Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Clinicogenomic Insights into Burkholderia pseudomallei Reveal Functional Variants Associated with Clinical Outcomes in Bangladesh

Lovely Barai
Bangladesh Institute of Research and Rehabilitation for Diabetes Endocrine and Metabolic Disorders
Ismat Ruh Jahan
North South University
Senzuti Sharmin
North South University
Mili Rani Saha
Bangladesh Institute of Research and Rehabilitation for Diabetes Endocrine and Metabolic Disorders
Tanjila Rahman
Bangladesh Institute of Research and Rehabilitation for Diabetes Endocrine and Metabolic Disorders
Md. Rokibul Hasan
Bangladesh Institute of Research and Rehabilitation for Diabetes Endocrine and Metabolic Disorders
Noshin Ibnat Rib
North South University
Fahad Khan
North South University
MSA Jilani
Ibrahim Medical College
Md Jakariya
North South University
Prosun Bhattacharya
KTH Royal Institute of Technology
Chowdhury Rafiqul Ahsan
University of Dhaka
Maqsud Hossain
muhammad.maqsud@northsouth.edu

North South University

Article

Keywords: Keywords: Burkholderia pseudomallei, melioidosis, antimicrobial resistance, virulence factors, host-pathogen interaction, mobile genetic elements

Posted Date: July 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9712675/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Burkholderia pseudomallei, the causative agent of melioidosis, is a Gram-negative environmental bacterium endemic to tropical regions and is frequently underdiagnosed. In this study, we performed comparative genomic analysis of 14 clinical *B. pseudomallei* isolates alongside 153 publicly available isolates from diverse sources.

Phylogenetic analysis placed the Bangladeshi isolates within the broader Asian lineage without geographic clustering, while core virulence systems remained conserved. An exploratory genome-wide scan of functional variants in isolates with known outcomes (fatal vs. non-fatal) identified a high mutational burden across all isolates. Functional mutations were enriched in fatal cases, particularly in LysR-family transcriptional regulators and molecular chaperones, suggesting that variation in regulatory, stress-response, and metabolic pathways may contribute to disease severity.

Predicted host–pathogen protein–protein interaction analysis highlighted bacterial proteins involved in metabolism, motility, stress response, and DNA repair, with corresponding host pathways enriched for adhesion, endocytosis, and immune regulation.

Collectively, our findings provide the first comparative genomic characterization of clinical *B. pseudomallei* isolates from Bangladesh and suggest that clinical outcome may be influenced by distributed functional variation rather than a single dominant virulence determinant. Given the limited sample size, these observations should be considered exploratory and require validation in larger cohorts.

Introduction

Melioidosis is an infectious disease caused by the Gram-negative environmental bacterium *Burkholderia pseudomallei*, commonly found in soil and surface water. Environmental exposure plays a crucial role in the transmission of the bacterium in tropical regions. Transmission typically occurs through percutaneous inoculation, inhalation, or ingestion (1), and is strongly associated with risk factors such as diabetes mellitus and chronic renal disease. The disease is endemic to Northern Australia and Southeast Asia, with an increasing number of reported cases in Bangladesh. In Bangladesh, seasonal flooding, waterlogging, and agricultural soil exposure facilitate repeated exposure to *B. pseudomallei*, particularly in rural areas (2). Melioidosis remains underdiagnosed, particularly in resource-limited settings where misidentification is common, due to its phenotypic similarity to other pathogens (3).

Clinical manifestations range from localized infections involving a single organ to fulminant, multi-organ disease, and can mimic various other conditions, such as tuberculosis. (4). Management is complicated by the intrinsic resistance of *B. pseudomallei* to multiple classes of antibiotics, including penicillin, ampicillin, early-generation cephalosporins, and aminoglycosides, requiring prolonged combination antibiotic therapy (5).

Genomic technologies provide tools for distinguishing *B. pseudomallei* from closely related species. Whole-genome sequencing allows characterization of antimicrobial resistance determinants. They also allow characterization of key virulence genes, including those involved in capsular polysaccharide synthesis, Type III and Type VI secretion systems, quorum sensing, motility, and intracellular survival, and detect mobile genetic elements that shape pathogenicity and resistance. To date, most studies have focused on presence-absence variation in virulence gene content, and it remains unclear whether disease severity is driven by classical virulence factors or by genome-wide functional variation.

We hypothesize that disease severity in *B. pseudomallei* is driven by genome-wide functional variation rather than differences in classical virulence gene content. Here, we performed whole-genome characterization of 14 clinical *B. pseudomallei* isolates from Bangladesh with 153 publicly available isolates. We examined phylogenetic relationships, virulence, and antimicrobial resistance gene profiles, and performed an unbiased genome-wide analysis of functional variants to identify associations with clinical outcomes. Collectively, these analyses provide new insights into the genomic features of clinical *B. pseudomallei* strains circulating in Bangladesh and identify potential genetic determinants that may contribute to disease severity.

Methods

Ethical Assessment

Ethical consent was obtained from the Ethical Review Committee (ERC) of Bangladesh Diabetic Somiti (BADAS) and in accordance with the Helsinki Declaration [Memo no.BADAS-ERC/EC/25/52, dated May 13, 2025] [Source 9]

Patient Population and Study Setting

The cross-sectional study was conducted at the Department of Microbiology, BIRDEM General Hospital, and NSU Genome Research Institute (NGRI), North South University, Dhaka, Bangladesh. Patients with *B. pseudomallei* isolated from any body fluid culture between September 2014 and August 2021 were included in this study.

Bacterial Strains and Sample Collection

A total of 14 archived *B. pseudomallei* strains were subcultured on blood agar, MacConkey agar, and Ashdown agar media. Phenotypic characterization of the isolates was performed based on Gram staining, culture characteristics, biochemical tests, and monoclonal antibody-based latex agglutination test. Antimicrobial susceptibility testing was performed using the Vitek-2 AST automated system and the Kirby-Bauer disk diffusion method to determine antibiotic susceptibility. Clinical data regarding demographic profiles, clinical diagnoses, signs and symptoms, and laboratory parameters were collected from BIRDEM record-keeping books and databases.

Bacterial DNA Extraction and Sequencing

Genomic DNA of the strains was extracted using an optimized DNA extraction protocol. DNA extraction and purification were performed with the automated Maxwell RSC instrument (Promega Corporation, USA) followed by quantification with the Quantus Fluorometer (Promega Corporation, USA). Library preparation was carried out using the NextSeq system following the protocol in the Illumina DNA Prep reference guide.

Purified isolates were sent to Invent Technologies Ltd., Dhaka, for whole-genome sequencing using the Illumina NextSeq platform (Illumina Inc., San Diego, CA). Paired-end sequencing was performed with the NextSeq Mid Output Reagent Kit (Illumina Inc., San Diego, CA).

Genomic Analysis

The raw reads were quality-checked using FastQC (6) and Fastp (7), followed by *de novo* assembly using SPAdes v3.15.0 (8) with 10x coverage and options to reduce the number of mismatches and short indels. The assembled contigs were scaffolded with MeDuSa v.1.6.0 (6) using a reference genome (Accession number: GCF003927255.1). The scaffolds were aligned using NUCmer (7) to detect SNPs, insertions, deletions, inversions, and rearrangements. The genes were annotated with Prokka v1.14.6 (8) In addition, to evaluate genome similarity and phylogenomic relationship, the average nucleotide identity (ANI) was calculated using FastANI (9). Sequence typing was performed using MLST v2.23.0 (10)

Pangenome analysis was performed using Roary, orthologous gene clusters were identified across all isolates (11), and the gene presence–absence matrix was used for subsequent analysis. Pan- and core-genome accumulation were calculated using a custom R script, an implementation of the (Tettelin et al., 2005) model (12), with genomes added in random order and averaged over 200 permutations. Parameters were estimated by nonlinear least squares in R, with a nonzero asymptote indicating an open pangenome.

Core genome decay was described as:

$Fc(n)=\kappa ce - n/\tau c + \Omega$,

where Ω represents the extrapolated strict core genome size.

The number of novel genes contributed by each additional genome was modeled as:

$Fs(n)=\kappa se - n/\tau s + \gamma$, where γ denotes the asymptotic gene discovery rate.

The total pan-genome size was calculated as:

$P(n) = i=1\sum nFs(i)$.

Model parameters were estimated using nonlinear least-squares regression implemented in R.

Phylogenetic analysis

Complete genomes of all global (n = 153) *B. pseudomallei* isolates were downloaded from the NCBI database. The selected comparative dataset represented 17 countries with the largest contribution from Australia (n = 30), USA (n = 22), Thailand (n = 15), India (n = 13) and Malaysia (n = 10). Most of the isolates in the dataset originated from human sources (35.4%). Unknown sources accounted for 19.3%, while environmental samples made up 11.6% of the total. A small proportion, 2.5%, came from animal sources. Among the clinical samples, blood was the most common source, representing 13% of all entries, followed by sputum samples (3.9%).

Single-nucleotide polymorphisms (SNPs) were called with Snippy v3.1 (13) using default parameters. The core-genome SNPs obtained were aligned with all the isolates that are publicly available on NCBI to infer a phylogeny. Core gene alignment was produced using Snippy, and GUBBINS (14) was run to produce a phylogeny, which was then visualized using iTOL (15).

Detection of AMR, Virulence Determinants, and Profiling of Mobile Genetic Elements

Antibiotic resistance genes were identified using the Comprehensive Antibiotic Resistance Database (CARD) (16). Virulence factors were identified using the Virulence Factor Database (VFDB)(17). Both CARD and VFDB were run using ABRICATE v1.0.1(18). Type VI secretion systems were identified using SecReT6 v3.0 (19). Heatmaps were generated using the packages Tidyverse (20) and pheatmap (21) in RStudio. In addition, mobile genetic elements were identified using Mobile Element Finder (22) and Plasmid Finder (23). Insertion sequences (ISs) were detected using ISEScan v1.7.2.3 (24), and TnCompFinder (25) was used for transposon detection. Prophage detection was done using geNomad (26) and VirSorter (27). Phage defense systems were identified across the 14 *Burkholderia pseudomallei* isolates using Defense-Finder (28). All bioinformatics analyses were performed on the scaffolded contigs using default parameters unless otherwise specified.

Genome-wide screen of functional variants and clinical outcome

An unbiased genome-wide scan using a custom R script and annotated SNP tables from the 14 Bangladeshi *B. pseudomallei* isolates was performed to identify coding sequence variants associated with clinical outcome without restricting the analysis to predefining virulence pathways. In this analysis, only predicted functional mutations were retained, including missense (nonsynonymous) variants, stop-gained variants, stop-lost variants, and indels. The synonymous and intergenic variants were excluded. Each mutation was assigned to its locus tag, gene name, and product description, and gene-level matrices were generated for each isolate based on (i) the presence or absence of at least one functional mutation in a gene and (ii) the number of functional mutations observed in that gene.

For the primary analysis, the 14 isolates were grouped as fatal or surviving, with the unknown outcomes excluded. This resulted in 10 isolates, i.e., 5 surviving and 5 fatal, with usable clinical outcome and SNP data. Associations between gene-level mutation presence and outcome were assessed using Fisher's exact test, and Haldane-corrected odds ratios were reported to account for sparse cell counts. Due to the small sample size and the large number of tested genes, this analysis was considered exploratory and hypothesis-generating rather than predictive. Following an unbiased scan, the highest ranking outcome-associated loci were analyzed.

Host-Pathogen Protein-Protein Interaction (HP-PPI) Network Construction

Amino acid sequences of 153 *B. pseudomallei* isolates were collected. Some isolates were obtained from blood, while others were collected from different body fluids of patients who suffered from fever. To reduce redundancy, all protein sequences were clustered using MMseqs2 (29) (version 13.45111) in easy-cluster mode. The parameters were set as follows: minimum sequence identity 90%, coverage mode 1, and cluster coverage 80%.

