



## OPEN ACCESS

## EDITED BY

Min Yue,  
Zhejiang University, China

## REVIEWED BY

Eunice M. Machuka,  
International Livestock Research Institute (ILRI),  
Kenya  
Said Oulghazi,  
Moulay Ismail University, Morocco

## \*CORRESPONDENCE

Mingsong Kang  
✉ mingsong.kang@inspection.gc.ca

## PRESENT ADDRESS

Ruimin Gao,  
National Microbiology Laboratory,  
Public Health Agency of Canada, Winnipeg,  
MB, Canada

<sup>†</sup>These authors have contributed equally to this work

RECEIVED 04 May 2023

ACCEPTED 20 November 2023

PUBLISHED 06 December 2023

## CITATION

Charron P, Gao R, Chmara J, Hoover E, Nadin-Davis S, Chauvin D, Hazelwood J, Makondo K, Duceppe M-O and Kang M (2023) Influence of genomic variations on glanders serodiagnostic antigens using integrative genomic and transcriptomic approaches. *Front. Vet. Sci.* 10:1217135. doi: 10.3389/fvets.2023.1217135

## COPYRIGHT

© 2023 Charron, Gao, Chmara, Hoover, Nadin-Davis, Chauvin, Hazelwood, Makondo, Duceppe and Kang. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Influence of genomic variations on glanders serodiagnostic antigens using integrative genomic and transcriptomic approaches

Philippe Charron<sup>‡</sup>, Ruimin Gao<sup>†‡</sup>, John Chmara, Emily Hoover, Susan Nadin-Davis, Danielle Chauvin, Jennifer Hazelwood, Kennedy Makondo, Marc-Olivier Duceppe and Mingsong Kang\*

Ottawa Laboratory-Fallowfield, Canadian Food Inspection Agency, Ottawa, ON, Canada

Glanders is a highly contagious and life-threatening zoonotic disease caused by *Burkholderia mallei* (*B. mallei*). Without an effective vaccine or treatment, early diagnosis has been regarded as the most effective method to prevent glanders transmission. Currently, the diagnosis of glanders is heavily reliant on serological tests. However, given that markedly different host immune responses can be elicited by genetically different strains of the same bacterial species, infection by *B. mallei*, whose genome is unstable and plastic, may result in various immune responses. This variability can make the serodiagnosis of glanders challenging. Therefore, there is a need for a comprehensive understanding and assessment of how *B. mallei* genomic variations impact the appropriateness of specific target antigens for glanders serodiagnosis. In this study, we investigated how genomic variations in the *B. mallei* genome affect gene content (gene presence/absence) and expression, with a special focus on antigens used or potentially used in serodiagnosis. In all the genome sequences of *B. mallei* isolates available in NCBI's RefSeq database (accessed in July 2023) and in-house sequenced samples, extensive small and large variations were observed when compared to the type strain ATCC 23344. Further pan-genome analysis of those assemblies revealed variations of gene content among all available genomes of *B. mallei*. Specifically, differences in gene content ranging from 31 to 715 genes with an average of 334 gene presence-absence variations were found in strains with complete or chromosome-level genome assemblies, using the ATCC 23344 strain as a reference. The affected genes included some encoded proteins used as serodiagnostic antigens, which were lost due mainly to structural variations. Additionally, a transcriptomic analysis was performed using the type strain ATCC 23344 and strain Zagreb which has been widely utilized to produce glanders antigens. In total, 388 significant differentially expressed genes were identified between these two strains, including genes related to bacterial pathogenesis and virulence, some of which were associated with genomic variations, particularly structural variations. To our knowledge, this is the first comprehensive study to uncover the impacts of genetic variations of *B. mallei* on its gene content and expression. These differences would have significant impacts on host innate and adaptive immunity, including antibody production, during infection. This study provides novel insights into *B. mallei* genetic variants, knowledge which will help to improve glanders serodiagnosis.

## KEYWORDS

*Burkholderia mallei*, genomic variation, serodiagnostic antigen, transcriptome, virulence factors

## Introduction

Glanders is an infectious and usually fatal zoonosis caused by *Burkholderia mallei* (*B. mallei*). Compared to other species in the family Burkholderiaceae, most of which are soil residents, *B. mallei* is a mammalian intracellular pathogen (1). Horses are highly susceptible to *B. mallei* infections and have been regarded as the natural reservoir (2). Other solipeds, such as mules and donkeys, are also susceptible to *B. mallei* infection, as well as humans (2). As *B. mallei* is highly infectious through aerosols, it was used as a biological weapon during World War I and has since been classified as a category B bioterrorism agent by the Centers for Disease Control and Prevention. Glanders has now been eradicated from most developed countries, however, it remains endemic in a number of Asian, European, African, Middle Eastern, and South American countries where sporadic infections have been detected, including Germany, Russia, Brazil, India, Turkey, and China, and significantly affects regional economic activities and international trade (3, 4). These incidents highlight the potential for importing the glanders agent to glanders-free countries or areas, thus posing a significant risk to equine and public health with huge economic loss and potentially fatal consequences globally.

So far, due to the absence of licensed vaccines or preventive treatment against glanders (5), a fast and accurate early diagnosis remains the most efficient way to prevent *B. mallei* transmission and infection. Diagnosis of glanders is mainly conducted using serodiagnostic assays for surveillance and international trade of Equidae. The complement fixation test (CFT), a serological test officially recommended by the World Organization for Animal Health (WOAH, founded as OIE) with decades of utilization in the equine industry, represents a practical method and is currently the internationally accepted approach for glanders diagnosis. However, the specificity of CFT may vary and its standardization has been questioned for years (4, 6). In order to improve the accuracy of glanders diagnosis, the immunoblot (IB) assay using Lipopolysaccharides (LPS) antigen (7) has been used to confirm CFT “positive” results. However, such a subjective test is still quite technically cumbersome, especially the production of the LPS antigen. In addition, IB may also generate false-positive results due to similar LPS structures between *Burkholderia* species, especially those acting as opportunistic pathogens (8, 9). Therefore, it would be beneficial to establish a simpler, more objective, and quantifiable serological test, such as an ELISA. ELISAs using single or double recombinant protein antigens have been developed and evaluated, and the results are promising (6, 10, 11).

Theoretically, *B. mallei* is believed to have evolved from an ancestral *B. pseudomallei* environmental isolate following an animal infection (12), by the expansion of insertion sequence (IS) elements, prophage elimination, and genome rearrangements, specifically deletions (13). The *B. mallei* genome (5.8 Mb) is 20% smaller than the *B. pseudomallei* genome (7.2 Mb), and many genes needed for environmental survival were lost from *B. mallei*, while those critical for host survival were maintained (13, 14). It has also been shown that the genome of *B. mallei* is significantly less stable than that of *B. pseudomallei*, and large-scale genomic alterations, such as insertions/deletions (INDELs), and structural variants (SVs) related to chromosomal rearrangements, have been observed in many *B. mallei* genomes (15–17). These dynamic processes may impact gene

content (gene presence/absence) and expression through various mechanisms (18).

Recently, the WOA reference laboratory for glanders in France reported that a genetic variant of *B. mallei* compromised the molecular diagnosis of glanders (19). Since interspecies differences in genome sequences, especially the accessory genome, also elicit distinctly different host adaptive immune responses (20), the high plasticity of *B. mallei* genomes may cause more variations in immunological responses that potentially disrupt the reliability of serodiagnosis, a vital tool in detecting glanders. In this study, using integrative genomic and transcriptomic analyses, we characterized the effects of genetic variation on gene content and expression to evaluate currently used and potentially useful serodiagnostic biomarkers to help improve glanders serodiagnosis.

