected tropical diseases. 2015;9(3):e0003635. Epub 2015/03/25.

74. Wuthiekanun V, Limmathurotsakul D, Chantratita N, Feil EJ, Day NP, Peacock SJ. Burkholderia Pseudomallei is genetically diverse in agricultural land in Northeast Thailand. PLoS neglected tropical diseases. 2009;3(8):e496. Epub 2009/08/05.

75. Gee JE, Gulvik CA, Elrod MG, Batra D, Rowe LA, Sheth M, et al. Phylogeography of Burkholderia pseudomallei Isolates, Western Hemisphere. Emerging infectious diseases. 2017;23(7):1133-8. Epub 2017/06/20.

76. Sarovich DS, Garin B, De Smet B, Kaestli M, Mayo M, Vandamme P, et al. Phylogenomic Analysis Reveals an Asian Origin for African Burkholderia pseudomallei and Further Supports Melioidosis Endemicity in Africa. mSphere. 2016;1(2). Epub 2016/06/16.

77. U'Ren J M, Hornstra H, Pearson T, Schupp JM, Leadem B, Georgia S, et al. Fine-scale genetic diversity among Burkholderia pseudomallei soil isolates in northeast Thailand. Applied and environmental microbiology. 2007;73(20):6678-81. Epub 2007/08/28.

78. Chen YL, Yen YC, Yang CY, Lee MS, Ho CK, Mena KD, et al. The concentrations of ambient Burkholderia pseudomallei during typhoon season in endemic area of melioidosis in Taiwan. PLoS neglected tropical diseases. 2014;8(5):e2877. Epub 2014/05/31.

79. Cheng AC, Jacups SP, Gal D, Mayo M, Currie BJ. Extreme weather events and environmental contamination are associated with case-clusters of melioidosis in the Northern Territory of Australia. International journal of epidemiology. 2006;35(2):323-9. Epub 2005/12/06.

80. 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. Epub 2015/10/22.

81. NOAA. Tropical Cyclone Climatology. National Oceanic and Atmospheric Administration, National Hurrican Center; 2018 [cited 2018 November 1, 2018]; Available from: https://www.nhc.noaa.gov/climo/.

82. Stewart S. 2012 Atlantic Hurricane Season. In: NOAA, editor. National Weather Service: NOAA; 2014.

83. Calle-Bellido J, Rivera-Vargas LI, Alameda M, Cabrera I. Bacteria occurring in onion (Allium cepa L.) foliage in Puerto Rico. The Journal of Agriculture of the University of Puerto Rico. 2012:21.

84. International Diabetes Federation2017 [November 1, 2018]; Available from: https://www.idf.org/our-network/regions-members/south-and-central-america/members/90-puerto-rico.html

CHAPTER 3 For the first time *Burkholderia thailandensis* is isolated from the environment in the Western Hemisphere including the United States

Abstract

Burkholderia thailandensis, an environmental bacterium, can cause disease under extreme conditions. Despite clinical cases in the Western Hemisphere, *B. thailandensis* has never been described from the environment in the Western Hemisphere until this study isolated it from water in Texas and Puerto Rico demonstrating a potential public health risk.

Introduction

Burkholderia thailandensis, an environmental Gram-negative bacterium, is of interest from a public health standpoint due to its potential to cause infection as a fortuitous pathogen (1-3). *B. thailandensis* is found in tropical regions of the world, such as Southeast Asia and northern Australia, similar to where its more pathogenic relative, *B. pseudomallei*, is found. *B. thailandensis* is the closest phylogenetic relative to *B. pseudomallei*, the Select Agent and causative agent of melioidosis. *B. thailandensis* is often used as a surrogate for *B. pseudomallei* (4-6) due to the similarities between the closely related bacteria and that *B. thailandensis* has reduced restrictions as a non-Select Agent, BSL-2 organisms (7). *B. thailandensis* is difficult to distinguish from *B. pseudomallei* in the clinic due to the same biochemical phenotypes, except for the ability to assimilate L-arabinose (3, 7, 8). *B. thailandensis* was first described after researches observed reduced virulence in what they thought was *B. pseudomallei* but the 16S rRNA gene phylogeny revealed a novel *Burkholderia* species that was later named *B. thailandensis* after the origin of the type strain, E264^{T}, from Thailand (7).

Human infections of *B. thailandensis* are not common but do occur (1-3). There have been three reported clinical cases of *B. thailandensis* in the Western Hemisphere, all from the southern U.S. in Louisiana in 1997, Texas in 2003 (3), and Arkansas in 2017 (2). Even with environmental sampling related to these clinical cases, *B. thailandensis* was never recovered from the environment. *B. thailandensis* has primarily been described from the environment in Southeast Asia, Australia (7, 9, 10), and recently from Africa, specifically in Sierra Leone (11). To date, the natural occurrence of *B. thailandensis* in the Western Hemisphere is poorly understood. It is important to note that until this study *B. thailandensis* had never been isolated from the environment in the Western Hemisphere.

We used a systematic approach for the detection and isolation of *B. thailandensis* from soil and water samples collected in Texas in November of 2019 and 2020 and Puerto Rico from December 2018 to March 2020. The information gained from this study is vitally important to understand where and how *B. thailandensis* persists in the environment and how the genetic diversity in the environment compares to the strains isolated from the clinic.

Results

Of the 1,960 environmental samples (370 from Texas and 1,590 from Puerto Rico) collected for this study, a *B. thailandensis* DNA signal was detected using TaqMan PCR in 10 of our complex broth samples with four from Texas and six from Puerto Rico (see Technical Appendix). Culturing efforts were used for the isolation of pure culture using previously described methods (12) resulting in *B. thailandensis* isolates from five of the samples (one from Texas and four from Puerto Rico). The environmental samples consisted of soil, water, or environmental scrapes with 280 soil, 80 water, and 10 environmental scrapes collected from Texas and 1,170 soil and

420 water samples collected from Puerto Rico. Samples were collected from 99 sites (25 from Texas and 74 from Puerto Rico) with *B. thailandensis* detected from seven sites (two from Texas and five from Puerto Rico) and isolated from five sites (one from Texas and four from Puerto Rico) (Table). Interestingly, *B. thailandensis* was only detected in water samples despite water samples consisting only 26% of our total samples collected.

Whole-genome analysis of these novel isolates allowed us to place them within a larger context of *B. thailandensis* from other global locations and to previous U.S. clinical isolates (Figure and Table). As expected, the environmental *B. thailandensis* from Texas grouped in the same clade as the other clinical isolates from the Western Hemisphere, supporting that the previous human cases were obtained from the environment in the U.S. Remarkably, the Texas environmental isolate was more closely related to clinical isolates from Louisiana (1,821 SNPs) and Arkansas (2,636 SNPs) than from Texas (9,212 SNPs). Conversely, to the genetic diversity observed in the Southern U.S., little diversity was found between the four 2019-2020 Puerto Rico *B. thailandensis* isolates. The four *B. thailandensis* isolates from Puerto Rico differed only between 260 SNPs.

In-silico MLST analysis revealed a novel ace allele (106) in all four *B. thailandensis* isolates from Puerto Rico and a novel gltB allele (175) in the single isolate from Texas resulting in novel STs for Puerto Rico (ST1772) and Texas (ST1785).

Discussion

Our study confirms the presence of *B. thailandensis* in the environment, specifically in water, from the Western Hemisphere for the first time. We also show that where we detected *B. thailandensis*, it appears to be rare in occurrence and abundance. Detection of *B. thailandensis*

required extensive sampling for detection with only 7.7% of sites we sampled resulting in detection of *B. thailandensis* suggesting its rarity. Once *B. thailandensis* was detected, substantial culturing efforts were required for isolation of a live culture suggesting a low abundance or that *B. thailandensis* is easily out competing by other environmental bacteria. Further supporting its rarity and low abundance, no environmental source was ever detected for the three previous human cases from the southern U.S (2, 3).

Interestingly, *B. thailandensis* was only detected in water samples despite the collection of both soil and water at each of the *B. thailandensis* sites. This is quite different from endemic regions, such as Thailand, where *B. thailandensis* is commonly isolated from the soil (13). It is also important to note that all three of the U.S. clinical cases were associated with traumatic events and two of the three cases were associated with water (2, 3). This demonstrates the public health risk of *B. thailandensis* since it appears traumatic water events can lead to disease with this pathogen. This information is especially useful in Puerto Rico where *B. thailandensis* was detected within neighborhoods in San Juan, the most populated city on the island, and Puerto Rico is prone to hurricanes resulting in flooding.

The genetic diversity observed between *B. thailandensis* isolated from Puerto Rico and from the continental U.S. was strikingly different. There were only 260 SNPs found between the strains from Puerto Rico despite being collected from three separate municipalities (Carolina, Fajardo, and San Juan) and over one year apart. The lack of diversity found between these strains suggests the possibility of a recent single introduction of *B. thailandensis* to Puerto Rico. Additional sampling in Puerto Rico would help support or reject this hypothesis. On the contrary, thousands of SNPs were observed between the various *B. thailandensis* strains isolated from Texas, Arkansas, and Louisiana. A possible explanation for the large amount of genetic diversity

observed is that *B. thailandensis* has been established in the southern U.S. for a long period of time and that it is widely present in the environment, potentially circulating in the water. Although it is unknown how long *B. thailandensis* can persist in water, its closest phylogenetic relative, *B. pseudomallei*, can survive in water for at least 16 years with no nutrients (14). Intriguingly, the greatest number of SNPs of *B. thailandensis* from the southern U.S. strains was observed between the two *B. thailandensis* strains from Texas, a clinical isolate from 2003 and environmental isolate from 2019, suggesting that *B. thailandensis* is also widespread in the environment in Texas.

This study provides valuable information about the potential of water to serve as a source of infection for a fortuitous pathogen, *B. thailandensis*, in the southern U.S. and in Puerto Rico.

Acknowledgments

We would like to thank all the Animal Health Technicians from USDA-APHIS-VS in Puerto Rico that assisted with sampling locations in the field. We would also like to thank Nicole Bratsch, Kim Celona, Kieston Guidry, Ashley Jones, Karisma Kocos, Amalee Nunnally, Bryce Schmidt, Kathleen Soria, and Mackenzie Roberts for their assistance in the laboratory.

Tables and Figures

Chapter 3 Tables

Table. *Burkholderia thailandensis* genomes used in whole genome phylogeny.

Isolate	Alternative ID	Country (State/Territory)	Sample type (source)	Year	MLST	Accession number
2.1		Vietnam	environmental	2017	696	GCA_002803565.1
82172	34; 2002721621	France	veterinary (horse)	1982	73	GCA_001555485.1
E555		Cambodia	environmental	2005	696	GCF_000179515.1
TXDOH	CDC3015869;2003015869	United States (TX)	clinical (human)	2003	101	GCA_002888425.1
Bt4	49639	Australia	environmental	unkno	699	GCA_000170395.1
MSMB59		Australia	environmental (soil)	2006	669	GCA_001718595.1
MSMB60		Australia	environmental (soil)	2006	669	GCA_001524345.1
Bp10009*	165-01_P1_S7	United States (TX)	environmental (water)	2019	1758	**TBD**
Bp10013*	203-09_P1_S27	United States (PR)	environmental (water)	2020	1772	**TBD**
Bp9795*	61_10_S54_S1_copy3	United States (PR)	environmental (water)	2018	1772	**TBD**
Bp9920*	89-06_P1_S1	United States (PR)	environmental (water)	2018	1772	**TBD**
Bp9942*	91-08_P2_S1	United States (PR)	environmental (water)	2018	1772	**TBD**
BtAR2017		United States (AR)	clinical (human)	2017	101	GCA_004684955.1
E1		Papua New Guinea	environmental	1995	669	GCA_001524325.1
E254		Thailand	environmental	1992	345	GCA_000765375.1
E264	ATCC 700388	Thailand	environmental (soil)	1994	80	GCA_003568605.1
E444	E0444	Thailand	environmental	2002	79	GCA_000567945.1
H0587	2002721121	United States (LA)	clinical (human)	1997	101	GCA_000567905.1
Phuket 4W-1		Thailand	environmental	2002	80	GCA_000877335.1
USAMRU Malaysia	Malaysia #20; 2002721744	Malaysia	unknown	unkno	80	GCA_000706745.1

* Isolated from this study. MLST, multi-locus sequence type.

Chapter 3 Figures

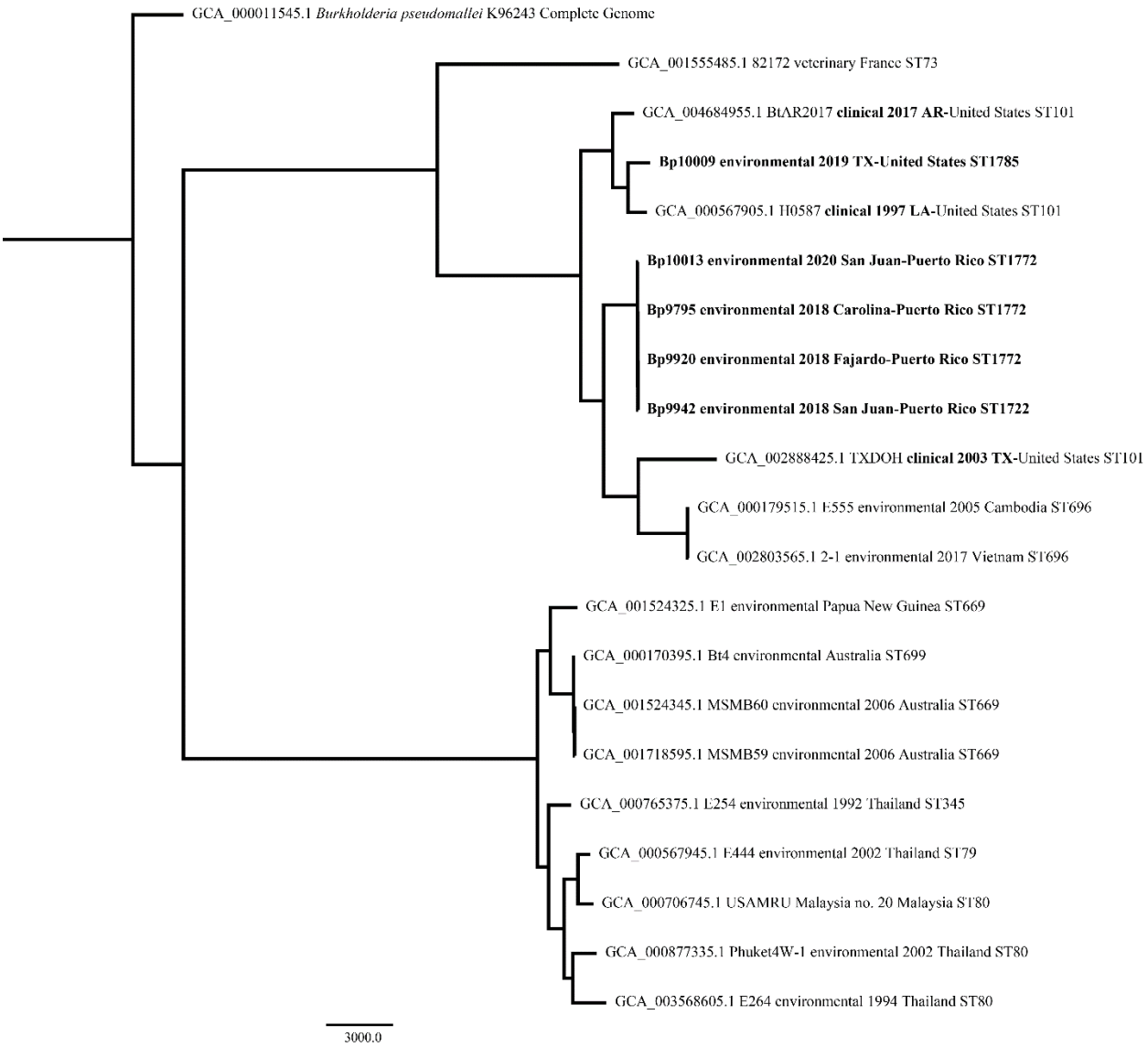


Figure. Whole genome phylogeny of *B. thailandensis* rooted with *B. pseudomallei*. *B. thailandensis* genomes used to generate this phylogeny are listed in the Table.