To construct the host–pathogen protein–protein interaction (HP-PPI) network, a dataset of 9,333 experimentally validated human–bacterial protein interactions was downloaded from the PHISTO database (30) (accessed on 22 March 2025). After removing duplicate entries which were due to multiple detection methods, 9,031 unique interactions were retained, and the corresponding bacterial protein sequences were retrieved from UniProt (31)

The *B. pseudomallei* proteins were aligned with bacterial proteins in the PHISTO data set using BLASTp (32). (NCBI BLAST+ v2.5.0). The alignment used the following parameters: query coverage %, e-value ≤1e − 4, sequence identity ≥ 30%, bit score > 50 and a maximum of one target hit per query (-outfmt 6) Based on these matches, *B. pseudomallei* proteins were linked to their corresponding human partners from PHISTO, resulting in a predicted host–pathogen interaction network.

The resulting network was visualized and analyzed using the igraph package (version 2.1.4) in R (33). Degree centrality was calculated to identify the top 10 *B. pseudomallei* proteins with the highest number of interactions. These were considered hub proteins for further functional analysis.

Functional Annotation of Hub Proteins and KEGG pathway analysis

The identified hub proteins were functionally annotated using EggNOG-mapper v2 (34).

In order to explore the host biological response, KEGG pathway enrichment analysis was performed on the human proteins interacting with the top 10 *B. pseudomallei* hub proteins based on data from the Kyoto Encyclopedia of Genes and Genomes database (52). The analysis was conducted in R using the clusterProfiler package (version 4.12.6) (35). The Benjamini–Hochberg (BH) method was used to adjust for multiple testing, and the top 15 significantly enriched pathways were selected based on adjusted p-values (FDR < 0.05). Furthermore, the pathways were visualized using ggplot2 (version 3.5.2) (36).

Results

Clinical and Demographic Characteristics of Patients

Fourteen patients with culture-confirmed *B. pseudomallei* infection were included (Supplementary Table 1), all of whom were admitted to BIRDEM General Hospital, Dhaka. The median age was 52 years (range, 30–65 years), and most patients were male (12/14). Diabetes mellitus was the main underlying risk factor, and fever was the most consistent presenting symptom. Isolates were most often obtained from blood (10/14), followed by pus/abscess, urine, sputum, and synovial fluid.

Septicemia was the most common clinical manifestation. Other presentations included septic arthritis, genitourinary and pulmonary involvement, and skin and soft tissue infections, while central nervous system involvement was rare (1/14).

Overall, outcomes were poor: five patients died, five survived, and outcomes were unavailable for four (Table 1)

Table 1
Demographic and Clinical characteristics of *B. pseudomallei* cases (n = 14)

Characteristics	
Age (median, range)	52 (30yrs – 65yrs)
Gender	
Male	12
Female	2
Exposure to Soil	
Yes	14
No	0
Risk Factors	
Diabetes Mellitus	13
Chronic Kidney Disease	3
Alcoholism	1
Source of isolate	
Blood	10
Pus	5
Urine	4
Sputum	2
Synovial Fluid	1
Clinical Presentations	
Septicemia	10
Joint involvement (septic arthritis)	5
Genitourinary involvement (UTI, Prostatic abscess)	4
Pulmonary Involvement (lung abscess, pneumonia)	3
Skin and soft tissue infection	3
CNS involvement	1
Multisystem involvement	12
Fever	14
Specimen Source	
Blood	10
Pus/abscess	5
Urine	4
Sputum	2
Synovial Fluid	2
Prognosis	
Survival	5

Characteristics	
Death	5
Others/Unknown	4

Genomic Characterization and pan-genome analysis of B. pseudomallei strains

Whole-genome sequencing (WGS) confirmed all isolates as *Burkholderia pseudomallei*. Assembly statistics indicated high-quality draft genomes (Table 2). The GenBank accession numbers and metadata are listed in Supplementary Table 2 and Table 2.

Multilocus sequence typing (MLST) showed limited overlap with known sequence types, with most isolates differing by a single allele (Table 2).

Table 2
Assembly Statistics of Draft Genomes for the 14 *B. pseudomallei* Isolates

Assembly	#Scaffolds	Largest contig	Total length	GC (%)	N50	N90	ANI	MLST
1-6643-5556	13	3920451	6912162	67.64	3920451	2877292	98.2628	1124
12-3568	42	3704885	6677007	67.35	3704885	2766741	98.2583	190
29-3504	15	3529528	6126352	67.30	3529528	2561742	98.1851	61
36-5689	27	3756239	6660753	67.26	3756239	2787992	98.2174	190
41-6495	34	3329403	5997591	66.90	3329403	2549413	97.8629	56
42-6983	45	3428482	6208041	67.07	3428482	2620316	97.9142	1637
43-5556	97	2956976	5397360	66.49	2956976	2194551	97.5226	unassigned ST
44-5337	15	3324217	5992868	66.82	3324217	2593165	98.0466	310
45-3087	45	3432358	6176519	66.74	3432358	2550902	97.8229	90
46-5373	19	3993686	7212368	67.96	3993686	3123330	98.667	486
4799	57	3548255	6330928	67.39	3548255	2610782	98.4817	211
49-4667	49	3271756	5854053	66.62	3271756	2401134	97.8746	756
5018	43	3481286	6269910	69.05	3481286	2748770	98.4959	unassigned ST
5411	29	3765973	6678206	67.63	3765973	2781552	98.2622	49
GCF003927255.1	2	3937683	7081111	68.25	3937683	3143428	100	51

Pan-genome analysis identified 40,119 gene clusters. Among these, 1,124 were classified as core genes, 2,216 as soft-core genes, and 3,344 as shell genes. The majority (33,435) were classified as cloud genes, representing rare or strain-specific genes present in fewer than 15% of isolates.

Analysis of genome accumulation patterns showed that core genome size decreased as additional genomes were incorporated, whereas the pan-genome continued to expand (Fig. 1). The pan-genome increased steadily with no indication of reaching a plateau, consistent with an open pan-genome structure.

(a) Core genome size as a function of the number of genomes analyzed. Grey points represent empirical mean values across genome permutations, and the blue curve represents the fitted exponential decay model $Fc(n)=\kappa ce - n/\tau c + \Omega$, extrapolating a strict core genome size of $\Omega = 647$ genes. (b) Pan-genome accumulation curve. Grey points indicate the empirical pan-genome size, and the red curve represents the modeled pan-genome size derived from the fitted function for new gene discovery $Fs(n)=\kappa se - n/\tau s + \gamma$. The non-zero asymptote ($\gamma \approx 197$ genes per genome) indicates an open pangenome.

Phylogenetic Analysis Indicates Regional Relatedness and zoonotic potential of Burkholderia pseudomallei Isolates with Limited Association to Clinical Phenotypes.

Whole-genome phylogenetic analysis was performed on the 14 clinical isolates and 153 publicly available genomes from NCBI (Supplementary Table 5), The reference genome GCF003927255.1 was used for phylogenetic anchoring.

Isolates within the same phylogenetic clade generally showed low SNP divergence (< 10,000 SNPs) while isolates from differing clades were highly divergent (> 15,000 SNPs), indicating substantial evolutionary divergence.

All Bangladeshi isolates fell within the same major clade in the maximum-likelihood phylogenetic tree (Fig. 2). However, no geographic clustering was observed within Bangladesh, even when isolates originated from the same district (Fig. 3). Isolates 5018 and 45-3087 clustered within the same subclade of Clade A in both the phylogenetic tree (Fig. 2) and hierarchical clustering (Fig. 3), despite originating from distant districts. The two isolates differed by 12,889 core-genome SNPs, but were both recovered from pus, and both patients survived (Supplementary Table 1). In addition, these isolates clustered with the Chinese animal-derived clinical isolate BPHN11 (Fig. 2), suggesting possible zoonotic linkage.

The fatal blood isolates 1-6643-5556 and 43-5556 also clustered within the same subclade of Clade A (Fig. 2) and differed by 5,251 core-genome SNPs, with no evident geographic association (Fig. 3). Globally, they clustered closely with an Indian isolate VB29710 (Fig. 2).

In contrast, isolates 5411 and 12-3568 were assigned to different phylogenetic groups (Fig. 2) and were geographically distinct in the hierarchical clustering (Fig. 3), with a pairwise distance of 17,121 SNPs. Despite this genetic divergence, both were recovered from blood and urine and were associated with fatal infection, uncontrolled diabetes, multi-organ involvement, dysuria, and prostatic abscesses (Supplementary Table 1).

More broadly, Bangladeshi isolates were interspersed among clinical isolates from Thailand, India, Japan, and Pakistan in the phylogenetic tree (Fig. 2), with no environmental isolates in immediately adjacent branches. Consistently, no geographic structuring was detected among the 14 Bangladeshi isolates in the hierarchical clustering (Fig. 3). For example, five Gazipur isolates (41-6495, 46-5372, 42-6983, 4799, and 12-3568) were distributed across multiple clades and showed pairwise SNP distances of 8,399–18,405, indicating pronounced local genetic heterogeneity (Fig. 3).

Overall, phylogenetic clustering revealed limited geographic structure (Fig. 3). However, some isolates, such as 1-6643-5556 and 43-5556, clustered with global clinical isolates, suggesting circulation of clinical isolates in neighbouring regions. (Fig. 2; Supplementary Table 1).

Variation in Secretion Systems and Other Virulence-Associated Loci Reveals Distinct and Suggestive Clinicogenomic Patterns

B. pseudomallei isolates harbor a broad repertoire of virulence-associated genes involved in host colonization, immune evasion, intracellular survival, and environmental persistence (Fig. 4). Core virulence genes encoding the capsular protein (*wcbO*), motility proteins (*motA* and *motB*), flagellar structural components (*fli* and *flg*), and the chemotaxis regulator (*cheW*) were detected in all analysed isolates.

The type III secretion system (T3SS) was broadly conserved across all isolates, with regulatory genes (*bsaN, bicA*) present in all genomes. The effector gene *bopE* was detected in 13/14 isolates but was absent in 43-5556. In 43-5556, bopC was detected, giving this isolate a *bopE*-negative/*bopC*-positive profile. Clinically, this isolate was recovered from a fatal bloodstream infection with septicemia, organ abscesses, and CNS involvement. In contrast, *bopA* was absent in isolates 44-5337, 45-3087, and 49-4667, which differed in sample source and clinical outcomes. Additional variation was observed in other T3SS-associated genes, including *bicP*, which was absent in 12-3568 and 43-5556, both fatal cases.

By comparison, the type VI secretion system (T6SS) showed greater variability, although most isolates retained a largely complete set of T6SS structural genes (Fig. 5). Variation was concentrated in *tssH, tssI*, and *tssL*, which separated the isolates into three broad groups. Isolate 5018 formed a distinct outlying cluster due to a markedly increased *tssI* copy number (five copies). A second cluster, comprising 44-5337, 46-5372, 12-3568, 5411, 41-6495, and 42-6983, showed similar *tssH/tssI* profiles, generally with two copies, and was enriched for invasive bloodstream infection: all 6/6 isolates were blood-positive, 5/6 were associated with septicemia, and 4/6

were fatal. In contrast, a third cluster, comprising 43-5556, 49-4667, 4799, 45-3087, 1-6643-5556, 29-3504, and 36-5689, showed increased *tssL* copy number (approximately two to five copies) and a more heterogeneous clinical distribution, including both blood and non-blood isolates and both fatal and nonfatal outcomes. Within this group, the highest *tssL* copy numbers were observed in 29-3504 and 36-5689.

Accessory virulence systems, including the *bsp, pml, bpr*, and *bap* families, showed heterogeneous distribution without clear linkage to clinical outcome. Although certain genes (e.g., *bapC, cdpA, pilP, pilV*) were restricted to individual isolates, these patterns were isolate-specific rather than cohort-associated.

Core flagellar structural genes were conserved across all isolates, whereas variation was observed in chemotaxis-associated loci (*cheD, cheR, cheW, cheY, cheY1*, and *cheZ*). Although no uniform clinical grouping was identified, the absence of *cheR* was observed in 5 of 6 fatal isolates, and 43-5556, a fatal bloodstream isolate with CNS involvement, additionally lacked *cheY* and *cheZ*.