## Materials and methods

### *Burkholderia mallei* genome assemblies from NCBI reference sequence database

All available *Burkholderia mallei* genome assemblies ( $n = 108$ ) and corresponding metadata, including NCBI RefSeq ID, level of assembly, contig count, L50, N50, size, and geographic location (Supplementary File S1), were downloaded from the NCBI Reference Sequence database (as of July 2023) using Bit v1.8.35<sup>1</sup> and EDirect v15.6.<sup>2</sup> All *B. mallei* genome assemblies and annotation completeness were evaluated using BUSCO with a completeness score cut-off value of 90% (21–23).

### Bacterial strains and growth condition

*B. mallei* type strain ATCC 23344 was transferred from The National Centre for Foreign Animal Disease (NCFAD) in Winnipeg (24) and strains Zagreb, Mukteswar and Bogor were kindly provided by the Friedrich Loeffler Institute, Federal Research Institute for Animal Health in Germany (7). Live *B. mallei* strains were handled in Containment Level 3 (CL3) facilities. Strains were cultured on a blood agar plate with 3% glycerol for 48 h at 37°C. Cultures were harvested for RNA purification or heat-killed at 85°C for 30 min and the sterility was verified by our standard operating procedure prior to further processing.

### Genomic DNA purification, library preparations and whole genome sequencing using Illumina MiSeq and Nanopore MinION

Genomic DNA was extracted from heat-killed *B. mallei* cultures using the PureLink Genomic DNA Mini Kit (Invitrogen, United States) or NanoBind CBB Kit (PacBio, United States) following

1 <https://github.com/AstroBioMike/bit#bioinformatics-tools-bit>

2 [https://astrobiomike.github.io/unix/ncbi\\_eutils](https://astrobiomike.github.io/unix/ncbi_eutils)

the manufacturer's instructions. The quality and the quantity of extracted DNA were determined by NanoDrop™ One<sup>C</sup> Spectrophotometer and Qubit™ 4 Fluorometer, respectively. The extracted genomic DNA samples have a 260/280 ratio of 1.8–2.0 and a 260/230 ratio of 2.0–2.3 with a concentration above 20 ng/μl. Library preparations and whole genome sequencing were performed as described previously (25). Briefly, Illumina (short-read) libraries were prepared using an Illumina DNA Prep (M) Tagmentation kit (Cat.No: 20060059, Illumina, United States), and sequenced on the Illumina MiSeq platform to generate 300-bp paired-end reads. MinION (long-read) libraries were established using Ligation Sequencing Kits (SQK-LSK108 or SQK-LSK109) with Native Barcoding Expansion (EXP-NBD103 or EXP-NBD104) (Oxford Nanopore Technologies, UK) without shearing and were sequenced using an FLO-MIN106 (R9.4.1) flow cell and a MinION Mk1B device.

## Genome assembly and annotation

Whole genome assembly and evaluation were performed as previously described with modifications (25). Briefly, the raw data (FASTQ files) generated from Illumina MiSeq were filtered and trimmed with BBTools v38.87 or Fastp v0.23.2 (26). Raw data (FAST5 files) generated from Nanopore MinION were base-called using Guppy v5.0.11 with the super accuracy mode, trimmed using Porechop v0.2.3,<sup>3</sup> and filtered using Filtlong v0.2.1.<sup>4</sup> Hybrid assembly of strain Zagreb using Illumina paired-end and MinION reads was performed by Unicycler v0.4.5 (27). Long read assembly of strains ATCC 23344, Bogor and Mukteswar was performed using Flye v2.9 (28), corrected using Medaka v1.4.4<sup>5</sup> and polished with Illumina MiSeq reads using a combination of NextPolish v1.4.0 (29) and Polypolish v0.5.0 (30). Nanopore long-read and Illumina short-read data were mapped to the assembled genome sequences using minimap2 (31), and the generated mapping bam files were quality checked. Qualimap (v2.2.1) (32) determined the sequencing coverage depth. The assembled genome was then annotated using the NCBI prokaryotic genome annotation pipeline (PGAP v6.2) (33).

## Genetic variant identification and analysis

Small variants, including SNPs and INDELS, were determined using Snippy v4.6.0 with default parameters,<sup>6</sup> and SVs were determined using MUM&Co v3.8.0<sup>7</sup> (34). The core genome SNP alignment obtained from Snippy was used for maximum-likelihood phylogenetic tree construction using IQtree v2.1.2<sup>8</sup> (35) with 1,000 ultrafast bootstraps. The phylogeny was partitioned into lineages using FastBaps v1.0.8<sup>9</sup> (36). The phylogenetic tree was visualized using iTOL

v6.6.<sup>10</sup> The distribution of SNPs, INDELS, and SVs was analyzed as described previously (37).

## Pan-genome analysis

Pan-genomes of 112 available *B. mallei* genome sequences were analyzed using PPanGGOLiN v1.2.74<sup>11</sup> (38), with PGAP-annotated Genbank files as input. In the clustering step, the default parameters were used. Further visualization and analysis of pan-genome results were performed using R v4.2.2, packages ggplot2 v3.4.0,<sup>12</sup> ggtree v3.2.1 (39), pheatmap v1.0.12,<sup>13</sup> UpSetR v1.4.0 (40) and vegan v2.6-4.<sup>14</sup>

## RNA isolation and library preparation

Total RNA was extracted from live *B. mallei* strains ATCC 23344 and Zagreb with three biological replicates for each strain in the Containment Level 3 (CL3) laboratory using the PureLink RNA mini kit with the On-column PureLink DNase Treatment (Invitrogen, United States) according to the manufacturer's instructions. The RNA samples were sent to Centre d'expertise et de service Génome Québec (Québec, Canada) for sequencing. Briefly, total RNA was quantified using a NanoDrop Spectrophotometer ND-1000 (Thermo Scientific, United States), and its integrity was assessed on a 2,100 Bioanalyzer (Agilent Technologies, United States). rRNA were depleted using QIAseq FastSelect -5S/16S/23S kits (QIAGEN, United States). cDNA synthesis was obtained using the NEBNext RNA First Strand Synthesis and NEBNext Ultra Directional RNA Second Strand Synthesis Modules (New England BioLabs, United States). The library preparation was performed using the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England BioLabs, United States), followed by quantification using the Kapa Illumina GA with Revised Primers-SYBR Fast Universal kit (Kapa Biosystems, Switzerland). The libraries were normalized and pooled for sequencing. After ExAMP was added, the pooled library was loaded on an Illumina cBot, and the flowcell was run on a HiSeq 4000 for 2 × 100 cycles (paired-end mode) to generate over 10 million reads for each sample (Supplementary Table S2). A PhiX library was used as a control and mixed with libraries at a 1% level. The Illumina control software was HCS HD 3.4.0.38, and the real-time analysis program was RTA v2.7.7. The bcl2fastq2 v2.20 tool was used to demultiplex samples and generate FASTQ reads.

## Prediction of virulence factors

The file (.faa file) containing coding sequences (CDS) of strain ATCC 23344 was used as an input of PathoFact v1.0 (ORF version) for the prediction of virulence factors (VFs) using default parameters (41). Only CDSs with predicted virulence confidence levels 1 (secreted

3 <https://github.com/rwick/Porechop>

4 <https://github.com/rwick/Filtlong>

5 <https://github.com/nanoporetech/medaka>

6 <https://github.com/tseemann/snippy>

7 <https://github.com/SAMtoBAM/MUMandCo>

8 <https://github.com/iqtree/iqtree2>

9 <https://github.com/gtonkinhill/fastbaps>

10 <https://itol.embl.de/>

11 <https://github.com/labgem/PPanGGOLiN>

12 <https://cran.r-project.org/web/packages/ggplot2/index.html>

13 <https://cran.r-project.org/web/packages/pheatmap/>

14 <https://github.com/vegandevs/vegan>

virulence factor) and 2 (non-secreted virulence factor) were used for further analysis.