References

1. Chang K, Luo J, Xu H, Li M, Zhang F, Li J, et al. Human Infection with Burkholderia thailandensis, China, 2013. Emerg Infect Dis. 2017;23(8):1416-8.

2. Gee JE, Elrod MG, Gulvik CA, Haselow DT, Waters C, Liu L, et al. Burkholderia thailandensis Isolated from Infected Wound, Arkansas, USA. Emerg Infect Dis. 2018;24(11):2091-4.

3. Glass MB, Gee JE, Steigerwalt AG, Cavuoti D, Barton T, Hardy RD, et al. Pneumonia and septicemia caused by Burkholderia thailandensis in the United States. J Clin Microbiol. 2006;44(12):4601-4.

4. Gallagher LA, Ramage E, Patrapuvich R, Weiss E, Brittnacher M, Manoil C. Sequence-Defined Transposon Mutant Library of Burkholderia thailandensis. mBio. 2013;4(6).

5. Morici LA, Heang J, Tate T, Didier PJ, Roy CJ. Differential susceptibility of inbred mouse strains to Burkholderia thailandensis aerosol infection. Microb Pathog. 2010;48(1):9-17.

6. West TE, Frevert CW, Liggitt HD, Skerrett SJ. Inhalation of Burkholderia thailandensis results in lethal necrotizing pneumonia in mice: a surrogate model for pneumonic melioidosis. Trans R Soc Trop Med Hyg. 2008;102 Suppl 1:S119-26.

7. Brett PJ, DeShazer D, Woods DE. Burkholderia thailandensis sp. nov., a Burkholderia pseudomallei-like species. Int J Syst Bacteriol. 1998;48:317-20.

8. Smith MD, Angus BJ, Wuthiekanun V, White NJ. Arabinose assimilation defines a nonvirulent biotype of Burkholderia pseudomallei. Infection and immunity. 1997;65(10):4319-21.

9. Ginther JL, Mayo M, Warrington SD, Kaestli M, Mullins T, Wagner DM, et al. Identification of Burkholderia pseudomallei Near-Neighbor Species in the Northern Territory of Australia. PLoS neglected tropical diseases. 2015;9(6).

10. Levy A, Merritt AJ, Aravena-Roman M, Hodge MM, Inglis TJJ. Expanded range of Burkholderia species in Australia. American Journal of Tropical Medicine and Hygiene. 2008;78(4):599-604.

11. Birnie E, van't Hof S, Bijnsdorp A, Mansaray Y, Huizenga E, van der Ende A, et al. Identification of Burkholderia thailandensis with novel genotypes in the soil of central Sierra Leone. PLoS neglected tropical diseases. 2019;13(6).

12. Hall CM, Jaramillo S, Jimenez R, Stone NE, Centner H, Busch JD, et al. Burkholderia pseudomallei, the causative agent of melioidosis, is rare but ecologically established and widely dispersed in the environment in Puerto Rico. PLoS Negl Trop Dis. 2019;13(9):e0007727.

13. Hantrakun V, Thaipadungpanit J, Rongkard P, Srilohasin P, Amornchai P, Langla S, et al. Presence of B. thailandensis and B. thailandensis expressing B. pseudomallei-like capsular polysaccharide in Thailand, and their associations with serological response to B. pseudomallei. Plos Neglect Trop D. 2018;12(1).

14. Pumpuang A, Chantratita N, Wikraiphat C, Saiprom N, Day NPJ, Peacock SJ, et al. Survival of Burkholderia pseudomallei in distilled water for 16 years. T Roy Soc Trop Med H. 2011;105(10):598-600.

Technical Appendix

Environmental sampling was conducted in Puerto Rico and Texas using methods previously described (1) and were adapted from the international consensus guidelines (2) with additional modifications developed by the Menzies School of Health and Research in Darwin, Australia (3). Permission was obtained from landowners to collect soil and/or water samples on their property and when necessary, permits were received to collect from reserve lands. Briefly, we collected 10 to 100 soil samples at a depth of 30 cm from holes spaced 2.5 meters apart in one to two 10-hole transects or in a grid for each site as previously described (1). Water samples (150 mL) were collected along a linear transect with 2.5 meters between each sampling location, when possible, using one of two sampling approaches: For approach A water samples were collected (20 samples per site in Puerto Rico and 10 samples per site in Texas) approximately 1 meter from the shoreline, avoiding flowing water while for approach B water samples were collected (10 samples per site in Puerto Rico) at five sampling locations with the first sample collected directly on the water's edge and another sample collected 1 meter from the shoreline. All water samples were filtered the same day as collection using a Sartorius water filtration manifold with 0.22 µm nitrocellulose filters as previously described (1). Environmental scrapes were only collected in Texas at one site from a partially empty residential 500-gallon water holding tank. Briefly, the bottom and sides of the tank were scraped, and the scrape contents were placed into a sterile 2 mL screw-cap tube.

A total of 1,800 environmental samples were collected throughout Puerto Rico during the months of December 2018, February and March 2019, and February and March 2020. These environmental samples consisted of 1,170 soil samples collected from 42 sites and 420 water samples from 32 sites (200 water samples were collected from 10 sites using sampling approach

A and 220 collected from 22 sites using sampling approach B). A total of 210 environmental samples were collected in Atascosa County in Texas during November 2019 including 120 soil samples collected from eight sites, 80 water samples from eight sites, and 10 environmental scrapes from one site. Another 160 soil samples were collected from six sites in Guadalupe, Goliad, and Wilson counties in Texas during November 2020.

All samples were kept from direct UV exposure, stored, and shipped at ambient temperature to NAU for further processing. Upon arrival, samples were stored in the dark at ambient temperature except for sampling approach B water filters which were stored at 4°C. To prepare the environmental scrapes for culturing the 2 mL tubes containing the scrapes were first vortex at high speed for 1 minute and then sonicated for 5 minutes using a Branson sonicator bath set to 70W, 42kHz at room temperature. All samples were processed in the same way for the detection and isolation of *Burkholderia* spp. as previously described (1) with the following modifications. Each water sample was filtered onto one filter which was cut in half and only one half was used for the inoculation of 30 mL of Ashdown's broth. The entire contents of the environmental scrape were transferred to the 30 mL of Ashdown's broth. The soil was processed in the same way as Hall *et al*., 2019 (1) except the soil was added straight into 30 mL of Ashdown's broth instead of first into water. Also, a *B. thailandensis* specific TaqMan assay was used for molecular detection of a *B. thailandensis* DNA signal within DNA extractions preformed on 1mL of the complex Ashdown's broth. Methods for the DNA extractions can be found in Hall *et al*., 2019 (1). The primers and probe used for the TaqMan assay were developed based on whole genome analysis of 1,130 *Burkholderia* genomes (4). Once a DNA signal was detected in an Ashdown's broth sample isolation efforts were focused on that sample. Culturing methods are also described in Hall *et al*., 2019 (1). Once a pure *B. thailandensis* isolate was

obtained after at least three isolation streaks long term glycerol stocks were created and high molecular weight gDNA was obtained using QIAGEN DNeasy kits. Whole genome sequencing was performed as previously described (1).

MLST

In-silico MLST analysis revealed a novel ace allele (106) in all four *B. thailandensis* isolates from Puerto Rico and a novel gltB allele (175) in the single isolate from Texas. The remaining MLST genes were previously characterized resulting in the following novel STs for Texas (ST1785; ace=106, gltB=5, gmhD=9, lepA=11, lipA=14, narK=20, ndh=14) and for Puerto Rico (ST1772; ace=106, gltB=5, gmhD=9, lepA=11, lipA=14, narK=20, ndh=14). The ST1785 and ST1772 all have the same alleles as ST101 (other U.S. B. thailandensis clinical samples) except for the gltB and ace genes respectively. It's also interesting to note that there was one SNPs between the 175 gltB and 5 gltB (from ST101) and 106 ace and 6 ace (from ST101) alleles.

Phylogenetic analysis

Genome assemblies were aligned against *B. thailandensis* E254 (GCA_000765375.1) with NUCmer v3.1 (5) and single nucleotide polymorphisms were called with NASP v1.2.0 (6). SNPs (*n*=54,617) that had a valid nucleotide call in all query genomes were extracted from the outgroup genome, *B. pseudomallei* K96243 (7), and concatenated into a single multifasta file. A maximum-likelihood phylogeny was inferred on the concatenated alignment with IQ-TREE v1.6.12 (8), using the TVM+F+ASC+G4 substitution model (9).

Technical Appendix References

1. Hall CM, Jaramillo S, Jimenez R, Stone NE, Centner H, Busch JD, et al. Burkholderia pseudomallei, the causative agent of melioidosis, is rare but ecologically established and widely dispersed in the environment in Puerto Rico. PLoS Negl Trop Dis. 2019;13(9):e0007727.

2. 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 Negl Trop Dis. 2013;7(3):e2105.

3. Kaestli M, Mayo M, Harrington G, Watt F, Hill J, Gal D, et al. Sensitive and specific molecular detection of Burkholderia pseudomallei, the causative agent of melioidosis, in the soil of tropical northern Australia. Appl Environ Microb. 2007;73(21):6891-7.

4. Sahl JW, Vazquez AJ, Hall CM, Busch JD, Tuanyok A, Mayo M, et al. The Effects of Signal Erosion and Core Genome Reduction on the Identification of Diagnostic Markers. MBio. 2016;7(5).

5. Delcher AL, Salzberg SL, Phillippy AM. Using MUMmer to identify similar regions in large sequence sets. Curr Protoc Bioinformatics. 2003;Chapter 10:Unit 10 3.

6. Sahl JW, Lemmer D, Travis J, Schupp JM, Gillece JD, Aziz M, et al. NASP: an accurate, rapid method for the identification of SNPs in WGS datasets that supports flexible input and output formats. Microb Genom. 2016;2(8):e000074.

7. Holden MT, Titball RW, Peacock SJ, Cerdeno-Tarraga AM, Atkins T, Crossman LC, et al. Genomic plasticity of the causative agent of melioidosis, Burkholderia pseudomallei. Proc Natl Acad Sci U S A. 2004;101(39):14240-5.

8. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Molecular Biology and Evolution. 2015;32(1):268-74.

9. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14(6):587-+.

CHAPTER 4 Expanding the *Burkholderia pseudomallei* complex with the addition of two novel species: *Burkholderia mayonis* sp. nov. and *Burkholderia savannae* sp. nov.

Abstract

Distinct *Burkholderia* strains were isolated from soil samples collected in tropical northern Australia (Northern Territory and the Torres Strait Islands, Queensland). Phylogenetic analysis of 16S rRNA and whole genome sequences revealed these strains were distinct from previously described *Burkholderia* species and assigned them to two novel clades within the *B. pseudomallei* complex (Bpc). Because average nucleotide identity and digital DNA-DNA hybridization calculations are consistent with these clades representing distinct species, we propose the names *Burkholderia mayonis* sp. nov. and *Burkholderia savannae* sp. nov. Strains assigned to *B. mayonis* sp. nov. include type strain BDU6^{T} (=TSD-80; LMG 29941; ASM152374v2) and BDU8. Strains assigned to *B. savannae* sp. nov. include type strain MSMB266^{T} (=TSD-82; LMG 29940; ASM152444v2), MSMB852, BDU18, and BDU19. Comparative genomics revealed unique coding regions for both putative species, including clusters of orthologous genes associated with phage. Type strains of both *B. mayonis* sp. nov. and *B. savannae* sp. nov. yielded biochemical profiles distinct from each other and other species in the Bpc, and profiles also varied among strains within *B. mayonis* sp. nov. and *B. savannae* sp. nov. Matrix-assisted laser desorption ionization–time of flight analysis revealed a *B. savannae* sp. nov. cluster separate from other species, whereas *B. mayonis* sp. nov. strains did not form a distinct cluster. Neither *B. mayonis* sp. nov. nor *B. savannae* sp. nov. caused mortality in mice when delivered via the subcutaneous route. The addition of *B. mayonis* sp. nov. and *B. savannae* sp. nov. results in eight species currently in the Bpc.

IMPORTANCE

Burkholderia species can be important sources of novel natural products and new species are of interest to diverse scientific disciplines. Although many *Burkholderia* species are saprophytic, *Burkholderia pseudomallei* is the causative agent of the disease melioidosis. Understanding the genomics and virulence of the closest relatives to *B. pseudomallei* (*i.e.*, the other species within the Bpc) is important for identifying robust diagnostic targets specific to *B. pseudomallei* and understanding evolution of virulence in *B. pseudomallei*. Two proposed novel species, *B. mayonis* sp. nov. and *B. savannae* sp. nov., were isolated from soil samples collected from multiple locations in northern Australia. The two proposed species belong to the Bpc but are phylogenetically distinct from all other members of this complex. The addition of *B. mayonis* sp. nov. and *B. savannae* sp. nov. results in a total of eight species within this significant complex of bacteria that are available for future studies.

Introduction

The genus *Burkholderia* was recently divided into *Burkholderia sensu stricto*, *Paraburkholderia*, *Caballeronia*, *Robbsia*, and *Pararobbsia*. Together, these taxonomic groups comprise over 100 described species (http://www.bacterio.net/) that can have pathogenic, mutualistic, and/or commensal relationships with plants, animals, and/or humans (1-3). This division resulted in *Burkholderia sensu stricto* containing most of the opportunistic pathogens belonging to one of two groups of species: the *Burkholderia pseudomallei* complex (Bpc) and the *Burkholderia cepacia* complex (Bcc). New species are regularly described in *Burkholderia sensu stricto* (4-9), and the majority of species within it are naturally found in the environment, primarily in soil and water (10).

Diverse niche adaptation is exhibited by members of the Bpc. *B. pseudomallei* has adapted to opportunistic pathogenicity, *B. mallei* to obligate pathogenicity, and *B. thailandensis* (11), *B. oklahomensis* (12), *B. humptydooensis* (13), and *B. singularis* (6) to environmental saprophytism with (except *B. humptydooensis*) occasional pathogenicity. *B. pseudomallei* is the causative agent of the serious human disease melioidosis and is commonly isolated from soil and water in endemic areas (14). *B. mallei* is a clone within *B. pseudomallei* that has undergone host adapted reductive niche specialization toward obligate pathogenicity in the form of the disease glanders (15). Given these niche differences, the ongoing study of the Bpc can provide insights into evolutionary mechanisms driving bacterial virulence and niche adaptation. Moreover, the classification of pathogenic members of the Bpc (*B. pseudomallei* and *B. mallei*) as U.S. Tier 1 Select Agents due to their potential to be aerosolized and used as biowarfare agents (14, 16), and the suggestion that global melioidosis cases may be severely underestimated (17), means that closely related species are of great interest due to their potential for cross-reactivity in diagnostic/detection technologies used across defense, health, and environmental applications. In addition, novel *Burkholderia* species are of significant interest to multiple scientific fields because previously described members of this genus, including members of the Bpc, have been shown to be important sources of new natural products (18, 19).

In this study we propose the addition of two additional members of the Bpc: *B. mayonis* sp. nov. and *B. savannae* sp. nov. We used a polyphasic approach, including bioinformatic and biochemical analyses, to confirm that they are distinct species and to investigate their unique coding region sequences, as well as those shared with other members of the Bpc, to better understand diversification and evolution within this group.

Materials and Methods

Strain isolation

The two *B. mayonis* sp. nov. strains (BDU6^T, BDU8) and the four *B. savannae* sp. nov. strains (MSMB266^T, MSMB852, BDU18, BDU19) were all isolated from soil collected in tropical northern Australia (Table 1; Fig. S1). A subset of the strains (BDU6^T, BDU8, BDU18, BDU19) was collected by James Cook University from a single soil sample collected from approximately 30 cm depth on Badu Island, in the Torres Strait Islands, Queensland, Australia in late October 2011, near the end of the dry season. The soil sample was moist, sandy, and collected less than a meter from stagnant water within an exposed root system of trees. Strains MSMB266^T and MSMB852 were collected by investigators from the Menzies School of Health Research from two different locations in the tropical "Top End" of the Northern Territory, Australia in 2006 and 2010, respectively. The BDU strains were recovered using a two-stage culture technique (20), and the MSMB strains were cultured from soil using standard *Burkholderia* culturing techniques (21); all strains were presumptively identified as *Burkholderia* based upon colony morphology but confirmed to not be *B. pseudomallei* via PCR (22). The proposed novel species, *B. mayonis* sp. nov. and *B. savannae* sp. nov., were previously reported as putative species 2 and putative species 3, respectively, by Sahl *et al* (4) based upon a whole genome analysis.