Variation was also noted in other virulence-associated loci, including capsule-related genes. Most isolates retained the *wcb*-associated genes examined, although partial reduction was observed in a small number of isolates, including 5018. In addition, the total number of virulence-associated genes detected varied across the collection, ranging from 66 in 43-5556 to 110 in 46-5372, but these totals did not clearly separate fatal from nonfatal cases. Taken together, these results identify T6SS copy number variation as a clinicogenomic feature in this cohort, whereas, variation in T3SS-associated genes, accessory regulators, capsule-related genes, pili, and chemotaxis loci was more limited or isolate-specific.

Unbiased functional SNP scan identified a small number of novel candidate loci.

An unbiased genome-wide analysis of functional variants identified a restricted set of candidate loci associated with clinical outcome. Across the 10 isolates with both SNP and clinical outcome data available, the overall burden of protein-altering mutations was high and comparable between fatal and survivor groups (mean per isolate: missense, 13,765.9; stop-gained, 184.6; stop-lost, 84.6; indels, 28,450.6), suggesting that mutation burden alone does not account for differences in clinical severity.

At the gene level, a limited number of loci showed enrichment in fatal isolates. The strongest signals were observed in a molecular chaperone (GTC254T_14250) and a LysR-family transcriptional regulator (GTC254T_36690), each harboring protein-altering mutations in all fatal isolates (5/5) and none of the survivors (0/5) (Fisher's exact test, P = 0.0079). Additional loci associated with metabolism, stress response, and transcriptional regulation were similarly enriched in fatal isolates (4/5 vs. 0/5), suggesting that outcome-associated variation may preferentially affect regulatory and adaptive pathways rather than canonical structural virulence systems (Table 4).

To further characterize these signals, a position-specific analysis of protein-altering variants was performed. This analysis revealed two complementary patterns of outcome-associated variation (Table 3).

First, recurrent substitutions at defined coding positions were identified within regulatory genes. Several LysR-family substitutions (A72G, A231G, and A1G) were present in all fatal isolates (5/5) but observed in only one survivor isolate (1/5), representing a previously unrecognized enrichment pattern. In addition, two substitutions, C184T (QUF21_RS19095) and A246C (QUF21_RS29730), were detected in 4/5 fatal isolates and absent in all survivors.

Second, locus-level accumulation of independent mutations was observed in metabolic genes. In the gluconate 2-dehydrogenase gene (QUF21_RS03700), the substitution C50T was identified in 3/5 fatal isolates and absent in survivors, while 16 additional distinct substitutions were individually detected in fatal isolates only. This pattern suggests accumulation of independent mutations within the same locus rather than recurrence of a single dominant variant.

A similar gene-level pattern was observed in several additional loci, including the molecular chaperone, H-NS family protein, ThiS sulfur carrier protein, and NAD(P)H oxidoreductase genes. These loci contained multiple protein-altering variants distributed across distinct positions without recurrent shared substitutions, consistent with gene-level functional perturbation rather than selection at a single recurrent site.

Detailed variant-level data, including isolate-specific substitutions and positional distributions, are provided in Supplementary Table 9.

Together, these findings suggest that clinical outcome may be associated less with overall mutation burden or variation in canonical virulence genes, and more with convergent mutation patterns affecting regulatory, metabolic, and stress-response pathways. Specifically, outcome-associated variation appears to involve both recurrent position-specific substitutions in regulatory loci and gene-level accumulation of diverse protein-altering mutations within adaptive pathways, consistent with convergent functional disruption across isolates.

Table 3
Position-specific protein-altering mutations enriched in fatal isolate

Gene/ Product	Locus_tag	Specific Substitutions	Position (AA)	Fatal Isolates (n = 5)	Survivor isolates (n = 5)	Classification	Interpretation
LysR family transcriptional regulator	QUF21_RS19095	C184T	184	4/5	0/5	Fatal-specific	Strong regulatory candidate
LysR family transcriptional regulator	QUF21_RS29730	A246C	246	4/5	0/5	Fatal-specific	Strong regulatory candidate
LysR substrate-binding domain protein	QUF21_RS20370	A72G	72	5/5	1/5	Enriched	Regulatory modulation
LysR family transcriptional regulator	QUF21_RS07610	A231G	231	5/5	1/5	Enriched	Regulatory modulation
LysR substrate-binding domain protein	QUF21_RS04850	A1G	1	5/5	1/5	Enriched	Likely structural/regulatory effect
Gluconate 2-dehydrogenase subunit 3	QUF21_RS03700	C50T	50	5/5	0/5	Fatal-associated	Metabolic adaptation
Gluconate 2-dehydrogenase subunit 3	QUF21_RS03700	Multiple single substitutions*	various	1/5	0/5	Isolate-specific	Local variation within locus
Molecular chaperone	GTC254T_14250	Multiple variants (missense + indel)	multiple	5/5	0/5	Gene-level fatal-specific	Stress-response alteration
H-NS histone family protein	GTC254T_46860	Missense substitutions	multiple	4/5	0/5	Enriched	Chromatin regulation
ThiS sulfur carrier protein	GTC254T_31620	Missense + indel	multiple	4/5	0/5	Enriched	Stress/metabolism
NAD(P)H oxidoreductase	GTC254T_30610	Missense substitutions	multiple	4/5	0/5	Enriched	Redox adaptation

Table 4
Top candidate loci identified by unbiased SNP scan

Gene / product	Locus tag	Fatal isolates with mutation	Surviving isolates with mutation	Odds ratio	P value
molecular chaperone	GTC254T_14250	5/5	0/5	121.0	0.0079
LysR family transcriptional regulator	GTC254T_36690	5/5	0/5	121.0	0.0079
gluconate 2-dehydrogenase subunit 3 family protein	GTC254T_07310	4/5	0/5	33.0	0.0476
MliC family protein	GTC254T_10930	4/5	0/5	33.0	0.0476
LysR substrate-binding domain-containing protein	GTC254T_14460	4/5	0/5	33.0	0.0476
NAD(P)H-dependent oxidoreductase	GTC254T_30610	4/5	0/5	33.0	0.0476
sulfur carrier protein ThiS	GTC254T_31620	4/5	0/5	33.0	0.0476
HPr family phosphocarrier protein	GTC254T_46200	4/5	0/5	33.0	0.0476
H-NS histone family protein	GTC254T_46860	4/5	0/5	33.0	0.0476
chorismate lyase	GTC254T_56370	4/5	0/5	33.0	0.0476

Antimicrobial Resistance Profile and AST Correlation

All isolates contained genes encoding components of resistance–nodulation–cell division (RND) family efflux systems, including *opcM*, *ceoA*, *ceoB*, and *amrB* encoding the *CeoAB–OpcM* and *AmrAB–OprM* efflux complexes, known to mediate resistance to aminoglycosides, fluoroquinolones, and macrolides. Consistently, all isolates exhibited phenotypic resistance to aminoglycosides (amikacin, gentamicin) and demonstrated resistance to colistin.

All isolates additionally carried OXA-type β-lactamase genes (*OXA-43*, *OXA-57*, and *OXA-59)*, and outer membrane porin gene (*omp38)*, associated with reduced membrane permeability (Fig. 6). Despite this, all isolates remained phenotypically susceptible to ceftazidime and carbapenems (meropenem and imipenem) in addition to amoxicillin–clavulanate, doxycycline, and trimethoprim–sulfamethoxazole. Minor variation was observed for trimethoprim–sulfamethoxazole susceptibility, with isolates 36-5689, 17-6048, and 5018 showing reduced susceptibility; however, this was not associated with distinct resistance gene profiles.

Importantly, no isolate-specific differences in resistance gene content or AST profiles were observed that could explain variation in clinical outcome. Fatal cases (associated with isolates 1-6643-5556, 5411, 41-6495, 12-3568, 44-5337, and 43-5556) and surviving cases were caused by isolates with indistinguishable resistomes and susceptibility patterns, indicating that antimicrobial resistance was not a determinant of disease severity or mortality in this cohort.

Burkholderia pseudomallei Isolates Exhibit Variable Mobile Genetic Element Burden with No Consistent Relationship to Clinical or AMR Phenotypes

MGE burden ranged from 3 to 14 elements per isolate, comprising insertion sequences, transposons, and prophages. The highest burden was observed in isolate 43-5556 (14 elements), whereas the lowest was observed in 5018 (3 elements).

Overall, no consistent pattern was observed between MGE burden and clinical manifestations. Fatal isolates exhibited a median MGE burden of 7.5 (range, 5–14), compared with 8 (range, 3–11) among isolates with survival or cure.

Fatal invasive cases with blood and urine positivity, including 5411, 12-3568, and 1-6643-5556, showed variable MGE burdens of 5, 6, and 8, whereas localized non-blood surviving cases such as 45-3087 and 5018 had totals of 11 and 3, respectively. Similarly, MGE burden did not correlate with disease presentation, as both bloodstream-associated fatal cases and localized infections were observed across the full range of MGE counts.

Prophage carriage was not enriched in fatal cases and was more frequently detected in non-fatal isolates, further supporting the lack of association between MGEs and disease severity.

Comparison with antimicrobial resistance (AMR) gene profiles showed no consistent association with mobile genetic element (MGE) burden. Isolates within the OXA-57-associated group exhibited a slightly higher mean MGE burden than those in the OXA-59-dominant group (9.8 vs 7.8 elements), but this difference was driven by a single high-burden outlier and showed substantial overlap between groups. The isolate with the lowest MGE burden (5018) carried a distinct AMR gene profile (*opcM*, *amrB*, and *OXA-43*), but similar MGE ranges were observed across other AMR backgrounds.

Similarly, no association was observed between MGE burden and phenotypic antimicrobial susceptibility. Isolates with higher MGE counts did not display distinct antibiotic susceptibility patterns, and resistance remained restricted to a limited subset of antibiotic classes across the cohort.

Phage Defense System Diversity Shows Partial Association with Clinical Severity

Phage defense systems identified by Defense-Finder were heterogeneous across the 14 *B. pseudomallei* isolates (Figs. 8a and b). Restriction–modification systems were the most prevalent class, whereas accessory systems, including Gabija and Wadjet, were restricted to two isolates (43-5556 and 1-6643-5556), both of which were associated with invasive bloodstream infection and multi-organ involvement.

Abortive infection systems were more widely distributed (6/14 isolates) and were present in both severe and non-severe cases, indicating no clear segregation with clinical outcome. Three isolates lacked complete detectable defense systems; however, these displayed heterogeneous clinical presentations, further suggesting no consistent association between defense system presence and disease phenotype.

No consistent geographic structuring of defense system profiles was observed. Among isolates from the same district, defense repertoires varied substantially, ranging from absent to expanded system content. In addition, phage defense profiles did not show concordance with antimicrobial resistance groups or total mobile genetic element burden.

Overall, phage defense system diversity was high but showed no consistent association with geographic origin, AMR genotype, or clinical outcome, aside from limited isolate-specific overlap with invasive bloodstream-associated disease.

Predicted host–pathogen protein interaction network

Following BLASTp alignment of clustered Burkholderia pseudomallei proteins against bacterial proteins in the PHISTO database, 1,996 homologous *B. pseudomallei* proteins were identified. Based on their correspondence to known human–bacterial interactions in PHISTO, 2,806 putative protein–protein interactions (PPIs) were predicted between *B. pseudomallei* and human proteins (Fig. 9). The host–pathogen PPI network was analyzed using igraph in R, and degree centrality was used to identify the most highly connected bacterial nodes. The top 10 *B. pseudomallei* hub proteins had degree values ranging from 15 to 31 and were selected for functional annotation using EggNOG-mapper v2 (Table 5).