## RNA-seq data analysis

Raw reads from the sequenced libraries were quality checked with fastp v0.23.2 (26). The reference cDNA sequences (CDS from genomic FASTA file) of *B. mallei* type strain ATCC 23344 were downloaded from the NCBI RefSeq database. Transcript abundance was quantified using kallisto v0.46.1 (42) with 100 bootstraps. Differential expression analysis was performed using the R package, sleuth v0.30.0 (43). Differential gene expressions (DEGs) were identified based on a fold change >2 and a  $q$ -value <10<sup>-3</sup> for further filtering using the Wald test in sleuth. The transcript abundance data were imported into R v4.2.2, and principal coordinates analysis (PCoA) was performed using vegan v2.6-4<sup>15</sup> and visualized using ggplot2 v3.4.0 (<https://cran.r-project.org/web/packages/ggplot2/index.html>).

## Kyoto encyclopedia of genes and genomes pathway based on transcriptomic analyzed differentially expressed genes

The identified differentially expressed genes (DEGs) were applied to explore the potentially affected Kyoto encyclopedia of genes and genomes (KEGG) pathways. The corresponding KO (KEGG ORTHOLOGY) table was obtained by GhostKOALA v2.2 (44) using the genus\_prokaryotes database. The obtained KO terms corresponding to the DEGs were fed to the KEGG Mapper reconstruction tool (last updated: July 1, 2021)<sup>16</sup> (45), and the affected pathways were evaluated.

## Effect of genomic variation on gene expression

Based on the simple model that the destruction of operon structure or disruption of the non-coding region upstream of operons can interfere with gene expression (46), variant calls (SNPs, INDELs, and SVs) were compared with RNA-seq data to link DEGs to specific genomic variation. Specifically, we identified the operons using operonSEQer v1.0 (47) and investigated changes in gene expression in relation to an alteration in the operon structure mediated by SVs, as well as changes in the non-coding region upstream of the operon, including promoter and regulators, mediated by SNPs or INDELs. This was achieved using custom scripts modified from a previously described method (48).

## Results

### Genetic variation of *Burkholderia mallei* strains

Due to genetic variability after laboratory or host passages, four *B. mallei* isolates, ATCC 23344, Bogor, Mukteswar, and Zagreb, were re-sequenced using both Nanopore long-read and Illumina short-read sequencing technologies to obtain complete or chromosome-level whole genome assemblies (Supplementary Table S1). In combination with 108 published genomes available on RefSeq (Supplementary File S1), we identified a total of 3,806 core SNPs by using strain ATCC 23344 (RefSeq accession number: GCF\_000011705.1) as a reference. The density and distribution of SNPs are similar among chromosome I (CH1) and chromosome II (CH2) (Figure 1A). In addition to SNPs, a large number of INDELs and SVs were identified. Several dense spots of variants indicating common SNPs, INDELs, and SVs are noticeable in multiple sites on both CH1 and CH2 (Figure 1A). To further explore this observation, the distribution and number of SNPs, INDELs, and SVs for 38 strains with complete and chromosome-level genome assembly were examined, respectively (Figures 1B,C). Multiple typical “hotspots” of SNPs and INDELs exist in all genomes, including derivatives of ATCC 23344 as shown in the five inner circles (Figure 1B). Around 70% of small variants, including SNPs, INDELs, MNPs (multiple nucleotide polymorphisms), and complex variants (combination of SNPs and MNPs), were located within coding sequences (CDS) (Figure 2A), some of which introduce early stop codons for several genes, including a gene encoding the serodiagnostic antigen ImpJ (TssK) shown in Table 1. The majority of small variants are still SNPs and INDELs (Figure 2A). Similarly, multiple types of SVs, especially large deletions, in *B. mallei* isolates were identified using strain ATCC 23344 as a reference, except for some ATCC 23344 derivatives (Figures 1C, 2B). These SVs were scattered in both chromosomes, and multiple hotspot regions were observed in most isolates on both chromosomes (Figure 1C).

### Variation in *Burkholderia mallei* gene content (presence/absence of genes)

All 112 available genome assemblies of *B. mallei* were used to identify the core and accessory genomes. Upon successive addition of each genome, the number of gene families in the pan-genome increased to 5,332 and seemed to reach a plateau with a very low  $\gamma$ -value (0.03) of a Heaps' law fitting, which reflects that *B. mallei*'s pan-genome is nearly closed (38) (Supplementary Figures S1A,B). Overall, 5,332 gene families were identified (Supplementary Figure S1B), in which there were 3,464 persistent genes present in over 95% of the *B. mallei* genomes, 1,386 shell genes that were present at intermediate frequencies, and 482 cloud genes that were found in one, or very few individuals. The gene frequencies tend to show a strongly asymmetric U-shaped distribution with a larger proportion of persistent genes (Supplementary Figure S1B), indicating a lower level of gene gain achieved through horizontal gene transfer (HGT) within *B. mallei* (53). The pan-genome matrix shown in Figure 3A demonstrated substantial variations of gene content