Bacterial growth and characteristics

All strains were cultivated at temperatures of 25°C, 37°C, and 42°C for 24, 48, 72, and 96 hours on Ashdown's selective agar, Columbia blood agar, MacConkey agar, and Luria-Bertani agar. Biochemical data were obtained for the two strains of *B. mayonis* sp. nov. (BDU6^T and BDU8) and the four strains of *B. savannae* sp. nov. (MSMB266^T, MSMB852, BDU18, BDU19) using the API 20NE and API Zym (bioMérieux) systems according to the manufacturer's instructions.

These data were compared to data generated for *B. thailandensis* strain E264^{T} and *B. oklahomensis* strain C6786, as well as previous data generated for *B. pseudomallei* strain K96243 (23). MALDI-TOF MS analysis also was performed for all *B. mayonis* sp. nov. and *B. savannae* sp. nov. strains listed above (see text in the supplemental material for a detailed description of the methods).

Antimicrobial susceptibility screening

The minimum inhibitory concentration (MIC) was determined using the broth microdilution method in biological duplicate using 96-well microtiter custom Micronaut-S plates (Merlin, Bornheim-Hersel, Germany) following manufacturer instructions. In total, 20 antimicrobials were tested with a 2-fold serial dilution at the following concentrations: amoxicillin/clavulanic acid (4/2-128/64 mg/L), azithromycin (4-64 mg/L), carbenicillin (4-512 mg/L), ceftazidime (4–128 mg/L), ceftazidime/avibactam (0.5/4-256/4 mg/L), chloramphenicol (4-128 mg/L), ciprofloxacin (0.5-16 mg/L), doripenem (0.5-16 mg/L), doxycycline (1-32 mg/L), gentamicin (2-64 mg/L), imipenem (1-32 mg/L), kanamycin (8-256 mg/L), meropenem (1-64 mg/L), piperacillin (8-256 mg/L), piperacillin/tazobactam (8-256/4 mg/L), polymyxin B (1-2048 mg/L), sulfamethoxazole (1-512 mg/L), tigecycline (0.25-32 mg/L), trimethoprim (1-32 mg/L), and trimethoprim/sulfamethoxazole (1/19-16/304 mg/L). Two broth and growth controls containing no antimicrobials were included on each plate and each strain was screened twice using biological duplicates on separate days. Briefly, for each strain individual colonies were mixed in 3 mL of sterile saline solution (0.85% NaCl) to achieve a 0.5 McFarland Standard. The suspension (0.2 mL) was added to 20 mL of cation-adjusted Mueller-Hinton II broth (catalog number B12322; Fisher Scientific). Then 100 µL was added into each well for a particular strain, excluding the growth control wells. Plates were incubated at 37°C for 20 hours and then

measured using an accuSkan FC plate spectrophotometer (Fisher Scientific) at a wavelength of 620 nm.

Virulence gene screening

Peptide sequences for genes associated with *bimA* (BPSS1492), the type III secretion system (BPSS1390-BPSS1410), and the type VI secretion system 5 (BPSS0091-BPSS0117) were screened against all *B. mayonis* sp. nov. and *B. savannae* sp. nov. genomes (Table 1) with LS-BSR v1.2.3 (24) in conjunction with tblastn v2.9.0 (25). The blast score ratio (BSR) (26) was calculated for each gene across each genome assembly.

Virulence testing in mouse models

The pathogenic potential of *B. mayonis* sp. nov. strain BDU6^T and *B. savannae* sp. nov. strain MSMB266^T were investigated in a BALB/c mouse model using methods previously reported (13); *B. thailandensis* strain E264^T also was included as a comparison. Briefly, live culture was cultivated to logarithmic phase (OD_{620} ~ 1.0) in Luria-Bertani (LB) broth as previously described (24). Sterile 1xPBS was used to wash cells twice before making dilutions for injecting mice. Viability counts of the final inocula were made on LB agar plates. BALB/c mice 6–8-week-old in treatment groups of 5 mice per cage were utilized; food and water were provided ad libitum. All mice in a single cage received the same infectious dose (*B. mayonis* sp. nov.: 3.82x 10^4, 10^5, or 10^6 CFU; *B. savannae* sp. nov.: 0.92x 10^4, 10^5, or 10^6 CFU; *B. thailandensis*: 3.4x 10^4, 10^5, or 10^6 CFU) via a single subcutaneous injection in the scruff of the neck. Mice were monitored daily for health status. All mice were euthanized on day 21 post-injection. This work was conducted under approved protocols from the Northern Arizona University's Institutional Animal Care and Use Committee (Protocol 14-011) and the US Department of Defense's Animal Care and Use Review Office (HDTRA1-12-C-0066_Wagner).

16S rRNA gene analysis

16S rRNA genes were extracted from genome assemblies for the two *B. mayonis* sp. nov. strains (BDU6^{T}, BDU8) and the four *B. savannae* sp. nov. strains (MSMB266^{T}, MSMB852, BDU18, BDU19) as previously described (11). We investigated the number of 16S rRNA operons present in the *B. mayonis* sp. nov. and *B. savannae* sp. nov. genomes using the publicly available rapid ribosomal RNA prediction tool barrnap v0.9 (https://github.com/tseemann/barrnap). A maximum likelihood phylogeny was inferred with IQ-TREE v2.0.3 (25) and the HKY+F+I substitution model (26) using 16S rRNA sequences, and was rooted with *B. ubonensis*. The number of pairwise SNPs between unique 16S rRNA gene copies was calculated with snp-dists v0.7.0 (https://github.com/tseemann/snp-dists).

Genome assembly and core genome phylogeny

Genomes for the two *B. mayonis* sp. nov. (BDU6^{T}, BDU8) and four *B. savannae* sp. nov. (MSMB266^{T}, MSMB852, BDU18, BDU19) strains were previously sequenced on the PacBio platform (4, 13). To construct the core genome phylogeny, assemblies were aligned against the genome of *B. pseudomallei* strain K96243 (GCA_000011545.1) (27) using NUCmer (28). The reference K96243 genome also was aligned against itself with NUCmer to identify duplicated regions, which were masked from subsequent analyses; these methods were wrapped by NASP v1.1.2 (29). A maximum-likelihood phylogeny was inferred from an alignment of 434,216 SNPs with IQ-TREE v1.6.10, using the TVM+F+ASC+R3 substitution model and 1,000 bootstrap replicates.

Multi-locus sequence typing (MLST)

Genes for the seven MSLT loci in the *B. pseudomallei* pubMLST typing scheme (15) were extracted *in silico* from the genomes of the two *B. mayonis* sp. nov. strains (BDU6^T, BDU8) and the four *B. savannae* sp. nov. strains (MSMB266^T, MSMB852, BDU18, BDU19) using blastn v2.5.0 (30). The seven genes in this MLST typing scheme are *ace*, *gltB*, *gmhD*, *lepA*, *lipA*, *narK*, and *ndh*. As of 21 June 2021, a total of 1,934 sequence types (STs) had been identified in *B. pseudomallei* and closely related species by MLST (http://pubmlst.org).

Average nucleotide identity values and digital DNA-DNA hybridization

Average nucleotide identity (ANI) and digital DNA-DNA hybridization (dDDH) were calculated using complete genome assemblies for *B. mayonis* sp. nov. strains BDU6^T and BDU8 and *B. savannae* sp. nov strains MSMB266^T and MSMB852, and genome assemblies with four contigs for *B. savannae* sp. nov. strains BDU18 and BDU19 (NCBI accession numbers listed in Table 1). These assemblies were compared to genome assemblies (using complete genome assemblies when available) of the following Bpc strains: *B. humptydooensis* MSMB43^T, *B. mallei* ATCC 23344^T, *B. oklahomensis* C6786^T, *B. pseudomallei* K96243, *B. singularis* MSMB175, and *B. thailandensis* E264^T (NCBI accession numbers listed in Table 1).

For ANI, all assemblies were uploaded to JSpecies WS and analyzed using the ANIb algorithm (31); the authors of JSpecies suggested that ANI values <95% suggest separate species. The digital DNA-DNA hybridization (dDDH) values were produced by the genome-to-genome distance calculator (GGDC), which correlates with values obtained by conventional DDH and also provides a confidence-interval estimation (32). Briefly, with this approach two strains are considered as belonging to different species if DNA-DNA relatedness between them is less than 70%. The dDDH values were calculated using formula 2 in the GGDC, which summed the

identities found in high-scoring segment pairs (HSP) and then divided the sums by the overall HSP length (32).

Comparative genomics

To better understand the composition of the genomes of the putative new species, annotated locus tags were obtained from GenBank for each genome. For both putative species, combined locus tags were de-replicated with cd-hit v4.8.1 (33) at an ID of 0.8 and the pan genome for each species was defined by the total number of cluster representatives. Unique locus tags were screened with LS-BSR v1.2.2 (34) against a set of 3,273 *Burkholderia* genome assemblies downloaded with the ncbi-genome-download tool (https://github.com/kblin/ncbi-genome-download). Any locus with a blast score ratio (BSR) value (35) of <0.4 in all non-target genomes was identified to be unique to that species. The functional profile of each unique region was identified with eggnog mapper v2.0.1 (36) and regions suspected to contain phage sequence were further classified using PHAST (37). The core genome for each putative species was distinguished by identifying coding regions with a BSR value of ≥0.8 across all target genomes.

To understand the overlap of the *B. pseudomallei* core genome with other species in the Bpc, including *B. mayonis* sp. nov. and *B. savannae* sp. nov., a set of 1,744 *B. pseudomallei* genomes were annotated with Prokka v1.14.6 (38) and the pan-genome was calculated with Panaroo v1.2.3 (39). The amount of overlap was determined for a coding region if it had a BSR value ≥0.8 in any genome from another species in the Bpc.

Results

Bacterial growth and characteristics

Growth of both type strains, BDU6^{T} (*B. mayonis* sp. nov.) and MSMB266^{T} (*B. savannae* sp. nov.), was observed on all media types tested in plate format (Ashdown's, Columbia Blood, MacConkey, and Luria-Bertani) after 24 hours when incubated at 25°C and 37°C, with the optimal growth for both strains observed at 37°C on all media types after at least 48 hours of incubation. Incubation at 25°C for at least 48 hours resulted in the optimal growth only on Columbia blood agar and for all other media types after at least 72 hours of incubation. Limited to no growth was observed at 42°C for all strains on the four media types. Colony morphology varied depending on media type (Fig. S2 and Fig. S3). Unless otherwise noted, Luria-Bertani agar was the medium used during various analyses and strains were stored long term in cryovials containing Luria-Bertani broth with 20% glycerol at -80°C.

Biochemical differentiation of the type strain of *B. mayonis* sp. nov. (BDU6^{T}) from its closest genetic near neighbor, *B. oklahomensis* (Figure 1), was observed in the inability of *B. mayonis* sp. nov. to hydrolyze esculin and assimilate arabinose. Biochemical differentiation of the type strain of *B. savannae* sp. nov. (MSMB266^{T}) from *B. oklahomensis* was observed in the inability of *B. savannae* sp. nov. to hydrolyze esculin and assimilate both arabinose and maltose. Type strains of all three of these species were positive for arginine, adipate, caprate, citrate, gelatin, gluconate, glucose, malate, mannitol, mannose, nitrate, N-acetylglucosamine, and phenylacetate. All three type strains were negative for glucose (acidification), tryptophan, urea, and PNPG (Table 2).

Matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) of the two *B. mayonis* sp. nov. and four *B. savannae* sp. nov. strains revealed that they cluster with other members of the Bpc and *B. ubonensis*. Within the MALDI-TOF MS cluster containing the species in the Bpc and *B. ubonensis*, the four *Burkholderia savannae* sp. nov.

strains form a cluster separate from other species, whereas *Burkholderia mayonis* sp. nov. strains did not form a distinct cluster (Fig. S4).

Antimicrobial susceptibility screening

All six *B. mayonis* sp. nov. and *B. savannae* sp. nov. strains were susceptible *in vitro* to amoxicillin/clavulanate, ceftazidime, doxycycline, imipenem, and trimethoprim/sulfamethoxazole based on CLSI breakpoints for *B. pseudomallei* (M45) (40). All of these strains were susceptible *in vitro* to meropenem, and were susceptible or intermediate to chloramphenicol with the exception of BDU6^{T}, which displayed resistance based on the CLSI breakpoints for *B. cepacia* complex (M100) (41). All minimal inhibitory concentrations (MICs) are reported in Table 3, including for other antimicrobials for which no breakpoints are established.

Virulence screening

Although none of the *B. pseudomallei* virulence genes (*bimA*, the complete type III secretion system, and the complete type VI secretion system 5) were conserved in any of the *B. mayonis* sp. nov. or *B. savannae* sp. nov. genomes there was a homolog to the type VI secretion system in the *B. savannae* sp. nov. genomes (Table S1). *B. mayonis* sp. nov. strain BDU6^{T}, *B. savannae* sp. nov. strain MSMB266^{T}, and *B. thailandensis* strain E264^{T} did not cause mortality in any mice at any of the doses when delivered via the subcutaneous route, nor did any mice show outward signs of illness. In comparison, subcutaneous infections of fully virulent *B. pseudomallei* results in 50% mortality within 10 days at a dose of 10^{3} CFU (42). It remains unknown if delivery via the inhalation route might increase the pathogenicity of these species; *B. thailandensis* E264^{T} can cause high mortality in mice at doses of 10^{4} - 10^{6} CFU when delivered as an aerosol (24, 43, 44).

The lack of mortality in mice suggests that *B. mayonis* sp. nov. and *B. savannae* sp. nov. are likely environmental saprophytes, similar to most other members of the Bpc.

Genetic and genomic comparative analysis

The 16S rRNA phylogeny revealed two novel clades for *B. mayonis* sp. nov. and *B. savannae* sp. nov. that were distinct from each other and from the other closely related *Burkholderia* species in the Bpc (Fig. S5). Similar to *B. pseudomallei*, *B. thailandensis*, *B. humptydooensis*, *B. oklahomensis*, and *B. singularis* (6), four rRNA operons are present in all examined *B. mayonis* sp. nov. and *B. savannae* sp. nov. strains with the exception of *B. savannae* sp. nov. strain BDU19, which has six rRNA operons. *B. mayonis* sp. nov. strains BDU6^{T} and BDU8 and *B. savannae* sp. nov. MSMB266^{T} each had two unique versions among the four copies of 16S rRNA, whereas the four copies within *B. savannae* sp. nov. strains MSMB852 and BDU18 and the six copies within BDU19 were all identical (Fig. S5). A pairwise similarity matrix shows the percent identity and number of SNPs between each of the unique16 rRNA sequences (Table S2). Briefly, within *B. mayonis* sp. nov. and *B. savannae* sp. nov. percent identity of the 16S rRNA sequences ranged from 99.1-99.9% (1-13 SNPs) and 99.7-100% (0-4 SNPs), respectively (Table S2). The most closely related species to *B. mayonis* sp. nov. in the 16S rRNA phylogeny (Fig. S5) was *B. thailandensis* (strain E264), with a percent identity ranging from 99.1-99.9% (12-14 SNPs; Table S2), depending on the *B. mayonis* sp. nov. strain. The most closely related species to *B. savannae* sp. nov. in the 16S rRNA phylogeny was *B. mayonis* sp. nov., with a percent identity ranging from 98.6-99.0% (16-21 SNPs; Table S2), depending on the strain (Fig. S5).

Each strain in this study was assigned a distinct sequence type (ST) using the *B. pseudomallei* complex MLST system (Table 1), demonstrating the significant genetic diversity found within both species. This is especially the case considering that four of the strains (*B.*

mayonis sp. nov., BDU6^{T} and BDU8; *B. savannae* sp. nov., BDU18 and BDU19) were collected from the same single soil sample. Although BDU18 and BDU19 appear closely related on the core genome phylogeny (Figure 1), there are 4,962 SNPs differentiating these two isolates.