The top bacterial hub proteins were associated mainly with metabolism, motility, stress response, and DNA repair (Table 5). These included *PepP* and *PurD*, linked to protein and nucleotide metabolism, *FliN*, associated with flagellar function, *RecN*, involved in DNA repair, and *TypA*, a GTP-binding protein implicated in stress-responsive translational regulation. Although these are not classical secretion-system effectors, their predicted centrality in the network suggests that core bacterial functions involved in adaptation, motility, and intracellular survival may contribute substantially to host interaction. This interpretation is consistent with previous studies showing that *B. pseudomallei* pathogenesis depends not only on dedicated virulence factors but also on broader physiological systems that support intracellular survival and persistence within host cells (37)

To examine host functions associated with these bacterial hubs, KEGG pathway enrichment analysis was performed on the corresponding human interacting proteins (Fig. 9). The top enriched pathways included Adherens junction, Endocytosis, Phagosome, Antigen processing and presentation, Apoptosis, and Bacterial invasion of epithelial cells. These pathways are notable because they overlap closely with experimentally supported stages of *B. pseudomallei* infection, including host-cell entry, phagosomal trafficking and escape, intracellular survival, and intercellular spread. In particular, prior work has shown that *B. pseudomallei* uses T3SS-3-

associated mechanisms to promote invasion and escape from phagosomal compartments, while T6SS-5 supports intracellular survival, multinucleated giant cell formation, and cell-to-cell spread in mammalian hosts.

These enrichments therefore, complement the virulence analysis presented above, in which T3SS was broadly conserved and T6SS showed the clearest clinicogenomic association, particularly among severe bloodstream-associated isolates. The predicted PPI and KEGG results support the biological relevance of these virulence systems by highlighting host pathways involved in cell entry, intracellular trafficking, epithelial interaction, and immune regulation, all of which are central to melioidosis pathogenesis. At the same time, because these interactions were inferred computationally from homology-based prediction rather than measured directly, they should be interpreted as supportive rather than definitive evidence linking the identified bacterial proteins to clinical severity.

Table 5
Functional annotation of the top 10 *B. pseudomallei* hub proteins.

Serial No.	Uniprot ID	Degree	Gene Name	Description
1	Q5NF74	31	-	Catalyzes the transfer of a two-carbon ketol group from a ketose donor to an aldose acceptor, via a covalent intermediate with the cofactor thiamine pyrophosphate
2	Q7CGS9	26	$pepP$	Aminopeptidase P
3	Q81LD0	24	$hemL$	Glutamate-1-semialdehyde aminotransferase
4	Q81S14	23	$mtaD$	Catalyzes the deamination of 5-methylthioadenosine and S-adenosyl-L-homocysteine into 5-methylthioinosine and S-inosyl-L- homocysteine, respectively. Is also able to deaminate adenosine
5	Q7CHZ0	20	$fliN$	FliN is one of three proteins (FliG, FliN, FliM) that form the rotor-mounted switch complex (C ring), located at the base of the basal body. This complex interacts with the CheY and CheZ chemotaxis proteins, in addition to contacting components of the motor that determine the direction of flagellar rotation
6	Q5NGF1	20	$purD$	Belongs to the GARS family
7	Q81VA8	16	-	formate dehydrogenase, alpha subunit
8	Q5NFP9	16	-	GTP-binding protein TypA
9	Q8CZZ7	15	$recN$	May be involved in recombinational repair of damaged DNA
10	Q8ZA76	15	$glgX$	Removes maltotriose and maltotetraose chains that are attached by 1,6-alpha-linkage to the limit dextrin main chain, generating a debranched limit dextrin

Discussion

This study provides genomic characterization of clinical *B. pseudomallei* isolates from Bangladesh and places these strains within the broader global population framework of this pathogen. *B. pseudomallei* is an environmental bacterium that usually persists in soil and surface water. Human infection develops following environmental exposure rather than through person-to-person transmission. The Bangladeshi clinical isolates analyzed here likely represent only a subset of the strains capable of causing human infection (2).

The Bangladeshi isolates exhibit high genomic diversity, with an open pangenome consistent with a high recombination rate and genome plasticity (38). Core virulence genes and the antimicrobial resistance genes are conserved in these isolates. Genome-wide SNP analysis also revealed that clinical outcome is not explained by key virulence or resistance genes; instead, a small number of regulatory and stress-response genes were enriched in fatal isolates, potentially contributing to differences in disease severity.

Consistent with large-scale comparative genomic analyses from Southeast Asia and Northern Australia, the isolates analyzed represent a genetically heterogeneous population embedded within the Asian lineage rather than a clonal expansion (39). The interspersion of Bangladeshi isolates among Indian and Thai strains, and the absence of geographic clustering, support the interpretation that melioidosis in Bangladesh is primarily environmentally acquired through exposure to contaminated water and soil

rather than driven by healthcare-associated transmission. (2) These findings support the need for integrated clinical and environmental surveillance to better understand environmental reservoirs and transmission pathways. From a One Health perspective, understanding melioidosis in Bangladesh involves combining human health data with environmental monitoring. *B. pseudomallei* naturally persists in soil and surface water, meaning that the rise of melioidosis depends on interactions among environmental factors such as seasonal flooding and waterlogging, farming practices, climate conditions, and human exposure behaviors. The lack of geographic clustering among Bangladeshi isolates further highlights the role of environmental reservoirs in transmission. Therefore, combining clinical genomics with ongoing sampling of soil, water, livestock-related environments, and hydrological systems will offer a clearer picture of how the pathogen circulates and exposes individuals in Bangladesh.

The close clustering of isolates 1-6643-5556 and 43-5556 with the Indian isolate VB29710 is an important finding. Despite not being geographically linked, both came from patients with severe disease and fatal outcomes. In addition, their grouping with VB29710 suggests that they may belong to a wider South Asian lineage of clinical importance. However, VB29710 is a rare gentamicin-susceptible harboring insertion of R962 and L963 into AmrB, the transmembrane protein of the AmrAB-OprA efflux pump (40). This indicates that antibiotic susceptibility profiles may not correlate with phylogenetic clustering and warrants further genomic and phenotypic studies.

Our pangenome analysis confirms that *B. pseudomallei* maintains an open pangenome, a defining feature previously described in global datasets (39). The dual-chromosome architecture of *B. pseudomallei* provides an important structural basis for genomic plasticity and reflects the species' high recombination rate and ecological adaptability (39). Our findings from Bangladeshi isolates reinforce the idea that the disease-causing ability comes from a shared set of core virulence genes within a flexible accessory genome that can keep changing over time.

The core virulence genes of *B. pseudomallei* mostly remained conserved, including the capsular polysaccharide operon (*wcb*) and T3SS, consistent with studies from Thailand, Malaysia, and Australia (2, 39, 41). Functional evidence indicates that the capsule protects against complement-mediated killing and supports bloodstream persistence (42), while T3SS regulators such as *bsaN* coordinate virulence gene expression and promote intracellular survival (5). Previous studies report that inter-strain variation was confined to accessory regulators and effector repertoires rather than to the core virulence clusters (41, 42). In our cohort, this pattern was most evident in the T6SS: isolates with similar *tssH/tssI* profiles were predominantly blood-associated and included several fatal cases. However, these patterns were not exclusive to specific outcomes and were not formally tested for statistical association.

This observation is noteworthy because prior work has shown that genomic variation in *B. pseudomallei* can modulate virulence-related traits. One study reported that some environmental isolates exhibited genomic loss associated with reduced intracellular replication and decreased multinucleated giant cell formation, supporting a link between genomic differences and pathogenic behavior (43). Another study identified sequence variation in *hcp1* and related T6SS proteins among clinical isolates, indicating that the T6SS locus is not completely uniform across strains (43).

In addition, a recent transcriptomics study of *B. pseudomallei* cultured under environmental and clinical conditions demonstrated differential gene expression. In this study, a total of 455 genes were differentially regulated. Genes involved in energy metabolism and cellular processes were upregulated, whereas downregulated genes included genes involved in fatty acid & phospholipid metabolism, amino acid biosynthesis, and regulatory function proteins (44). However, we did not find previous reports specifically linking copy number variation in *tssL, tssH*, or *tssI* with clinical source or outcome. Together, these findings suggest that the fundamental pathogenic mechanisms of *B. pseudomallei* remain stable, while finer-scale variation, particularly within the T6SS, may influence host–pathogen interactions and clinical disease presentation.

All 14 isolates showed the expected intrinsic resistance profile, characterized by RND-family efflux components (opcM, ceoA, ceoB, and amrB) and OXA-type β-lactamases (including omp38). Consistent with this, the strains were phenotypically resistant to aminoglycosides and colistin but remained susceptible to first-line therapeutics, including ceftazidime, meropenem, amoxicillin–clavulanate, doxycycline, and trimethoprim–sulfamethoxazole. The maintenance of susceptibility to ceftazidime and carbapenems, despite the presence of *OXA-43, OXA-57, OXA-59*, and *omp38*, indicates that these isolates carried baseline intrinsic resistance determinants rather than additional mutational or regulatory changes typically associated in treatment failure. This is notable because, although resistance to frontline melioidosis therapy is uncommon, it is clinically important when it emerges. (45).

In this context, the lack of any clear association between AMR profile and fatal outcome is important because it indicates that, unlike the partial clinicogenomic pattern observed for T6SS and the weaker trends seen for some defense-system profiles, antimicrobial resistance was unlikely to be a major driver of disease severity in this cohort. Instead, clinical outcome more likely reflected host susceptibility and virulence-associated differences than variation in baseline resistance genotype.

In contrast to the relative stability of virulence gene presence–absence profiles, genome-wide functional variant analysis identified a small number of candidate loci that may contribute to disease severity. Functional mutations in genes encoding molecular chaperone–associated proteins and *LysR* family transcriptional regulators were consistently observed in fatal isolates but were absent in survivor isolates within the subset analyzed. Additional candidate loci were associated with metabolic and stress-response functions.

One of the most notable findings of this study was the enrichment of functional variation in *LysR*-family regulatory genes among fatal isolates. Most of the identified *LysR*-associated loci encoded transcriptional regulators, a conserved family known to control diverse bacterial processes including virulence, metabolism, quorum sensing, motility, stress adaptation, and intracellular survival (46). Unlike canonical virulence determinants, *LysR*-type regulators function as global regulatory nodes capable of coordinating broad transcriptional responses to environmental and host-associated stress. In *Brucella abortus*, disruption of a *LysR*-family regulator reduced intracellular survival and attenuated virulence in both in vitro and in vivo models (47). Similarly, alteration of the BCAS02225 *LysR*-type regulator in *B. cenocepacia* impaired motility, reduced biofilm formation, and diminished virulence-associated phenotypes (48). In *Listeria monocytogenes*, deletion of a *LysR*-family regulator reduced bacterial adhesion and invasion capacity (49).

Importantly, our analysis identified recurrent, position-specific *LysR*-family substitutions enriched among fatal *B. pseudomallei* isolates. Multiple amino acid substitutions, including A72G, A231G, and A1G, were detected in fatal isolates but were uncommon in survivors, suggesting convergent alteration of regulatory loci associated with severe disease. Although functional validation was beyond the scope of this study, the recurrence of substitutions at defined coding positions raises the possibility that these variants affect conserved functional domains involved in transcriptional regulation or environmental sensing.

In addition to recurrent regulatory substitutions, accumulation of multiple independent mutations was observed within metabolic and stress-response genes. Unlike the recurrent *LysR*-family variants, these loci contained diverse protein-altering mutations distributed across multiple coding positions, suggesting gene-level functional changes rather than selection of a single dominant mutation. These findings support a model in which severe melioidosis may be associated less with variation in canonical virulence systems and more with changes in regulatory, metabolic, and stress-response pathways.

