scientific reports

OPEN



Analysis of fine-scale phylogeny of *Burkholderia pseudomallei* in relation to regional geography and drug susceptibility in Thailand

Yothin Hinwan^{1,2}, Pisit Chareonsudjai^{2,3}, Pipat Reungsang^{2,4}, Ratthaphol Kraiklang^{2,5}, Ploenchan Chetchotisakd^{2,6}, Sorujsiri Chareonsudjai^{1,2}, Auttawit Sirichoat^{1,2}, Arnone Nithichanon^{1,2}, Lumyai Wonglakorn⁷, Rasana W. Sermswan⁸, David Blair⁹ & Kiatichai Faksri^{1,2}

Melioidosis caused by *Burkholderia pseudomallei* (*Bp*) is a public health threat. Genomic-epidemiology research on this deadly disease is scarce. We investigated whole-genome sequences of *Bp* isolates in relation to environmental source and drug susceptibility. In total, 563 *Bp* isolates were collected from 11 Northeast Thai provinces during the period 2004–2021. Patients (n = 530 isolates), infected animals (n = 8), and environmental sources (n = 25) provided samples. Phylogenetic analysis revealed genetic diversity among the *Bp* isolates, including numerous well-supported clusters of varying sizes. Through in-depth analysis of 38 monophyletic clades (MCs), we found eleven associated with province of origin (*p*-value < 0.001). Closely related clusters (CRCs) within MCs resembled MLST-identified "sequence types" (STs). We found 102 known and 52 novel STs. ST-70 was the most prevalent in this area (n = 78; 13.85%). Sample type (human/environmental) and sampling time intervals were not correlated with genetic distance among clonal *Bp* isolates. Some members of 12 CRCs had acquired resistance to co-trimoxazole and one against amoxicillin-clavulanic acid. Within Northeast Thailand, there is an association between *Bp* genotype and geographical origin.

Keywords Next-generation sequencing, Whole-genome sequencing, *Burkholderia pseudomallei*, Genetic diversity, Drug resistance, Phylogenetic analysis

Melioidosis is a significant public health concern, particularly in the northeastern region of Thailand where it is endemic^{1,2}. Infection with *Burkholderia pseudomallei* (*Bp*) causes this disease. In Thailand, approximately 4000–5000 patients are diagnosed annually with melioidosis³. According to a 2016 report by the Thai Bureau of Epidemiology, melioidosis has the highest mortality rate (up to 40%) of all infectious diseases in Thailand, including dengue, malaria, and leptospirosis^{3,4}. *Burkholderia pseudomallei* is a saprophytic bacterium that resides in the environment^{1,2,5}, particularly on the surface of soil, or at depths of up to 10 cm, as well as in water³. Infection can be acquired via the respiratory system, the epidermis, and the digestive system and may involve any organ, including the skin, liver, lungs and even the bloodstream. Patients may have a fever and acute pulmonary symptoms. The United States Centers for Disease Control and Prevention (CDC) regards this pathogen as a potential biological weapon². Despite its severity and the fact that the disease is preventable, melioidosis has been neglected by health policy makers^{6,7}. As a consequence, epidemiological and drug-resistance data for *Bp*

¹Department of Microbiology, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand. ²Research and Diagnostic Center for Emerging Infectious Diseases (RCEID), Khon Kaen University, Khon Kaen, Thailand. ³Department of Environmental Science, Faculty of Science, Khon Kaen University, Khon Kaen, Thailand. ⁴Department of Computer Science, College of Computing, Khon Kaen University, Khon Kaen, Thailand. ⁵Faculty of Public Health, Nutrition for Health Program, Khon Kaen University, Khon Kaen, Thailand. ⁶Department of Medicine, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand. ⁸Department of Biochemistry, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand. ⁸Department of Biochemistry, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand. ⁸Department of Biochemistry, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand. ⁸Department of Biochemistry, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand. ⁸Department of Biochemistry, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand. ⁸Department of Biochemistry, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand. ⁸Department of Biochemistry, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand. ⁹College of Science and Engineering, James Cook University, Townsville, Australia. [⊠]email: kiatichai@ kku.ac.th

are limited, affecting control and treatment efforts. Approximately 6.7–8.3 million USD per year is consumed for melioidosis treatment and management in Thailand alone⁴.

Humans most at risk of infection with *Bp* are those whose occupations expose them to soil and water⁵ and notably agricultural workers, especially farmers⁸. Northeast Thailand is a predominantly agricultural region. Nonetheless, there is little information regarding the prevalence and distribution of *Bp* in this region. Moreover, *Bp* has intrinsic resistance to many kinds of antibiotics including penicillin, ampicillin, and first- and second-generation cephalosporins, macrolides, rifamycin, colistin, and aminoglycoside. The drugs suitable for patient treatment in Thailand are few but include ceftazidime (CAZ) which is the first-line drug, meropenem (MEM), imipenem (IPM), trimethoprim-sulfamethoxazole or co-trimoxazole (SXT), and amoxicillin-clavulanic acid (AMC)^{3,9,10}.