Finished assemblies were completed for both *B. mayonis* sp. nov. strains (BDU6^{T} and BDU8) and two of the four *B. savannae* sp. nov. strains (MSMB266^{T} and MSMB852) using PacBio sequencing. The assemblies for *B. mayonis* sp. nov. strains BDU6^{T} and BDU8 consist of two contigs, corresponding to the two chromosomes typical of *Burkholderia* spp.; the chromosomes 1 and 2 of BDU6^{T} are 3,838,800 bp and 2,752,114 bp, respectively, whereas chromosomes 1 and 2 of BDU8 are 4,439,942 bp and 2,917,588, respectively. The assemblies for *B. savannae* sp. nov. strains MSMB266^{T} and MSMB852 consist of three contigs each, corresponding to two chromosomes and one plasmid each; chromosome 1, chromosome 2, and the plasmid of MSMB266^{T} (pMSMB0266) are 4,228,278 bp, 2,824,254 bp, and 375,023 bp, respectively, whereas chromosome 1, chromosome 2, and the plasmid of MSMB852 (pMSMB0852) are 4077888 bp, 2934072 bp, and 69213 bp, respectively. The PacBio assemblies for the other two *B. savannae* sp. nov. strains, BDU18 and BDU19, consist of four contigs each consisting of contig sizes of 4,097,543 bp, 249,544 bp, 66,284 bp, and 2,746,170 bp for BDU18 and 2,833,644 bp, 2,161,131 bp, 1,648,896 bp, and 215,161 bp for BDU 19 (Table 1). The PacBio whole-genome sequence NCBI accession numbers for BDU6^{T} are CP013386.1 for chromosome 1 and CP013387.1 for chromosome 2; and for MSMB266^{T} are CP013417.1 for chromosome 1, CP013418.1 for chromosome 2, and CP013419.1 for pMSMB0266. The PacBio whole-genome assembly NCBI accession numbers for all strains are listed in Table 1.

The core genome phylogeny revealed the phylogenetic positions of *B. mayonis* sp. nov. and *B. savannae* sp. nov. in relation to each other and to other species in the Bpc (Fig. 1). *B.*

savannae sp. nov. forms a distinct clade that is separate from all other species in the Bpc. Although *B. mayonis* sp. nov. is most closely related to *B. oklahomensis,* it also forms a distinct and separate clade with >35,000 core genome SNPs separating it from *B. oklahomensis*.

The ANI and dDDH values calculated among the *B. mayonis* sp. nov. and *B. savannae* sp. nov. strains, and between them and strains from other species in the Bpc, supports our proposal that the *B. mayonis* sp. nov. and *B. savannae* sp. nov. strains belong to their corresponding species and that *B. mayonis* sp. nov. and *B. savannae* sp. nov. are distinct from all other Bpc species. Although the two *B. mayonis* sp. nov. strains have a dDDH value of 68.5 ± 2.9, which is slightly below the similarity threshold defining members of the same species, the ANI value (95.63%) supports these two strains belonging to the same species. The amount of genetic diversity observed between these two *B. mayonis* sp. nov. strains is quite intriguing, especially given that both strains were collected from not only the same location but also the same soil sample. Isolating additional *B. mayonis* sp. nov. strains from soil collected in other locations will shed important new insights on overall levels of genetic diversity within this novel species. The ANI and dDDH values for the four *B. savannae* sp. nov. strains (ANI: 98.98% to 99.31%, dDDH: 92.6 ± 1.8 to 93.5 ± 1.7; Table 4) clearly support that these strains are members of the same species. Collectively, ANI values above 95% and/or dDDH values above 70 indicate that each set of strains belongs to its corresponding single species, including the proposed *B. mayonis* sp. nov. type strains BDU6^{T} and the proposed *B. savannae* sp. nov. type strain MSMB266^{T}. As expected, ANI values between *B. pseudomallei* and its host-adapted clone, *B. mallei*, were >95%, as previously shown (4, 13, 15). However, the remaining ANI values <95% and dDDH values <70% indicate separate species for *B. mayonis* sp. nov., *B. savannae* sp. nov., and the other Bpc species, with ANI values ranging from 83.73% to 94.67% and dDDH values ranging

from 29.3 ± 2.4 to 59.8 ± 2.8 (Table 4). This confirms that the *B. mayonis* sp. nov. strains comprise a distinct species from *B. oklahomensis* and the other species in the Bpc, as do the *B. savannae* sp. nov. strains.

The sizes of the pan-genomes in *B. mayonis* sp. nov. and *B. savannae* sp. nov. were 7,460 and 7,804 coding DNA sequences (CDSs), respectively, with core-genome sizes of 4,448 and 5,435 CDSs, respectively. There were 223 CDSs within *B. mayonis* sp. nov. and 159 CDSs within *B. savannae* sp. nov. that share no close homolog to those within all other examined public *Burkholderia* genome assemblies (*n*=3,269). An analysis based on clusters of orthologous genes (COGs) identified the broad functional categories of some of these unique genes (Figure 2), although the majority of CDSs could not be classified or the function was unknown. Many unique CDSs in both *B. mayonis* sp. nov. and *B. savannae* sp. nov. were identified in clusters. For example, a number of unique coding regions in a contiguous cluster were associated with phage (*B. mayonis* sp. nov., in strain BDU8 WS71_RS21930 to WS71_RS22315; *B. savannae* sp. nov., in strain BDU18 WS72_RS13230 to WS72_RS13570), suggesting these regions are mobile genetic elements associated with phage integration into the chromosome. Although other phages have been associated with virulence in *Burkholderia* (45), the function of these particular phages is not known and could be the focus of future study.

The ability to distinguish between *B. mayonis* sp. nov. or *B. savannae* sp. nov. and other commonly isolated species of the Bpc, such as *B. pseudomallei* and *B. thailandensis*, in environmental and, less likely, clinical samples is important. Obviously, this could be achieved via whole genome sequencing of isolates, but this often is not possible, particularly in developing areas of the world. Different colony morphologies on Ashdown's agar should provide a clear distinction between these two novel species and *B. pseudomallei* and *B. thailandensis* but there

could be morphological differences within species based on differences among strains, across geographic locations, and among different laboratories. Fortunately, distinguishing *B. mayonis* sp. nov. or *B. savannae* sp. nov. from other *Burkholderia* spp. can be achieved with biochemical tests. *B. mayonis* sp. nov. and *B. savannae* sp. nov. can be distinguished from *B. pseudomallei* with tryptophan and from *B. thailandensis* with arginine. Of course, the most definitive way to distinguish among any of the Bpc species would be to use whole genome sequencing (4) or species-specific PCR assays, if available.

There are several reasons why members of the Bpc, including *B. mayonis* sp. nov. and *B. savannae* sp. nov., are of interest to the wider scientific community. The Bpc includes the U.S. Tier 1 Select Agents *B. pseudomallei* and *B. mallei*. Previously, we demonstrated the importance of including near-neighbor genomes when designing sensitive and specific diagnostics for *B. pseudomallei* (4, 46). *B. mayonis* sp. nov. and *B. savannae* sp. nov. share seven and 23 CDSs, respectively, with the *B. pseudomallei* core genome that are not shared by other species in the Bpc (Figure 3). Thus, the addition of genomes from these novel species further constrains CDSs in the *B. pseudomallei* core genome that can be used as diagnostic targets for that species and, as such, the *B. mayonis* sp. nov. and *B. savannae* sp. nov. whole genome sequences provided here should be utilized when designing DNA-based assays specific for *B. pseudomallei*. Members of the Bpc, and *Burkholderia* species in general, also can be sources of novel natural products (18, 19). Indeed, *Burkholderia* species have been demonstrated to be useful for bioremediation (47, 48), biocontrol (49), and as potential sources of novel antibiotics (50). The detailed genomic data generated in this study, and the deposition of the type strains in public strain collections, will hopefully facilitate detailed bioprospecting studies of *B. mayonis* sp. nov. and *B. savannae* sp. nov.

Description of *Burkholderia mayonis* sp. nov.

Burkholderia mayonis sp. nov. (may.o.e.i.) L. gen. adj. *mayonis*, pertaining to Mark Mayo, an experienced and highly respected *Burkholderia* scientist in Australia whose family is linked culturally to Badu Island, an island located in the Torres Strait archipelago of Queensland, Australia, where the first group of members of this species was isolated. Mark Mayo was present on Badu Island when the strain was collected, and he serves as a mentor for local indigenous and non-indigenous scientists in northern Australia and elsewhere.

The organism is Gram-negative, rod-shaped, and non-spore forming. Growth is observed at 25°C and 37°C within 24 hours on Ashdown's selective agar, Columbia blood agar, MacConkey agar, and Luria-Bertani agar. Optimal growth at 37°C for 48-72 hours and at 25°C for 72-96 hours aerobically. No hemolysis on Columbia blood agar.

Assimilation (API 20NE) was found for arginine, adipate, caprate, citrate, gluconate, glucose, malate, mannitol, mannose, nitrate, N-acetylglucosamine, and phenylacetate, whereas it is negative for arabinose, glucose (acidification), urea, 4-nitrophenyl-β D-galactopyranoside (PNPG), and tryptophan. Gelatin is hydrolyzed. Assimilation of maltose and esculin hydrolysis is strain-dependent (Table 2).

Positive (API ZYM) for acidic phosphatase, alkaline phosphatase, esterase, esterase lipase, lipase, leucine arylamidase, and naphthol-AS-BI-phosphohydrolase. Enzymes absent on API ZYM are trypsin, α-chymotrypsin, α- and β-galactosidase, β-glucuronidase, α- and β-glucosidase, α-mannosidase, and α-frucosidase with inconsistent results for cystin arylamidase, N-acetyl-β-glucosaminidase, and valine arylamidase. This species is aerobic, oxidase positive, and catalase negative with no immediate bubbling.

B. mayonis sp. nov. strains are resistant to gentamicin and polymyxin B, have resistance or immediate resistance to chloramphenicol, but are susceptible to amoxicillin/clavulanic acid, ceftazidime, doxycycline, imipenem, meropenem, and trimethoprim/sulfamethoxazole.

The type strain is BDU6^{T}, which has been deposited to the American Type Culture Collection as TSD-80 and the Belgian Co-ordinated Collections of Microorganisms as LMG 29941.

Description of *Burkholderia savannae* sp. nov.

Burkholderia savannae sp. nov. (sa.van.na. TAINO or Spanish origin. noun. *savannae*, pertaining to grassy plains with scattered trees in tropical regions with distinct wet and dry seasons, where the first group of members of this species was isolated).

The organism is Gram-negative, rod-shaped, and non-spore forming. Growth is observed at 25°C and 37°C within 24 hours on Ashdown's selective agar, Columbia blood agar, MacConkey agar, and Luria-Bertani agar. Optimal growth at 37°C for 48-72 hours aerobically. No hemolysis on Columbia blood agar. Colony morphology varied between strains.

Assimilation (API 20NE) was found for arginine, adipate, caprate, citrate, gluconate, glucose, malate, mannitol, mannose, N-acetylglucosamine, and phenylacetate, whereas it is negative for glucose (acidification), urea, 4-nitrophenyl-β D-galactopyranoside (PNPG), and tryptophan. Hydrolysis of gelatin and esculin and the assimilation of arabinose, maltose, and nitrate are strain-dependent (Table 2).

Positive (API ZYM) for acidic phosphatase, alkaline phosphatase, cystin arylamidase, esterase, esterase lipase, lipase, leucine arylamidase, naphthol-AS-BI-phosphohydrolase, and valine arylamidase. Enzymes absent on API ZYM are trypsin, α-chymotrypsin, α- and β-galactosidase, β-glucuronidase, α- and β-glucosidase, α-mannosidase, and α-frucosidase, with

inconsistent results for N-acetyl-β-glucosaminidase. This species is aerobic, oxidase positive, and catalase negative with no immediate bubbling.

B. savannae sp. nov. strains are resistant to gentamicin and polymyxin B, but are susceptible to amoxicillin/clavulanic acid, ceftazidime, doxycycline, imipenem, meropenem, and trimethoprim/sulfamethoxazole; immediate resistance or susceptibility to chloramphenicol is strain dependent.

The type strain is MSMB266^{T}, which was deposited to the American Type Culture Collection as TSD-82 and the Belgian Co-ordinated Collections of Microorganisms as LMG 29940.

Acknowledgments

The authors declare no conflicts of interest. This work was funded by DOD | Defense Threat Reduction Agency (DTRA; HDTRA1-12-C-0066 and HDTRA1-17-1-0051), the Australian National Health and Medical Research Council, and the Australian Research Council.

Tables and Figures

Chapter 4 Tables

Table 1. Whole-genome sequence, sequence type (ST), and epidemiology data for *B. pseudomallei* complex species, including *B. mayonis* sp. nov. and *B. savannae* sp. nov. Two chromosomes are present for all genomes and some also include a single plasmid. All *B. mayonis* sp. nov. and *B. savannae* sp. nov. strains originated from Australia, and all were isolated from soil.

Species and strain	GC content (%)	Genome size (Mb)	No. of CDS[d]	NCBI assembly accession number	ST[f]	Region of isolation or country	Year	Originating lab
B. mayonis sp. nov. BDU6[T]	66.25	6.6[a]	5,672	ASM152374v2[e]	1003	Badu Island QLD	2011	James Cook University
B. mayonis sp. nov. BDU8	66.47	7.4[a]	6,368	ASM152263v2[e]	962	Badu Island QLD	2011	James Cook University
B. savannae sp. nov. MSMB266[T]	67.05	7.4[a,c]	6,408	ASM152444v2[e]	646	Acacia Hills NT	2006	Menzies School of Health and Research
B. savannae sp. nov. MSMB852	67.32	7.1[a,c]	6,024	ASM152462v2[e]	1773	Robin Falls NT	2010	Menzies School of Health and Research
B. savannae sp. nov. BDU18	67.25	7.2[b]	6,056	ASM154691v1	963	Badu Island QLD	2011	James Cook University
B. savannae sp. nov. BDU19	67.49	6.9[b]	5,785	ASM154695v1	964	Badu Island QLD	2011	James Cook University
B. singularis MSMB175	64.8	5.7	4,715	ASM171887v1[e]	n/a	Australia	2004	Menzies School of Health and Research
B. humptydooensis MSMB43	67.14	7.3[c]	6,324	ASM151374v1[e]	318	Australia	1995	Menzies School of Health and Research
B. thailandensis E264	67.6	6.7	5,652	ASM1236v1[e]	80	Thailand	1994	n/a external genome
B. oklahomensis C6786	66.9	7.1	6,097	ASM17037v1	81	United States	1973	n/a external genome
B. pseudomallei K96243	68.05	7.2	5,948	ASM1154v1[e]	10	Thailand	1998	n/a external genome
B. mallei ATCC 23344	68.5	5.8	5,006	ASM1170v1	40	Burma	1944	n/a external genome

[a] PacBio sequencing from this study resulting in a complete genome.

[b] PacBio sequencing from this study resulting in four contigs.

[c] One plasmid present.

[d] CDS = coding DNA sequences

[e] Complete genome assembly available from NCBI.

[f] Based on the *B. pseudomallei* MLST (https://pubmlst.org).

Table 2. Differential phenotypic characteristics of strains of *B. mayonis* sp. nov., *B. savannae* sp. nov., as well as representative strains from closely related species within the *B. pseudomallei* complex. Species: Bp, *Burkholderia pseudomallei*; Bt, *B. thailandensis*; Bo, *B. oklahomensis*; Bm, *B. mayonis* sp. nov.; Bs, *B. savannae* sp. nov.. +, positive reaction; -, negative reaction. All strains were positive for the assimilation of adipate, caprate, citrate, gluconate, glucose, malate, mannitol, mannose, N-acetylglucosamine, and phenylacetate; and all strains were negative for glucose (acidification), urea, and PNPG (these data not shown).

	Characteristic (compound present in medium or assimilated by strain)								
Biochemical reaction	**Bp*** K96243	**Bt** E264	**Bo** C6786	**Bm** BDU6^T	**Bm** BDU8	**Bs** MSMB266^T	**Bs** MSMB852	**Bs** BDU18	**Bs** BDU19
Nitrate	+	+	+	+	+	+	+	-	+
Tryptophan	+	-	-	-	-	-	-	-	-
Arginine	+	-	+	+	+	+	+	+	+
Esculin	-	+	+	-	+	-	-	-	+
Gelatin	+	+	+	+	+	+	-	+	+
Arabinose assimilation	-	+	+	-	-	-	+	-	-
Maltose assimilation	-	+	+	+	-	-	+	-	-

* Data obtained from a previous study (23).

