STUDY PROTOCOL

REVISED BurkHostGEN: a study protocol for evaluating variations in the *Burkholderia pseudomallei* and host genomes associated with melioidosis infection [version 2; peer review: 2 approved]

Kesorn Angchagun[1], Phumrapee Boonklang[1], Chalita Chomkatekaew[1], Sukritpong Pakdeerat[1], Gumphol Wongsuwan[1], Premjit Amornchai[1], Vanaporn Wuthiekanun[1], Salwaluk Panapipat[1], Thatsanun Ngernseng[1], Naomi Waithira[1], Steve Walton[2], Direk Limmathurotsakul[1], Anoree Surawong[3], Suwatthiya Siriboon[3], Parinya Chamnan[3], Narisara Chantratita[4], Susie Dunachie[5], Jukka Corander[2,6], Emma E. Davenport[2], Julian Knight[7], Julian Parkhill[8], Sharon J. Peacock[9], Nicholas R. Thomson[2], Nicholas P.J. Day[1,5], Claire Chewapreecha[1,2,10]

[1]Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, Salaya, Nakhon Pathom, 10400, Thailand
[2]Wellcome Sanger Institute, Hinxton, England, CB10 1SA, UK
[3]Sunpasitthiprasong Regional Hospital, Ubon Ratchathani, 34000, Thailand
[4]Department of Microbiology and Immunology, Faculty of Tropical Medicine, Mahidol University, Salaya, Nakhon Pathom, 10400, Thailand
[5]Nuffield Department of Medicine, Centre of Tropical Medicine and Global Health, University of Oxford, Oxford, England, OX3 7LG, UK
[6]Institute of Basic Medical Sciences, Faculty of Medcine, University of Oslo, Oslo, Oslo, 0317, Norway
[7]Wellcome Centre for Human Genetics, University of Oxford, Oxford, England, OX3 7BN, UK
[8]Department of Veterinary Medicine, University of Cambridge, Cambridge, England, CB3 0ES, UK
[9]Department of Medicine, University of Cambridge, Cambridge, England, CB2 0SP, UK
[10]Department of Clinical Medicine, Faculty of Tropical Medicine, Mahidol University, Salaya, Nakhon Pathom, 10400, Thailand

v2 **First published:** 17 Aug 2023, **8**:347
https://doi.org/10.12688/wellcomeopenres.19809.1
Latest published: 02 Nov 2023, **8**:347
https://doi.org/10.12688/wellcomeopenres.19809.2

Abstract

Background

Melioidosis is a frequently fatal disease caused by an environmental bacterium *Burkholderia pseudomallei*. The disease is prevalent in northeast Thailand, particularly among rice field farmers who are at

Open Peer Review

Approval Status ✓ ✓

	1	2
version 2 (revision) 02 Nov 2023	✓ view	
version 1 17 Aug 2023	?	✓

risk of bacterial exposure through contact with contaminated soil and water. However, not all exposure results in disease, and infection can manifest diverse outcomes. We postulate that genetic factors, whether from the bacterium, the host or the combination of both, may influence disease outcomes. To address this hypothesis, we aim to collect, sequence, and analyse genetic data from melioidosis patients and controls, along with isolates of *B. pseudomallei* obtained from patients. Additionally, we will study the metagenomics of the household water supply for both patients and controls, including the presence of *B. pseudomallei.*

Methods

BurkHostGEN is an ongoing observational study being conducted at Sunpasitthiprasong Hospital, Ubon Ratchathani, Thailand. We are obtaining consent from 600 melioidosis patients and 700 controls, spanning both sexes, to collect 1 mL of blood for host DNA analysis, 3 mL of blood for RNA analysis, as well as 5 L of household water supply for metagenomic analysis. Additionally, we are isolating *B. pseudomallei* from the melioidosis patients to obtain bacterial DNA. This comprehensive approach will allow us to identify *B. pseudomallei* and their paired host genetic factors associated with disease acquisition and severity. Ethical approvals have been obtained for BurkHostGEN. Host and bacterial genetic data will be uploaded to European Genome-Phenome Archive (EGA) and European Nucleotide Archive (ENA), respectively.

Conclusions

BurkHostGEN holds the potential to discover bacterial and host genetic factors associated with melioidosis infection and severity of illness. It can also support various study designs, including biomarker validation, disease pathogenesis, and epidemiological analysis not only for melioidosis but also for other infectious diseases.

Keywords
melioidosis, genomics, diagnostics, epidemiology, microbiology, bioresources

This article is included in the Mahidol Oxford Tropical Medicine Research Unit (MORU) gateway.

	1	2
	view	view

1. **Sheila Nathan**, National University of Malaysia, Bangi, Malaysia
2. **Chiranjay Mukhopadhyay**, Manipal Academy of Higher Education, Manipal, India
   Ranita Ghosh Dastidar, Kasturba Medical College, Manipal, India

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Claire Chewapreecha (claire@tropmedres.ac)

Author roles: Angchagun K: Data Curation, Investigation, Project Administration, Writing – Original Draft Preparation; **Boonklang P**: Data Curation, Formal Analysis, Investigation, Writing – Original Draft Preparation; **Chomkatekaew C**: Data Curation, Formal Analysis, Investigation; **Pakdeerat S**: Data Curation, Investigation; **Wongsuwan G**: Data Curation, Methodology; **Amornchai P**: Methodology; **Wuthiekanun V**: Methodology, Writing – Review & Editing; **Panapipat S**: Project Administration; **Ngernseng T**: Data Curation; **Waithira N**: Software; **Walton S**: Methodology; **Limmathurotsakul D**: Resources; **Surawong A**: Resources; **Siriboon S**: Resources; **Chamnan P**: Resources; **Chantratita N**: Resources; **Dunachie S**: Resources; **Corander J**: Conceptualization, Methodology, Software; **Davenport EE**: Methodology, Software; **Knight J**: Conceptualization, Methodology; **Parkhill J**: Conceptualization, Methodology; **Peacock SJ**: Conceptualization, Methodology, Writing – Review & Editing; **Thomson NR**: Conceptualization, Methodology, Resources; **Day NPJ**: Conceptualization, Resources; **Chewapreecha C**: Conceptualization, Formal Analysis, Funding Acquisition, Investigation, Supervision, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome [216457, https://doi.org/10.35802/216457, an International Intermediate Fellowship award to CC; and 220540, a Sanger Review award to EED].
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Angchagun K, Boonklang P, Chomkatekaew C *et al.* **BurkHostGEN: a study protocol for evaluating variations in the *Burkholderia pseudomallei* and host genomes associated with melioidosis infection [version 2; peer review: 2 approved]** Wellcome Open Research 2023, **8**:347 https://doi.org/10.12688/wellcomeopenres.19809.2

First published: 17 Aug 2023, **8**:347 https://doi.org/10.12688/wellcomeopenres.19809.1

REVISED **Amendments from Version 1**

In the updated version of the study protocol, we have provided a more complete rationale for collecting household water samples and explained our choice of the plate-sweep sequencing approach to capture *Burkholderia pseudomallei* populations that were previously obscured by fast-growing species. This clarifies our message about capturing previously missing data with improved study resolution. We are also grateful to the reviewers for highlighting potential sources of bias in our study design, which we have addressed. Furthermore, we have included essential information on shipping infectious samples between study sites, a step that was missing in the previous version but is now part of the protocol. Additionally, we have worked on enhancing the overall language to improve clarity and readability.

Any further responses from the reviewers can be found at the end of the article

Introduction

Melioidosis is a public health burden yet an often neglected tropical disease, causing an estimated 4.64 million disability-adjusted life-years (DALYs), primarily due to years of life lost (YLL)[1,2]. The disease is caused by *Burkholderia pseudomallei,* a Gram-negative bacterium found in soil and contaminated water in disease-endemic areas. Infections can occur through inoculation, inhalation, or ingestion of the bacterium, resulting in either subclinical or clinical infection. Fatalities are observed in 10 to 40% of cases[2]. Epidemiological studies conducted in northeast Thailand, an area endemic for melioidosis, has provided valuable insights into the disease. The reported prevalence of culture-confirmed clinical infection in this region is approximately 8.73 cases per 100,000 population per year[3]. However, serological studies among healthy population in the same area revealed a much higher rate of exposure, with an estimated 1 in 4,600 antibody-producing exposures resulting in clinical infection[4]. These findings suggest that not all bacterial exposure leads to melioidosis. Clinical manifestations are commonly observed in individuals aged 45 years and above, particularly those with one or more risk factors[5,6]. The primary risk factor is regular unprotected contact with *B. pseudomallei*, which is prevalent among agricultural workers in Southeast Asia and highlights the importance of occupational health campaigns. Another common risk factor is diabetes mellitus, which is present in approximately half of melioidosis cases. The prevalence of diabetes in Thailand has increased over the past decade[7], further increasing the population at risk of melioidosis.