Whole-genome sequencing (WGS) is a high-resolution tool that can generate comprehensive genomic data with application to numerous research questions. WGS analysis has been used to trace the origin of Bp infection from patient samples and environmental sources¹¹, document genetic diversity¹², and investigate the mechanisms of drug resistance⁹. WGS data, that might throw light on the relationship between the bacterium and its geography, patterns of drug resistance and site of infection, are few in Thailand^{13,14}. This study aims to investigate the genomic diversity and relationships among Bp isolates from human, animal, and environmental samples collected in Northeast Thailand. The association between each sample and its source (human/animal/environmental), province of origin, and drug-resistance status were investigated using WGS. A phylogenetic tree was constructed and used as the basis for further analyses. We examined the relationship between drug susceptibility of an isolate and its clustering position in the tree.

Results

Characteristics of the samples

Of the 563 *Bp* isolates included in this study, 530 (94.14%) were from human cases, 8 (1.42%) from animals, 3 (0.53%) from water and 22 (3.91%) from soil samples. These samples were collected between 2004 and 2021, covering eleven of the twenty provinces in Northeast Thailand (Fig. 1). Khon Kaen Province yielded the most samples (n = 295, 52.77%), followed by Nong Khai (n = 58, 10.38%) and Nakhon Ratchasima (n = 48, 8.55%). Among the 563 isolates, a total of 102 known sequence types (STs) were found (Supplementary Table 1) as well as 52 novel STs (Supplementary Table 2). The most prevalent ST was ST-70 (n = 78 isolates), followed by ST-10 (n = 31) and ST-34 (n = 23).



Fig. 1. Geographical location of the study region in Northeast Thailand and collection sites of *Burkholderia pseudomallei* isolates (n = 563) used. (A) Location of Northeast Thailand, indicating provinces from which samples were obtained and (B) expanded view of the study region with number of samples from each province indicated by level of color shading. Numbers in the figure represent the name of each province; 1 Nong Khai, 2 Nakhon Phanom, 3 Mukdahan, 4 Amnat Charoen, 5 Ubon Ratchathani, 6 Sisaket, 7 Buri Ram, 8 Nakhon Ratchasima, 9 Chaiyaphum, 10 Khon Kaen, and 11 Maha Sarakham.

Types of samples (human/environmental) and sampling time intervals were not associated with genetic distances among *Bp* isolates within individual patients or soil collection sites

The SNP-distance analysis was done using the isolates of *Bp* taken serially from individual patients or soil sites as same-clone controls. There were 35 control sets (94 isolates) of which 31 (83 isolates) were from patients and 4 (11 isolates) were from soil samples. The number of isolates in each series ranged from two to nine. Pairwise difference among members of a same-clone set never exceeded six SNPs (of the 122,010 SNPs used in the phylogenetic analysis) (Supplementary Tables 3–4). We therefore set six as the maximum allowed number of differences in assigning isolates (other than the control sets) to closely related clusters (CRCs). There was no significant difference in SNP variation within each control set when sampling interval and genetic distance were evaluated (p-value = 0.7224), and type (human subdivided into anatomical/pathological source of sample, soil) of sample (p-value = 0.5673) (Fig. 2A,B).

Geographical association of Bp clusters

The phylogenetic tree is shown in Fig. 3 and illustrates the substantial diversity among genomes of the sampled isolates.

Approximately one-third of the *Bp* isolates (n = 198/563, 35.17%) fell into 38 monophyletic clades (MCs) based on members differing by a maximum of 76 SNPs (the SNP cutoff that was significantly different relative to the total dataset; *p*-value < 0.001). In addition, all MCs were supported by bootstrap values of 100%, with the exception of one, which had 95% support. Each MC contained from 2 to 18 members. Most MCs contained isolates of a single sequence type. Three MCs (indicated by orange arrows in Fig. 4A) were only found in blood samples (*p*-value = 0.0001). We have also identified 11 MCs associated with particular provinces (Fig. 4, panels B-L) (*p*-value < 0.001).

Excluding the 94 serial same-clone control isolates there was a total of 68 closely related clusters (CRCs) (containing 175/469 isolates, 47.78%) nested within 22 of the 38 MCs (based on a maximum of six SNPs difference among isolates). Thus, it is possible that many of these CRCs represent a single clone sampled several times and at time intervals ranging from 6 days to 17 years. Each CRC included from 2 to 10 isolates (Fig. 5) and all isolates within a CRC shared the same sequence type (ST). Two sets of serial controls formed a single CRC (CRC55, purple). Nearly half of the CRCs (32/68 CRCs, 72/469 isolates) showed a significant geographical association, members of each all originating from the same province (*p*-value = 0.0001).



Fig. 2. Analysis of types (sources) of samples plotted against sampling time intervals within individual patients, soil collection sites and closely related clusters (CRCs) of *Burkholderia pseudomallei*. (**A**-**B**) Box plots represent genetic distance, sampling time intervals, and types of samples (animal, human subdivided into anatomical/ pathological source of sample, soil, water). Analysis of the same type (**A**) or different type of samples (**B**) of the same-clone control isolates (35 sets comprising 94 isolates). There was no significant association between the SNP differences and sample type or time interval between serial isolates. (**C**) Box plots represent genetic distance and sampling time intervals within CRCs: again, no significant association between these was found.