Table 3. Summary of minimal inhibitory concentrations (MICs) of antimicrobials determined in duplicate by the microdilution method for *B. mayonis* sp. nov. (Bm) and *B. savannae* sp. nov. (Bs).

Antimicrobial substance	Resistance breakpoint (mg/liter) if available	MIC (mg/liter)					
		Bm BDU6^T	Bm BDU8	Bs MSMB266^T	Bs MSMB852	Bs BDU18	Bs BDU19
Amoxicillin/clavulanic acid[1]	≥32/16 (40)	8/4	≤4/2	8/4	8/4	8/4	8/4
Azithromycin		>64	>64	>64	>64	>64	>64
Carbenicillin		128	64	64	32	64	64
Ceftazidime	≥32 (40)	≤4	≤4	≤4	≤4	≤4	≤4
Ceftazidime/avibactam[2]		4/4	1/4	≤0.5/4	≤0.5/4	≤0.5/4	≤0.5/4
Chloramphenicol	≥32 (41)	32	16	8	16	8	16
Ciprofloxacin		2	≤0.5	≤0.5	≤0.5	1	1
Doripenem		≤0.5	≤0.5	≤0.5	≤0.5	≤0.5	≤0.5
Doxycycline	≥16 (40)	≤1	≤1	≤1	≤1	≤1	≤1
Gentamicin		32	>64	32	32	64	64
Imipenem	≥16 (40)	≤1	≤1	≤1	≤1	≤1	≤1
Kanamycin		16	32	16	16	16	32
Meropenem	≥16 (41)	≤1	≤1	≤1	≤1	≤1	≤1
Piperacillin		≤8	≤8	≤8	≤8	≤8	≤8
Piperacillin/tazobactam		≤8/4	≤8/4	≤8/4	≤8/4	≤8/4	≤8/4
Polymyxin B		512	>2048	512	>2048	>2048	>2048
Sulfamethoxazole		>512	256	>512	>512	>512	>512
Tigecycline		1	0.5	2	1	1	1
Trimethoprim		4	4	2	≤1	4	2
Trimethoprim/sulfamethoxazole	≥4/76 (40)	2/38	≤1/19	≤1/19	≤1/19	≤1/19	≤1/19

[1] For amoxicillin/clavulanic acid, clavulanic acid was maintained at 4 mg/ml in all wells.

[2] For ceftazidime/avibactam, avibactam was maintained at 4 µg/ml in all wells.

Table 4. ANI and dDDH values for whole-genome sequences similarities. Assemblies used for analyses are listed in Table 1. Species: Bma, *B. mallei*; Bp, *Burkholderia pseudomallei*; Bt, *B. thailandensis*; Bo, *B. oklahomensis*; Bh, *B. humptydooensis;* Bm, *B. mayonis* sp. nov.; Bs, *B. savannae* sp. nov.; Bsi, *B. singularis*.

	ANIb or dDDH value for comparison with genome of[a]:											
Species and strain	**Bma** ATCC 23344^T	**Bp** K96243	**Bt** E264	**Bo** C6786	**Bh** MSMB 43^T	**Bm** BDU8	**Bm** BDU6^T	**Bs** MSMB 266^T	**Bs** MSMB 852	**Bs** BDU18	**Bs** BDU19	**Bsi** MSMB 175
Bma ATCC23344		92.7 ± 1.8	48 ± 2.6	42.6 ± 2.6	50.2 ± 2.7	40.9 ± 2.5	41 ± 2.5	42.6 ± 2.5	42.6 ± 2.6	42.6 ± 2.5	42.6 ± 2.5	30.4 ± 2.5
Bp K96243	98.09		47.2 ± 2.6	40.1 ± 2.5	48.6 ± 2.6	38.8 ± 2.5	39.5 ± 2.5	40.1 ± 2.5	40.4 ± 2.5	40.3 ± 2.5	40.5 ± 2.5	29.3 ± 2.4
Bt E264^T	91.54	92.17		42.9 ± 2.6	52.9 ± 2.7	41.9 ± 2.6	41.8 ± 2.5	43.6 ± 2.5	43.6 ± 2.6	43.5 ± 2.6	43.6 ± 2.6	30.4 ± 2.5
Bo C6786^T	89.42	89.61	89.78		43.5 ± 2.6	59.8 ± 2.8	59.1 ± 2.8	42.8 ± 2.6	43.1 ± 2.6	42.9 ± 2.6	43.1 ± 2.6	30.1 ± 2.5
Bh MSMB43^T	91.44	91.92	92.22	90.53		42.3 ± 2.6	42.4 ± 2.6	44.2 ± 2.5	44.5 ± 2.6	44.7 ± 2.6	44.4 ± 2.6	29.8 ± 2.5
Bm BDU8	89.04	89.23	89.47	94.67	90.17		68.5 ± 2.9	41.8 ± 2.6	41.7 ± 2.5	41.5 ± 2.5	41.6 ± 2.6	29.5 ± 2.5
Bm BDU6^T	89.15	89.47	89.77	94.50	90.32	95.63		41.9 ± 2.6	41.9 ± 2.5	41.7 ± 2.5	41.9 ± 2.6	29.4 ± 2.5
Bs MSMB266^T	89.82	89.97	90.31	90.88	90.86	90.64	90.45		92.6 ± 1.8	92.8 ± 1.8	93 ± 1.75	29.5 ± 2.5
Bs MSMB852	89.94	90.20	90.66	91.18	91.06	90.75	90.65	99.06		92.8 ± 1.8	92.7 ± 1.8	29.6 ± 2.5
Bs BDU18	89.81	90.00	90.36	91.05	91.03	90.63	90.46	99.02	98.98		93.5 ± 1.7	29.6 ± 2.5
Bs BDU19	89.98	90.25	90.73	91.25	91.19	90.83	90.67	99.25	99.31	99.31		29.6 ± 2.5
Bsi MSMB175	83.73	83.84	84.13	84.32	84.64	84.25	84.34	84.13	84.23	84.22	84.22	

[a] Average nucleotide identity (ANIb) are shown in the bottom left half of the matrix (below the line of identity, i.e., the line formed by blank cells for comparison of strains with themselves); digital DNA-DNA hybridization (dDDH) (with confidence intervals) are shown in the top right half of the matrix. Values in shaded boxes represent values above the similarity threshold that defines members of the same species.

Chapter 4 Figures

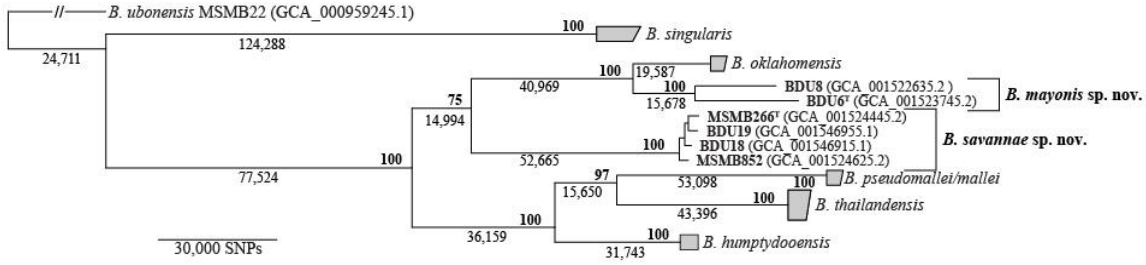


Figure 1. Core genome phylogeny of 66 strains (Table S2) in the *B. pseudomallei* complex, including two *B. mayonis* sp. nov. strains and four *B. savannae* sp. nov. strains. This maximum-likelihood phylogeny was created using core genome SNPs shared by all strains and rooted on *B. ubonensis* strain MSMB22 as an outgroup. Bold numbers at nodes indicate bootstrap support values and non-bolded numbers indicate the number of core SNPs defining that node. Collapsed nodes are shown in gray. The type strains are reflected with a T superscript in the strain name.

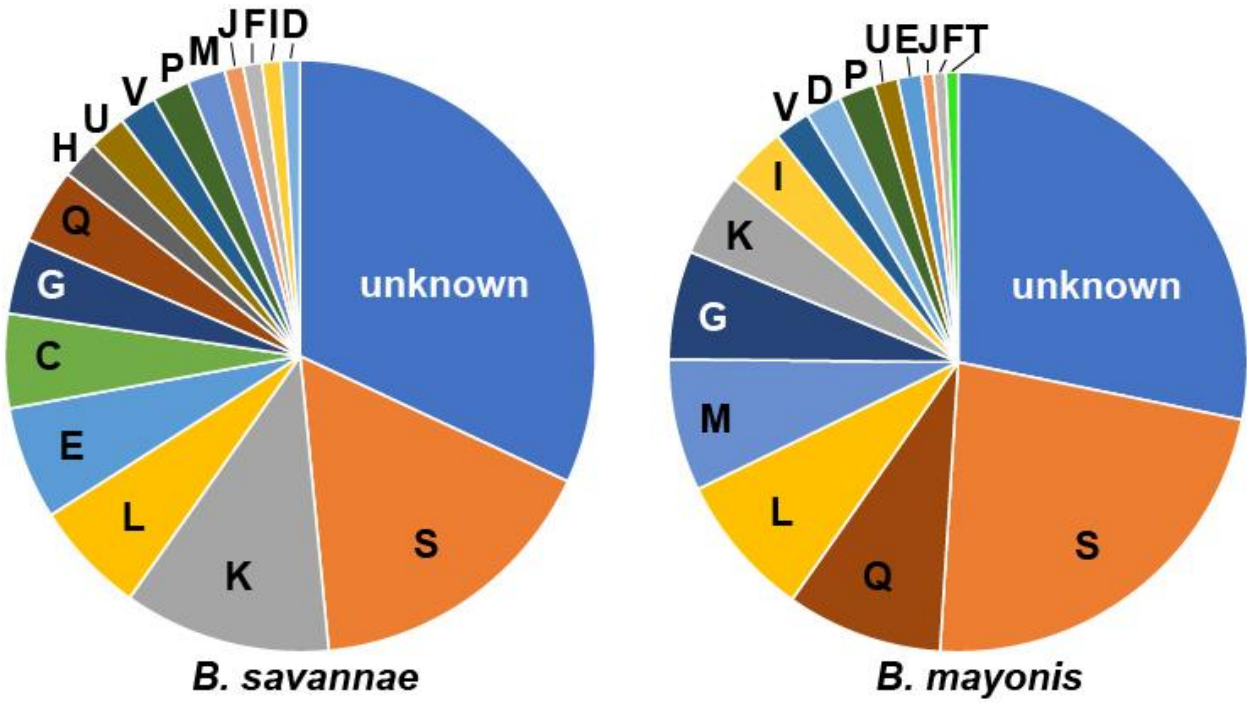


Figure 2. Cluster of orthologous genes (COG) classification (*n*=18) of unique coding DNA sequences (CDSs) in *B. savannae* sp. nov. strains (*n*=97 unique CDSs) and *B. mayonis* sp. nov. strains (*n*=149 unique CDSs) with some CDSs that have no homolog (unknown) resulting in 31 unique CDSs for *B. savannae* sp. nov. and 42 unique CDSs for *B. mayonis* sp. nov. The COG categories are as follows with the number of unique CDSs for *B. savannae* sp. nov. and *B. mayonis* sp. nov. listed respectively after each COG category: C) energy production and conversion (5; 0), D) cell cycle control and mitosis (1; 3), E) amino acid metabolism and transport (6; 2), F) nucleotide metabolisms and transport (1; 1), G) carbohydrate metabolism and transport (4; 9), H) coenzyme metabolism (2; 0), I) lipid metabolism (1; 5), J) translation (1; 1), K) transcription (11; 7), L) replication, recombination and repair (6; 12), M) cell wall/membrane/envelop biogenesis (2; 11), P) inorganic ion transport and metabolism (2; 3), Q) secondary structure (4; 13), S) function unknown (16; 34), T) signal transduction (0; 1), U)

intracellular trafficking and secretion (2; 2), V) defense mechanisms (2; 3). All classifications were performed with the eggnog-mapper.

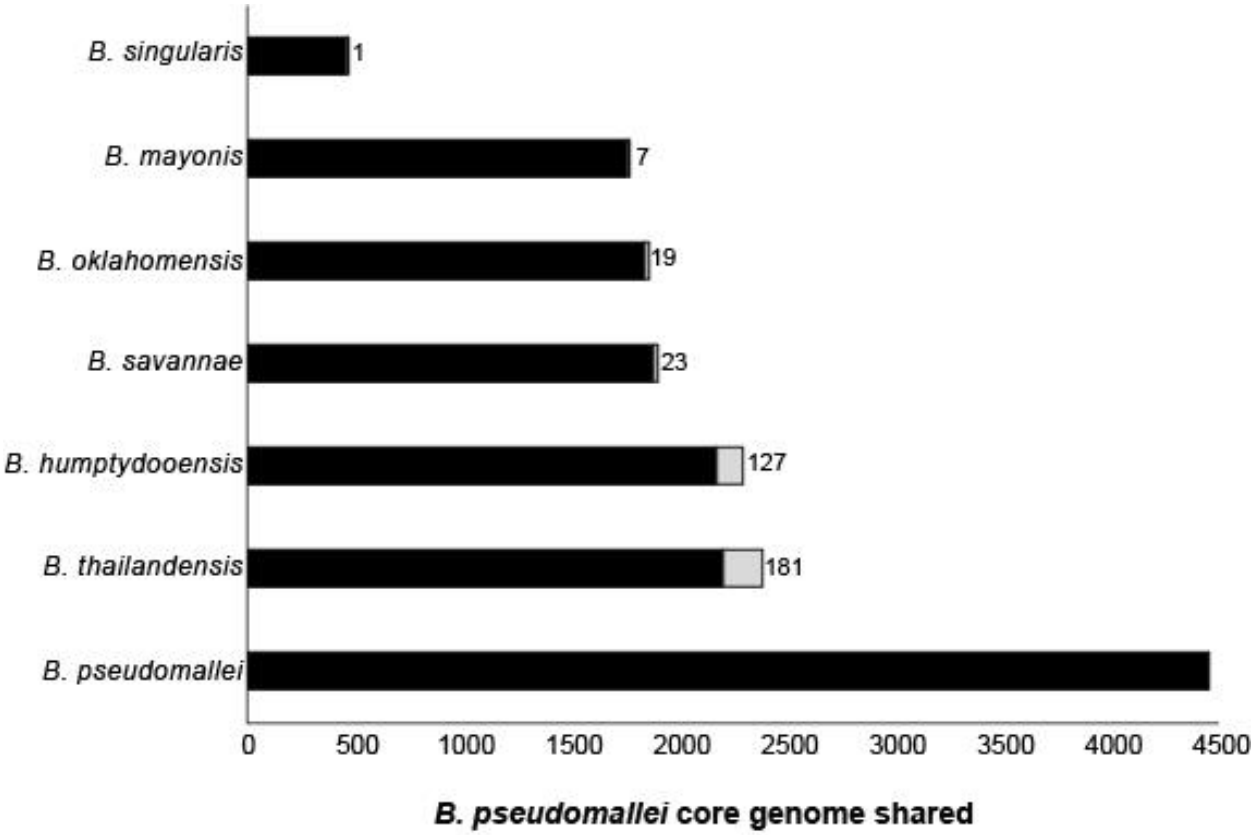


Figure 3. Overlap of the *B. pseudomallei* core genome (*n*=4,452 CDSs) with pan-genomes from other species in the *B. pseudomallei* complex (Bpc). Included *B. pseudomallei* CDSs have a blast score ratio (BSR) value >0.8 in at least one genome from the near-neighbor species. Gray regions for each bar represent CDSs that are uniquely covered by at least one genome from that species; the number at the end of the bar corresponds to these CDSs. Black bars represent *B. pseudomallei* core CDSs found in the indicated species and other species in the Bpc.