Although melioidosis is commonly regarded as an opportunistic infection, this does not preclude the possibility that certain *B. pseudomallei* strains or individuals may possess genetic factors that increase their susceptibility to developing melioidosis or having more severe outcomes. Previous studies have used additive heritability scores (h^2) to systematically quantify the proportion of variations in the infection outcomes that can be explained by genetic variations in the pathogens and the hosts. For instance, bacterial genetic factors accounted for 70% and 36.5% of invasiveness potential in *Streptococcus pneumoniae*[8] and *Neisseria meningitidis*[9], respectively. These findings justified conducting a microbial genome-wide association study (mGWAS) to identify bacterial genetic factors associated with transition from carriage to disease, offering insights into disease pathology. On the host side, h^2 estimation and human genome-wide association study (hGWAS) have been reported for sepsis[10,11], as well as specific pathogen infections, such as SARS-CoV-2[12], HIV[13], hepatitis[14], *Salmonella* infection[15], tuberculosis[16,17], and malaria[18]. Although sample sizes are still limited, joint host-pathogen studies have emerged for invasive pneumococcal disease[8], tuberculosis[19] and hepatitis C[20]. Collectively, these genetic studies shed light on the underlying biological mechanisms involved in heterogeneous infectious responses and facilitate the identification of specific genetic markers in pathogens and hosts. This knowledge is pertinent for future advancements in treatment stratification and vaccine development.

Currently, a comprehensive genetic study dedicated to melioidosis, along with a genetic database encompassing both bacterial and host genetic variations are lacking. To address this gap, we established a *Burkholderia pseudomallei* and Host Genetic cohort (BurkHostGEN). This initiative aims to collect and consolidate existing genetic data on both the bacterium and host to identify markers associated with the disease. BurkHostGEN will also create a community database that can be shared among researchers and healthcare professionals. The resulting data and analyses will facilitate the design of more effective clinical intervention by targeting the most harmful bacteria and the population at highest risks. The first data collection started in October 2019, with data and analyses expected to be complete in August 2025.

Ethical approvals: Dissecting the genetic basis of melioidosis infection

The study received ethical approval from the Sunpasitthiprasong Hospital Ethical Review Board (015/62C) and the Oxford Tropical Research Ethics Committee (OxTREC, 25-19). The initial approval from the Sunpasitthiprasong Hospital Ethical Review Board was granted on 7th August 2019, followed by an amended approval on 23rd July 2020. The Oxford Tropical Research Ethics Committee initially approved the protocol on 30th May 2019, and subsequently approved an amendment on 7th September 2020. The amendment was to include household water sampling activities.

Protocol

Objectives and purposes

Primary objective

To identify bacterial and host biomarkers that are associated with melioidosis acquisition and disease outcome.

Secondary objectives

- To understand how host diabetic status modulates the effect of the bacterial and patient biomarkers in disease.
- To establish a genetic database of pathogen and host dedicated to melioidosis.

Study design

This is an observational study that collects and analyses linked host and bacterial genetics samples (Figure 1). Firstly, DNA and RNA blood samples are collected from both melioidosis patients and controls to explore the host factors contributing to melioidosis susceptibility. Secondly, DNA from *B. pseudomallei* isolated from respective melioidosis patients is collected to investigate the bacterial factors involved in disease acquisition and infection outcomes. Thirdly, DNA of the microbiome of the household water supply of both melioidosis patients and controls is collected to examine their environmental exposure. It is important to note that although we initially contemplated and even endorsed a comprehensive soil sampling study, the constraints on available manpower within our project compelled us to focus our environmental sampling efforts primarily on the household water supply. Nevertheless, this approach offers an additional dimension to our investigation, providing insights into water treatment practices and the presence of *B. pseudomallei*. Additionally, demographic data including age, self-reported sex and ancestry, alongside clinical data linked to these samples are collected to provide context for genetic findings.

To ensure statistical power as well as budget and timeline feasibility, we plan to recruit a sample size of 600 melioidosis patients, 550 healthy controls and 150 patients with other infections as additional controls. Considering that 50% of melioidosis patients have diabetes mellitus, we recruited individuals with and without diabetes mellitus in equal proportions for the control groups to match the melioidosis group. BurkHostGEN Case Report Forms[21], Patient information sheet (PIS) and informed consent forms (ICFs)[22] can be found as *Extended data*.

Study site and population

The study is being conducted at Sunpasitthiprasong Hospital, a regional hospital located in Ubon Ratchathani, Thailand. The hospital serves as a vital healthcare provider for the

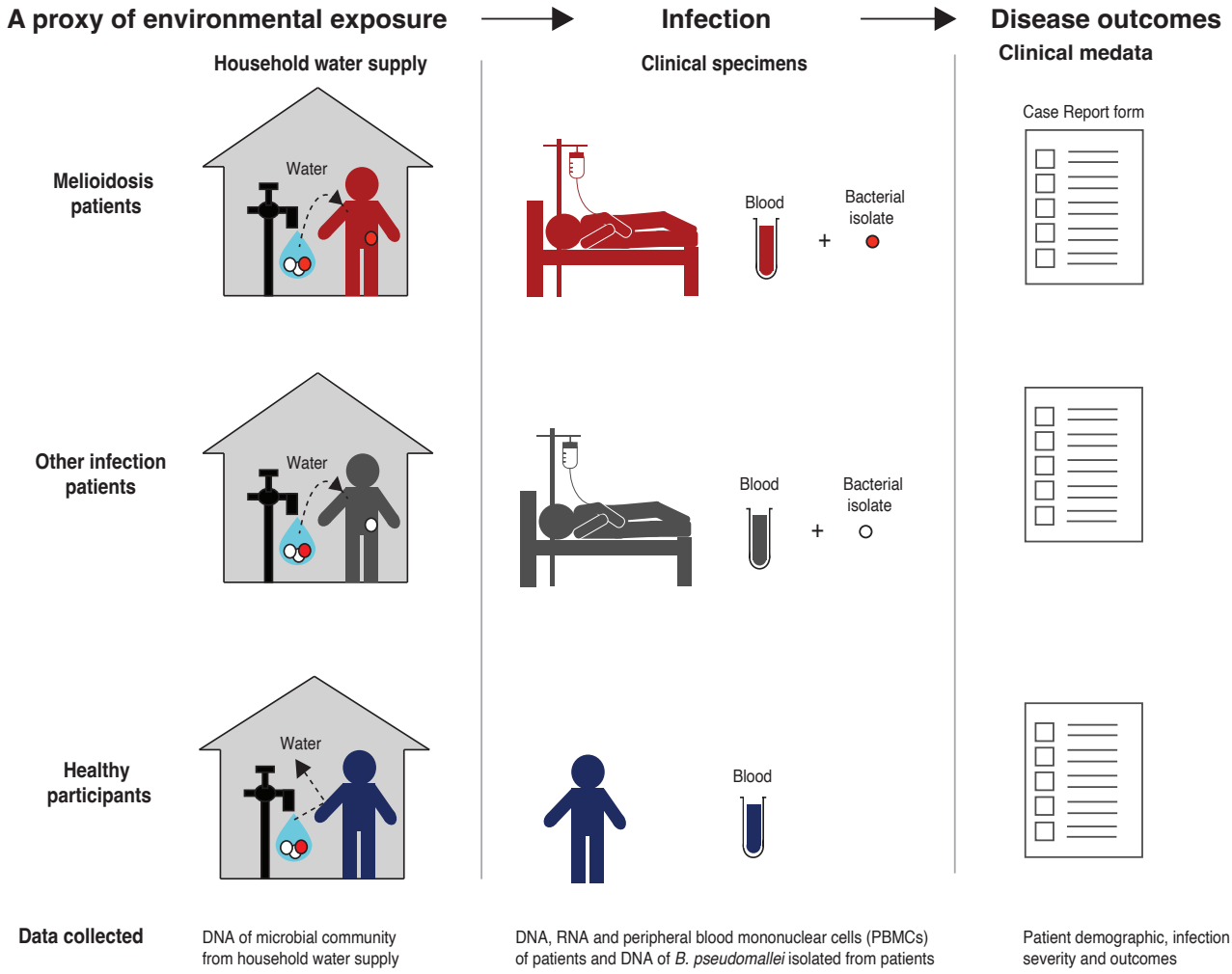

Figure 1. Study design. BurkHostGEN aims to collect, sequence and investigate the genetics of melioidosis patients, the bacterium that cause the disease in patients, as well as the bacteria recovered from patient's household water supply.

local community and surrounding areas and has gained recognition for its expertise in diagnosing and treating melioidosis, given the high incidence of disease in the region. For the study, the melioidosis, and other infection groups are being recruited from the infectious department of the hospital. The healthy group are being sourced from two different hospital departments. Healthy controls without diabetes mellitus are being recruited from the hospital blood donor clinic at the hospital. Healthy controls with diabetes are being recruited from diabetes outpatient clinic. During the recruitment process, effort is being made to ensure representation across all age groups and both sexes.