The predicted host–pathogen PPI network and KEGG enrichment results are consistent with the established intracellular lifestyle of *B. pseudomallei*, which enters host cells, survives intracellularly, escapes phagosomal compartments, and spreads cell-to-cell during infection (37). The pathway enrichment is also concordant with the virulence architecture identified in our isolates, particularly the broad conservation of T3SS and the stronger clinicogenomic signal associated with T6SS, because these systems are central to phagosomal escape and intracellular dissemination. In particular, (50) demonstrated the contribution of *T3SS/BopA* to intracellular trafficking and escape, whereas (51) reviewed the role of T6SS-5 in intracellular survival, multinucleated giant cell formation, and cell-to-cell spread. Together, these results suggest that the inferred PPI network captures host processes mechanistically linked to invasion, intracellular persistence, and tissue-invasive disease in melioidosis, although these predictions remain computational and require experimental validation (37).

While this study provides important insights into the genomic diversity and clinicogenomic characteristics of clinical *B. pseudomallei* isolates in Bangladesh, it is limited by the relatively small sample size and single-center study design. In particular, the outcome-associated functional variant analysis was exploratory and based on a limited number of isolates with available clinical outcome data. Functional validation of the identified LysR-family regulatory variants and other candidate adaptive loci was beyond the scope of this study. A key limitation of the study is the lack of environmental isolates, which prevents the characterization of the environmental reservoirs, hydrological pathways and soil-water transmission dynamics of *B. pseudomallei.* Future studies integrating larger cohorts, environmental sampling, transcriptomics, and experimental validation will be important to determine whether the regulatory and metabolic alterations identified here contribute directly to intracellular survival, host adaptation, and severe disease. Furthermore, a One Health surveillance framework that integrates clinical genomics, environmental sampling, geospatial risk mapping, hydrological data, climatic variables, agricultural land use, and animal health surveillance is crucial for identifying high-risk landscapes and transmission hotspots. Such an approach would enable early detection of environmental reservoirs, enhance understanding of the

seasonal and ecological drivers of disease, and support targeted public health interventions for melioidosis prevention in Bangladesh.

Overall, our findings suggest that clinical outcome in melioidosis may be influenced not only by canonical virulence determinants, but also by genome-wide adaptive variation affecting regulatory and stress-response pathways, highlighting an underexplored dimension of *B. pseudomallei* pathogenesis.

Declarations

Data Availability

Whole Genome Sequencing Data generated in this study has been deposited in the European Nucleotide Archive (ENA) under BioProject Accession Number PRJEB102501. All data supporting the findings of this study are available within the paper and its Supplementary Information.

Author statements

Author contributions

Conceptualization: LB, MSAJ, MJ, PB, CRA

Data curation: IRJ, MRS, TR, MRH

Formal analysis: LB, IRJ, SS, NI, FK, MH

Funding acquisition: LB, MJ, PB, MH

Investigation: LB, MRS, TR, MRH

Methodology: IRJ, SS, FK, MJ, MJ, PB, PB

Project administration: LB, MH

Resources: LB, MRS, TR, MRH, MJ, PB, MSAJ

Supervision: LB, MSAJ, MJ, PB, CRA MH

Validation: IRJ, MSAJ, MJ, PB, CRA, MH

Visualization: IRJ, SS, NI, FK

Writing – original draft: LB, IRJ, MJ, PB, MH

Writing – review & editing: LB, MSAJ, MJ, PB, CRA, MH

Conflicts of interest

The authors declare that they have no conflicts of interest.

Funding information

No funding was available for conducting the experiments. IRJ received stipend support from NSU

CTRG (grant No. 323630_221864/1) for the analysis.

Ethical approval

Ethical consent was taken from the Ethical Review Committee (ERC) of Bangladesh

Diabetic Somiti (BADAS) and in accordance with the Helsinki Declaration [Memo no.BADAS-

ERC/EC/25/52, dated May 13, 2025].

Consent for publication

The study was approved by the Institutional Ethics Review Committee of Bangladesh Diabetic Somiti

(BADAS) (Approval No. BADAS-ERC/EC/25/52, and written informed consent was obtained from

all participants before enrolment in the study.

Acknowledgements

We thank the staff of BIRDEM General Hospital for assistance with sample collection and

clinical data. We also acknowledge the staff of the Department of Microbiology, BIRDEM

General Hospital for providing laboratory resources and technical support. We acknowledge MJ and PB for environmental interpretation and Bangladesh-specific soil–water exposure framing of *B. pseudomallei*.

References

1. Mohapatra, P. R. & Mishra, B. Burden of melioidosis in India and South Asia: Challenges and ways forward. *Lancet Reg. Health - Southeast. Asia*. **2**, 100004. 10.1016/j (2022).
2. Cheng, A. C. & Currie, B. J. Melioidosis: Epidemiology, pathophysiology, and management. *Clin. Microbiol. Rev.* 383–416. 10.1128/CMR.18.2.383-416 (2005). 2005 PubMed PMID: 15831829.
3. Peera, H., D, B. J. & Risa, H. Burkholderia pseudomallei: Challenges for the Clinical Microbiology Laboratory. *J. Clin. Microbiol.* **54** (12), 2866–2873. 10.1128/jcm.01636-16 (2016).
4. Edwardes, C. & Jarratt-Barnham, I. Neonatal patients with melioidosis: a systematic review. *BMC Infect. Dis. BioMed. Cent. Ltd.* 10.1186/s12879-025-11343-1 (2025).
5. Chen, Y. et al. Characterization and analysis of the Burkholderia pseudomallei BsaN virulence regulon. *BMC Microbiol.* **14** (1). 10.1186/s12866-014-0206-6 (2014). PubMed PMID: 25085508.
6. Bosi, E. et al. MeDuSa: a multi-draft based scaffolder. *Bioinformatics* **31** (15), 2443–2451. 10.1093/bioinformatics/btv171 (2015).
7. Marçais, G. et al. MUMmer4: A fast and versatile genome alignment system. *PLoS Comput. Biol.* **14** (1). 10.1371/journal.pcbi.1005944 (2018). PubMed PMID: 29373581.
8. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30** (14), 2068–2069. 10.1093/bioinformatics/btu153 (2014).
9. Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* **9** (1). 10.1038/s41467-018-07641-9 (2018). PubMed PMID: 30504855.
10. 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** 10.12688/wellcomeopenres.14826.1 (2018).
11. Page, A. J. et al. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31** (22), 3691–3693. 10.1093/bioinformatics/btv421 (2015).
12. Tettelin, H. et al. Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: Implications for the microbial 'pan-genome' [Internet]. (2005). Available from:
13. Seemann, T. & Snippy Rapid Haploid Variant Calling and Core Genome Alignment.
14. Croucher, N. J. et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res.* **43** (3), e15–e15. 10.1093/nar/gku1196 (2015).
15. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49** (W1), W293–W296. 10.1093/nar/gkab301 (2021).

16. Alcock, B. P. et al. CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Res.* **51** (1 D), D690–D699. 10.1093/nar/gkac920 (2023). PubMed PMID: 36263822.
17. Chen, L. VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res.* **33** (Database issue), D325–D328. 10.1093/nar/gki008 (2004).
18. Seemann, T. https://github.com/tseemann/abricate [Internet]. [cited 2026 Jan 16]. ABRicate: Mass screening of contigs for antimicrobial and virulence genes. (2016). Available from: https://github.com/tseemann/abricate
19. Li, J. et al. SecReT6: a web-based resource for type VI secretion systems found in bacteria. *Environ. Microbiol.* **17** (7), 2196–2202. 10.1111/1462-2920.12794 (2015).
20. Wickham, H. et al. Welcome to the Tidyverse. *J. Open. Source Softw.* **4** (43), 1686. 10.21105/joss.01686 (2019).
21. Kolde, R. & pheatmap Pretty Heatmaps. *CRAN: Contributed Packages*. 10.32614/CRAN.package.pheatmap (2010).
22. Johansson, M. H. K. et al. Detection of mobile genetic elements associated with antibiotic resistance in Salmonella enterica using a newly developed web tool: MobileElementFinder. *J. Antimicrob. Chemother.* **76** (1), 101–109. 10.1093 (2021). /JAC/DKAA390 PubMed PMID: 33009809.
23. Carattoli, A. & Hasman, H. PlasmidFinder and In Silico pMLST: Identification and Typing of Plasmid Replicons in Whole-Genome Sequencing (WGS). In. pp. 285–94. (2020). 10.1007/978-1-4939-9877-7_20
24. Xie, Z. & Tang, H. ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. *Bioinformatics* **33** (21), 3340–3347. 10.1093/bioinformatics/btx433 (2017). PubMed PMID: 29077810.
25. Ross, K. et al. TnCentral: a Prokaryotic Transposable Element Database and Web Portal for Transposon Analysis. *mBio* **12** (5). 10.1128/ (2021). mBio.02060-21 PubMed PMID: 34517763.
26. Camargo, A. P. et al. Identification of mobile genetic elements with geNomad. *Nat. Biotechnol.* **42** (8), 1303–1312. 10.1038/s41587-023- (2024). 01953-y PubMed PMID: 37735266.
27. Roux, S., Enault, F., Hurwitz, B. L. & Sullivan, M. B. VirSorter: mining viral signal from microbial genomic data. *PeerJ* **3**, e985. 10.7717/peerj.985 (2015).
28. Tesson, F. et al. Systematic and quantitative view of the antiviral arsenal of prokaryotes. *Nat. Commun.* **13** (1). 10.1038/s41467-022-30269-9 (2022). PubMed PMID: 35538097.
29. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35** (11), 1026–1028. 10.1038/nbt.3988 (2017).
30. Durmuş Tekir, S. et al. PHISTO: pathogen–host interaction search tool. *Bioinformatics* **29** (10), 1357–1358. 10.1093/bioinformatics/btt137 (2013).
31. Bateman, A. et al. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.* **51** (D1), D523–D531. 10.1093/nar/gkac1052 (2023).
32. Camacho, C. et al. BLAST+: Architecture and applications. *BMC Bioinform.* **10** 10.1186/1471-2105-10-421 (2009). PubMed PMID: 20003500.
33. Csárdi, G. et al. igraph: Network Analysis and Visualization. CRAN: Contributed Packages. (2006). 10.32614/CRAN.package.igraph
34. Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol. Biol. Evol.* **38** (12), 5825–5829. 10.1093/molbev/msab293 (2021).
35. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS* **16** (5), 284–287. 10.1089/omi.2011.0118 (2012).
36. Wickham, H. & Cham ggplot2. Springer International Publishing; (2016). 10.1007/978-3-319-24277-4
37. Allwood, E. M., Devenish, R. J., Prescott, M., Adler, B. & Boyce, J. D. Strategies for intracellular survival of Burkholderia pseudomallei. Front Microbiol. ;2(AUG). (2011). 10.3389/fmicb.2011.00170 PubMed PMID: 22007185.
38. Spring-Pearson, S. M. et al. Pangenome analysis of burkholderia pseudomallei: Genome evolution preserves gene order despite high recombination rates. *PLoS One*. **10** (10). 10.1371/journal.pone.0140274 (2015). PubMed PMID: 26484663.