Fig. 3. Phylogenetic analysis of *Burkholderia pseudomallei* isolates. Phylogenetic tree of 563 samples showing province of origin, source (human/animal/environmental), type/source of specimens (animal, human subdivided into anatomical/pathological source of sample, soil, water) and results of drug-resistance tests (where applicable) for each isolate. *Burkholderia thailandensis* was used as an outgroup (SRR1609233). All branch lengths are proportional to the corresponding SNP distances.

There were 18/68 CRCs (39/469 isolates, 8.32%) in which all members were exclusively found from a single type of sample (type of sample includes source and/or infection site). Ten of these CRCs (n = 10/18; 55.56%) contained only isolates from blood samples. These ten CRCs (23 isolates) were significantly different in terms of SNP distances from isolates from blood elsewhere on the tree (n = 88) (p-value = 0.0011). A further eight CRCs contained only isolates from a single type of sample, including pus (n = 5/18, 27.79%), sputum (n = 1/18, 5.55%), body fluid (n = 1/18, 5.55%) and animal (n = 1/18, 5.55%).

We also analyzed the genetic distance and time interval of samples among CRCs. Regardless of sampling intervals between isolates within a CRC (up to 17 years, but mostly within a single year), pairwise differences never exceeded 6 SNPs (p-value = 0.2995) (Fig. 2C and Supplementary table 6).

Phenotypic drug resistance and genetic diversity of Bp strains

Out of the 563 isolates, at least one DST result was available for 218 isolates (38.72%). In four MCs, all members were resistant to a specific drug. Three of these MCs (n = 6 isolates) were associated with SXT resistance (MC2, MC10 and MC32) (Fig. 4, panels A-L). However, no significant association (p-value = 0.2881) was observed overall between MCs and DST patterns.

Among the 17 CRCs with SXT-resistant members, all tested members of four (n = 10 isolates from CRC37, CRC53, CRC54 and CRC69) were resistant to SXT. Additionally, two CRCs (two isolates each from CRC3 and CRC44) showed resistance to CAZ (Fig. 5). However, no significant association (p-value = 0.2267) was observed overall between CRCs and DST patterns. DST analysis of CRCs revealed 12 CRCs with acquired resistance; 11 CRCs for SXT and one CRC for AMC (Supplementary table 6).

Analysis of STs in relation to DSTs revealed that all members (n = 56 isolates) representing 31 STs were resistant to trimethoprim-sulfamethoxazole (SXT) (Supplementary table 7). One ST, ST-177 (n = 2 isolates), consisted of *Bp* strains resistant to ceftazidime (CAZ) but to no other drug. Because of the rather limited number of STs in which drug resistance was found, we were unable to reach strong conclusions about any association between ST and resistance (Table 1 and Supplementary table 7).





Among the 11 provinces sampled, AMC resistance was mostly found in Buri Ram (n = 5 isolates; 100% of isolates found to have resistance to any drug in that province), CAZ resistance in Ubon Ratchathani (n = 4; 100%), SXT resistance in Nong Khai (n = 58; 49.15%). No isolates resistant to IPM were found. For seven provinces, there were significant correlations between geographical location (province) and drug resistance (*p*-value < 0.001) (Supplementary table 5).

Discussion

The SNP-based phylogenetic tree of all isolates demonstrated a high genetic diversity, with a total of 154 STs in our studied population. We used highly stringent analysis to obtain 122,010 high-confidence SNPs from the 563 Thai Bp strains, compared with 320,000 SNPs from 1654 strains, or 469 strains collected among countries in other studies^{15,16}. The lower number of SNPs in our study was supported by the internal controls of serial isolates from the same patient. The higher number of SNPs found by previous studies might be explained by the greater number of Bp strains used or by application of lower stringency in identifying SNPs.

We analyzed *Bp* control strains derived from the same clone isolated serially from the same patient or the same soil site (50-m radius). The type of sample and the sampling interval (up to 16 days following the first sampling) were not associated with the number of SNPs differing among serial samples. This indicates that the genome within a clone of *Bp* is stable for at least a few weeks within a single patient or a small area of soil surface. The SNP cutoff values (≤ 6) obtained from the same-clone controls therefore give us confidence that our CRCs represent clonal lineages. Apart from serial control samples, we also analyzed the collection time intervals of isolates in the same CRC, ranging from 6 days to 17 years. The number of SNP differences was not associated with the sampling time interval in the CRCs. Although most isolates within a CRC were collected within the same year, some had members that were isolated more than 3 years apart. This might be explained by CRC members originating as independent human infections from the same environmental clone at different times. Clones persisting in the



Fig. 5. Distribution of closely related clusters (CRCs). (**A**) A total of 103 CRCs are identified by colored radial bars, (**B–K**) Geographical distribution of each CRC. In our study, 31 same-clone serial control sets (the same clone sampled from patients at different time points) comprising 83 samples were used; alternating (for visual contrast) dark- and light-blue bars indicate clusters exclusively containing serial same-clone controls, similarly, dark- and light-purple bars indicate serial same-clone controls along with other *Bp* isolates, the pink bar indicates a mixture of serial same-clone controls and a CRC (CRC55), red and light-red bars indicate sample CRCs, and the gray bar indicates a CRC (CRC54 in **G**) sampled in more than one province. All branch lengths are proportional to the corresponding SNP distances.