Participant inclusion criteria

Melioidosis group

- Age ≥ 18 years old.
- Culture-confirmed infection of *B. pseudomallei* from any clinical samples.
- Willingness to participate in the study and written, informed consent obtained from patient or their relative.
- Resident in northeast Thailand for at least the previous two years.

Healthy group

- Age ≥ 18 years old.
- Currently well (non-diabetic) or outpatient of diabetic clinic (diabetic group) but have no other medical problems requiring hospital supervision.
- Willingness to participate in the study and written, informed consent obtained from patient or their relative.
- Resident in northeast Thailand for at least the previous two years.

Other infection group

- Age ≥ 18 years old.
- Culture-confirmed infection by bacterial pathogens other than *B. pseudomallei*.
- Willingness to participate in the study and written, informed consent obtained from patient or their relative.
- Resident in northeast Thailand for at least the previous two years.

Participant exclusion criteria

Melioidosis group

- Current tuberculosis (TB) or treatment of TB within the last six months.
- Documented human immunodeficiency virus (HIV) infection, chemotherapy, or other immunosuppressant therapy in the last 12 months.
- Pregnancy.

Healthy group

- Previous history of melioidosis.
- Significant acute intercurrent illness.
- Current TB or treatment of TB within the last six months.
- Documented HIV infection, chemotherapy, or other immunosuppressant therapy in the last 12 months.
- Pregnancy.

Other infection group

- Previous history of melioidosis.
- Significant acute intercurrent illness.
- Current TB or treatment of TB within the last six months.
- Documented HIV infection, chemotherapy, or other immunosuppressant therapy in the last 12 months.
- Pregnancy.

Participant recruitment and sample collection

Screening and enrolment

Upon receiving a culture-confirmed diagnosis, eligible patients at Sunpasitthiprasong Hospital are approached by a member of study staff. A screening process based on participant inclusion criteria is conducted. If the patients meet the criteria, they are presented with a written participant information (PIS) and an informed consent form (ICF) in Thai language. In cases where the patient lacks capacity, the information is provided to their relative.

The participants receive a clear explanation of the study, including its nature, protocol implications, known side effects and any risks. They are explicitly informed of their right to withdraw from the study at any time, without prejudice for future care and without the need to provide a reason for withdrawal. Sufficient time, at least 10 minutes are given for consideration, along with an opportunity to ask questions to the investigator or other independent parties.

Written informed consent is obtained through dated signatures of the participant or relative, as well as the person presenting and obtaining consent. In cases where the participant or relative is illiterate, a thumbprint is obtained. A copy of the signed informed consent forms is provided to the participant or responsible relative, while the original signed forms are retained at the study site.

After the participant has been enrolled into the study, they are assigned a unique study number. These numbers consist of a letter indicating the group they belong to, with “M”, “H” and “I” representing melioidosis, healthy, and other infection groups, respectively.

Sample collection

For each participant, two to three types of samples are collected to support the study objectives.

Blood samples (Table 1) are collected once at enrolment for all studied groups. Participants are informed about the blood drawing procedure, and preferred sites for phlebotomy are the median antecubital and basilica veins in the upper extremity. Alternative sites such as veins on the dorsum on the hand and other forearm veins are used if necessary. The collected blood is used for DNA and RNA profiling, addressing the primary study objective. Additionally, glycated haemoglobin (HbA1c) levels are measured to assess the impact of chronic hyperglycaemia on infection across all groups. Patient peripheral blood mononuclear cells (PBMC) are also collected to serve as cell models for experimental validation.

A single colony pick of *B. pseudomallei,* cultured from the specimens of respective melioidosis patients, is obtained from the Sunpasitthiprasong hospital Central Lab, where routine bacterial cultures are performed for diagnostic purposes.

The water supply samples from both melioidosis patients and controls are collected within three months after participant enrolment in the project, with efforts made to minimise seasonal fluctuations.

Sample processing

Sample processing adheres to Biosafety laboratory 3 (BSL3) guidelines for handling infectious materials (suspected *B. pseudomallei* presence) and Biosafety Laboratory 2 (BSL2) guidelines for handling sterile samples (processed DNA or RNA).

Host blood samples

Blood samples are collected and stored at -80°C in Ubon Ratchathani until the recruitment phase is completed. EDTA tubes, and Tempus™ RNA blood tube are used to preserve blood DNA and RNA, respectively. Notably, previous study[23] has demonstrated the stability of blood cell RNA Tempus™ tubes for up to six years, well beyond our study's collection period, ensuring sample quality for future quantitative analysis.

The DNA extraction from whole blood will be performed using QIAamp® DNA blood kits. The extracted DNA will then undergo whole genome sequencing to achieve 15x coverage using Novaseq with PE 150 bp read length. Similarly, RNA extraction will be performed from whole blood using Tempus™ Spin RNA isolation kit. The RNA samples will undergo globin depletion and will be sequenced using Novaseq with PE 150 bp read length.

PMBC are processed on-site in Ubon Ratchathani. Due to the low average of *B. pseudomallei* count (1.1 cfu/mL) in direct blood samples and the antibiotic treatment of patients during blood collection, we expect the amount of live bacterial cells from processed samples to be very low. However, we take all necessary precautions by handling samples within biosafety cabinets. PMBCs are isolated from blood samples and are promptly frozen in liquid nitrogen to facilitate downstream analysis at various locations. We also assess the PMBC count before and after freezing to determine the viability of PBMCs for future application.

Bacterium isolated from melioidosis patients

B. pseudomallei isolated from melioidosis patients recruited for this study are stored at -80°C in Ubon Ratchathani. Subsequently, they will be transported in batches to Bangkok for DNA extraction under biosafety level 3 (BSL3) laboratory. The DNA extraction will be performed using QIAamp® DNA mini kit, followed by sequencing using Novaseq with 150 bp read length.

Bacterial community recovered from host household water supply

Participants' household water supply may encompass pond-, borehole-, well-, and pipe-water, and may vary in treatment status. We collected and processed water samples according to described methodology[6], as well as recording their water treatment methods. To facilitate the growth of *B. pseudomallei* while inhibiting fungal growth, selective plates are used. Following growth on selective plates, all microbes are plate-swept for further investigation. As *B. pseudomallei* colonies can sometimes go unnoticed due to competition with faster-growing microbes, the use of plate-sweep method ensures that all bacterial

Table 1. Different types and amounts of samples collected are tabulated below.

Studied group	Ranges of specimens from host blood sample					DNA of clinical bacterial sample	DNA of environmental microbial community
	DNA	RNA	HbA1c	PBMC*	total		
Melioidosis	1 mL	3 mL	1 mL	12 mL	17 mL	A single colony of *B. pseudomallei*	A plate-sweep culture from concentrated 5 L of water
Other infection	1 mL	3 mL	1 mL	12 mL	17 mL	N/A	A plate-sweep culture from concentrated 5 L of water
Healthy individual	1 mL	3 mL	1 mL	12 mL	17 mL	N/A	A plate-sweep culture from concentrated 5 L of water

*If participant in melioidosis and other infection groups are very ill and attending physician rules out that > 5 mL of blood withdrawal could impact the condition, PBMC collection may be omitted. HbA1c, haemoglobin; PBMC, peripheral blood mononuclear cells; N/A, not applicable.

growth, regardless of their visibility on plate, is included for DNA extraction and analysis. This minimises bias associated with visual inspection. Furthermore, this approach accounts for the potential presence of multiple *B. pseudomallei* clones or lineages in the sample, preventing the oversight of minor populations being unsampled. For visible *B. pseudomallei* colonies, conventional single-colony picks are also performed. Glycerol stocks of samples from colony-picks and plate-sweep approaches are stored at -80°C in Ubon Ratchathani before batch transportation to Bangkok. DNA extraction for plate-sweep microbial community will be performed using QIAGEN Genomic tips, which minimise DNA shearing to facilitate long-read sequencing. The microbial DNA will then undergo sequencing using GridION with ligation library preparation on a pool of up to 10 samples per flow cell. As for *B. pseudomallei* single colony picks from the environment, they will be sequenced using Novaseq with the same protocol used for *B. pseudomallei* isolated from patients.