References

1. Eberl L, Vandamme P. 2016. Members of the genus Burkholderia: good and bad guys. F1000Res 5.
2. Beukes CW, Palmer M, Manyaka P, Chan WY, Avontuur JR, van Zyl E, Huntemann M, Clum A, Pillay M, Palaniappan K, Varghese N, Mikhailova N, Stamatis D, Reddy TBK, Daum C, Shapiro N, Markowitz V, Ivanova N, Kyrpides N, Woyke T, Blom J, Whitman WB, Venter SN, Steenkamp ET. 2017. Genome Data Provides High Support for Generic Boundaries in Burkholderia Sensu Lato. Front Microbiol 8:1154.
3. Lin QH, Lv YY, Gao ZH, Qiu LH. 2020. Pararobbsia silviterrae gen. nov., sp. nov., isolated from forest soil and reclassification of Burkholderia alpina as Pararobbsia alpina comb. nov. International Journal of Systematic and Evolutionary Microbiology 70:1412-1420.
4. Sahl JW, Vazquez AJ, Hall CM, Busch JD, Tuanyok A, Mayo M, Schupp JM, Lummis M, Pearson T, Shippy K, Colman RE, Allender CJ, Theobald V, Sarovich DS, Price EP, Hutcheson A, Korlach J, LiPuma JJ, Ladner J, Lovett S, Koroleva G, Palacios G, Limmathurotsakul D, Wuthiekanun V, Wongsuwan G, Currie BJ, Keim P, Wagner DM. 2016. The Effects of Signal Erosion and Core Genome Reduction on the Identification of Diagnostic Markers. MBio 7.
5. Tuanyok A, Mayo M, Scholz H, Hall CM, Allender CJ, Kaestli M, Ginther J, Spring-Pearson S, Bollig MC, Stone JK, Settles EW, Busch JD, Sidak-Loftis L, Sahl JW, Thomas A, Kreutzer L, Georgi E, Gee JE, Bowen RA, Ladner JT, Lovett S, Koroleva G, Palacios G, Wagner DM, Currie BJ, Keim P. 2017. Burkholderia humptydooensis sp

nov., a New Species Related to Burkholderia thailandensis and the Fifth Member of the Burkholderia pseudomallei Complex. Applied and Environmental Microbiology 83.

6. Vandamme P, Peeters C, De Smet B, Price EP, Sarovich DS, Henry DA, Hird TJ, Zlosnik JEA, Mayo M, Warner J, Baker A, Currie BJ, Carlier A. 2017. Comparative Genomics of Burkholderia singularis sp nov., a Low G plus C Content, Free-Living Bacterium That Defies Taxonomic Dissection of the Genus Burkholderia. Frontiers in Microbiology 8.
7. Martina P, Leguizamon M, Prieto CI, Sousa SA, Montanaro P, Draghi WO, Stammler M, Bettiol M, de Carvalho C, Palau J, Figoli C, Alvarez F, Benetti S, Lejona S, Vescina C, Ferreras J, Lasch P, Lagares A, Zorreguieta A, Leitao JH, Yantorno OM, Bosch A. 2018. Burkholderia puraquae sp. nov., a novel species of the Burkholderia cepacia complex isolated from hospital settings and agricultural soils. Int J Syst Evol Microbiol 68:14-20.
8. Ong KS, Aw YK, Lee LH, Yule CM, Cheow YL, Lee SM. 2016. Burkholderia paludis sp. nov., an Antibiotic-Siderophore Producing Novel Burkholderia cepacia Complex Species, Isolated from Malaysian Tropical Peat Swamp Soil. Front Microbiol 7:2046.
9. Takeshita K, Tamaki H, Ohbayashi T, Meng XY, Sone T, Mitani Y, Peeters C, Kikuchi Y, Vandamme P. 2018. Burkholderia insecticola sp. nov., a gut symbiotic bacterium of the bean bug Riptortus pedestris. Int J Syst Evol Microbiol 68:2370-2374.
10. Compant S, Nowak J, Coenye T, Clement C, Barka EA. 2008. Diversity and occurrence of Burkholderia spp. in the natural environment. Fems Microbiology Reviews 32:607-626.
11. Brett PJ, Deshazer D, Woods DE. 1997. Characterization of Burkholderia pseudomallei and Burkholderia pseudomallei-like strains. Epidemiol Infect 118:137-48.

12. Glass MB, Steigerwalt AG, Jordan JG, Wilkins PP, Gee JE. 2006. Burkholderia oklahomensis sp nov., a Burkholderia pseudomallei-like species formerly known as the Oklahoma strain of Pseudomonas pseudomallei. International Journal of Systematic and Evolutionary Microbiology 56:2171-2176.

13. Tuanyok A, Mayo M, Scholz H, Hall CM, Allender CJ, Kaestli M, Ginther J, Spring-Pearson S, Bollig MC, Stone JK, Settles EW, Busch JD, Sidak-Loftis L, Sahl JW, Thomas A, Kreutzer L, Georgi E, Gee JE, Bowen RA, Ladner JT, Lovett S, Koroleva G, Palacios G, Wagner DM, Currie BJ, Keim P. 2017. Burkholderia humptydooensis sp. nov., a New Species Related to Burkholderia thailandensis and the Fifth Member of the Burkholderia pseudomallei Complex. Appl Environ Microbiol 83.

14. Currie BJ. 2015. Melioidosis: Evolving Concepts in Epidemiology, Pathogenesis, and Treatment. Seminars in Respiratory and Critical Care Medicine 36:111-125.

15. 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 41:2068-79.

16. Cheng AC, Currie BJ. 2005. Melioidosis: Epidemiology, pathophysiology, and management. Clinical Microbiology Reviews 18:383-+.

17. Limmathurotsakul D, Golding N, Dance DA, Messina JP, Pigott DM, Moyes CL, Rolim DB, Bertherat E, Day NP, Peacock SJ, Hay SI. 2016. Predicted global distribution of Burkholderia pseudomallei and burden of melioidosis. Nat Microbiol 1.

18. Kunakom S, Eustaquio AS. 2019. Burkholderia as a Source of Natural Products. Journal of Natural Products 82:2018-2037.

19. Liu XY, Cheng YQ. 2014. Genome-guided discovery of diverse natural products from Burkholderia sp. Journal of Industrial Microbiology & Biotechnology 41:275-284.

20. Baker A, Tahani D, Gardiner C, Bristow KL, Greenhill AR, Warner J. 2011. Groundwater seeps facilitate exposure to Burkholderia pseudomallei. Appl Environ Microbiol 77:7243-6.

21. Limmathurotsakul D, Dance DA, Wuthiekanun V, Kaestli M, Mayo M, Warner J, Wagner DM, Tuanyok A, Wertheim H, Yoke Cheng T, Mukhopadhyay C, Puthucheary S, Day NP, Steinmetz I, Currie BJ, Peacock SJ. 2013. Systematic review and consensus guidelines for environmental sampling of Burkholderia pseudomallei. PLoS Negl Trop Dis 7:e2105.

22. Novak RT, Glass MB, Gee JE, Gal D, Mayo MJ, Currie BJ, Wilkins PP. 2006. Development and evaluation of a real-time PCR assay targeting the type III secretion system of Burkholderia pseudomallei. Journal of Clinical Microbiology 44:85-90.

23. Wuthiekanun V, Smith MD, Dance DA, Walsh AL, Pitt TL, White NJ. 1996. Biochemical characteristics of clinical and environmental isolates of Burkholderia pseudomallei. J Med Microbiol 45:408-12.

24. Morici LA, Heang J, Tate T, Didier PJ, Roy CJ. 2010. Differential susceptibility of inbred mouse strains to Burkholderia thailandensis aerosol infection. Microb Pathog 48:9-17.

25. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Molecular Biology and Evolution 37:1530-1534.

26. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nature Methods 14:587-+.

27. 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 101:14240-5.

28. Delcher AL, Phillippy A, Carlton J, Salzberg SL. 2002. Fast algorithms for large-scale genome alignment and comparison. Nucleic Acids Res 30:2478-83.

29. Sahl JW, Lemmer D, Travis J, Schupp JM, Gillece JD, Aziz M, Driebe EM, Drees KP, Hicks ND, Williamson CHD, Hepp CM, Smith DE, Roe C, Engelthaler DM, Wagner DM, Keim P. 2016. NASP: an accurate, rapid method for the identification of SNPs in WGS datasets that supports flexible input and output formats. Microb Genom 2:e000074.

30. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421.

31. Richter M, Rossello-Mora R, Glockner FO, Peplies J. 2016. JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. Bioinformatics 32:929-931.

32. Meier-Kolthoff JP, Auch AF, Klenk HP, Goker M. 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics 14:60.

33. Fu LM, Niu BF, Zhu ZW, Wu ST, Li WZ. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28:3150-3152.

34. Sahl JW, Caporaso JG, Rasko DA, Keim P. 2014. The large-scale blast score ratio (LS-BSR) pipeline: a method to rapidly compare genetic content between bacterial genomes. Peerj 2.

35. Rasko DA, Myers GSA, Ravel J. 2005. Visualization of comparative genomic analyses by BLAST score ratio. Bmc Bioinformatics 6.

36. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, Bork P. 2017. Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. Molecular Biology and Evolution 34:2115-2122.

37. Zhou Y, Liang YJ, Lynch KH, Dennis JJ, Wishart DS. 2011. PHAST: A Fast Phage Search Tool. Nucleic Acids Research 39:W347-W352.

38. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068-2069.

39. Tonkin-Hill G, MacAlasdair N, Ruis C, Weimann A, Horesh G, Lees JA, Gladstone RA, Lo S, Beaudoin C, Floto RA, Frost SDW, Corander J, Bentley SD, Parkhill J. 2020. Producing polished prokaryotic pangenomes with the Panaroo pipeline. Genome Biology 21.

40. Institute CaLS. 2016. Methods for antimicrobial dilution and disk susceptibility testing of infrequently isolated or fastidious bacteria. M45, 3rd edition, Wayne, PA.

41. Institute CaLS. 2020. Performance standards for antimicrobial susceptibility testing. M100, 30th edition, Wayne, PA.

42. Barnes JL, Ketheesan N. 2005. Route of infection in melioidosis. Emerg Infect Dis 11:638-9.

43. West TE, Frevert CW, Liggitt HD, Skerrett SJ. 2008. Inhalation of Burkholderia thailandensis results in lethal necrotizing pneumonia in mice: a surrogate model for pneumonic melioidosis. Trans R Soc Trop Med Hyg 102 Suppl 1:S119-26.

44. Wiersinga WJ, de Vos AF, de Beer R, Wieland CW, Roelofs JJ, Woods DE, van der Poll T. 2008. Inflammation patterns induced by different Burkholderia species in mice. Cell Microbiol 10:81-7.

45. Summer EJ, Gill JJ, Upton C, Gonzalez CF, Young R. 2007. Role of phages in the pathogenesis of Burkholderia, or 'Where are the toxin genes in Burkholderia phages?'. Current Opinion in Microbiology 10:410-417.

46. Williamson CHD, Wagner DM, Keim P, Sahl JW. 2018. Developing Inclusivity and Exclusivity Panels for Testing Diagnostic and Detection Tools Targeting Burkholderia pseudomallei, the Causative Agent of Melioidosis. Journal of Aoac International 101:1920-1926.

47. Afzal M, Khan S, Iqbal S, Mirza MS, Khan QM. 2013. Inoculation method affects colonization and activity of Burkholderia phytofirmans PsJN during phytoremediation of diesel-contaminated soil. International Biodeterioration & Biodegradation 85:331-336.

48. Yang ZH, Zhang Z, Chai LY, Wang Y, Liu Y, Xiao RY. 2016. Bioleaching remediation of heavy metal-contaminated soils using Burkholderia sp. Z-90. Journal of Hazardous Materials 301:145-152.

49. Sarli DA, Sanchez LA, Delgado OD. 2021. Burkholderia gladioli MB39 an Antarctic Strain as a Biocontrol Agent. Current Microbiology 78:2332-2344.

50. Depoorter E, Bull MJ, Peeters C, Coenye T, Vandamme P, Mahenthiralingam E. 2016. Burkholderia: an update on taxonomy and biotechnological potential as antibiotic producers. Applied Microbiology and Biotechnology 100:5215-5229.

Supplemental information

MALDI-TOF

Preparations of bacterial isolates (BDU6^{T}, BDU8, MSMB266^{T}, MSMB852, BDU18, BDU19) for matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) were done according to the ethanol/formic acid extraction protocol recommended by the manufacturer (Bruker Daltonics) and as previously described (1). Briefly, a loopful of bacterial material was evenly dissolved in 300 µl analytical grade water, and 900 µl pure ethanol was added. The cell suspension was centrifuged at 13,000 × g for 2 minutes and the supernatant was discarded. The centrifugation was repeated, and the residual ethanol was discarded. The pellet was air dried and thoroughly resuspended in 5-50 µl 70% formic acid depending on the size, and, finally, an equal volume of acetonitrile was added. After centrifugation at 13,000 × g for 2 minutes, 1 µl of the supernatant was transferred to the MALDI target plate. After air-drying at room temperature 1 µl of matrix solution (saturated solution of α-cyano-hydroxy-cinnaminic acid in 50% aqueous acetonitrile containing 2.5% trifluoroacetic acid) per spot was was applied. MALDI-TOF MS was conducted using a Microflex LT mass spectrometer (Bruker Daltonics) equipped with a nitrogen laser. All spectra were recorded in linear, positive ion mode across a mass/charge ratio (m/z) of 2,000 to 20,000. The acceleration voltage was 20 kV. Spectra were collected as a sum of 240 shots across a spot. Main spectra

were calculated from 8 spectra per strain and used for construction of a score oriented dendrogram by using the BioTyper software (version 2.3, Bruker Daltonics). In these analyses *B. mayonis* sp. nov. and *B. savannae* sp. nov. strains grouped together with other members of the *B. pseudomallei* complex and with *B. ubonensis* from the *B. cepacia* complex and (Fig. S4).

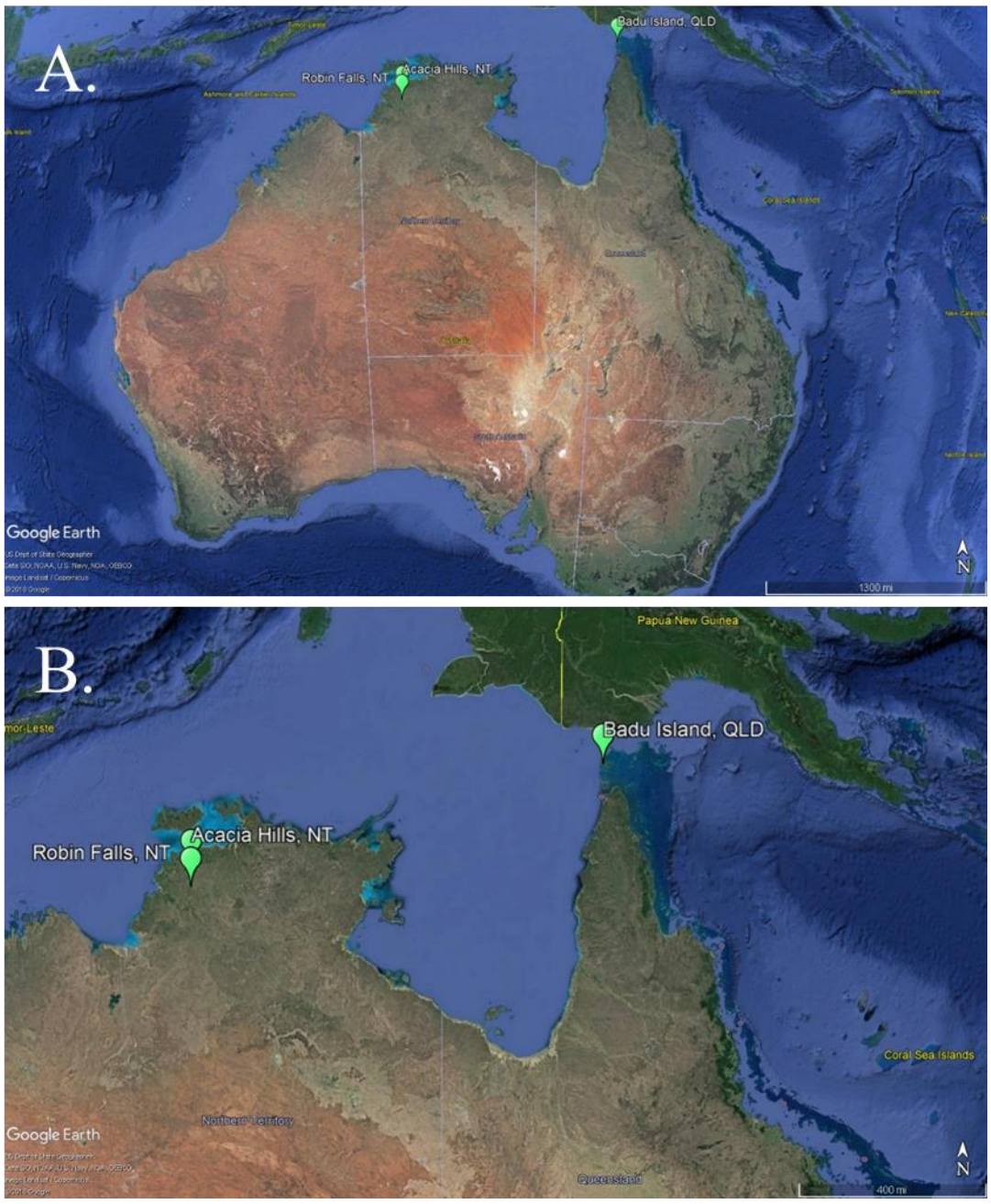


Fig. S1. Geographic origin of *B. mayonis* sp. nov. (Badu Island) and *B. savannae* sp. nov. (Acacia Hills, Badu Island, and Robin Falls) isolates from the environment (Google Earth, US Dept of State Geographer, Data SIO, NOAA, US Navy, NGA, GEBCO, Image Landset/Coperricus, ©2018 Google). Panel B is a zoomed in version of panel A.