39. Holden, M. T. G. et al. Genomic plasticity of the causative agent of melioidosis, Burkholderia pseudomallei [Internet]. (2004). Available from:
40. Bakthavatchalam, Y. D. et al. Genomics and structural insight into the masking of gentamicin-resistance in clinical Burkholderia pseudomallei strain VB29710 from India. *Diagn. Microbiol. Infect. Dis.* **105** (3), 115878. 10.1016/j.diagmicrobio.2022.115878 (2023).
41. Meumann, E. M., Limmathurotsakul, D., Dunachie, S. J., Wiersinga, W. J. & Currie, B. J. Burkholderia pseudomallei and melioidosis. *Nat. Reviews Microbiol. Nat. Res.* 155–169. 10.1038/s41579-023-00972-5 (2024). PubMed PMID: 37794173.
42. Reckseidler-Zenteno, S. L., DeVinney, R. & Woods, D. E. The capsular polysaccharide of Burkholderia pseudomallei contributes to survival in serum by reducing complement factor C3b deposition. *Infect. Immun.* **73** (2), 1106–1115 (2005). .2005 PubMed PMID: 15664954.
43. Tandhavanant, S. et al. Genetic variation of hemolysin co-regulated protein 1 affects the immunogenicity and pathogenicity of Burkholderia pseudomallei. *PLoS Negl. Trop. Dis.* **19** (1). 10.1371/journal.pntd.0012758 (2025). PubMed PMID: 39761280.
44. Kong, C. et al. Transcriptional landscape of Burkholderia pseudomallei cultured under environmental and clinical conditions. *Microb. Genom*. **9** (4). 10.1099/mgen.0.000982 (2023). PubMed PMID: 37018040.
45. Bugrysheva, J. V. et al. Antibiotic Resistance Markers in Burkholderia pseudomallei Strain Bp1651 Identified by Genome Sequence Analysis [Internet]. Available from: https://doi.org/10 (2017).
46. Maddocks, S. E. & Oyston, P. C. F. Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. *Microbiology* 3609–3623. 10.1099/mic.0.2008/022772 (2008). -0 PubMed PMID: 19047729.
47. Sheehan, L. M., Budnick, J. A., Blanchard, C., Dunman, P. M. & Caswell, C. C. A LysR-family transcriptional regulator required for virulence in Brucella abortus is highly conserved among the α-proteobacteria. *Mol. Microbiol.* **98** (2), 318–328. 10.1111/mmi.13123 (2015). PubMed PMID: 26175079.
48. Bernier, S. P., Nguyen, D. T. & Sokol, P. A. A LysR-type transcriptional regulator in Burkholderia cenocepacia influences colony morphology and virulence. *Infect. Immun.* **76** (1), 38–47. 10.1128/IAI.00874-07 (2008). PubMed PMID: 17967860.
49. Abdelhamed, H. et al. Contributions of a LysR transcriptional regulator to Listeria monocytogenes virulence and identification of its regulons. *J. Bacteriol.* **202** (10). 10.1128/JB.00087-20 (2020). PubMed PMID: 32179628.
50. Gong, L. et al. The Burkholderia pseudomallei type III secretion system and BopA are required for evasion of LC3-associated phagocytosis. *PLoS One*. **6** (3). 10.1371/journal.pone.0017852 (2011). PubMed PMID: 21412437.
51. Lennings, J., West, T. E. & Schwarz, S. The Burkholderia Type VI secretion system 5: Composition, regulation and role in virulence. *Front. Microbiol. Front. Media S A*. 10.3389/fmicb.2018.03339 (2019).
52. Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M. & Tanabe, M. KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res.* **49** (D1), D545–D551. https://doi.org/10.1093/nar/gkaa970 (2020).

Figures

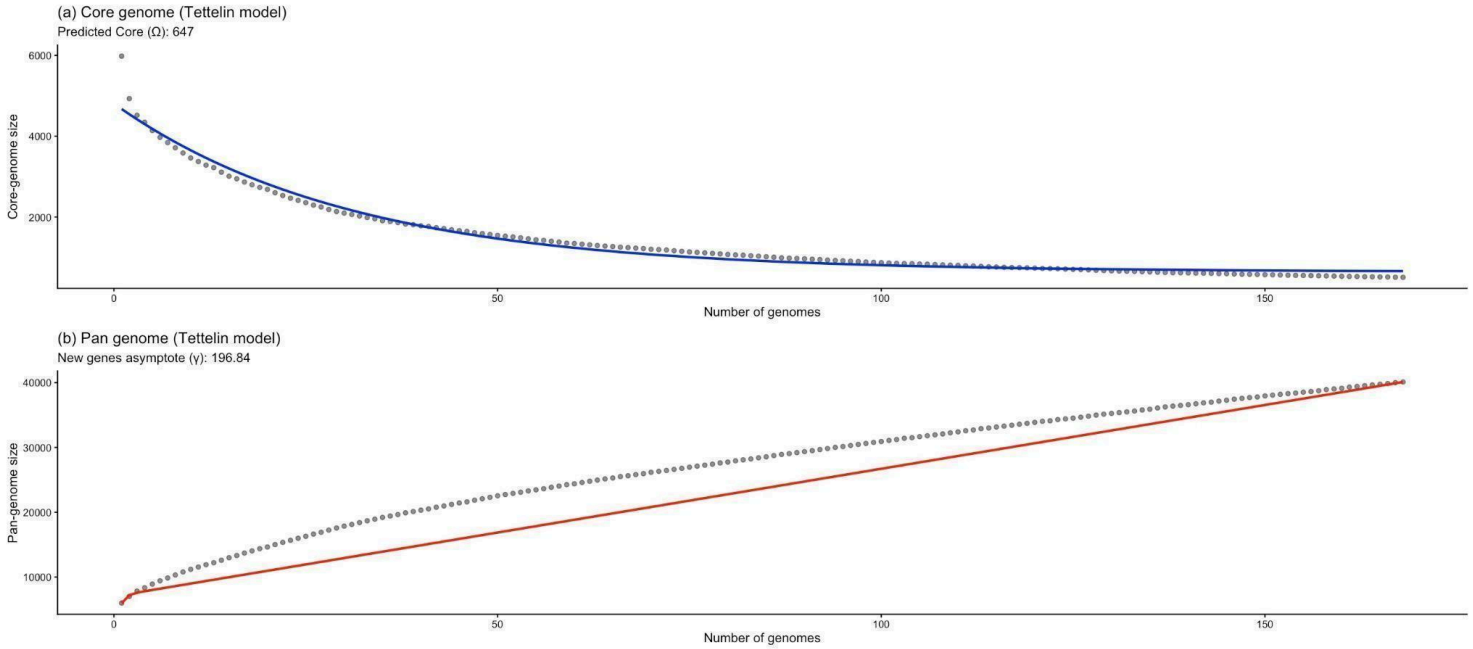


Figure 1

Pan- and core-genome modeling based on the Tettelin (15) framework

(a) Core genome size as a function of the number of genomes analyzed. Grey points represent empirical mean values across genome permutations, and the blue curve represents the fitted exponential decay model $F_c(n)=\kappa_c e^{-n/\tau_c}+\Omega$, extrapolating a strict core genome size of Ω = 647 genes. (b) Pan-genome accumulation curve. Grey points indicate the empirical pan-genome size, and the red curve represents the modeled pan-genome size derived from the fitted function for new gene discovery $F_s(n)=\kappa_s e^{-n/\tau_s}+\gamma$. The non-zero asymptote (γ ≈ 197 genes per genome) indicates an open pangenome.

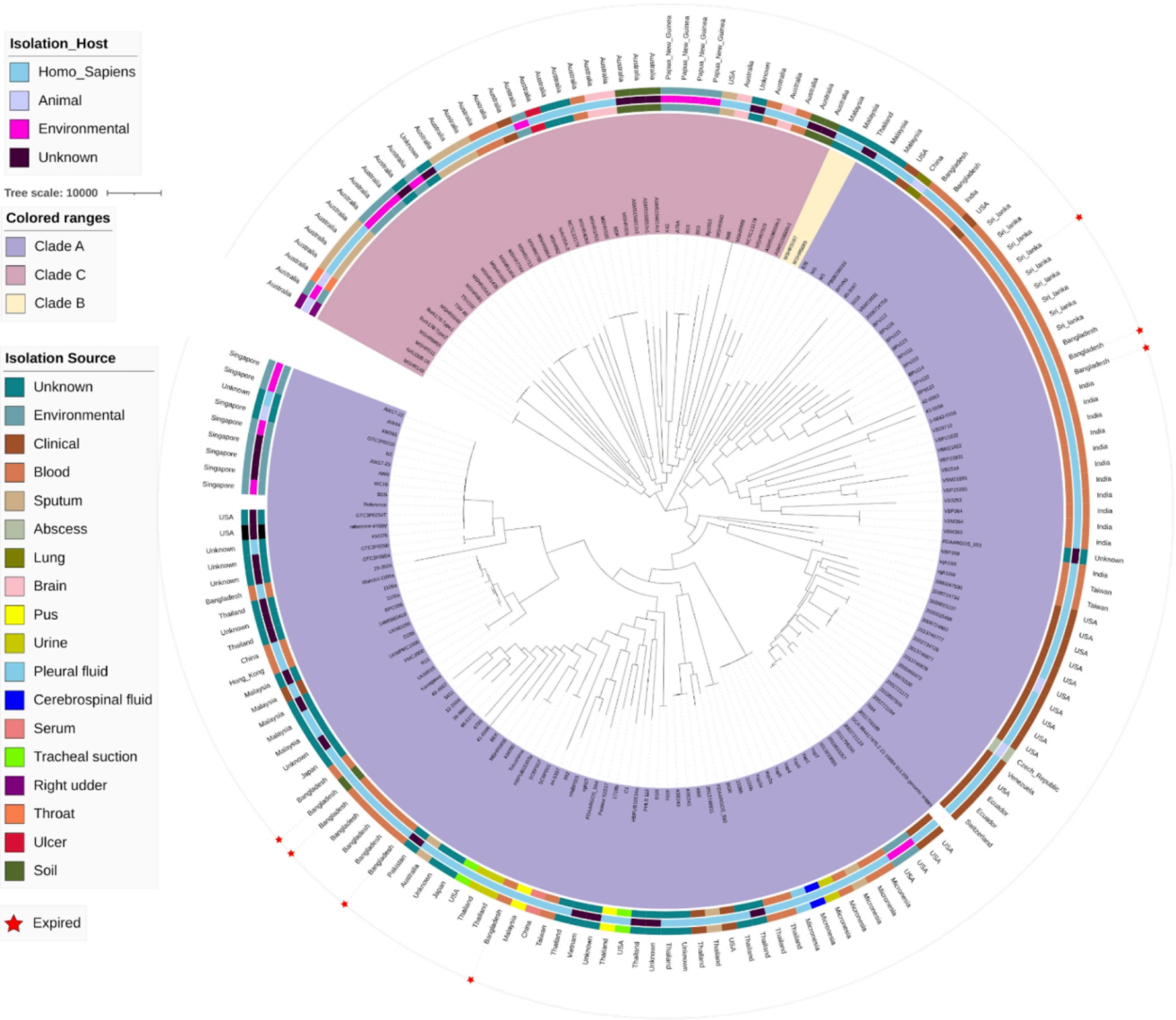


Figure 2

Phylogenetic Analysis of *Burkholderia pseudomallei*isolates. Phylogenetic tree consisting of the 14 Bangladeshi *B. pseudomallei*isolates and the publicly available isolates from NCBI. The host (e.g., humans, animals, or environmental), Isolation source, and country of origin are shown. The reference genome was obtained from NCBI (GCF003927255.1).