Scientific Reports | (2024) 14:19961 |

Factors	AMC	CAZ	IPM	SXT
Available DST data	214	218	139	214
	(S=209, R=5)*	(S=214, R=4)	(All susceptible)	(S=96, R=118)
Top 5 STs found only in susceptible isolates	ST512 (n = 15)	ST70 (n=23)	ST70 (n=17)	ST34 (n=7)
	ST10 (n = 13)	ST512 (n=15)	ST512 (n=15)	ST492 (n=4)
	ST54 (n = 8)	ST306 (n=14)	ST306 (n=10)	ST298 (n=3)
	ST34 (n = 7)	ST10 (n=13)	ST54 (n=7)	ST498 (n=2)
	ST167 (n = 7)	ST54 (n=8)	ST34 (n=7)	ST500 (n=2)
Top 5 STs found only in resistant isolates	No STs found	Only ST177 (n = 2)	No STs found	ST167 (n=7) ST93 (n=5) ST520 (n=3) ST203 (n=3) ST511 (n=3)

Table 1. Drug-susceptibility test results and sequence type (ST) of relevant isolates. *S sensitive and *R* resistant.

environment are likely to accumulate mutations slowly because of the lack of strong selection pressure in their typical environment, coupled with low replication rates. This view is supported by a previous study reporting that two environmental isolates collected 17 years apart had only 2 SNP differences¹². Another study reported identical *Bp* strains collected 8 years apart¹⁷.

Our analysis using many SNPs yielded far better phylogenetic resolution than would a tree constructed only using ST data. Nevertheless, each CRC in our study represented a single ST, but a particular ST could be shared with other CRCs. ST analysis is the most common approach used to determine and describe strains of $Bp^{18,19}$, so we will focus on STs for comparison with other studies. ST-70 was the most common ST in Northeast Thailand (n = 78 isolates), followed by ST-10 (n = 31) and ST-34 (n = 23). This result is consistent with high levels of diversity reported in previous studies, with 35 different STs among 135 northern Australian isolates¹⁸ and 43 STs among 84 clinical isolates in Sri Lanka²⁰. The most prevalent STs reported from Australia and Sri Lanka were ST-36 and ST-1137, respectively, whereas in our study, ST-70 predominated. This finding indicates that genetic variants of Bp can be associated with the geographical location. There are 2,066 STs listed in the MLST database for Bp^{21} , and we found an additional 52 novel STs in this study, indicating the high genetic diversity of Bp in our region.

From the phylogenetic analysis, 38 MCs were found. Thirteen MCs were significantly associated with single provinces, and one was associated only with soil isolates. Previous studies have reported that the same strains of Bp (based on WGS and ST analysis) can be found within a radius of 50 km, especially in flood-prone areas^{15,18}. The fact that we found 13 MCs that were significantly associated with single provinces could in part be due to the low occurrence of flooding in our region¹⁸.

Cluster analysis of bacterial isolates in this study identified 103 distinct CRCs comprising 269 isolates. Almost half of these CRCs contained isolates from the same province, predominantly originating from human infections (Fig. 5). This suggests a connection between clusters of human infections and local environmental strains, as human-to-human transmission has not been confirmed. However, transmission from sources other than the environment cannot be ruled out. Intriguingly, 36 CRCs included isolates from more than one province, potentially explained by the dispersal of *Bp* clones through mechanical means (e.g. dispersal along water courses, transportation of plants and soil for agricultural purposes or on birds' feet), leading to infections by the same clone in patients from different provinces. Furthermore, workers carrying the disease may migrate between provinces. Migration is common in Northeast Thailand, where agricultural workers grow rice during the wet season (an activity that can lead to infection) and then work in other provinces for the rest of the year. Previous studies^{15,22,23} have reported geographically specific alleles and STs associated with specific regions, consistent with our findings. While clonal infection by *Bp* among animals in the same zoo has been documented²⁴, no such clonal infection has been reported among humans in the same area. This analysis emphasizes the clustering of *Bp* infection among humans across the provinces of Northeast Thailand.

Ten CRCs (23 isolates) were obtained only from blood samples via blood culture from sepsis-diagnosed patients. This was highly significant (*p*-value = 0.0011) when compared with the non-clustered (non-CRC) blood isolates, suggesting a possible association between certain specimen types and certain *Bp* clusters. However, *Bp* infection typically disseminates throughout the body^{3,25}, so there is no a priori reason to expect such an association. Clearly, further studies are required. Other confounding factors, such as underlying comorbidities and treatment complications, may also have contributed to the severity of the patients' illness and thus we should interpret our results with caution.