Sample shipment

To safely transport infectious materials, including bacterial and host blood samples within Thailand, we strictly adhere to IATA Dangerous Goods Regulation (DGR) packing guidelines. The entails a triple packaging system, comprising a primary receptacle, secondary packaging, and rigid outer packaging. The primary receptacle and secondary packaging must withstand an internal pressure of 95 kPa with a temperature range of -40 C to 55 C. Each primary receptacle is securely sealed with a lid or screwed cap, further protected by stretch wrapping (e.g. parafilm), and cushioned with absorbent material to manage potential breakage or leakage. The secondary packaging is a durable, watertight, leak-proof, designed to enclose and protect the primary receptacles. Adequate absorbent padding is positioned between the primary receptacles and the secondary packaging to handle potential leaks or provide cushioning. Dry ice is placed outside of the secondary packaging, and the rigid outer packaging must have sufficient strength to protect the contents from physical damage and water during transportation while allowing carbon dioxide gas to vent from the dry ice.

Prior to shipping sterile DNA and RNA derived from bacterial or host samples, a random set of samples is inoculated in media enriched for *B. pseudomallei* growth to detect any potential bacterial contamination, ensuring safety and compliance during transport.

For all sample categories, each package is clearly labelled with the sample ID, pathogen name, specimen type, production date, and the contact details of both the shipper and consignee.

Data analysis and statistical consideration

Identification of differential host gene expression signatures based on infection severity

Our hypothesis is that gene expression signatures differ between individuals with different infection severity. We will categorise the groups as follows: a healthy population with no or subclinical symptoms where individuals did not seek healthcare, a group with less severe melioidosis infection that required healthcare but recover, and a group with severe infection resulting in mortality within 28 days. To identify genes that vary among and between these groups, we will perform multi-group differential expression analysis using ANOVA while controlling for common comorbidities and covariates such as participant age and sex. We will also test for consistency with the previously defined sepsis response signature (SRS)[24,25]. The SRS1, SRS2, and SRS3 profiles will define patients with an immunosuppressed profile, immunocompetent profile, and healthy individuals, respectively. We will compare the signatures obtained from melioidosis patients to those obtained from the other infection group to determine if the identified signatures are unique to melioidosis infection.

Exploring host regulatory markers affecting gene expression in melioidosis

We hypothesise that part of the variation in gene expression is influenced by regulatory genetic determinants. To investigate this, we will perform expression quantitative trait loci (eQTL) analysis. Using a linear model, we will test the correlation between each single nucleotide polymorphism (SNP) and the nearby gene expression. Principal components that define the population structure will be included as covariates. We will search for enrichment of transcription factor binding motifs within these expression-associated SNPs to identify potential regulatory markers affecting gene expression.

Examining the impact of host diabetic status or anti-diabetic treatment on differential host gene expression

While individuals with diabetes have an increased risk of developing melioidosis, they have reduced risk of mortality compared to non-diabetic patients[26]. We aim to investigate how the diabetic status of patients may modulate the outcomes in melioidosis. For each gene showing differential expression, we will compare two nested logistic regression models: one assessing the association with outcomes alone, and the other including host diabetic status and an interaction terms. By doing so, we will identify differentially expressed genes whose effects may be influenced, either dampened or enhanced, by the host's diabetic status or anti-diabetic treatment.

Bacterial factors influencing disease severity

Our goal is to examine whether specific genetic variants in *B. pseudomallei* influence the outcomes of melioidosis. Using a genome-wide association study (GWAS), we will identify associations between bacterial genetic variants (genes, unitigs, and core SNPs) and clinical phenotypes, including 28-day mortality, SOFA scores, and patient diabetic status. To account for *B. pseudomallei* population structure, a phylogeny and principal component analysis (PCA) defining the bacterial population structure will be used as covariates in the GWAS. The expected outcome will be a list of *B. pseudomallei* variants associated with melioidosis outcomes. Additionally, we will provide a summary statistic to facilitate future meta-analysis in *B. pseudomallei* research.

Impact of host diabetic status or anti-diabetic treatment on bacterial factors

We also hypothesise that the severity of melioidosis, influenced by bacterial genetic variants, may be modulated by the host's diabetic status and/or anti-diabetic treatment. To explore this

hypothesis, we will compare two nested models: one assessing the association with outcomes alone, and the other incorporating host diabetic status and an interaction terms. The findings are expected to yield a catalogue of *B. pseudomallei* disease-severity variants, highlighting their potential influence, whether attenuated or amplified, by the host's diabetic status or anti-diabetic treatment.

Diversity of B. pseudomallei in the household water supply of melioidosis and control groups

To better understand the transition from environmental *B. pseudomallei* to infection, we proposed to compare genome assemblies of *B. pseudomallei* detected in patient's household water supply to those causing disease in the patients. Previous studies relied on culture-based approaches to identify *B. pseudomallei*. Due to its relatively slow growth, other microbes in the environment may outcompete *B. pseudomallei*, resulting in invisible colonies and negative culture results. To overcome this limitation, we employ a new plate-sweep followed by sequencing technique to allow us to capture previously undetected *B. pseudomallei* population (see sample processing). We will use a probabilistic model to quantify microbial taxa and *B. pseudomallei* lineages in each sample.

Sample size consideration

Analysis on host differential gene expression

For two sample RNA-seq experiment design, we used a linear model[27] to estimate sample size with predicted proportion of differentially expressed genes for each possible condition (healthy *vs.* recovered melioidosis, recovered melioidosis *vs.* fatal melioidosis, and healthy *vs.* fatal melioidosis). Assuming that all genes have homogenous read counts, we estimated that a group size of 41 is needed to detect a 1.5-fold change in gene expression, with false discovery rate (FDR) at 0.05 with a statistical power of 80%. With an estimated melioidosis mortality rate at 30%, a group size of 550 healthy individuals, 420 recovered melioidosis cases, and 180 fatal melioidosis cases which could be achieved in the study time frame should allow sufficient statistical power (Figure 2a).

Analysis on host expression quantitative loci (eQTL)

Given a SNP and a continuous gene expression data, we used a linear regression[28] to calculate the sample size with different minor allele frequencies (MAF). Using Bonferroni correction for 200,000 hypotheses based on 20,000 genes and an average of 10 SNPs in proximity of each gene as in 29, we estimated that the size of 600 samples, which is the feasible size collected through this project and compiled with previous reports[30,31], will give 80% power for detecting eQTL with a MAF at 0.03 (Figure 2b).

Analysis of bacterial genetic variants

Given different units of genetic variations (gene presence, unitig, and SNP), and a binary disease outcomes (estimated 30% mortality), we used an additive model[32] to calculate the sample size with different MAF. Using a genome-wide significant cut-off at 5×10^{-8} and genotype relative risk of 1.5, our combined size of over 2,400 *B. pseudomallei* genomes collected