<sup>15</sup> <https://github.com/vegandevs/vegan>

<sup>16</sup> <https://www.kegg.jp/kegg/mapper/reconstruct.html>

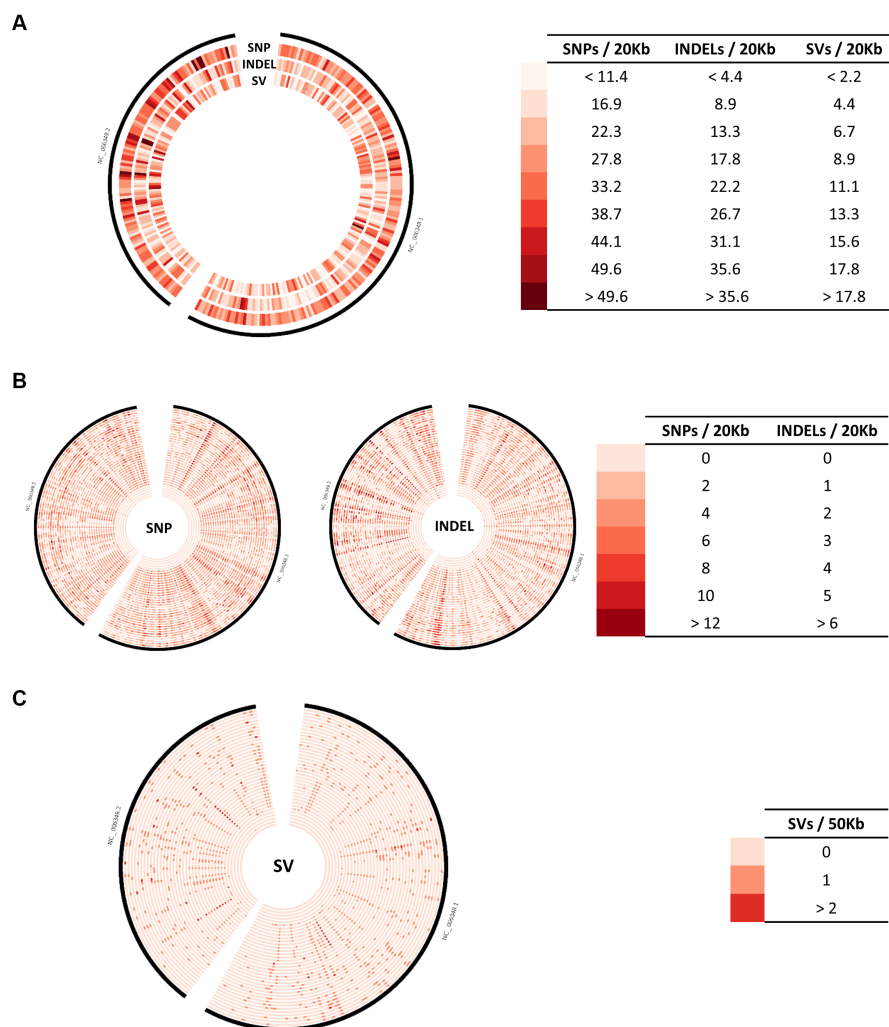


FIGURE 1

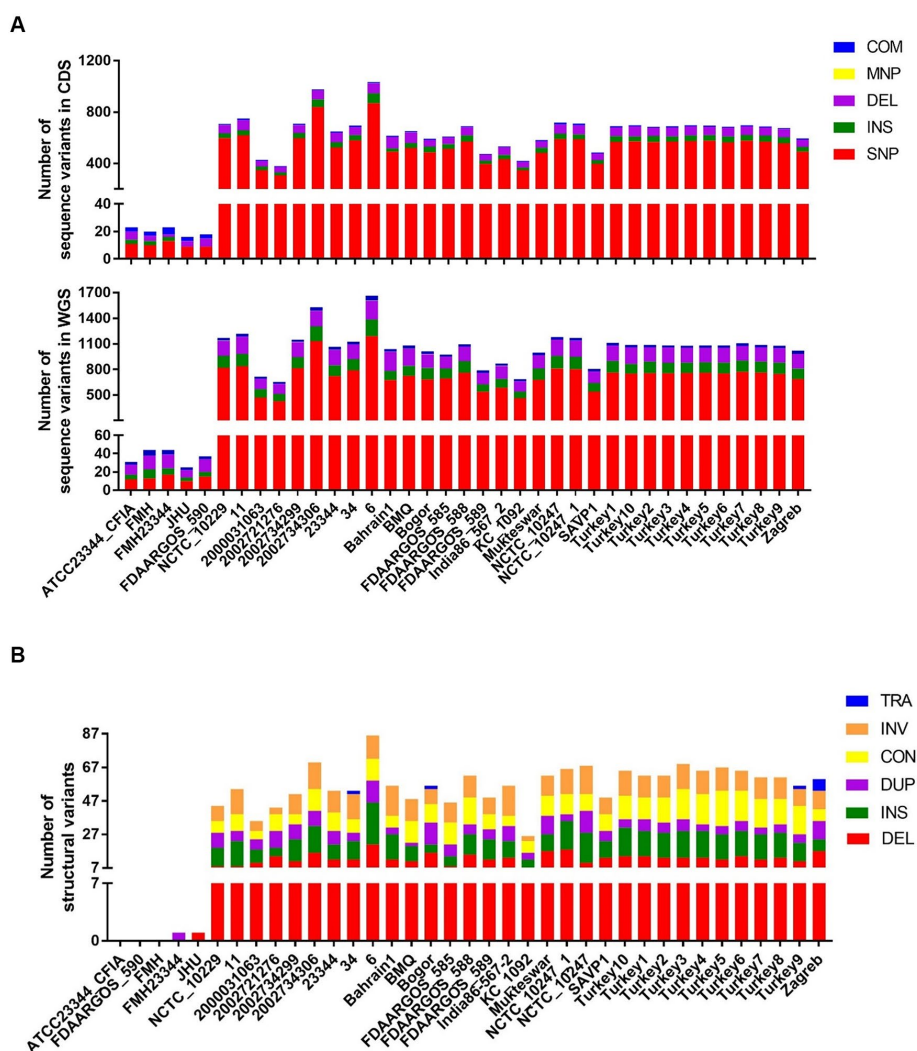
Distribution of genomic variations across the genome of *B. mallei*. (A) Distribution of SNPs, INDELS, and SVs across the genome for 112 *B. mallei* strains. (B) Distribution of SNPs and INDELS for 38 isolates with complete or chromosome-level genome assembly. (C) Distribution of SVs for 38 isolates with complete or chromosome-level genome assembly. The red gradient represents the density of each type of genomic variation.

among all the genomes of *B. mallei* available in this study. These data indicate gene loss is a significant consequence of *B. mallei*'s genetic variation as limited gene gain was observed.

Gene content variations of a bacterial pathogen have been demonstrated to contribute to various acute and adaptive immune responses elicited by different strains of the same bacterial species (20). Due to the highly plastic genome of *B. mallei* (14), it is highly possible that long term evolution of this species may magnify this phenomenon thereby increasing the complexity of glanders serodiagnosis. To achieve greater accuracy in analyzing gene content variations and avoid unreliable or unavailable metadata mentioned previously (54), we implemented hierarchical Bayesian clustering via FastBaps, which further divided the worldwide phylogeny based on core SNPs into strain-level clusters (Supplementary Figure S2). In total, we identified eight clusters corresponding to different lineages (L1-8) (Supplementary Figure S2). L1 was built up from genomes of the type strain ATCC 23344 and derivatives of ATCC 23344 (US as geographic origin). Other Chinese isolates, such as strain China 5 and ATCC 10399, were found in L5. Interestingly, strains FDAARGOS\_587,

2000031063, and KC\_1092 were also clustered in L5, an observation which is consistent with previous publications that have pointed out inaccurate metadata for these isolates (15, 54, 55). The majority of isolates from Turkey were grouped into L2. Within the Indian strains in L3, the Budapest strain from Hungary was surprisingly close to strain SAVP1, as previously reported (15). Five genome sequences of possible Russian isolates with SCPM Identifier comprised L4. An exception was the strain SCPM-O-B-7093, which was most closely related to the strains Bahrain1, Bogor, Zagreb, and Mukteswar isolated in the 1990s or 2010s within L6. The genomes of UK strains 2002734306, FDAARGOS\_586, and NCTC 120 were grouped with the Turkey strain 6 in L7. The L8 group consisted of strains isolated from different geographical locations and hosts.

Comparing gene contents between different lineages, isolates from the same lineage or close geographic location tended to have rather similar gene contents (Figures 3A,B), except for those of L3 and L8 which overlapped with other lineages. To investigate this finding further, using binary data of the presence/absence of genes, we performed Principal Coordinates Analysis (PCoA) to explore the



**FIGURE 2** Various types of genomic variations identified in each *B. mallei* strain with complete or chromosome-level genome assembly. **(A)** The number of SNP, INDEL, MNP, and complex variants identified in the coding regions and whole genome sequences. **(B)** The number of multiple types of SVs identified in WGS. SNP, Single Nucleotide Polymorphism; MNP, Multiple Nucleotide Polymorphism; INS, Insertion; DEL, Deletion; COM, Complex (combination of SNP/MNP); TRA, Translocation; INV, Inversion; CON, Contraction; DUP, Duplication.

dissimilarity of gene content between lineages and further confirmed the distinct gene presence-absence variations within L3 and L8 (Figure 3C). Additionally, there were clear differences in gene content between L2 or L7 and the rest of the lineages (Figure 3C), demonstrating that the gene content of the Turkey or UK strains is quite different from those of strains isolated from other geographic locations. Variations in each lineage-core genome (genes present in over 95 percent of the genomes in each lineage) were evident (Figure 3D). All these data demonstrate that *B. mallei* strains from different lineages or isolated from various geographic locations have different gene contents, suggesting various adaptive immune responses can possibly be elicited by strains from different lineage or geographic locations.