Burkholderia pseudomallei is naturally resistant to numerous antibiotics, resulting in few options for effective medication⁹. We examined DST results for *Bp* isolates in relation to ST, MC, and CRC. ST is the strain-identification method usually used for outbreak tracking or transmission tracking^{17,25}. There are studies correlating ST with drug resistance in other bacterial species, for instance in *Streptococcus suis*²⁶, *Klebsiella pneumoniae*²⁷, and *Escherichia coli*²⁸, but such analyses have not been performed with *Bp*. Here, we found all isolates representing some STs of *Bp* were resistant to co-trimoxazole and ceftazidime. Some STs (Supplementary Table 7) showed significant association. Furthermore, we found geographical association of *Bp* with drug resistance in seven out of eleven provinces. On one hand, this result might indicate the emergence of drug resistance associated with a particular province which may be related to clinical or agricultural factors. An example of this has been seen in *Mycobacterium tuberculosis*, with transmission of drug-resistant strains associated with specific geographic regions^{29,30}. On the other hand, the apparent association may be due to

sampling bias because we gave preference to drug-resistant isolates when selecting samples. The relationships among geography, ST and drug resistance remain unclear and require further study.

We identified only four MCs associated with drug-resistant isolates: three associated with SXT resistance and one with AMC resistance. Among these MCs, only four CRCs were associated with SXT resistance and two CRCs with resistance to CAZ. Different members of some CRCs had different drug resistance-associated mutations. This indicates parallel, independent acquisition of drug resistance by different isolates (termed acquired resistance). This was seen in eleven CRCs, which exhibited a range of phenotypic drug susceptibility for SXT; CRC5, CRC8, CRC18, CRC25, CRC43, CRC68, CRC82, CRC83, CRC85, CRC89, and CRC103 and only CRC54 for AMC (Supplementary Table 6). Thus, isolates derived from a single clone can differ in drug susceptibility. This might arise as a consequence of the use of antibiotics in agriculture³¹. Alternatively, acquired resistance might occur due to long-term antibiotic treatment in a melioidosis patient³. The scenario of acquired drug resistance needs to be further elucidated for the better control and surveillance of emerging drug-resistant *Bp*.

Limitations of our study should be noted. We analyzed the association between each Bp strain and its source, province of origin, and drug-resistance status. Our Bp strains covered only eleven provinces out of twenty provinces in Northeast Thailand. We focused on Bp isolates with known resistance from human patients, resulting in a higher proportion of drug-resistant isolates, particularly for co-trimoxazole, compared to some other studies^{32,33}. This may have impacted our interpretations. The limited number of animal samples relative to clinical and soil samples hindered association analysis. A more comprehensive approach, incorporating additional samples from animal Bp strains, environmental sources, and clinical cases across all provinces in Northeast Thailand, would offer a more comprehensive understanding of Bp epidemiology in the region. The absence of detailed, documented epidemiological links among patients is a significant limitation in elucidating socio-economic associations with Bp genetic clusters. To overcome this limitation and facilitate disease control and patient management, a fully controlled study design is recommended for a clearer interpretation.

In conclusion, we have reported phylogenomic relationships among *Bp* isolates from human and environmental samples in Northeast Thailand based on SNPs and STs. We identified MCs associated with particular provinces. Fifty-two novel STs were found in the region. We also reported that members of some CRCs exhibited acquired resistance to co-trimoxazole and amoxicillin-clavulanic acid. No association was found between the time interval separating collection of isolates and genetic distance among isolates.

Materials and methods

Bacterial populations, phenotypic identification, and drug-susceptibility testing

Bacteria were isolated from human (n=530), animal (n=8), water (n=3) and soil (n=22) samples. Of the human samples, 111 were from patients at Srinagarind Hospital, Khon Kaen, Thailand collected during 2020 and 2021. A further 419 human samples were stock cultures collected during regional surveillance from 11 provinces in Northeast Thailand in years spanning 2004–2021, with numbers from each province proportional to reported incidence of melioidosis there. We gave preference to isolates exhibiting resistance to any drug.

Sequential sampling from the same sources yielded a total of 94 isolates (from 31 individual humans and four soil sites). Note that these isolates were included in the grand total of 563 isolates. Genomic differences within a single clone over time were estimated using data from these 35 datasets and these are termed the "same-clone controls". Sequential soil samples were taken from the same rice field within a 50-m radius.

Unprocessed soil samples were plated on Ashdown's agar and grown in Luria–Bertani (LB) broth. Phenotypic identification of *Bp* was confirmed by positive latex-agglutination results. Drug-susceptibility testing employed broth microdilution following protocols in CLSI m45 3rd edition³⁴. Briefly, fresh colonies were taken after 24 h of incubation at 37 °C on blood agar with 5% sheep blood (Clinag Co., Limited, Thailand), resuspended in normal saline, and adjusted to 0.5 McFarland. This suspension was then added to 10 μ L of cation-adjusted Mueller–Hinton broth (Clinag Co., Limited, Thailand) and mixed thoroughly. An automated inoculation machine (Sensititre AIM, Thermo Scientific, USA) dispensed 50 μ L of the mixture into a 96-well plate (THAN1F, Sensititre TM, Thermo Fisher Scientific, USA), which also automatically read the plates and interpreted the drug susceptibility test.

This study used leftover specimens without any information that could lead to the identification of any participants; no informed consent was required. All the above processes were conducted at the biosafety level 2 enhanced laboratory (BSL-2 enhanced) and were carried out under the Pathogens and Animal Toxin Act, Section 18, 2018, Ministry of Public Health, Thailand. Based on the Declaration of Helsinki and the ICH Good Clinical Practice Guidelines, the Khon Kaen University Ethics Committee for Human Research approved the study's protocol, including the waiver of informed consent (approval no. HE641201).