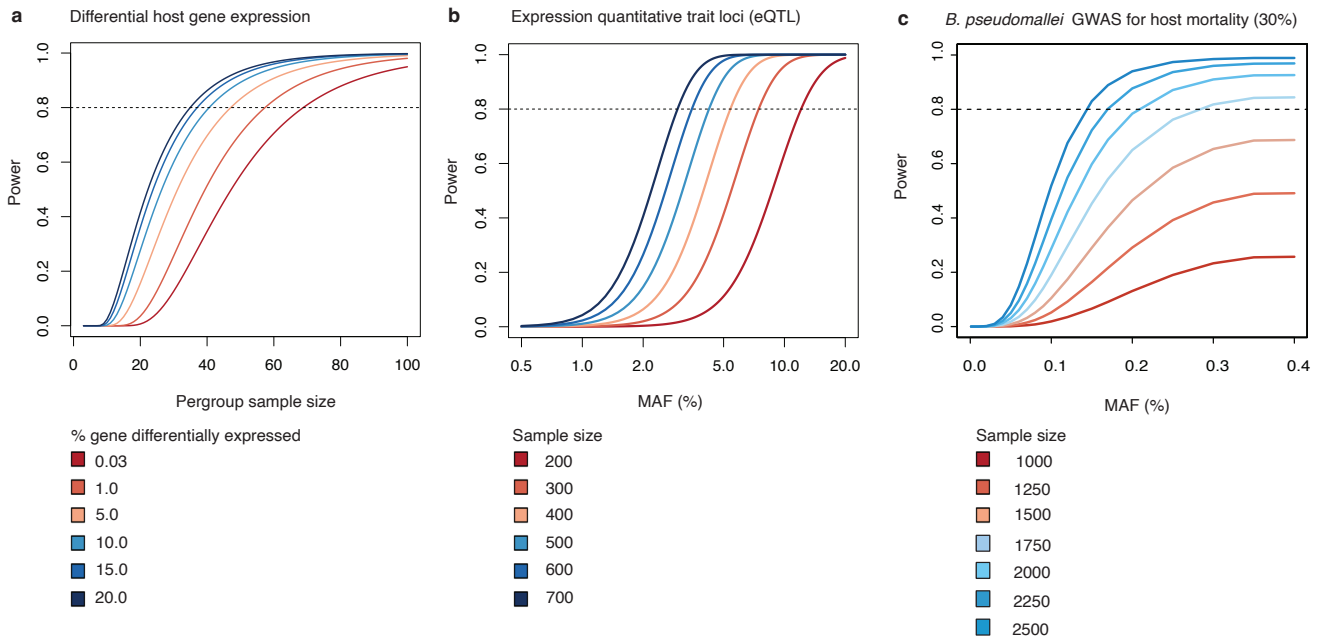

Figure 2. Power calculation. (**a**) Sample size required to detect host differential host gene expression. (**b**) sample size required for eQTL analysis. (**c**) sample size required for bacterial GWAS for host 28-day mortality. eQTL, expression quantitative trait loci; GWAS, genome-wide association study; MAF, minor allele frequencies.

in the same endemic area under this study and those reported in 6,30,33 has sufficient power to detect associations with MAF of 0.15 (Figure 2c).

Considering the exploratory nature of the remaining objectives, the sample size was not based on power calculations.

Ethics

This study is conducted in compliance with the conditions stipulated by the sponsor and local Ethics Committee/Institutional Review Board (EC/IRB), as well as applicable regulatory requirements and ICH Good Clinical Practice (GCP) guidelines.

Risks

This study is categorised as minimal risk since it does not involve experimental interventions or investigational new drugs. The total amount of blood collected from patients (up to 17 mL) is considered safe. Stored samples are used for diagnostic purposes relevant to the study aims, and consent is obtained from patients for the storage of all clinical specimens.

Participant confidentiality

All study-related information is securely stored at the study site. Subject information is kept in locked file cabinets, with restricted access limited to study staff. The data of participants' home addresses are used for the one-off collection of water sample and are strictly limited to the water collection team. Specimens, reports, data collection forms, and administrative records are identified by anonymised codes. Personal identifiers such as names are stored separately from study records. The study database has password-protected access systems. Subject information will not be released without written permission. Research dataset will be pseudonymised. This allows for eventualities such as feedback of individual data or withdrawal of consent. Subject names or identities will not be disclosed. The risk of re-identification from the published data is considered very low.

Data handling and record keeping

Clinical data recorded on Case Report Forms (CRF) are entered into the MACRO EDC, a GCP-compliant data management system. The database is password-protected and incorporates internal quality checks to ensure data consistency, completeness, and accuracy. Study participants are identified by unique participant numbers in the database.

Participant records are stored in binders or scanned and stored electronically. Given participant consent, anonymised data and results from blood and water sample analyses stored in the database maybe shared with other researchers. Personal information is anonymised to protect privacy.

Genetic data, reuse of data and data sharing

The project's outputs, including sequencing data and associated metadata hold value for other researchers. To ensure reproducibility and comply with the Wellcome Trust policy, the sequenced data will be stored in two separate archives. The European Genome-Phenome Archive (EGA), a managed access database, will host the genomic and transcriptomic data of human participants. The bacterial genomic data will be stored in the European Nucleotide Archive (ENA), an open access database.

To comply with the General Data Protection regulation, pseudonymised participant data and microbiological data will be shared with researchers. To safeguard participant privacy, stringent measures will be implemented, such as the removal of rare genetic variants or traits unique to only a few individuals, ensuring that the study participants cannot be identified from any shared data. This approach effectively maintains participant confidentiality and privacy while enabling the dissemination and utilisation of research data among the scientific community. Access to human datasets will be carefully managed and granted transparently to appropriately qualified researchers.

Anonymised participant data submitted to EGA for data reuse will include sex, phenotype (disease and diabetic status), self-reported ancestry, and a study number. Additionally, the summary statistics to be released as part of the future publication will encompass variant identifiers (rsID, chromosome position), effect alleles, effect size, confident intervals, and associated statistics. The risk of re-identification of individual research participants is considered low.

Study status

Ethical approvals were granted for BurkHostGEN in August 2019 and recruitment began shortly afterwards. To date, over 1,000 participants have been successfully enrolled. The collected samples are scheduled to undergo processing, sequencing and analysis with an expected completion in August 2025.

A portion of the collected data has been reported in a preprint and can be found on medRxiv (doi: https://doi.org/10.1101/2023.05.06.23289616). The preprint relies on BurkHostGEN clinical data collected between October 2019 to December 2022. The dataset provides a rationale for the development of a rapid CRISPR diagnostic test and does not conflict with the publication of this study protocol. This prompted us to develop the rapid CRISPR-based diagnosis test reported in the preprint. Currently, the preprint is undergoing revisions within a peer-reviewed journal.

Conclusions

BurkHostGEN aims to link environmental metagenomic, pathogen genomic and host genomic in a less studied population of northeast Thailand. This approach will provide unique insights into the patients' environmental exposures, the infecting pathogen, *B. pseudomallei*, and the patient population. The genetic and metadata from BurkHostGEN can facilitate the discovery and validation of genetic factors associated with melioidosis development and outcomes in both the pathogen and the hosts. This comprehensive data allows for an initial exploration of the interplay between *B. pseudomallei* and host genetics, potentially revealing genetic factors associated

with melioidosis development and outcomes. Given the current sample size, the initial results may be confined to common alleles in the pathogen and regulatory alleles in the hosts. However, there are plans for future data expansion to capture rarer alleles. This dataset holds many potentials, enabling genetic investigations not only specific to melioidosis but also to other common infectious diseases prevalent in this less studied global region.

Data availability

Underlying data

No data are associated with this article.

Extended data

Figshare: BurkHostGEN Case Report Forms (CRF). https://doi.org/10.6084/m9.figshare.23908641[21]

This project contains the following extended data:

- BurkHostGEN_CRF_Melioid_V5.0_12May20_approved .doc
- BurkHostGEN_CRF_Healthy_V.5.0_12May20_approved.doc
- BurkHostGEN_CRF_Other-Infection_V5.0_12May20_approved.doc

Figshare: BurkHostGEN Patient information sheet (PIS) and informed consent forms (ICF). https://doi.org/10.6084/m9.figshare.23905611[22]

This project contains the following extended data:

- BurHostGEN_PIS-ICF_Healthy_V.5.1_22July2020_approved.docx
- BurHostGEN_PIS-ICF_Melioid_V.5.1_22July2020_approved.docx
- BurHostGEN_PIS-ICF_Other-Infection_V.5.1_22July2020_approved.docx