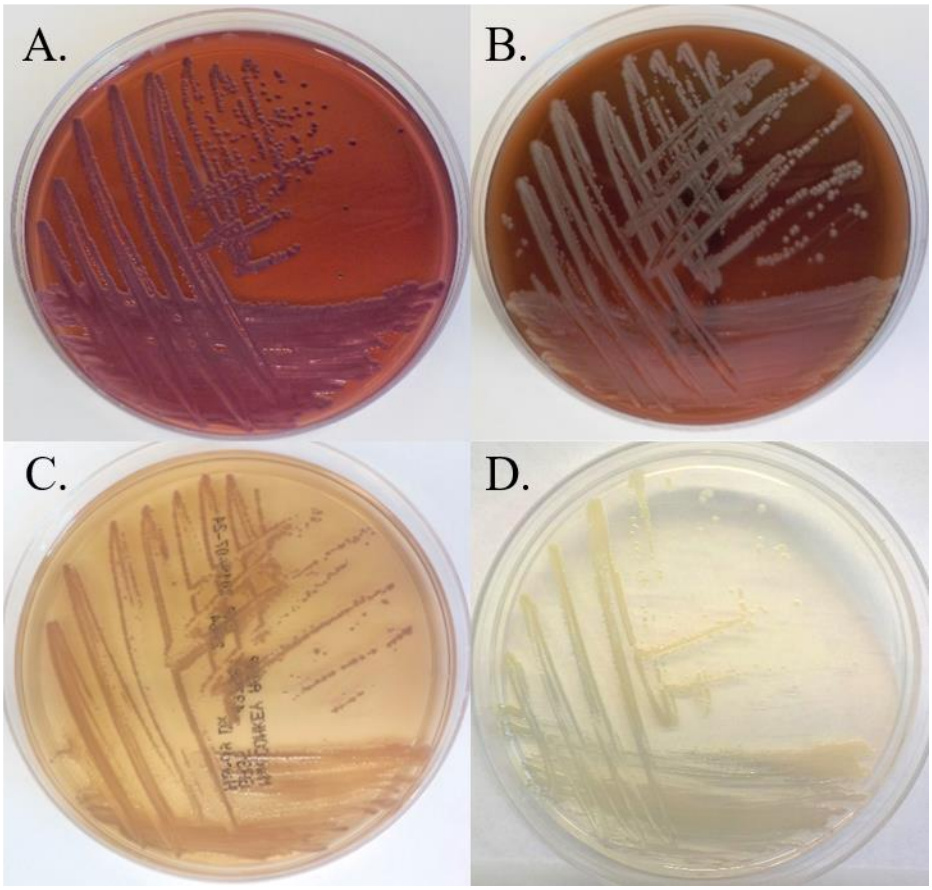


Fig. S2. *B. mayonis* sp. nov. (BDU6^{T}) on four media types after 96 hours of incubation at 25°C. Panels A. Ashdown's Agar; B. Columbia Blood Agar; C. MacConkey Agar; D. Luria-Bertani Agar.

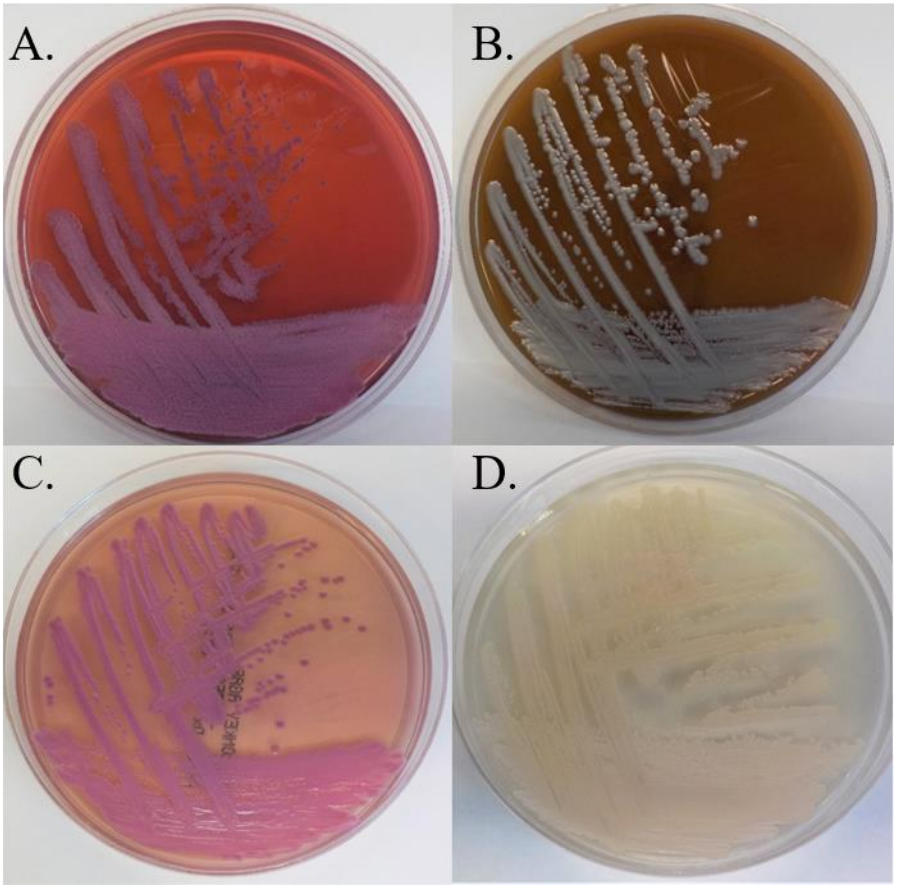


Fig. S3. *B. savannae* sp. nov. (MSMB266^{T}) on four media types after 96 hours of incubation at 25°C. Panels A. Ashdown's Agar; B. Columbia Blood Agar; C. MacConkey Agar; D. Luria-Bertani Agar.

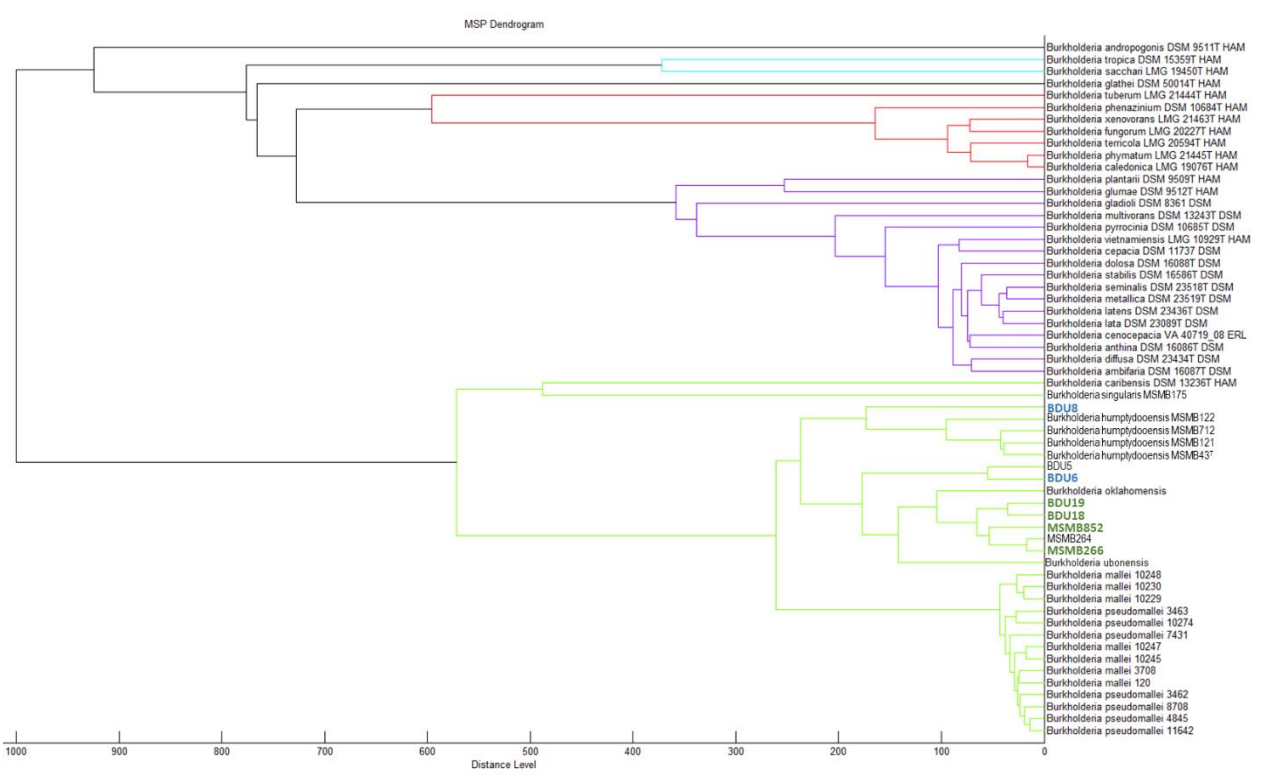


Fig. S4. Dendrogram demonstrating strain relatedness revealed by MALDI – TOF analysis. *B. mayonis* sp. nov. strains are in blue text while *B. savannae* sp. nov. strains are in green text. Although BDU5 was not described in this study the MALDI-TOF analysis suggests that this may be an additional *B. mayonis* sp. nov. strain and this is further supported by BDU5, BDU6, and BDU8 all reported as putative species 2 in a previous study (2). MSMB264 was also not described in this study but the MALDI-TOF analysis suggests that it may be an additional *B. savannae* sp. nov. strain.

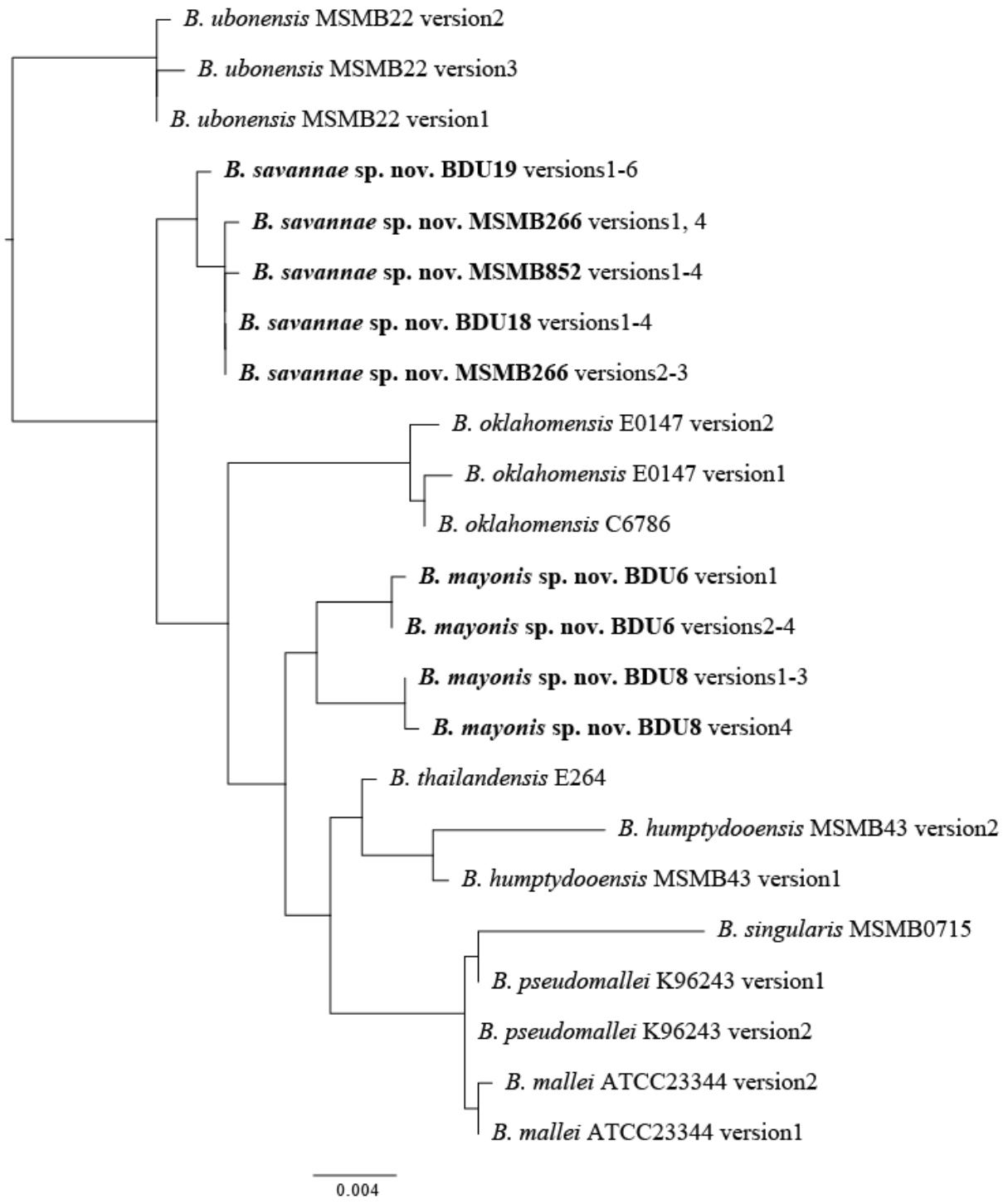


Fig. S5. 16S maximum likelihood phylogeny using unique copies of 16S from *B. mayonis* sp. nov. and *B. savannae* sp. nov. strains, bolded (1,529 bp and 23 sequences). Rooted with *B. ubonensis.*

Table S1. The blast score ratios of three virulence genes, *bimA*, the type III secretion system (T3SS), and the type VI secretion system 5 (T6SS-5) across the *B. mayonis* sp. nov. and *B. savannae* sp. nov. strains. The blast score ratios are show where "0" (white) represents the locus fully absent and "1" (red) represents the locus fully present for that strain.