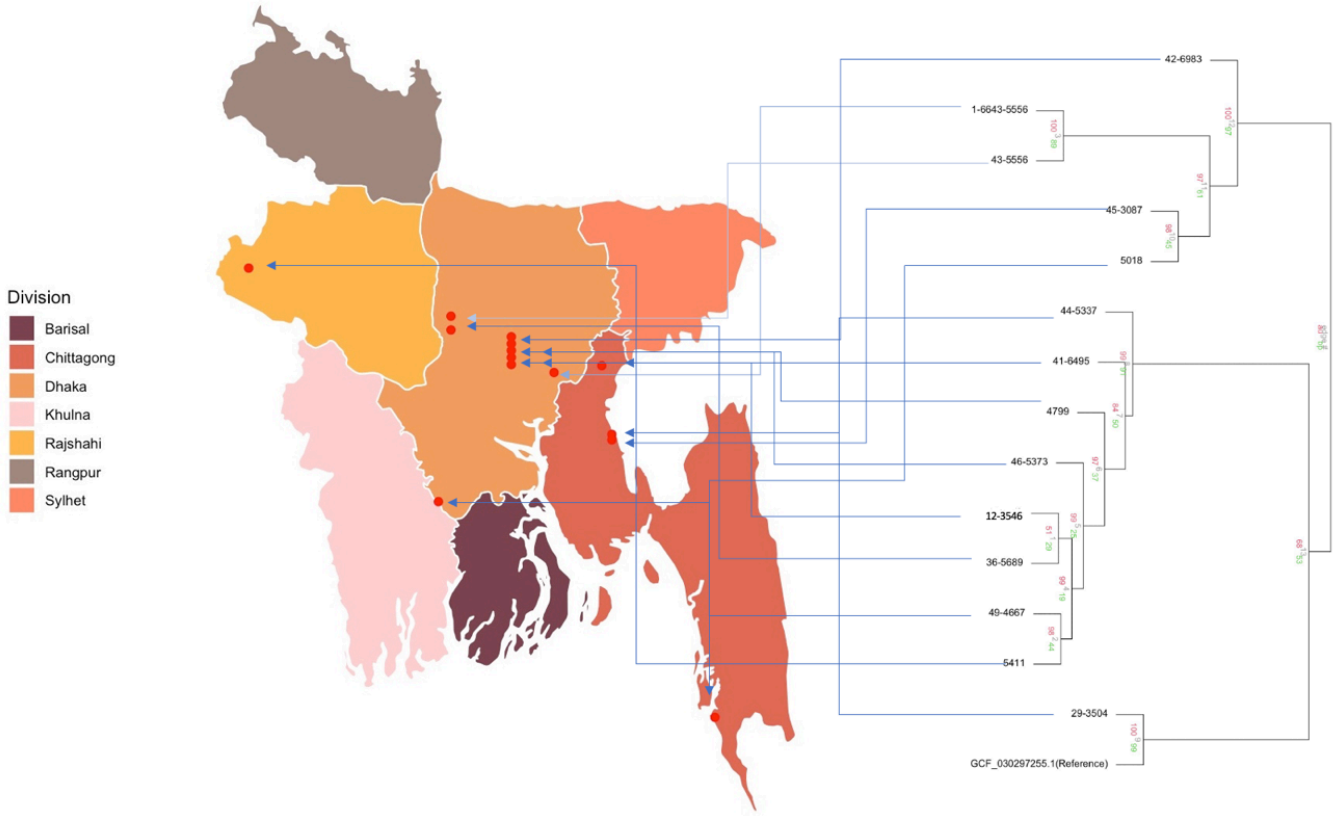


Figure 3

Geographic origin and phylogenetic relationship of the 14 *B. pseudomallei* isolates collected in Bangladesh. The map represents the city of origin of the patients from whom the 14 B. pseudomallei isolates were isolated. All isolates were obtained from clinical specimen of the patients admitted to BIRDEM General Hospital, Dhaka, Bangladesh. The dendrogram depicts the genetic relatedness of the isolates showing clustering and geographical correlation.

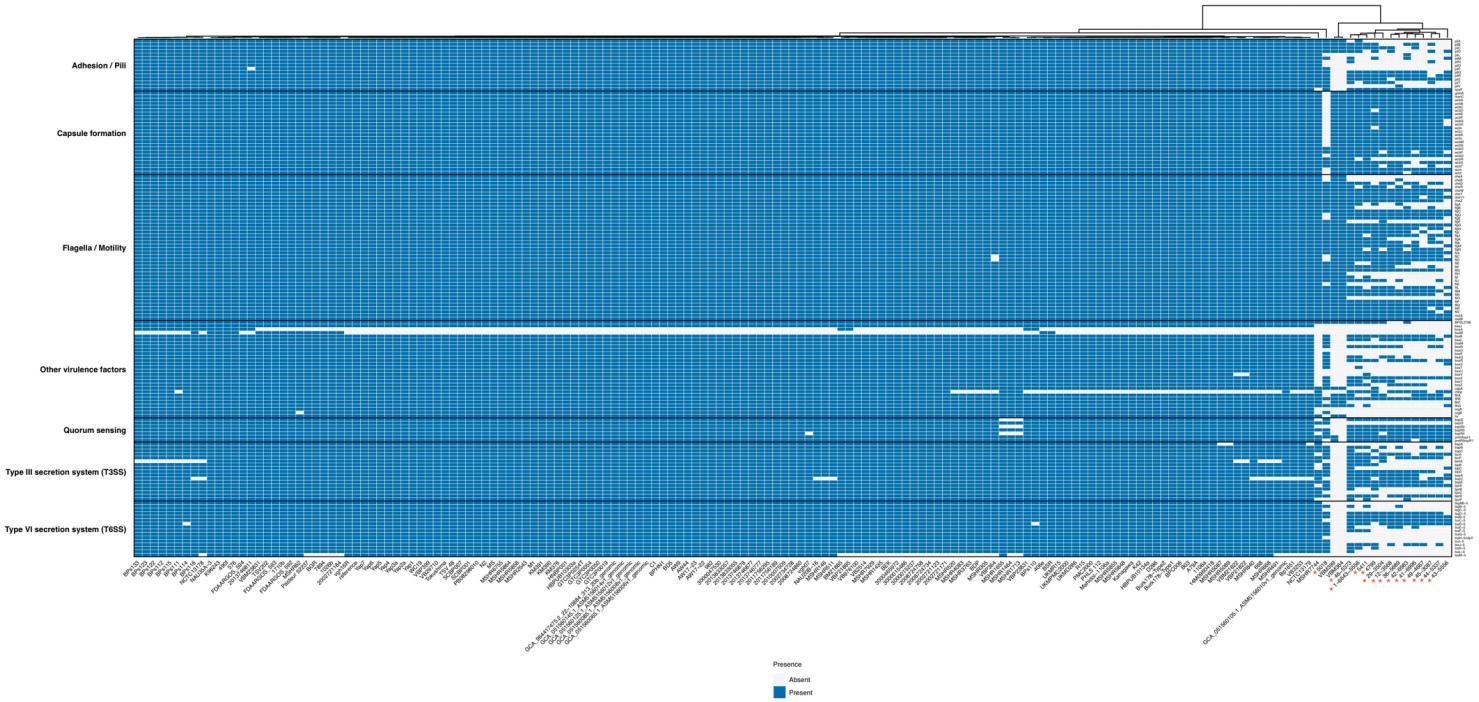


Figure 4

Virulence Profiling of the isolates. Heatmap of virulence genes identified across all publicly available isolates and the 14 Bangladeshi *B. pseudomallei* isolates (indicated using red asterisk *). Rows represent isolates, and columns represent the virulent genes. Gene presence is indicated in blue, and absence is indicated in white.

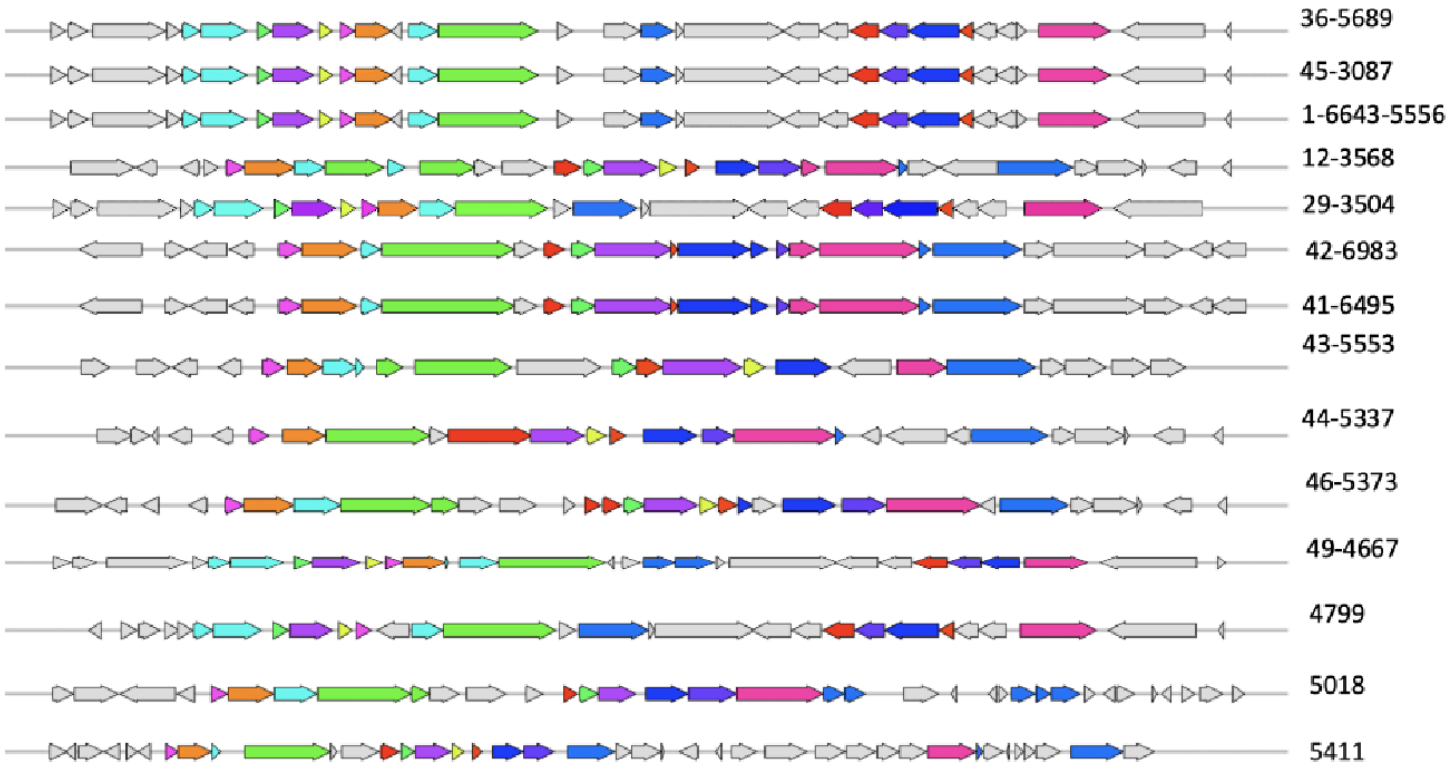


Figure 5

Organization of Type VI secretion system (T6SS) loci in multiple *Burkholderia pseudomallei* Predicted ORFs are shown as arrows indicating gene orientation. ORFs without a known T6SS assignment are shown in gray, while annotated T6SS genes are highlighted in color and labeled according to gene identity. The same color scheme is used across all loci to denote homologous genes.