References

1. Birnie E, Virk HS, Savelkoel J, *et al.*: **Global burden of melioidosis in 2015: a systematic review and data synthesis.** *Lancet Infect Dis.* 2019; **19**(8): 892–902.
PubMed Abstract | Publisher Full Text | Free Full Text
2. Wiersinga WJ, Virk HS, Torres AG, *et al.*: **Melioidosis.** *Nat Rev Dis Primers.* 2018; **4**: 17107.
PubMed Abstract | Publisher Full Text | Free Full Text
3. Hantrakun V, Kongyu S, Klaytong P, *et al.*: **Clinical Epidemiology of 7126 Melioidosis Patients in Thailand and the Implications for a National Notifiable Diseases Surveillance System.** *Open Forum Infect Dis.* 2019; **6**(12): ofz498.
PubMed Abstract | Publisher Full Text | Free Full Text
4. Cheng AC, Wuthiekanun V, Limmathurotsakul D, *et al.*: **Intensity of exposure and incidence of melioidosis in Thai children.** *Trans R Soc Trop Med Hyg.* 2008; **102**(Suppl 1): S37–39.
PubMed Abstract | Publisher Full Text
5. Suputtamongkol Y, Chaowagul W, Chetchotisakd P, *et al.*: **Risk Factors for Melioidosis and Bacteremic Melioidosis.** *Clin Infect Dis.* 1999; **29**(2): 408–13.
PubMed Abstract | Publisher Full Text
6. Limmathurotsakul D, Kanoksil M, Wuthiekanun V, *et al.*: **Activities of daily living associated with acquisition of melioidosis in northeast Thailand: a matched case-control study.** *PLoS Negl Trop Dis.* 2013; **7**(2): e2072.
PubMed Abstract | Publisher Full Text | Free Full Text
7. Aekplakorn W, Chariyalertsak S, Kessomboon P, *et al.*: **Prevalence of Diabetes and Relationship with Socioeconomic Status in the Thai Population: National Health Examination Survey, 2004-2014.** *J Diabetes Res.* 2018; **2018**: 1654530.
PubMed Abstract | Publisher Full Text | Free Full Text
8. Lees JA, Ferwerda B, Kremer PHC, *et al.*: **Joint sequencing of human and pathogen genomes reveals the genetics of pneumococcal meningitis.** *Nat Commun.* 2019; **10**(1): 2176.
PubMed Abstract | Publisher Full Text | Free Full Text
9. Earle SG, Lobanovska M, Lavender H, *et al.*: **Genome-wide association studies reveal the role of polymorphisms affecting factor H binding protein expression in host invasion by *Neisseria meningitidis*.** *PLoS Pathog.* 2021; **17**(10): e1009992.
PubMed Abstract | Publisher Full Text | Free Full Text
10. Rautanen A, Mills TC, Gordon AC, *et al.*: **Genome-wide association study of survival from sepsis due to pneumonia: an observational cohort study.** *Lancet Respir Med.* 2015; **3**(1): 53–60.
PubMed Abstract | Publisher Full Text
11. Kwok AJ, Mentzer A, Knight JC: **Host genetics and infectious disease: new tools, insights and translational opportunities.** *Nat Rev Genet.* 2021; **22**(3): 137–153.
PubMed Abstract | Publisher Full Text | Free Full Text
12. Pairo-Castineira E, Rawlik K, Bretherick AD, *et al.*: **GWAS and meta-analysis identifies 49 genetic variants underlying critical COVID-19.** *Nature.* 2023; **617**(7962): 764–768.
PubMed Abstract | Publisher Full Text | Free Full Text
13. Fellay J, Shianna KV, Ge D, *et al.*: **A Whole-Genome Association Study of Major Determinants for Host Control of HIV-1.** *Science.* 2007; **317**(5840): 944–7.
PubMed Abstract | Publisher Full Text | Free Full Text
14. Kamatani Y, Wattanapokayakit S, Ochi H, *et al.*: **A genome-wide association study identifies variants in the *HLA-DP* locus associated with chronic hepatitis B in Asians.** *Nat Genet.* 2009; **41**(5): 591–5.
PubMed Abstract | Publisher Full Text
15. Gilchrist JJ, Rautanen A, Fairfax BP, *et al.*: **Risk of nontyphoidal *Salmonella* bacteraemia in African children is modified by *STAT4*.** *Nat Commun.* 2018; **9**(1): 1014.
PubMed Abstract | Publisher Full Text | Free Full Text
16. African TB Genetics Consortium, The Wellcome Trust Case Control Consortium, Thye T, *et al.*: **Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11.2.** *Nat Genet.* 2010; **42**(9): 739–741.
PubMed Abstract | Publisher Full Text | Free Full Text
17. Mahasirimongkol S, Yanai H, Mushiroda T, *et al.*: **Genome-wide association studies of tuberculosis in Asians identify distinct at-risk locus for young tuberculosis.** *J Hum Genet.* 2012; **57**(6): 363–7.
PubMed Abstract | Publisher Full Text
18. Malaria Genomic Epidemiology Network, Band G, Le QS, *et al.*: **Insights into malaria susceptibility using genome-wide data on 17,000 individuals from Africa, Asia and Oceania.** *Nat Commun.* 2019; **10**(1): 5732.
PubMed Abstract | Publisher Full Text | Free Full Text
19. Phelan J, Gomez-Gonzalez PJ, Andreu N, *et al.*: **Genome-wide host-pathogen analyses reveal genetic interaction points in tuberculosis disease.** *Nat Commun.* 2023; **14**(1): 549.
PubMed Abstract | Publisher Full Text | Free Full Text
20. STOP-HCV Consortium, Ansari MA, Pedergnana V, *et al.*: **Genome-to-genome analysis highlights the effect of the human innate and adaptive immune systems on the hepatitis C virus.** *Nat Genet.* 2017; **49**(5): 666–673.
PubMed Abstract | Publisher Full Text | Free Full Text
21. Chewapreecha C: **BurkHostGEN Case Report Forms (CRF).** figshare. [Dataset], 2023.
http://www.doi.org/10.6084/m9.figshare.23908641.v2
22. Chewapreecha C: **BurkHostGEN Patient information sheet (PIS) and**

informed consent forms (ICF). figshare. [Dataset], 2023.
http://www.doi.org/10.6084/m9.figshare.23905611.v2

23. Duale N, Lipkin WI, Briese T, *et al.*: **Long-term storage of blood RNA collected in RNA stabilizing Tempus tubes in a large biobank--evaluation of RNA quality and stability.** *BMC Res Notes.* 2014; **7**: 633.
PubMed Abstract | Publisher Full Text | Free Full Text

24. Davenport EE, Burnham KL, Radhakrishnan J, *et al.*: **Genomic landscape of the individual host response and outcomes in sepsis: a prospective cohort study.** *Lancet Respir Med.* 2016; **4**(4): 259–71.
PubMed Abstract | Publisher Full Text | Free Full Text

25. Cano-Gamez E, Burnham KL, Goh C, *et al.*: **An immune dysfunction score for stratification of patients with acute infection based on whole-blood gene expression.** *Sci Transl Med.* 2022; **14**(669): eabq4433.
PubMed Abstract | Publisher Full Text | Free Full Text

26. Chantratita N, Phunpang R, Yarasai A, *et al.*: **Characteristics and one year outcomes of melioidosis patients in Northeastern Thailand: A prospective, multicenter cohort study.** *Lancet Reg Health Southeast Asia.* 2023; **9**: 100118.
PubMed Abstract | Publisher Full Text | Free Full Text

27. Ran B, Lieu P, Triche T: **ssizeRNA Sample Size Calculation for RNA-Seq Experimental Design.** 2017.

28. Dong X, Chang TW, Weiss ST, *et al.*: **powerEQTL: Power and Sample Size Calculation for eQTL Analysis.** 2017.

29. Lonsdale J, Thomas J, Salvatore M, *et al.*: **The Genotype-Tissue Expression (GTEx) project.** *Nat Genet.* 2013; **45**(6): 580–5.
PubMed Abstract | Publisher Full Text | Free Full Text

30. Dunachie SJ, Jenjaroen K, Reynolds CJ, *et al.*: **Infection with *Burkholderia pseudomallei* - immune correlates of survival in acute melioidosis.** *Sci Rep.* 2017; **7**(1): 12143.
PubMed Abstract | Publisher Full Text | Free Full Text

31. Yimthin T, Cliff JM, Phunpang R, *et al.*: **Blood transcriptomics to characterize key biological pathways and identify biomarkers for predicting mortality in melioidosis.** *Emerg Microbes Infect.* 2021; **10**(1): 8–18.
PubMed Abstract | Publisher Full Text | Free Full Text

32. Skol AD, Scott LJ, Abecasis GR, *et al.*: **Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies.** *Nat Genet.* 2006; **38**(2): 209–13.
PubMed Abstract | Publisher Full Text

33. Seng R, Tandhavanant S, Saiprom N, *et al.*: **Genetic diversity, determinants, and dissemination of *Burkholderia pseudomallei* lineages implicated in melioidosis in northeast Thailand.** *Microbiology.* 2023.

Open Peer Review

Current Peer Review Status: ✔ ✔

Version 2

Reviewer Report 03 November 2023

https://doi.org/10.21956/wellcomeopenres.22585.r69547

✔ **Sheila Nathan**
Faculty of Science and Technology, School of Biosciences and Biotechnology, National University of Malaysia, Bangi, Selangor, Malaysia

The authors have addressed the concerns I raised in my initial peer review report appropriately and I am happy to recommend indexing with no further comments or changes.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Molecular microbiology; host-pathogen interaction; Burkholderia pseudomallei genetics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 08 Nov 2023

Claire Chewapreecha

We are grateful for the reviewer's time and constructive comments. This helps enhance the clarity of our study protocol. Thank you immensely.