		B. mayonis		*B. savannae*				*B. pseudomallei*	
Locus	**Region**	BDU8	BDU6	MSMB 266	BDU19	BDU18	MSMB 852	K96243	1026b
BPSS1492	bimA	0.30	0.19	0.31	0.19	0.19	0.19	1.00	1.00
BPSS1390	T3SS	0.52	0.19	0.53	0.53	0.52	0.53	1.00	1.00
BPSS1391	T3SS	0.34	0.05	0.33	0.33	0.33	0.33	1.00	1.00
BPSS1392	T3SS	0.47	0.19	0.45	0.45	0.45	0.45	1.00	1.00
BPSS1393	T3SS	0.27	0.12	0.29	0.28	0.29	0.28	1.00	1.00
BPSS1394	T3SS	0.74	0.49	0.74	0.74	0.74	0.74	1.00	1.00
BPSS1395	T3SS	0.35	0.07	0.36	0.36	0.36	0.36	1.00	1.00
BPSS1396	T3SS	0.17	0.07	0.18	0.18	0.18	0.18	1.00	1.00
BPSS1397	T3SS	0.54	0.26	0.53	0.53	0.53	0.53	1.00	1.00
BPSS1398	T3SS	0.26	0.00	0.28	0.28	0.28	0.28	1.00	0.99
BPSS1399	T3SS	0.30	0.00	0.28	0.29	0.29	0.29	1.00	1.00
BPSS1400	T3SS	0.50	0.24	0.49	0.49	0.50	0.49	1.00	1.00
BPSS1401	T3SS	0.67	0.36	0.68	0.68	0.67	0.68	1.00	1.00
BPSS1402	T3SS	0.21	0.08	0.21	0.21	0.21	0.21	1.00	0.99
BPSS1403	T3SS	0.23	0.00	0.21	0.21	0.21	0.21	1.00	1.00
BPSS1404	T3SS	0.79	0.46	0.79	0.79	0.79	0.79	1.00	1.00
BPSS1405	T3SS	0.60	0.34	0.60	0.60	0.60	0.60	1.00	1.00
BPSS1406	T3SS	0.09	0.07	0.11	0.11	0.12	0.12	1.00	0.96
BPSS1407	T3SS	0.28	0.07	0.30	0.30	0.30	0.31	1.00	0.99
BPSS1408	T3SS	0.32	0.20	0.32	0.32	0.33	0.32	1.00	1.00
BPSS1409	T3SS	0.00	0.18	0.16	0.18	0.18	0.17	1.00	0.96
BPSS1410	T3SS	0.73	0.00	0.72	0.72	0.72	0.72	1.00	1.00
BPSS0091	T6SS-5	0.21	0.29	0.81	0.81	0.81	0.81	1.00	0.99
BPSS0092	T6SS-5	0.28	0.41	0.91	0.90	0.91	0.90	1.00	1.00
BPSS0093	T6SS-5	0.28	0.32	0.89	0.89	0.89	0.89	1.00	1.00
BPSS0094	T6SS-5	0.00	0.26	0.85	0.84	0.84	0.84	1.00	1.00
BPSS0095	T6SS-5	0.08	0.00	0.77	0.77	0.77	0.77	1.00	0.99
BPSS0096	T6SS-5	0.06	0.07	0.83	0.84	0.84	0.84	1.00	0.99

BPSS0097	T6SS-5	0.76	0.76	0.98	0.98	0.98	0.98	1.00	1.00
BPSS0098	T6SS-5	0.86	0.86	0.99	0.99	0.99	0.99	1.00	1.00
BPSS0099	T6SS-5	0.69	0.70	0.96	0.96	0.96	0.96	1.00	0.99
BPSS0100	T6SS-5	0.44	0.45	0.89	0.89	0.89	0.89	1.00	0.99
BPSS0101	T6SS-5	0.64	0.64	0.99	0.99	0.99	0.99	1.00	1.00
BPSS0102	T6SS-5	0.39	0.39	0.93	0.93	0.93	0.93	1.00	1.00
BPSS0103	T6SS-5	0.44	0.44	0.91	0.91	0.91	0.91	1.00	1.00
BPSS0104	T6SS-5	0.00	0.00	0.86	0.86	0.86	0.86	1.00	0.99
BPSS0105	T6SS-5	0.28	0.28	0.85	0.86	0.86	0.85	1.00	0.97
BPSS0106	T6SS-5	0.24	0.20	0.82	0.82	0.82	0.82	1.00	1.00
BPSS0108	T6SS-5	0.00	0.00	0.64	0.64	0.64	0.64	1.00	0.99
BPSS0109	T6SS-5	0.00	0.00	0.48	0.47	0.47	0.47	1.00	1.00
BPSS0110	T6SS-5	0.39	0.39	0.92	0.92	0.92	0.92	1.00	0.99
BPSS0111	T6SS-5	0.53	0.53	0.96	0.96	0.96	0.95	1.00	1.00
BPSS0112	T6SS-5	0.65	0.66	0.99	0.99	0.99	0.99	1.00	1.00
BPSS0113	T6SS-5	0.43	0.43	0.93	0.93	0.93	0.93	1.00	1.00
BPSS0114	T6SS-5	0.37	0.36	0.90	0.90	0.90	0.90	1.00	1.00
BPSS0115	T6SS-5	0.05	0.05	0.82	0.82	0.82	0.82	1.00	0.99
BPSS0116	T6SS-5	0.79	0.79	0.92	0.92	0.92	0.92	1.00	1.00
BPSS0117	T6SS-5	0.85	0.14	0.87	0.87	0.87	0.87	1.00	1.00

Table S2. Pairwise matrix of 16S rRNA sequence percent identity and SNPs. Top half of matrix lists the number of SNPs and the bottom half of the matrix lists the percent identity of the 16S rRNA gene sequences. Species: Bma, *Burkholderia mallei*; Bp, *B. pseudomallei;* Bt, *B. thailandensis*; Bo, *B. oklahomensis*; Bh, *B. humptydooensis*; Bm, *B. mayonis* sp. nov.; Bs, *B. savannae* sp. nov.; Bsi, *B. singularis*; Bu, *B. ubonensis*.

	Bo E0147 c1	Bo E0147 c2	Bo C6786	Bu MSMB 22 c3	Bu MSMB 22 c1	Bu MSMB 22 c2	Bs BDU1 9	Bs MSMB 266 c1, 4	Bs MSMB 852 c 1-4	Bs BDU1 8 c 1-4	Bs MSMB 266 c2-3	Bsi MSMB 715	Bm BDU6 c1	Bm BDU6 c2-4	Bm BDU8 c1-3	Bm BDU8 c4	Bma ATCC 23344 c2	Bma ATCC 23344 c1	Bp K9624 3 c1	Bp K9624 3 c2	Bh MSMB 43 c2	Bt E264	Bh MSMB 43 c1
Bo E0147 c1		5	2	33	31	32	22	22	22	23	23	30	24	23	21	22	25	24	24	23	26	23	27
Bo E0147 c2	99.7		3	31	29	30	20	20	20	21	21	30	24	23	19	20	24	23	23	22	25	21	25
Bo C6786	99.9	99.8		31	29	30	20	20	20	21	21	28	24	23	19	20	23	22	22	21	24	21	25
Bu MSMB22 c3	97.8	98.0	98.0		2	3	22	24	24	23	23	29	31	32	34	35	27	26	24	25	21	26	26
Bu MSMB22 c1	98.0	98.1	98.1	99.9		1	20	22	22	21	21	27	29	30	32	33	25	24	22	23	19	24	24
Bu MSMB22 c2	97.9	98.0	98.0	99.8	99.9		21	23	23	22	22	28	30	31	33	34	26	25	23	24	20	25	25
Bs BDU19	98.6	98.7	98.7	98.6	98.7	98.6		4	4	3	3	23	16	17	20	21	23	22	20	21	22	18	22
Bs MSMB266 c1, 4	98.6	98.7	98.7	98.4	98.6	98.5	99.7		2	1	1	21	20	21	16	17	25	24	22	23	24	20	24
Bs MSMB852 c 1-4	98.6	98.7	98.7	98.4	98.6	98.5	99.7	99.9		1	1	21	20	21	16	17	25	24	22	23	24	20	24
Bs BDU18 c 1-4	98.5	98.6	98.6	98.5	98.6	98.6	99.8	99.9	99.9		0	20	19	20	17	18	24	23	21	22	23	19	23
Bs MSMB266 c2-3	98.5	98.6	98.6	98.5	98.6	98.6	99.8	99.9	99.9	100		20	19	20	17	18	24	23	21	22	23	19	23
Bsi MSMB715	98.0	98.0	98.2	98.1	98.2	98.2	98.5	98.6	98.6	98.7	98.7		27	28	26	27	17	16	14	15	31	22	27
Bm BDU6 c1	98.4	98.4	98.4	98.0	98.1	98.0	99.0	98.7	98.7	98.8	98.8	98.2		1	12	13	20	19	17	18	29	14	19
Bm BDU6 c2-4	98.5	98.5	98.5	97.9	98.0	98.0	98.9	98.6	98.6	98.7	98.7	98.2	99.9		11	12	19	18	18	17	28	13	18
Bm BDU8 c1-3	98.6	98.8	98.8	97.8	97.9	97.8	98.7	99.0	99.0	98.9	98.9	98.3	99.2	99.3		1	22	21	21	20	27	12	19
Bm BDU8 c4	98.6	98.7	98.7	97.7	97.8	97.8	98.6	98.9	98.9	98.8	98.8	98.2	99.1	99.2	99.9		23	22	22	21	28	13	20
Bma ATCC23344 c2	98.4	98.4	98.5	98.2	98.4	98.3	98.5	98.4	98.4	98.4	98.4	98.9	98.7	98.8	98.6	98.5		1	3	2	26	14	17
Bma ATCC23344 c1	98.4	98.5	98.6	98.3	98.4	98.4	98.6	98.4	98.4	98.5	98.5	99.0	98.8	98.8	98.6	98.6	99.9		2	1	25	13	16
Bp K96243 c1	98.4	98.5	98.6	98.4	98.6	98.5	98.7	98.6	98.6	98.6	98.6	99.1	98.9	98.8	98.6	98.6	99.8	99.9		1	25	13	16
Bp K96243 c2	98.5	98.6	98.6	98.4	98.5	98.4	98.6	98.5	98.5	98.6	98.6	99.0	98.8	98.9	98.7	98.6	99.9	99.9	99.9		24	12	15
Bh MSMB43 c2	98.3	98.4	98.4	98.6	98.8	98.7	98.6	98.4	98.4	98.5	98.5	98.0	98.1	98.2	98.2	98.2	98.3	98.4	98.4	98.4		15	12
Bt E264	98.5	98.6	98.6	98.3	98.4	98.4	98.8	98.7	98.7	98.8	98.8	98.6	99.1	99.1	99.2	99.1	99.1	99.1	99.1	99.2	99.0		7
Bh MSMB43 c1	98.2	98.4	98.4	98.3	98.4	98.4	98.6	98.4	98.4	98.5	98.5	98.2	98.8	98.8	98.8	98.7	98.9	99.0	99.0	99.0	99.2	99.5	

Table S3. List of the 66 assemblies used to construct the WGS phylogeny for Figure 1.

NCBI assembly accession number	***Burkholderia species***	**Strain ID**	**Assembly type**
GCA_001522635.2	*B. mayonis* sp. nov.	BDU8	Complete
GCA_001523745.2	*B. mayonis* sp. nov.	BDU6	Complete
GCA_001462435.1	*B. humptydooensis*	MSMB122	Chromosome
GCA_002888035.1	*B. humptydooensis*	MSMB121	Contig
GCA_002888095.1	*B. humptydooensis*	MSMB122	Contig
GCA_000170355.1	*B. oklahomensis*	EO147	Contig
GCA_000755985.1	*B. oklahomensis*	EO147	Complete
GCA_001522105.2	*B. oklahomensis*	EO147	Complete
GCA_000170375.1	*B. oklahomensis*	C6786	Contig
GCA_001522135.2	*B. oklahomensis*	C6786	Complete
GCA_000959365.1	*B. oklahomensis*	C6786	Complete
GCA_900608545.1	*B. oklahomensis*	LMG_23618	Contig
GCA_900446275.1	*B. oklahomensis*	NCTC13388	Contig
GCA_000011545.1	*B. pseudomallei*	K96243	Complete
GCA_001524445.2	*B. savannae* sp. nov.	MSMB0266	Complete
GCA_001546955.1	*B. savannae* sp. nov.	BDU19	Contig
GCA_001546915.1	*B. savannae* sp. nov.	BDU18	Contig
GCA_001524625.2	*B. savannae* sp. nov.	MSMB0852	Complete
GCA_001523725.1	*B. singularis*	TSV85	Contig
GCA_900176645.1	*B. singularis*	LMG_28154	Contig
GCA_000012365.1	*B. thailandensis*	E264	Complete
GCA_000567925.1	*B. thailandensis*	2002721723	Complete
GCA_002588555.1	*B. thailandensis*	FDAARGOS_425	Contig
GCA_002888235.1	*B. thailandensis*	2002721723	Contig
GCA_000959425.1	*B. thailandensis*	2002721643	Complete
GCA_002891195.1	*B. thailandensis*	FDAARGOS_240	Contig
GCA_002891115.1	*B. thailandensis*	FDAARGOS_236	Contig
GCA_003568605.1	*B. thailandensis*	E264	Complete
GCA_000152285.1	*B. thailandensis*	E264	Chromosome
GCA_002888315.1	*B. thailandensis*	2002721643	Contig
GCA_000567945.1	*B. thailandensis*	E444	Complete
GCA_002888205.1	*B. thailandensis*	E0444	Contig
GCA_003614835.1	*B. thailandensis*	FDAARGOS	Complete
GCA_000706745.1	*B. thailandensis*	USAMRU_Malaysia20	Complete
GCA_001555555.1	*B. thailandensis*	Malaysia20	Contig
GCA_003614815.1	*B. thailandensis*	FDAARGOS_241	Complete
GCA_000877335.1	*B. thailandensis*	Phuket4W1	Scaffold

GCA_002888355.1	*B. thailandensis*	Phuket4W1	Contig
GCA_003020025.1	*B. thailandensis*	FDAARGOS_237	Complete
GCA_000765375.1	*B. thailandensis*	E254	Complete
GCA_002891095.1	*B. thailandensis*	FDAARGOS_244	Contig
GCA_002888335.1	*B. thailandensis*	E254	Contig
GCA_000170395.1	*B. thailandensis*	Bt4	Contig
GCA_000764595.1	*B. thailandensis*	MSMB59	Complete
GCA_001718595.1	*B. thailandensis*	MSMB59	Complete
GCA_001524345.1	*B. thailandensis*	MSMB60	Contig
GCA_001524325.1	*B. thailandensis*	E1	Contig
GCA_000170315.1	*B. thailandensis*	TXDOH	Contig
GCA_000808035.2	*B. thailandensis*	2003015869	Complete
GCA_001718295.1	*B. thailandensis*	2003015869	Complete
GCA_002888425.1	*B. thailandensis*	TXDOH	Contig
GCA_000179515.1	*B. thailandensis*	E555	Contig
GCA_002803565.1	*B. thailandensis*	2_1	Contig
GCA_000567905.1	*B. thailandensis*	H0587	Complete
GCA_002888395.1	*B. thailandensis*	H0587	Contig
GCA_003020185.1	*B. thailandensis*	FDAARGOS_238	Complete
GCA_001718615.1	*B. thailandensis*	2002721121	Complete
GCA_000959605.1	*B. thailandensis*	34	Complete
GCA_002891155.1	*B. thailandensis*	FDAARGOS_243	Contig
GCA_001718635.1	*B. thailandensis*	2002721643	Complete
GCA_001555485.1	*B. thailandensis*	2002721621	Contig
GCA_000170495.1	*B. thailandensis*	MSMB43	Contig
GCA_000266985.1	*B. thailandensis*	MSMB43	Scaffold
GCA_000385525.1	*B. thailandensis*	MSMB121	Complete
GCA_002386205.1	*B. thailandensis*	FDAARGOS_426	Complete
GCA_000959245.1	*B. ubonensis*	MSMB22	Complete

Supplemental Information References

1. Mellmann A, Cloud J, Maier T, Keckevoet U, Ramminger I, Iwen P, Dunn J, Hall G, Wilson D, LaSala P, Kostrzewa M, Harmsen D. 2008. Evaluation of matrix-assisted laser desorption ionization-time-of-flight mass spectrometry in comparison to 16S rRNA gene sequencing for species identification of nonfermenting bacteria. Journal of Clinical Microbiology 46:1946-1954.

2. Sahl JW, Vazquez AJ, Hall CM, Busch JD, Tuanyok A, Mayo M, Schupp JM, Lummis M, Pearson T, Shippy K, Colman RE, Allender CJ, Theobald V, Sarovich DS, Price EP, Hutcheson A, Korlach J, LiPuma JJ, Ladner J, Lovett S, Koroleva G, Palacios G, Limmathurotsakul D, Wuthiekanun V, Wongsuwan G, Currie BJ, Keim P, Wagner DM. 2016. The Effects of Signal Erosion and Core Genome Reduction on the Identification of Diagnostic Markers. mBio 7:e00846-16. https://doi.org/10.1128/mBio.00846-16.