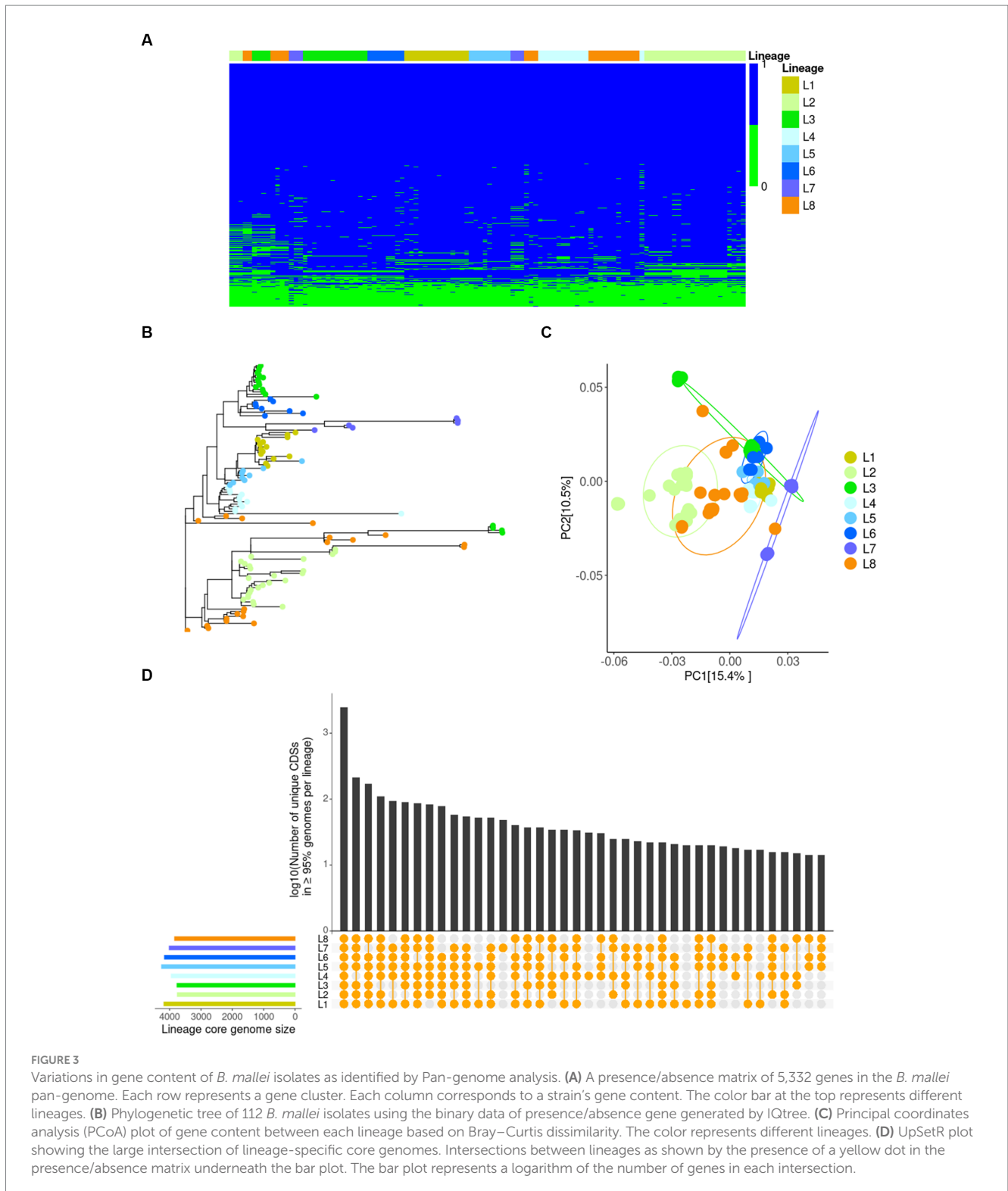
Moreover, we explored gene content diversity in the complete and chromosome-level genome assemblies to uncover gene presence-absence variations at the individual strain level using

strain ATCC 23344 as a reference. Gene loss for each isolate and missing genes in the ATCC 23344 genome (gene gain in other genomes) were identified (Figure 4A). Interestingly, some strains, such as strains 6, 2002734306, and Turkey 6, even lost many persistent genes, while strain ATCC 23344 also lost four persistent genes (Figure 4A). The absence of genes in each isolate was mapped to the chromosomes of ATCC 23344, and multiple hotspots on CH1 and CH2 were observed, especially on CH2 (Figure 4B). Notably, most of the genes encoding antigens, such as BimA, HCP1, TssA, TssB, TssM, A03050H, and 0376H, used in established or developing serodiagnostic tests for glanders or melioidosis are located within those CH2 hotspots and are thus missing from some *B. mallei* isolates (Figure 4B and Table 1). All these data strongly demonstrate that *B. mallei* genome plasticity affects gene content, including some genes which encode proteins used for glanders serodiagnosis.

TABLE 1 Gene-encoded purified proteins used for ELISA development or as potential serodiagnostic markers.

Antigen	Present in ATCC 23344	CH <sup>a</sup>	Gene location number <sup>b</sup>	Locus Tag <sup>c</sup>	Present in <i>B. mallei</i> genomes <sup>d</sup>	Present in <i>B. mallei</i> genomes (complete or chromosome level) <sup>d</sup>	Present in other non-pathogenic Burkholderia spp <sup>d</sup>	DEG (Zagreb vs. ATCC 23344)	References
FlgK	Yes	1	3143	BMA_RS15840	112/112 (100%)	38/38 (100%)	Y	N	Wagner et al. (49)
Lipoprotein	Yes	1	1320	BMA_RS06655	112/112 (100%)	38/38 (100%)	Y	N	Wagner et al. (49)
Putative exported protein	Yes	1	817	BMA_RS04115	112/112 (100%)	38/38 (100%)	Y	N	Wagner et al. (49)
Peroxisredoxin	Yes	1	1386	BMA_RS06985	112/112 (100%)	38/38 (100%)	Y	N	Wagner et al. (49)
DUF2059 domain-containing protein	Yes	1	407	BMA_RS02025	112/112 (100%)	38/38 (100%)	Y	N	Wagner et al. (49)
OmpA	Yes	1	409	BMA_RS02035	112/112 (100%)	38/38 (100%)	Y	N	Wagner et al. (49)
GroEL	Yes	1	1873	BMA_RS09440	112/112 (100%)	38/38 (100%)	Y	N	Wagner et al. (49)
GroES_1	Yes	1	1874	BMA_RS09445	112/112 (100%)	38/38 (100%)	Y	N	Wagner et al. (49)
O-antigen acetylase	Yes	1	1856	BMA_RS09350	110/112 (98.2%)	37/38 (97.4%)	Y	N	Stone et al. (9)
Flagellin	Yes	1	2696	BMA_RS13600	112/112 (100%)	38/38 (100%)	Y	N	Wagner et al. (49)
GroES_2	Yes	1	2279	BMA_RS11465	112/112 (100%)	38/38 (100%)	Y	N	Wagner et al. (49)
ImpJ (TssK) <sup>e</sup>	Yes	2	3623	BMA_RS26750	78/112 (69.6%)	23/38 (60.5%)	Y	N	Wagner et al. (49)
BopC	No	–	–	–	75/112 (67.0%)	24/38 (63.2%)	Y	N/A	Wagner et al. (49)
BopE	Yes	2	4613	BMA_RS23210	107/112 (95.5%)	35/38 (92.1%)	Y	N	Wagner et al. (49)
BipB	Yes	2	4620	BMA_RS23245	108/112 (96.4%)	36/38 (94.7%)	Y	N	Wagner et al. (49)
Malate dehydrogenase	Yes	2	4835	BMA_RS24325	108/112 (96.4%)	38/38 (100%)	Y	N	Wagner et al. (49)
ABC transporter permease subunit	Yes	2	3525	BMA_RS17765	103/112 (92.0%)	38/38 (100%)	Y	Y	Wagner et al. (49)
Hcp1	Yes	2	3889	BMA_RS19605	105/112 (93.8%)	36/38 (94.7%)	Y	N	Elschner et al. (10)
N-acetylmuramoyl-L-alanine amidase	Yes	2	3447	BMA_RS17365	111/112 (99.1%)	38/38 (100%)	Y	N	Wagner et al. (49)
TssB	Yes	2	3890	BMA_RS19610	105/112 (93.8%)	36/38 (94.7%)	Y	N	Elschner et al. (10)
Phage integrase family protein (0376TH)	Yes	2	4167	BMA_RS21000	111/112 (99.1%)	38/38 (100%)	Y	N	Pal et al. (50)
Hypothetical protein (0375H)	Yes	2	4168	BMA_RS21005	111/112 (99.1%)	38/38 (100%)	Y	N	Pal et al. (50)
BimA	Yes	2	3896	BMA_RS19640	83/112 (74.1%)	36/38 (94.7%)	Y	N	Kumar et al. (51)
TssM	Yes	2	3875	BMA_RS26795	45/112 (40.2%)	11/38 (28.9%)	Y	N	Shanks et al. (52)
TssA	Yes	2	3894	BMA_RS19630	103/112 (92.0%)	35/38 (92.1%)	Y	N	Elschner et al. (10)
Hypothetical protein (A03050H)	No	–	–	–	47/112 (42.0%)	21/38 (55.3%)	Y	N/A	Pal et al. (50)

<sup>a</sup>Chromosome of ATCC 23344.<sup>b</sup>Gene location number match with x axis label of Figure 4B.<sup>c</sup>Locus tag of genome assembly of ATCC 23344.<sup>d</sup>Using cutoff values: coverage > 80% and identity > 80%.<sup>e</sup>Pseudogene with early stop codon in *B. mallei* genomes when compared to *B. pseudomallei* protein.