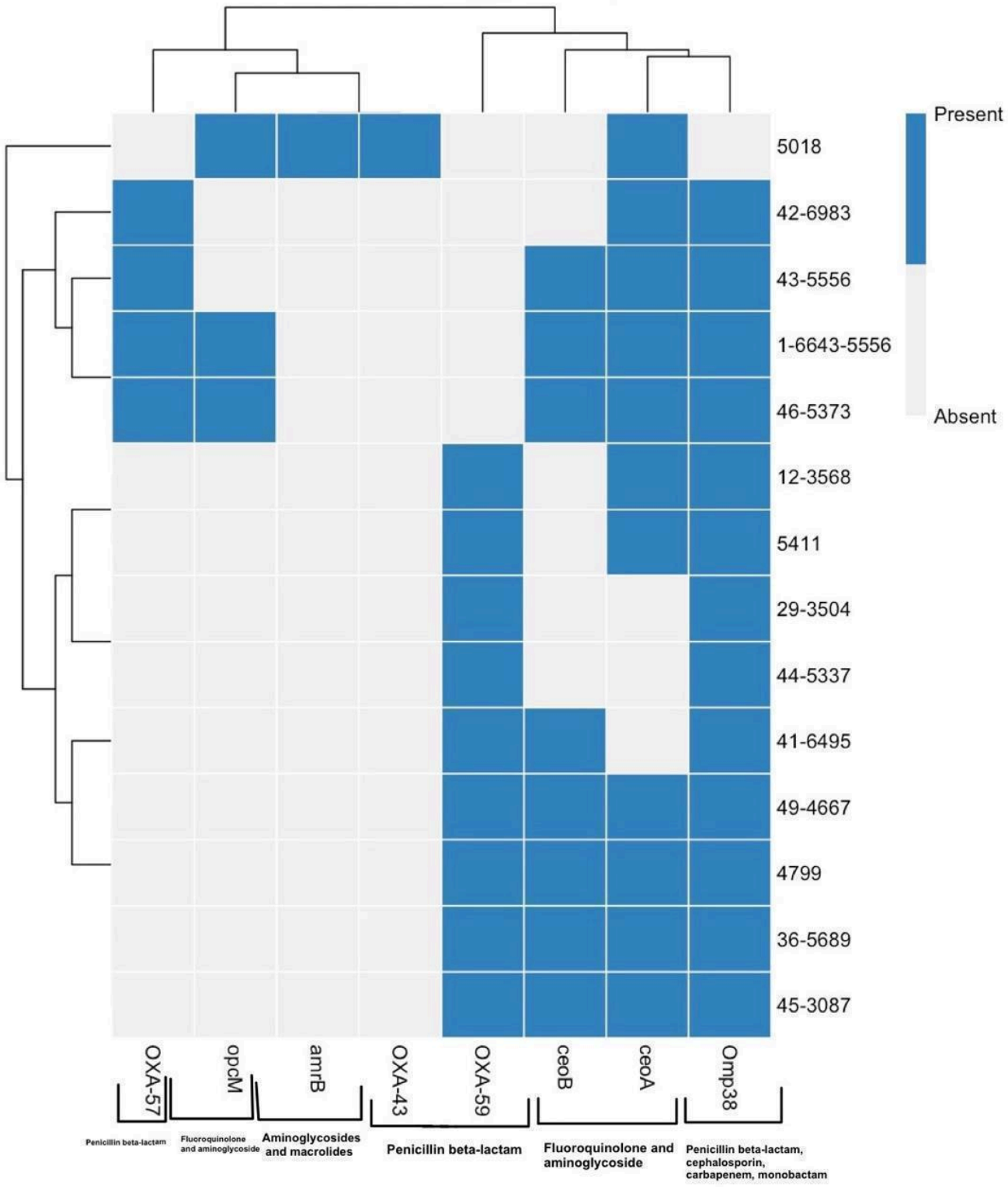


Figure 6

Distribution of Antimicrobial Resistance (AMR) Genes across 14 *B. pseudomallei* isolates. The heatmap depicts the distribution of Antimicrobial Resistance genes among the 14 *B. pseudomallei*isolates and the antibiotic classes they confer resistance to. The AMR genes were identified using CARD and ResFinder.

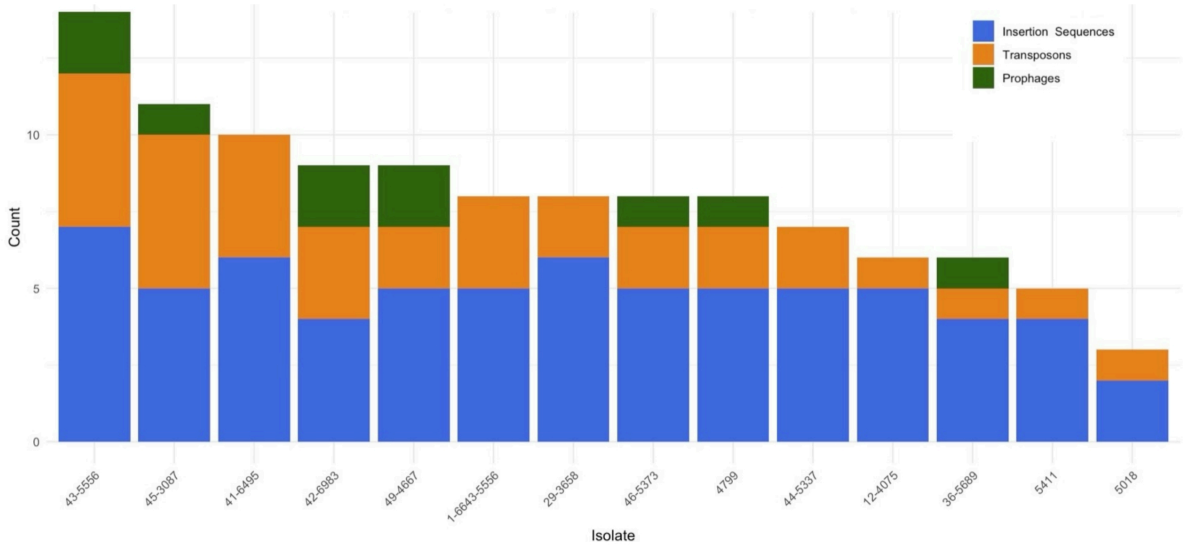


Figure 7

Mobile Genetic Element burden across the 14 analyzed *B. pseudomallei* isolates. The bars represent the predicted mobile genetic elements (MGEs) including insertion sequences, transposons and prophages across the 14 isolates. The X-axis represents the isolates, the Y-axis shows the number of isolates.



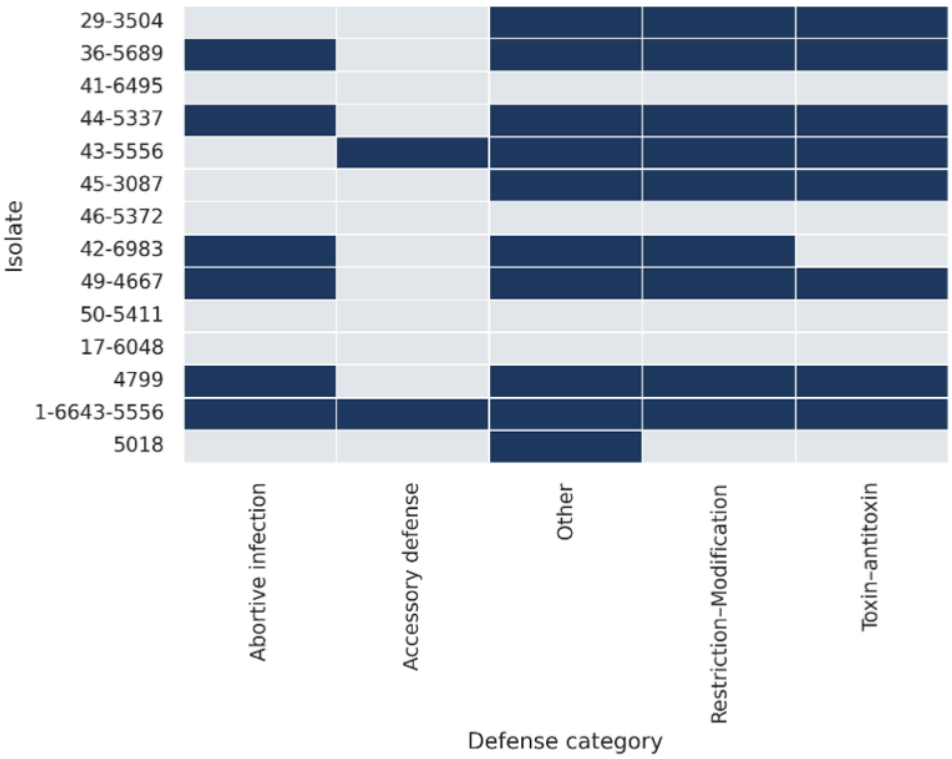


Figure 8

(A) System-level analysis reveals heterogeneous phage defense repertoires across 14 B. pseudomallei isolates, with expanded defense architectures enriched among isolates associated with severe disease. (**B)** Category-level clustering shows segregation into expanded-defense and streamlined-defense strategies independent of geographic origin and core-genome phylogeny

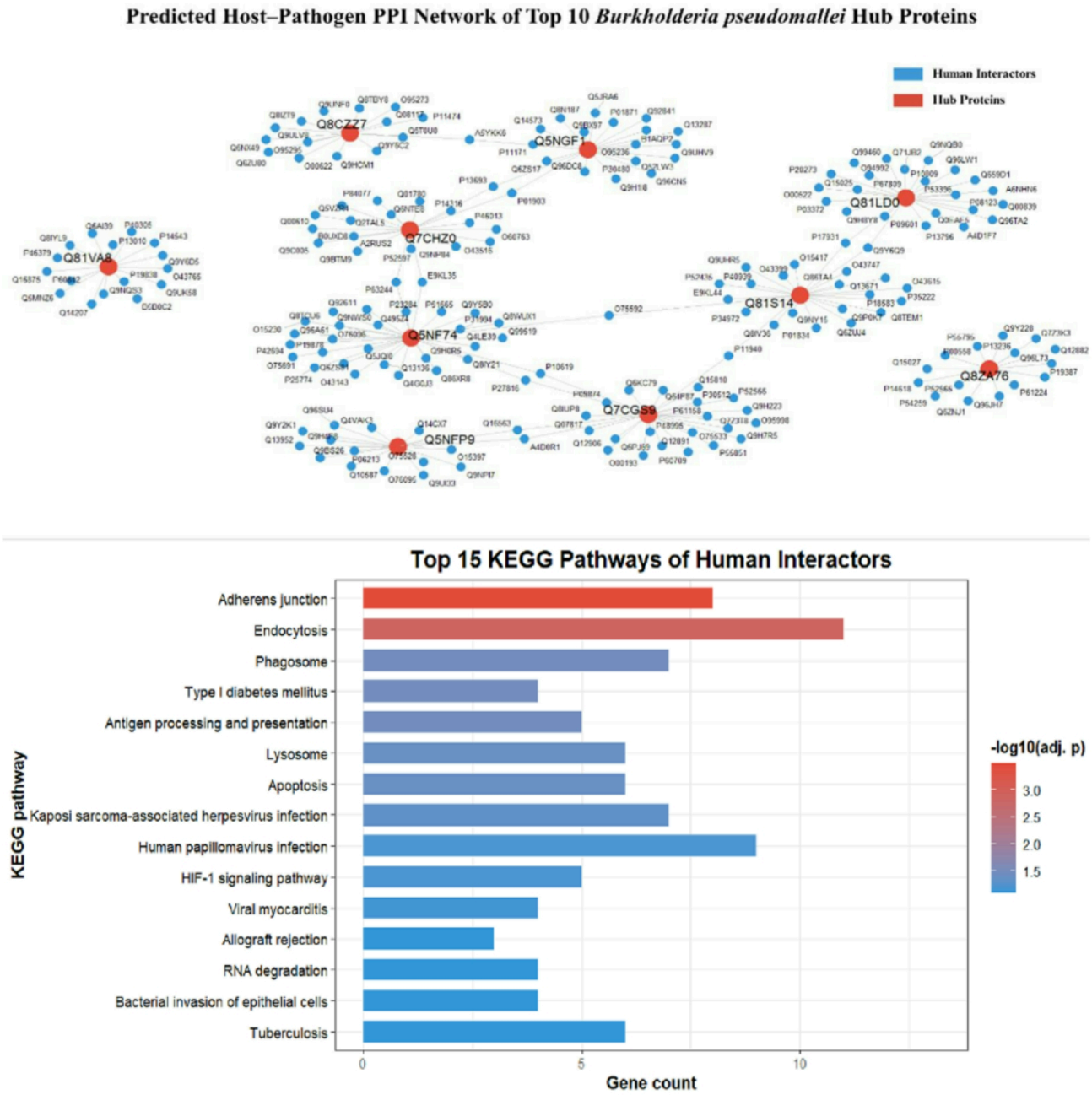


Figure 9

(a) Predicted subnetwork of the top 10 *Burkholderia pseudomallei* hub proteins and their human interactors. Red nodes represent the top 10 bacterial hub proteins by degree centrality, while blue nodes indicate their interacting human proteins. The network was visualized using igraph in R, showing the most connected bacterial proteins within the predicted host–pathogen interaction network. **(b)** KEGG pathway enrichment analysis of human proteins interacting with *Burkholderia pseudomallei*. The bar plot shows the top 15 significantly enriched pathways based on adjusted p-values (Benjamini–Hochberg correction). Bar length represents the number of human genes involved, and color intensity indicates pathway significance (-log10 adjusted p-value)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTable.pdf