Competing Interests: No competing interests were disclosed.

Version 1

Reviewer Report 12 October 2023

https://doi.org/10.21956/wellcomeopenres.21941.r65464

Chiranjay Mukhopadhyay
Department of Microbiology, Kasturba Medical College and Hospital, Manipal Academy of Higher Education, Manipal, Karnataka, India

Ranita Ghosh Dastidar
Department of Biochemistry, Kasturba Medical College, Manipal, Karnataka, India

In the study protocol entitled “BurkHostGEN: a study protocol for evaluating variations in the *Burkholderia pseudomallei* and host genomes associated with melioidosis infection” by Kesorn Angchagun *et al.*, the authors have emphasized that not all bacterial exposures lead to melioidosis. Acquisition and further the outcomes of melioidosis vary largely due to genetic factors of both the acquired bacteria and the genetic makeup of the individual. With this fact in focus, the authors have undertaken an initiative to perform a detailed genomic study hypothesizing that genetic factors of both bacteria and host (separately or combinatorially) determine the acquisition and outcome patterns of patients. The protocol includes comparative DNA and RNA level studies from the patient and control blood samples as well as comparing their HbA1c levels. The protocol also enables comparison of DNA from bacteria collected from patient sample and community water source in nearby areas of patients and controls by metagenomic analyses. It also collects and stores the patient PBMCs for future utilization as cell models. The sample collection procedure is already over but can continue further as well. The sequencing and analysis, when complete, will provide the opportunity to discover several novel bacterial and host genetic factors and also can be followed by all melioidosis-prone south eastern Asia and northern Australian regions and beyond for better understanding, prevention and cure of the disease. This study protocol can further serve as basic guideline for the study of many other infectious diseases and emerging tropical diseases and also help generate community database nationally and globally.

After careful review of the manuscript, these reviewers feel that there are following minor concerns regarding the manuscript, which after being addressed can allow the approval of this protocol for indexing.

1. For DNA extraction of environmental microbial community, emphasis has been on the plate-sweep culture from concentrated 5 litre of water (as shown in table 1). The disadvantages of single-colony pick method should be elaborated if any. Methods section, however, mentions the use of both plate-sweep and single-colony pick methods being used for sequencing using GridION (with ligation library preparation on a pool of up to 10 samples per flow cell) and Novaseq (as for patient sample isolation) respectively. Please clarify.

2. In section “Diversity of *B. pseudomallei* in the household water supply of melioidosis and control group”, it is mentioned that slow growth of *B. pseudomallei* can lead to overgrowth of other microbes and hence negative culture results due to invisible colonies. This section can be better aligned with the methods section for better clarification.

3. Please mention clearly - Are all sequencing procedures done under BSL3 laboratory facilities or only DNA/RNA extraction methods?

4. In data analysis section, one of the infection severity groups includes, “a healthy population with no or subclinical symptoms” – please elaborate subclinical symptoms.

5. The definition for ‘mild melioidosis’ may be more clear.

6. Since a portion of the collected data has been reported in another preprint awaiting peer-review, assure that this does not create any conflict of interest amongst participating individuals or groups. A statement to that effect will be preferred that states these two groups are identical or have no conflicts.

7. In sample collection section, preferred and alternative sites of phlebotomy are mentioned. Any information related to any expected differences in the HbA1c reports and amount of PBMC that can be collected that can vary based on the site of blood collection (for phlebotomy) if any should be mentioned (as a matter of fact if known or as an assumption).

Is the rationale for, and objectives of, the study clearly described?

Yes

Is the study design appropriate for the research question?

Yes

Are sufficient details of the methods provided to allow replication by others?

Yes

Are the datasets clearly presented in a useable and accessible format?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Infectious Diseases, Emerging Tropical Diseases, Melioidosis

We confirm that we have read this submission and believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 30 Oct 2023

Claire Chewapreecha

#Authors: We sincerely thank the reviewers for their feedback and constructive insights, which have enhanced the clarity and readability of our study protocol. We really appreciate your time and support, and hope that our work will make a meaningful contribution to the genetic progress in our melioidosis community. A point-by-point response is provided below

#Reviewers: In the study protocol entitled “BurkHostGEN: a study protocol for evaluating

variations in the Burkholderia pseudomallei and host genomes associated with melioidosis infection" by Kesorn Angchagun et al., the authors have emphasized that not all bacterial exposures lead to melioidosis. Acquisition and further the outcomes of melioidosis vary largely due to genetic factors of both the acquired bacteria and the genetic makeup of the individual. With this fact in focus, the authors have undertaken an initiative to perform a detailed genomic study hypothesizing that genetic factors of both bacteria and host (separately or combinatorially) determine the acquisition and outcome patterns of patients. The protocol includes comparative DNA and RNA level studies from the patient and control blood samples as well as comparing their HbA1c levels. The protocol also enables comparison of DNA from bacteria collected from patient sample and community water source in nearby areas of patients and controls by metagenomic analyses. It also collects and stores the patient PBMCs for future utilization as cell models. The sample collection procedure is already over but can continue further as well. The sequencing and analysis, when complete, will provide the opportunity to discover several novel bacterial and host genetic factors and also can be followed by all melioidosis-prone south eastern Asia and northern Australian regions and beyond for better understanding, prevention and cure of the disease. This study protocol can further serve as basic guideline for the study of many other infectious diseases and emerging tropical diseases and also help generate community database nationally and globally.After careful review of the manuscript, these reviewers feel that there are following minor concerns regarding the manuscript, which after being addressed can allow the approval of this protocol for indexing.For DNA extraction of environmental microbial community, emphasis has been on the plate-sweep culture from concentrated 5 litre of water (as shown in table 1). The disadvantages of single-colony pick method should be elaborated if any. Methods section, however, mentions the use of both plate-sweep and single-colony pick methods being used for sequencing using GridION (with ligation library preparation on a pool of up to 10 samples per flow cell) and Novaseq (as for patient sample isolation) respectively. Please clarify. The disadvantages and advantages of both plate-sweep and single-colony pick sequencing has been discussed.

#Authors: We apologise for the lack of the rationale for using the dual plate-sweep and single-colony pick approach in the previous version. We have now provided a detailed explanation. To briefly summarise, the plate-sweep method ensure that all bacterial growth, including potentially unnoticed B. pseudomallei colonies due to competition with faster-growing microbes, is included for DNA extraction and analysis. This minimises bias associated with visual inspection and accommodates the potential presence of multiple B. pseudomallei clones or lineages, thereby preventing the oversight of minor populations. Additionally, for visible B. pseudomallei colonies, we perform conventional single-colony picks to complement the analysis.

#Reviewers: 2. In section "Diversity of B. pseudomallei in the household water supply of melioidosis and control group", it is mentioned that slow growth of B. pseudomallei can lead to overgrowth of other microbes and hence negative culture results due to invisible colonies. This section can be better aligned with the methods section for better clarification.

#Authors: We thank the reviewers for this comment and have restructured the text to make the two sections aligned.

#Reviewers: 3. Please mention clearly - Are all sequencing procedures done under BSL3

laboratory facilities or only DNA/RNA extraction methods?

#Authors: We appreciate this comment and have specifically outlined the procedure involved in BSL2, and BSL3 at the beginning of the sample processing section as "sample processing adheres to Biosafety laboratory 3 (BSL3) guidelines for handling infectious materials (suspected B. pseudomallei presence) and Biosafety Laboratory 2 (BSL2) guidelines for handling sterile samples (processed DNA or RNA)."

#Reviewers: 4. In data analysis section, one of the infection severity groups includes, "a healthy population with no or subclinical symptoms" – please elaborate subclinical symptoms.

#Authors: We thank the reviewer for this feedback and have defined a healthy population with no or subclinical symptoms as those who were healthy and did not seek healthcare to make this clearer.

#Reviewers: 5. The definition for 'mild melioidosis' may be more clear.

#Authors: We appreciate this comment and have defined "mild melioidosis" as a group with less severe melioidosis infection that required healthcare but recovered.

#Reviewers: 6. Since a portion of the collected data has been reported in another preprint awaiting peer-review, assure that this does not create any conflict of interest amongst participating individuals or groups. A statement to that effect will be preferred that states these two groups are identical or have no conflicts

#Authors: We thank the reviewer for pointing this and have now explicitly mentioned in the text that there are no publication conflicts.

#Reviewers: 7. In sample collection section, preferred and alternative sites of phlebotomy are mentioned. Any information related to any expected differences in the HbA1c reports and amount of PBMC that can be collected that can vary based on the site of blood collection (for phlebotomy) if any should be mentioned (as a matter of fact if known or as an assumption).