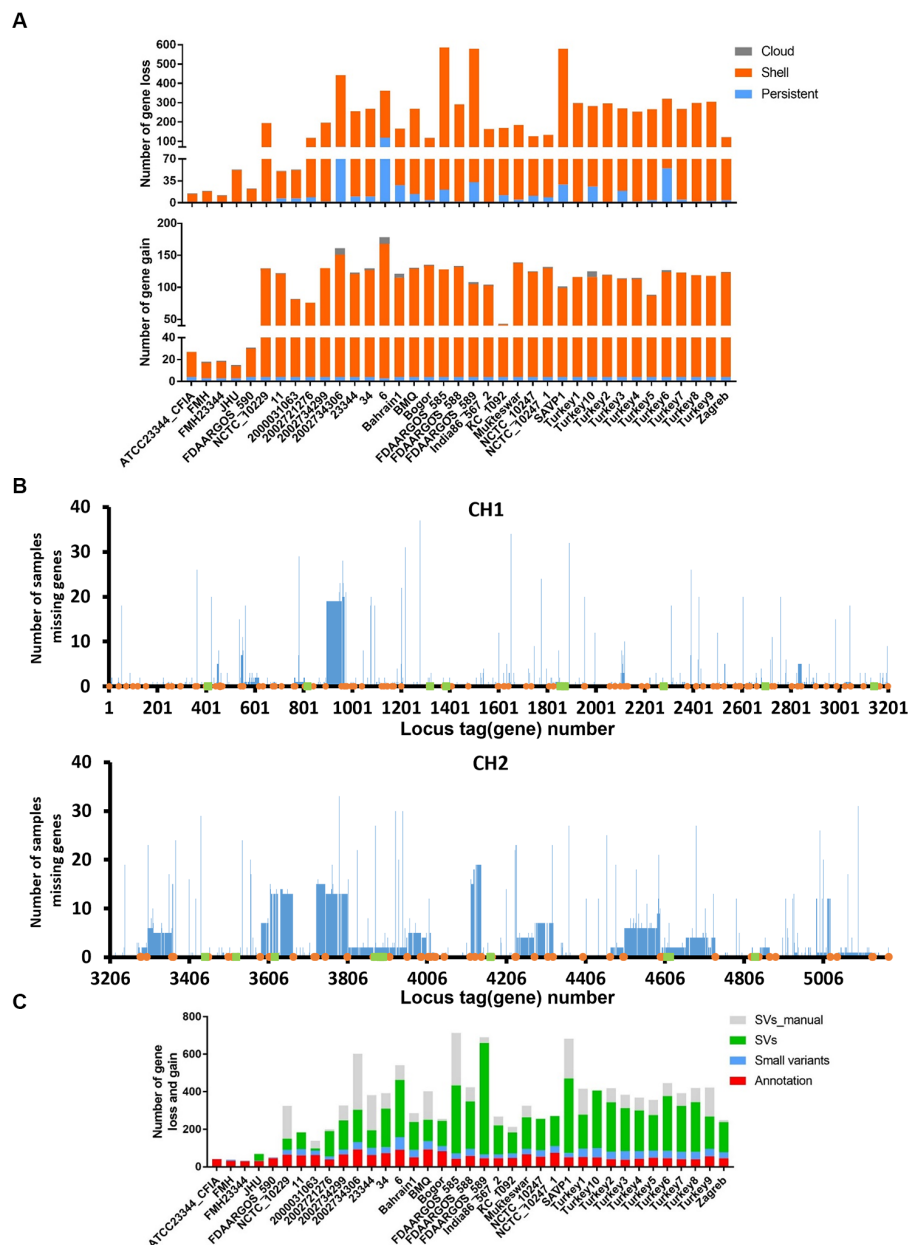


## Contribution of genomic variations to gene loss

To determine how genomic variations correlated with gene loss, we focused on the cross-comparison of genetic variation with gene content. Although a large number of small variants have been detected in CDS (Figure 2A), resulting in early stop codons and frameshifts,

most of the gene loss events were associated with SVs (Figure 4C), which further confirms the key role of SVs in genome plasticity and its role in the loss of genes encoding antigens related to serodiagnosis such as Hcp1, TssB, BimA, and TssA (Figure 4B and Table 1). Additionally, some gene losses are related to the annotation performed by PGAP, which are mainly labeled as pseudogenes due to frameshifts or genes not being annotated (Figure 4C).





**FIGURE 4**  
 Effects of genomic variations on gene content. **(A)** Number of genes lost and gained (loss in strain ATCC 23344) in each *B. mallei* strain with complete or chromosome-level genome assembly. **(B)** Distribution of gene losses in *B. mallei* strains with complete or chromosome-level genome assembly on the genome of ATCC 23344. The orange dot represents transposase, and the green square shows gene coding antigens shown in Table 1. **(C)** Number of gene losses and gains in each *B. mallei* strain with complete or chromosome level genome assembly, compared to strain ATCC 23344. The color represents the cause of the gene loss or gain. Annotation: genes were not predicted by PGAP; Small variants: gene losses are due to SNPs, INDELS, MNPs, and complex variants. SVs: gene loss due to SVs. SV\_manual: gene loss due to SVs that were not identified by MUM&Co.

## Effects of sequence variations on gene expression

To understand potential effects of sequence variations on gene expression, we compared the transcriptome profiles of strain Zagreb with ATCC 23344 and found slight, but not significant differences overall (PERMANOVA:  $R^2=0.415$   $Pr(> F)=0.1$ ) (Supplementary Figure S3A), due to a variation of gene expression of

less than 10% (Figure 5A and Supplementary File S2). To be specific, only 388 DEGs were identified, for which a total of 155 available KEGG Orthology (KO) were retrieved from the KEGG mapper reconstruction server. The results showed that a total of 82 pathways were affected. The key differences in affected gene functions involve various bacterial metabolic pathways, pathogenesis, and virulence, including ABC transporters, quorum sensing (QS), two-component system, and biofilm formation (Supplementary Figure S3B).