DNA extraction

From LB broth, the samples were cultured on Ashdown's agar. Two or three loopfuls of each colony were then collected into a 16×150 mm tube with a few drops of Tris–EDTA (TE) buffer and 6–8 sterile 5 mm glass beads. The cultivation and colony collection were conducted at BSL-2 enhanced. The extraction was performed using the cetyl-trimethyl-ammonium bromide-sodium chloride method³⁵. The DNA pellet was dried at room temperature overnight, then dissolved in 50 µL of TE buffer. DNA concentration was measured using a Nanodrop instrument (Thermo Scientific[™] NanoDrop[™]) and DNA stored at – 20 to – 80 °C until sequencing was done.

Whole-genome sequencing and data processing

DNA from each isolate was submitted to a sequencing-service company (OMICS DRIVE NGS Laboratory, Singapore) to generate 150-bp paired-end reads using the Illumina HiSeq platform. The quality of the WGS

data was assessed using FastQC version 0.11.9³⁶ and any low-quality regions were trimmed using Trimmomatic version 0.38³⁷. The remaining high-quality sequences were then mapped to the *Bp* K96243 reference genome (chromosome 1: NC_006350.1 and chromosome 2: NC_006351.1)³⁸ using BWA-MEM version 0.7.12³⁹ and converted to BAM format and sorted using SAMtools version 1.15.1⁴⁰. Realignment of reads was performed using GATK version 3.6.0⁴¹ and variants were called using both GATK version 3.6.0 and BCFtools version 1.2⁴², focusing on single-nucleotide polymorphisms (SNPs) and insertions/deletions (indels). Variants were called with a minimum mapping quality of 50 and a base alignment quality (Q score) of 20. The called variants were then filtered based on a minimum depth of coverage of 10 and a Q score of 20. The overlapping variants from SAMtools and GATK were subjected to further analysis. Mpile-up files were generated using SAMtools and coverage files were created from these. The SAMtools mpile-up files and variant-calling files (VCF) were used to create a combined nucleotide-frequencies file to include all isolates. Then, the variant present at each SNP position was analyzed. Whole genome sequencing data (.fastq files) have been deposit in the GenBank BioProject PRJNA1051349.

Phylogenetic analysis

The analysis included a total of 122,010 SNP variants from 563 isolates, along with one outgroup isolate (*Burkholderia thailandensis* MSMB121 SRA: SRP048783). The variant set was filtered based on criteria of a minimum of fivefold coverage and a minimum frequency of read proportion at 0.75. maximum-likelihood phylogenetic trees were inferred using IQ-TREE multicore version $1.6.9^{43}$, with 1000 bootstrap replicates and the GTR + G + ASC (General Time-Reversible with Discrete Gamma and ascertainment bias correction) model^{44,45}. The final circular depiction of the phylogenetic tree based on the entire dataset was prepared using iTOL⁴⁶. Monophyletic clades (MCs) were defined as consisting of isolates differing from one another by no more than 76 SNPs. Nested within most MCs, closely related clusters (CRCs) were defined based on a maximum of 6 SNPs differing between members, identified using snp-dists version $0.8.2^{47}$. Some CRCs were not associated with an MC. The visualization of both MCs and CRCs was performed using R studio with the phytools library.

Multilocus sequence-typing analysis (MLST)

All 563 isolates were assigned an MLST genotype through in silico analysis using the Bacterial Isolates Genome Sequence database (BIGSdb) tool, which can be accessed online on the MLST website (http://pubmlst.org/bpseudomallei/)⁴⁸. The MLST tool was used to scan contig files against traditional PubMLST typing schemes⁴⁹. StringMLST software was additionally used for cross validation of MLST analysis⁵⁰. To identify the sequence type (ST) of each isolate, the fastq files were submitted to StringMLST and BIGdb via (http://pubmlst.org/bpseu domallei/) online tools for ST identification.

Data and statistical analysis

Non-parametric data were analyzed using Fisher's exact tests for statistical determination in data relating to drug resistance, province of origin and relationship of STs with particular CRCs or MCs. For the same-clone control sets and within-CRC sets, comparisons were done using one-way ANOVA followed by Dunnett's multiple-comparisons test using GraphPad Prism (version 9.5.0 for Mac, GraphPad Software, San Diego, California USA).

Data availability

Whole genome sequencing data (.fastq files) have been deposit in the GenBank BioProject PRJNA1051349.