#Authors: Irrespective of the phlebotomy sites, PBMC and HbA1c were processed identically. However, we noted that the processing time between blood collection and storage are critical factor affecting the number of viable cells and have now included this specification in the protocol.

Again, a heartfelt thank you for your help reviewing this study protocol.

Competing Interests: No competing interests were disclosed.

Reviewer Report 04 September 2023

https://doi.org/10.21956/wellcomeopenres.21941.r65463

?

Sheila Nathan
Faculty of Science and Technology, School of Biosciences and Biotechnology, National University of Malaysia, Bangi, Selangor, Malaysia

This study is proposed to identify *Burkholderia pseudomallei* and host genetic factors that influence the outcome of an infection by this tropical bacterium. Additional factors included in this study are the risk factors of diabetes and environmental exposure to the pathogen. Overall, this is an ambitious but timely research undertaking, which expects to contribute new knowledge to benefit the design of diagnostic biomarkers as well potentially identify susceptibility markers to forecast disease severity.

The standard of the English used requires improvement in various sections. Some examples within the Abstract and Introduction are noted below:

Abstract

- 'infection can have different infection outcomes' - the word "infection" appears twice in the same sentence.
- 'acquisition and outcomes of melioidosis' - one does not acquire the disease.
- 'Weare obtaining' - should be "We are"

Intro

- 'amendment approval' - amended approval
- The use of past-, present- and future tense should be standardised.

General Comments

"The resulting data and analyses will pave the way for future progress in diagnosis, treatment stratification, and development of vaccines for this challenging infection." - what are the issues or bottlenecks currently faced?

What was the need to collect and analyse the metagenomic content of household water collected from patients' houses? What is the source of the water supply - treated or untreated? Why is the assumption made that a potential infection route is the consumed water and not from agriculture-based activities (soil sampling)?

"Primary objective - To identify bacterial and **patient** biomarkers that are associated with melioidosis acquisition and disease outcome." Does the study include those individuals who are antibody-positive but do not display symptoms? In this case, they would not be classified as "patients".

There are a large number of methods to be undertaken in this study. Nonetheless, the completeness of the protocols is lacking (subject to Wellcome Open Research's requirements for describing an ongoing study).

The blood samples (1400) will be stored until recruitment is completed. When is recruitment expected to be done? How long will it take to perform the downstream experiments and analysis? Do you expect any differences in samples stored for 12 months compared to samples obtained 1 month prior to the completion of recruitment?

"B. pseudomallei single colony picks from the environment" - no details related to the selected "environments" are provided as well as on how sampling will be done. If "environment" refers to the household water samples, I suggest retaining the use of a single term rather than inter-changing. Nonetheless, I also refer to my earlier comment on the source of the water - is it ground water in a well or is it treated / untreated pipe/tap water?

Risks - what are the containment measures to be taken when transporting the bacteria to Bangkok? Will the blood samples taken from melioidosis-confirmed patients be routinely screened for other infectious agents that might compromise the PBMC analysis?

Genetic Data, reuse of data and data sharing - last sentence "The risk of re-identification of individual research participants, even if **unlikely**, is considered low." - I think it should be "likely" instead of "unlikely".

Is the rationale for, and objectives of, the study clearly described?
Partly

Is the study design appropriate for the research question?
Yes

Are sufficient details of the methods provided to allow replication by others?
No

Are the datasets clearly presented in a useable and accessible format?
Not applicable

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Molecular microbiology; host-pathogen interaction; Burkholderia pseudomallei genetics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 30 Oct 2023

Claire Chewapreecha

#Authors: We are immensely grateful to the review for their constructive feedback and insights. This helps improve the clarity and readability of our study protocol. We deeply appreciate the reviewer's time and support, and sincerely hope that the work generated

under this study will contribute to progress in genetic research in melioidosis. A point-by-point response is provided below.

#Reviewer: This study is proposed to identify Burkholderia pseudomallei and host genetic factors that influence the outcome of an infection by this tropical bacterium. Additional factors included in this study are the risk factors of diabetes and environmental exposure to the pathogen. Overall, this is an ambitious but timely research undertaking, which expects to contribute new knowledge to benefit the design of diagnostic biomarkers as well potentially identify susceptibility markers to forecast disease severity. The standard of the English used requires improvement in various sections. Some examples within the Abstract and Introduction are noted.

#Authors: We apologise for any language issues. As English is not our native language, we had a native English-speaking co-author review the study protocol to improve its readability and language. We have revised all pertinent sections in the text and standardised the use of past, present, and future tenses to correspond to the protocol's historical, current, and future activities, respectively.

#Reviewer: "The resulting data and analyses will pave the way for future progress in diagnosis, treatment stratification, and development of vaccines for this challenging infection." - what are the issues or bottlenecks currently faced?

#Authors: We acknowledge that the previous version lacked clarity on how genetic data can enhance clinical intervention. To address this, we have rewritten the text as "The resulting data and analyses will facilitate the design of more effective clinical intervention by targeting the most harmful bacteria and the population at highest risks.", thus improving overall clarity.

#Reviewer: What was the need to collect and analyse the metagenomic content of household water collected from patients' houses? What is the source of the water supply - treated or untreated? Why is the assumption made that a potential infection route is the consumed water and not from agriculture-based activities (soil sampling)?

#Authors: We agree with the reviewer. Initially, we considered and supported a comprehensive soil sampling study. However, practical constraints related to available manpower within our project led us to primarily focus our environmental sampling on household water supplies. Participants' household water supply may include pond, borehole, well, and piped water, with varying treatment statuses. This approach adds depth to our investigation, offering insights into water treatment practices and the presence of B. pseudomallei. We have revised the text to align with this clarification.

#Reviewer: "Primary objective - To identify bacterial and patient biomarkers that are associated with melioidosis acquisition and disease outcome." Does the study include those individuals who are antibody-positive but do not display symptoms? In this case, they would not be classified as "patients".

#Authors: We thank the reviewer for pointing this out. We acknowledge that the term

"patient" was not appropriate in this context. We have replaced it with "host".

#Reviewer: There are a large number of methods to be undertaken in this study. Nonetheless, the completeness of the protocols is lacking (subject to Wellcome Open Research's requirements for describing an ongoing study).

#Authors: In light of the reviewer's comment, we have enhanced the clarity of the methods for sample processing in this protocol paper. However, since several downstream genetic analyses are yet to be completed and we are currently in the trial phase of our methods, we intend to provide a more detailed methodologies in future papers once the analysis is finalised and more robust.

#Reviewer: The blood samples (1400) will be stored until recruitment is completed. When is recruitment expected to be done? How long will it take to perform the downstream experiments and analysis? Do you expect any differences in samples stored for 12 months compared to samples obtained 1 month prior to the completion of recruitment?

#Authors: We appreciate the reviewer's comment and have included literature reference demonstrating RNA stability in the same storage environment for up to six years, thereby ensuring data quality.

#Reviewer:"B. pseudomallei single colony picks from the environment" - no details related to the selected "environments" are provided as well as on how sampling will be done. If "environment" refers to the household water samples, I suggest retaining the use of a single term rather than inter-changing. Nonetheless, I also refer to my earlier comment on the source of the water - is it ground water in a well or is it treated / untreated pipe/tap water?

#Authors: We thank the reviewer for this comment and have expanded our explanation on the rationale and strategies of household sampling. Importantly, we have explain the dual use of plate-sweep and single-colony picks. The plate-sweep method ensures the inclusivity of all B. pseudomallei colonies, reducing bias related to visual inspection, and accommodating the presence of multiple clones or lineages. In addition to this, we conduct conventional single-colony picks for visible B. pseudomallei colonies to complement the plate-sweep approach.

#Reviewer: Risks - what are the containment measures to be taken when transporting the bacteria to Bangkok? Will the blood samples taken from melioidosis-confirmed patients be routinely screened for other infectious agents that might compromise the PBMC analysis?

#Authors: We have substantially expanded the sample processing section to include these details. Specifically, we also added a new sub-section, "Sample shipment", to provide detailed instructions on transporting infectious materials which could be useful for researchers working in the field.

#Reviewer: Genetic Data, reuse of data, and data sharing - last sentence "The risk of re-identification of individual research participants, even if unlikely, is considered low." - I think it should be "likely" instead of "unlikely".

#Authors: We appreciate that the use of double negative could be confusing. We have now simplified the sentence to make it clearer.

Again, thank you immensely for your help reviewing our study protocol.

Competing Interests: No competing interests were disclosed.