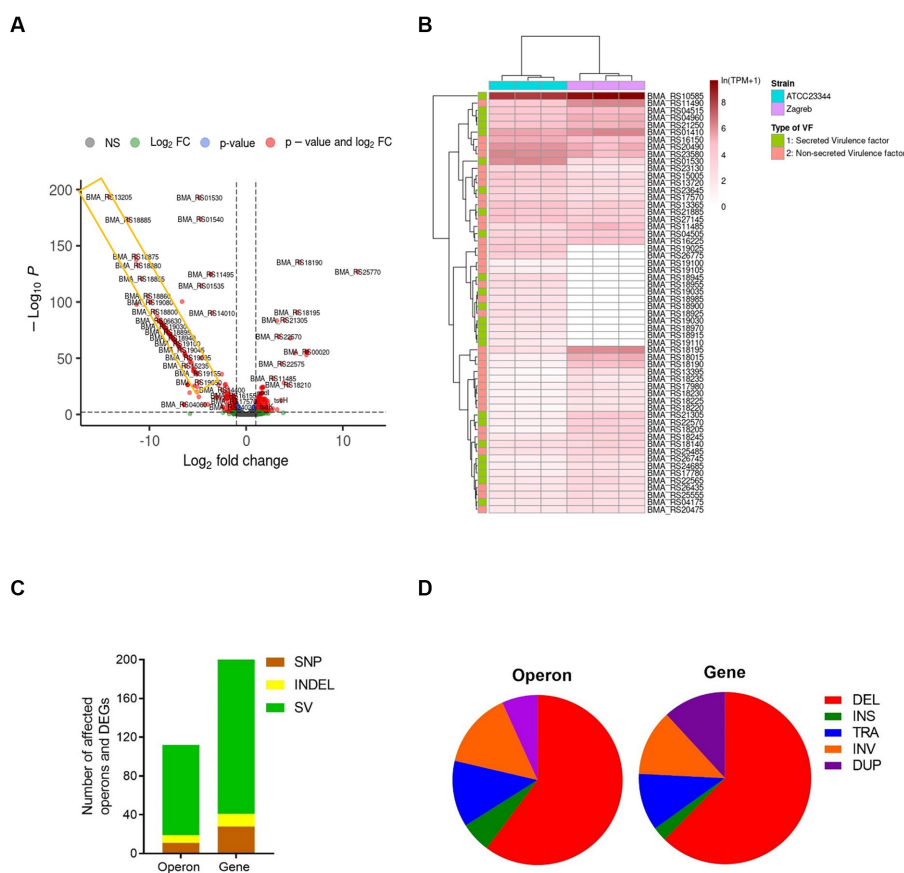


FIGURE 5

Effects of genomic variants on gene expression. (A) Volcano plot to display differentially expressed genes (DEGs) between type strain ATCC 23344 and strain Zagreb of *B. mallei*. The x and y axis represent Log<sub>2</sub>fold change and  $-\log_{10}P$ , respectively. The genes indicated in red dots represent the DEGs with  $|\log_2(\text{FC})| > 1$  and  $p < 0.001$ . (B) The expression level of 58 differentially expressed virulence genes coding secreted VF (green) and non-secreted VF (pink) between strain ATCC 23344 and Zagreb. The density of the red color represents the gene expression level of  $\ln(\text{TPM}+1)$ . Three biological replicates for each strain. (C) The number of SNP (brown), INDEL (yellow), and SV (green) related to affected operons and DEGs. (D) The percentage of different SVs related to affected operons and DEGs. The association of genomic variants on DEGs or affected operons was predicted using the simple predictive model that the destruction of the structure of operon and alteration of the non-coding regions, can interfere with gene expression. INS, Insertion; DEL, Deletion; TRA, Translocation; INV, Inversion; DUP, Duplication.

Specifically, the expression of ABC transporter permease subunit, which has been regarded as a potential serodiagnostic marker (56) was significantly different between these two strains (Table 1). Since proteins related to bacterial pathogenesis and virulence are typically used as serodiagnostic biomarkers, we investigated expression differences for virulence factors. A total of 842 virulence factors and 2,294 potential virulence factors were predicted using the PathoFact pipeline (Supplementary File S3). Fifty-eight (58) out of the 842 virulence genes, which code for 24 secreted and 34 non-secreted virulence factors, were among the identified DEGs, including some genes related to the type IV secretion system (Figure 5B and Table 2).

In an attempt to understand the effects of genomic variants on gene expression, we focused on not only changes in transcription related to the alteration of operon structures mediated by SVs, but also changes of non-coding regions upstream of operons, including promoter and regulator regions, mediated by SNPs or INDELS, respectively. Using transcriptomic data of strain ATCC 23344, OperonSEqer predicted a total of 2,455 operons (Supplementary Files S4, S5). One hundred and eighty-five of the DEGs were located within 108 operons disturbed by at least one type of genomic variant, especially SVs, with a large

number being caused by deletions (Figures 5C,D). A large deletion from locus\_tag BMA\_RS18805 to BMS\_RS19170 was observed in the genome of strain Zagreb, which was also reflected in transcriptomic data shown within the yellow rectangle in Figure 5A. Besides deletions, translocations also had a big impact on the expression of many genes, including several virulence genes (Figure 5D and Table 2).

## Discussion

Glanders diagnosis mainly relies on serological tests. Apart from different assay platforms, selection of appropriate antigens used in serological assays is critical to achieving adequate diagnostic sensitivity and specificity. Due to its simple standardization and objective analysis, ELISAs using recombinant proteins as antigens have shown very promising results with a potential replacement of CFT for glanders serodiagnosis in the near future (6). However, in the current study, we showed that, due to its high plasticity, the genome of *B. mallei* contains large-scale variants, resulting in differences in gene content and expression among each isolate. A number of

TABLE 2 DEGs associated with SVs.

Locus_Tag	Gene	Product	Operon	SV type
BMA_RS10585		Hypothetical protein	OPERON_1002	Translocation
BMA_RS18190	<i>tssC</i>	Type VI secretion system contractile sheath large subunit	OPERON_1710	Translocation
BMA_RS18195		Type VI secretion system tube protein Hcp	OPERON_1710	Translocation
BMA_RS18205	<i>tssF</i>	Type VI secretion system baseplate subunit TssF	OPERON_1710	Translocation
BMA_RS18900		Hypothetical protein	OPERON_1783	Deletion
BMA_RS18915		Porin	OPERON_1786	Deletion
BMA_RS18925		LuxR C-terminal-related transcriptional regulator	OPERON_1788	Deletion
BMA_RS18945		Hypothetical protein	OPERON_1790	Deletion
BMA_RS18955		LysR family transcriptional regulator	OPERON_1791	Deletion
BMA_RS18970		Porin	OPERON_1793	Deletion
BMA_RS18985		PucR family transcriptional regulator	OPERON_1794	Deletion
BMA_RS19025		Hypothetical protein	OPERON_1798	Deletion
BMA_RS19030		Cytochrome c peroxidase	OPERON_1799	Deletion
BMA_RS19035		Alkaline phosphatase family protein	OPERON_1799	Deletion
BMA_RS19105	<i>mhpT</i>	3-(3-hydroxy-phenyl)propionate transporter MhpT	OPERON_1809	Deletion
BMA_RS19110		Porin	OPERON_1810	Deletion
BMA_RS22565		Efflux transporter outer membrane subunit	OPERON_2149	Deletion
BMA_RS26775		Hypothetical protein	OPERON_1804	Deletion

gene-encoded proteins, including antigens used for the development of glanders ELISAs and virulence factors, are deleted, modified, or inactivated, in some isolates thereby compromising the efficacy of serodiagnostic tests using purified recombinant antigens.

Using all the *B. mallei* genome sequences available online and in-house, we undertook a comprehensive investigation of the nature of genetic variations of *B. mallei* isolates. SNPs are the most common type of genetic variation. The rate of single nucleotide alterations generated upon passage or infection has been reported as being typically very low (17). This suggests that SNPs may have a minor effect on genome plasticity, which is further supported by the small number of core SNPs identified in our study. Therefore, SNPs may provide a good level of discrimination and avoid genome alteration during passages. Indeed, the whole genome SNP-based phylogenetic analysis presented in this study demonstrates a high resolution of relationships among *B. mallei* isolates collected from diverse geographical locations (Supplementary Figure S2). Our phylogeny is quite similar to the phylogenetic relationships obtained from previous cgMLST (15, 54) or WGS-based SNP typing analyses (57). Compared to the previous three lineages (57) or 12 clusters (BHR, UAE, IND1/2/3, CHN1/2, TUR1/2/3, HUN, RUS) (15), hierarchical Bayesian clustering (FastBaps) used in this study subdivided the phylogeny into eight lineages (L1-8). The previously defined L1 is the same as our L7 (57), containing strain 6 and 20027344306 (NCTC120 or FDAARGOS\_586), while previously reported L2 was further subdivided into current L1 (cluster CHN2), L5 (cluster CHN1), L3 (cluster IND1), and L6 (clusters IND2 and BHR). The previously defined L3 consisted of our L2 (cluster TUR1), L4 (cluster RUS) and L8 (clusters TUR 2, TUR3, HUN, and IND3).