Received: 19 December 2023; Accepted: 19 August 2024 Published online: 28 August 2024

References

- 1. Palasatien, S., Lertsirivorakul, R., Royros, P., Wongratanacheewin, S. & Sermswan, R. W. Soil physicochemical properties related to the presence of *Burkholderia pseudomallei. Trans. R. Soc. Trop. Med. Hyg.* **102**(Suppl 1), S5-9 (2008).
- Wang-ngarm, S., Chareonsudjai, S. & Chareonsudjai, P. Physicochemical factors affecting the growth of *Burkholderia pseudomallei* in soil microcosm. Am. J. Trop. Med. Hyg. 90, 480–485 (2014).
- 3. Wiersinga, W. J. et al. Melioidosis. Nat. Rev. Dis. Primer 4, 17107 (2018).
- 4. Melioidosis Information. https://www.melioidosis.info/info.aspx?pageID=104&contentID=1040103#.
- 5. Dance, D. A. B. Pseudomonas pseudomallei: danger in the paddy fields. Trans. R. Soc. Trop. Med. Hyg. 85, 1-3 (1991).
 - 6. Chansrichavala, P. *et al.* Public awareness of melioidosis in Thailand and potential use of video clips as educational tools. *PLoS ONE* **10**, e0121311 (2015).
 - 7. Suntornsut, P. et al. Barriers and recommended interventions to prevent melioidosis in Northeast Thailand: A focus group study using the behaviour change wheel. PLoS Negl. Trop. Dis. 10, e0004823 (2016).
- Leelarasamee, A., Trakulsomboon, S., Kusum, M. & Dejsirilert, S. Isolation rates of Burkholderia pseudomallei among the four regions in Thailand. Southeast Asian J. Trop. Med. Public Health 28, 107–113 (1997).
- Fen, S. H. Y. *et al.* Antibiotic susceptibility of clinical Burkholderia pseudomallei isolates in northeast Thailand during 2015–2018 and the genomic characterization of β-lactam-resistant isolates. *Antimicrob. Agents Chemother. AAC* https://doi.org/10.1128/AAC. 02230-20 (2021).
- Khosravi, Y., Vellasamy, K. M., Mariappan, V., Ng, S.-L. & Vadivelu, J. Antimicrobial susceptibility and genetic characterisation of Burkholderia pseudomallei isolated from Malaysian patients. Sci. World J. 2014, e132971 (2014).
- 11. Sarovich, D. S. *et al.* Whole-genome sequencing to investigate a non-clonal melioidosis cluster on a remote Australian island. *Microb. Genomics* **3**, e000117 (2017).
- 12. Chapple, S. N. J. *et al.* Whole-genome sequencing of a quarter-century melioidosis outbreak in temperate Australia uncovers a region of low-prevalence endemicity. *Microb. Genomics* **2**, e000067 (2016).
- 13. Vuddhakul, V. et al. Epidemiology of Burkholderia pseudomallei in Thailand. Am. J. Trop. Med. Hyg. 60, 458-461 (1999).

- 14. Chierakul, W. et al. Melioidosis in 6 tsunami survivors in southern Thailand. Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am. 41, 982–990 (2005).
- 15. Chewapreecha, C. et al. Global and regional dissemination and evolution of Burkholderia pseudomallei. Nat. Microbiol. 2, 16263 (2017).
- 16. Zheng, H. *et al.* Genetic diversity and transmission patterns of *Burkholderia pseudomallei* on Hainan island, China, revealed by a population genomics analysis. *Microb. Genomics* 7, 000659 (2021).
- 17. Meumann, E. M. et al. Emergence of Burkholderia pseudomallei sequence type 562, Northern Australia. Emerg. Infect. Dis. 27, 1057–1067 (2021).
- Rachlin, A. *et al.* Whole-genome sequencing of *Burkholderia pseudomallei* from an urban melioidosis hot spot reveals a fine-scale population structure and localised spatial clustering in the environment. *Sci. Rep.* 10, 5443 (2020).
- 19. Zueter, A. R., Rahman, Z. A., Abumarzouq, M. & Harun, A. Multilocus sequence types of clinical *Burkholderia pseudomallei* isolates from peninsular Malaysia and their associations with disease outcomes. *BMC Infect. Dis.* **18**, 5 (2018).
- Jayasinghearachchi, H. S. et al. Biogeography and genetic diversity of clinical isolates of Burkholderia pseudomallei in Sri Lanka. PLoS Negl. Trop. Dis. 15, e0009917 (2021).
- Jolley, K. A., Bray, J. E. & Maiden, M. C. J. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. Wellcome Open Res. 3, 124 (2018).
- 22. Cheng, A. C. et al. Genetic diversity of Burkholderia pseudomallei isolates in Australia. J. Clin. Microbiol. 46, 249-254 (2008).
- Chapple, S. N. J. et al. Burkholderia pseudomallei genotype distribution in the Northern Territory, Australia. Am. J. Trop. Med. Hyg. 94, 68–72 (2016).
- Rachlin, A. et al. A cluster of melioidosis infections in hatchling saltwater crocodiles (Crocodylus Porosus) resolved using genomewide comparison of a common North Australian strain of Burkholderia Pseudomallei. Microb. Genomics 5, e000288 (2019).
- 25. Aziz, A. et al. Suspected cases of intracontinental Burkholderia pseudomallei sequence type homoplasy resolved using wholegenome sequencing. Microb. Genomics 3, e000139 (2017).
- Athey, T. B. T. et al. Population structure and antimicrobial resistance profiles of Streptococcus suis serotype 2 sequence type 25 strains. PLoS ONE 11, e0150908 (2016).
- Liu, L., Feng, Y., Long, H., McNally, A. & Zong, Z. Sequence type 273 carbapenem-resistant *Klebsiella pneumoniae* carrying blaNDM-1 and blaIMP-4. *Antimicrob. Agents Chemother.* 62, e00160-e218 (2018).
- 28. Pajand, O. *et al.* Arrangements of mobile genetic elements among virotype E subpopulation of *Escherichia coli* sequence type 131 strains with high antimicrobial resistance and virulence gene content. *Sphere* **6**, e0055021 (2021).
- Nonghanphithak, D. et al. Clusters of drug-resistant Mycobacterium tuberculosis detected by whole-genome sequence analysis of nationwide sample, Thailand, 2014–2017. Emerg. Infect. Dis. 27, 813–822 (2021).
- Kostyukova, I., Pasechnik, O. & Mokrousov, I. Epidemiology and drug resistance patterns of *Mycobacterium tuberculosis* in highburden area in Western Siberia, Russia. *Microorganisms* 11, 425 (2023).
- Bengtsson-Palme, J., Kristiansson, E. & Larsson, D. G. J. Environmental factors influencing the development and spread of antibiotic resistance. FEMS Microbiol. Rev. 42, fux053 (2018).
- 32. Wuthiekanun, V. et al. Trimethoprim/sulfamethoxazole resistance in clinical isolates of Burkholderia pseudomallei. J. Antimicrob. Chemother. 55, 1029–1031 (2005).
- 33. Lumbiganon, P., Tattawasatra, U., Chetchotisakd, P., Wongratanacheewin, S. & Thinkhamrop, B. Comparison between the antimicrobial susceptibility of *Burkholderia pseudomallei* to trimethoprim-sulfamethoxazole by standard disk diffusion method and by minimal inhibitory concentration determination. *J. Med. Assoc. Thail. Chotmaihet Thangphaet* 83, 856–860 (2000).
- M45 Ed3 Test Infrequently Isolated/Fastidious Bacteria. Clinical & Laboratory Standards Institute https://clsi.org/standards/produ cts/microbiology/documents/m45/.
- Larsen, M. H., Biermann, K., Tandberg, S., Hsu, T. & Jacobs, W. R. Jr. Genetic manipulation of *Mycobacterium tuberculosis. Curr.* Protoc. Microbiol. 6, 10A.2.1-10A.2.21 (2007).
- 36. Babraham Bioinformatics—FastQC A quality control tool for high throughput sequence data. https://www.bioinformatics.babra ham.ac.uk/projects/fastqc/.
- 37. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
- Holden, M. T. G. et al. Genomic plasticity of the causative agent of melioidosis, Burkholderia pseudomallei. Proc. Natl. Acad. Sci. U. S. A. 101, 14240–14245 (2004).
- Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 0 Bytes (2014) https://doi.org/10.6084/ M9.FIGSHARE.963153.V1.
- 40. Li, H. et al. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078-2079 (2009).
- McKenna, A. *et al.* The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303 (2010).
- 42. Danecek, P. et al. Twelve years of SAMtools and BCFtools. GigaScience 10, giab008 (2021).
- Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274 (2015).
- 44. Yang, Z. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. *J. Mol. Evol.* **39**, 306–314 (1994).
- 45. Lewis, P. O. A likelihood approach to estimating phylogeny from discrete morphological character data. *Syst. Biol.* **50**, 913–925 (2001).
- Letunic, I. & Bork, P. Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 49, W293–W296 (2021).
- 47. Seemann, T. Source code for snp-dists software. Zenodo https://doi.org/10.5281/zenodo.1411986 (2018).
- Jolley, K. A. & Maiden, M. C. BIGSdb: Scalable analysis of bacterial genome variation at the population level. BMC Bioinform. 11, 595 (2010).
- 49. GitHub-tseemann/mlst: Scan contig files against PubMLST typing schemes. https://github.com/tseemann/mlst.
- 50. Gupta, A., Jordan, I. K. & Rishishwar, L. stringMLST: A fast k-mer based tool for multilocus sequence typing. *Bioinformatics* 33, 119–121 (2017).

Acknowledgements

This research project was supported by the National Science, Research, and Innovation Fund (project number 4295682). This work was also supported by the National Research Council of Thailand (project numbers NRCT N41A640200 and NRC MHESI 483/2563). We would like to thank Melioidosis Research Center, Khon Kaen University for the source of the *Bp* strains used in this study.

Author contributions

Y.H.: Formal analysis, Methodology, Validation, Visualization, Writing—original draft, Writing—review & editing. P.C.: Data curation, Project administration, Resources, Writing—review & editing. P.R.: Data curation, Project administration, Resources, Software. R.K.: Methodology, Project administration, Supervision. P.C.: Methodology, Resources, Supervision, Validation. S.C.: Methodology, Supervision, Resources, Writing—review & editing. A.S.: Methodology, Project administration, Supervision, Validation. A.N.: Methodology, Project administration, Writing—review & editing. L.W.: Data curation, Methodology, Resources, Supervision, Writing—review & editing. D.B.: Data curation, Supervision, Writing—review & editing. D.B.: Data curation, Supervision, Writing—review & editing. K.F.: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing—original draft, Writing—review & editing.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/ 10.1038/s41598-024-70558-5.

Correspondence and requests for materials should be addressed to K.F.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2024