Furthermore, our eight lineages match clusters identified through high-resolution melting PCR (PCR-HRM) (58). Except for L8 containing assemblies from three subtypes: L3B2, L3B3sB1, and L3B3sB3, each lineage is comprised of only genomes from the same

PCR-HRM cluster, such as L1 (L2B2sB1Grp1), L2 (L3B3sB2), L3 (L2B2sB2), L4 (L3B1), L5 (L2B2sB1Grp2), L6 (L2B1), and L7 (L1). It is likely that L8 can be further subdivided using FastBap, which may match PCR-HRM cluster analysis. The lineage analysis in this study will provide new clues to enhance future molecular typing.

Besides SNPs, a relatively small number of INDELS were also identified across all *B. mallei* genomes available in RefSeq (Figure 1B). In contrast to consequences of SNPs, it is believed that there are some links between the accumulation of INDELS and the evolution of *B. mallei* (17). INDELS in both coding and non-coding regions were observed in the *B. mallei* genome upon passage *in vitro* and *in vivo*, thereby resulting in a few ORF frameshifts and alteration of gene expression, respectively, (17). Similar results can be seen in our data (Figures 2A, 4C, 5C).

It is generally believed that *B. mallei* evolved from an ancestral *B. pseudomallei* following serendipitous infection of a living host followed by ongoing evolution in which the genome underwent large-scale chromosome rearrangements and reduction (16, 17). Following this change, *B. mallei* continues to evolve with additional large genome reductions resulting from chromosomal rearrangements, which represent a significant contributor to genome plasticity (16, 17). Therefore, it is not surprising that a large number of SVs, specifically deletions, were observed in many *B. mallei* isolates (Figures 1C, 2B). So far, consequences, causes, and underlying mechanisms of SVs in *B. mallei* have not been well characterized. Our analysis revealed that SVs, in particular large deletions, are a significant cause for gene loss (Figure 4C) and altered operons, resulting in interference with structural gene expression (Figure 5D). Additionally, the strain JHU, obtained from a laboratory-acquired infection of *B. mallei* ATCC 23344, had evidence of an approximately 97 kb deletion starting from 1,226,811 to 1,324,421 on the CH2, compared to the reference strain ATCC 23344, which had not been reported in the previous study (17), thus suggesting that chromosome rearrangement could also occur in *B. mallei* during infection.

Consistent with this study, a previous report clearly indicated that the smaller secondary chromosome evolves more rapidly than the larger one, showing greater substitution rates and gene dispensability (59). This secondary chromosome usually serves as an evolutionary testing ground where genes are weakly preserved (18), which may explain the higher frequency of gene loss on the *B. mallei* CH2, especially *in vivo* (17). Based on previous transcriptional data, transposases on CH1 were down-regulated, and most of the over-expressed genes were located on CH2 *in vivo* (60). Over-expression of transposases on CH2 may be associated with instability of CH2 *in vivo*, which can be investigated and confirmed by future studies.

Genetically different strains of the same pathogenic bacterial species can elicit markedly diverse host immune responses during infection (20). The presence of large scale genetic changes in *B. mallei* may amplify this phenomenon during long-term evolution, particularly *in vivo*, as demonstrated in other *Burkholderia* species (61–64). Many genes in *B. mallei* CH2 have virulence or virulence-related functions, playing a pivotal role in pathogenesis (60). The extensive gene losses observed in CH2 (Figure 4B) could potentially lead to a reduction in its virulence during infection, thereby contributing to *B. mallei*'s ability to transmit and colonize between hosts. This serves as an example of optimized interaction between pathogenic bacteria and hosts to maximize their reproductive success, such as transmission, survival, and colonization (65). However, due to gene loss and alteration of gene expression, it is highly possible that antigens may be disrupted at the genomic or transcriptomic levels during infection, a phenomenon demonstrated in previous studies. For example, one spleen isolate of *B. mallei* obtained post-infection displayed a change in LPS phenotype, from smooth to rough, due to loss of O-polysaccharide during infection (66). Additionally, several *B. mallei* strains, such as *B. mallei* ATCC 23344 human and horse isolates, showed several gene deletions, including BMAA0367 (GNAT family N-acetyltransferase), BMAA0623, BMA2996, BMAA0729 (TssM), and BMA0685 (TonB) due to INDELS (17). TssM and TonB are well-known virulence factors of *B. mallei* (52, 67). Furthermore, TssM in *B. pseudomallei* has been investigated as a potential serodiagnostic biomarker (68) and used to develop a vaccine for melioidosis (69). BMAA0367, a GNAT family N-acetyltransferase, has not been well characterized in *Burkholderia* spp., but it was up-regulated during *B. mallei* infections (60), indicating that it is associated with *B. mallei*'s pathogenesis. Here, we found that genes encoding several antigens used to develop ELISA or potential serodiagnostic biomarkers, such as TssB, TssA, BimA, A03050H, and Hcp1, are not present in several *B. mallei* complete genome assemblies, due to genomic variations, but are present in other environmental *Burkholderia* spp. (Table 1). Most importantly, most of these genes are located within hotspots of gene loss on CH2 (Figure 4B). Loss of these gene-encoded antigens or virulence factors in *B. mallei* may be more likely to occur during natural infection compared to laboratory culturing since genome evolution in bacterial pathogens can occur faster and at higher frequency during the span of short host infections (13, 70–72). Based on the instability of CH2, it would be optimal to identify stable, unique, and conserved serodiagnostic biomarkers using genes located on CH1, such as GroEL (73). Moreover, by targeting multiple conserved serodiagnostic antigens from different lineages in a single serodiagnostic test, we could mitigate the effects of specific antigen loss and could cover different lineages, leading to a potential improvement in sensitivity and specificity.

Genomic variations, especially structural variations, have a significant impact on overall gene expression. Unfortunately, most SV

studies mainly focus on humans and other multicellular eukaryotic organisms. To study the effects of multiple types of variants on gene expression in bacteria, we focused on the *B. mallei* strain Zagreb. This strain was chosen due to its wide usage for diagnostic antigen production and the presence of a large number of inter-chromosomal translocations observed in its genome when compared to ATCC 23344 (data not shown). Using the simple predictive model that the destruction of the operon structure and alteration of the 5' flanking non-coding regions, including promoters and transcriptional regulators, can interfere with gene expression, this study has shown that genomic variations, specifically SVs, can alter gene expression, including some virulence genes (Figure 5B and Table 2). Interestingly, deletion of LuxI/LuxR-type QS systems observed in the strain Zagreb may also be indirectly altering the expression of many other genes, including those of virulence factors, toxins, and biofilm components (74).

In this study, we used all available *B. mallei* genome assemblies in the RefSeq database and assemblies of four additional in-house *B. mallei* strains. It is important to note that the quality of bacterial genome assemblies in the RefSeq database may vary. To address this issue, we assessed the quality of the 112 genome assemblies and annotations using BUSCO's completeness score (Supplementary File S1) for quality control. However, it should be acknowledged that the presence of sequencing and assembly errors, in addition to accuracy of aligners and variant callers, may affect the genetic variation identification performed in our study (75). Nevertheless, this limitation must be taken into consideration. Besides that, bacterial gene expression and regulation are complicated and have yet to be fully elucidated, so it is hard to comprehensively investigate the effect of genomic variations on gene expression using the limited sample size in this study. Furthermore, there is a significant difference in gene expression profiling of a *B. mallei* isolate between *in vivo* and *in vitro*, particularly with most of the over-expressed genes on CH2 *in vivo* (60). This indicates *in vitro* transcriptional data may not be instrumental in predicting *in vivo* gene expression and regulation, and the possibility that virulence genes or serodiagnostic markers located on CH2 show variable expression between different isolates during infection will be investigated further. However, it is likely that the disruption of operon structures by SVs, such as deletion and translocation, impacts the expression of genes within or near the operon during infection as well, since these co-regulated genes under a common control system are interrupted (46). The specific impacts of SVs on gene expression needs to be further investigated using more transcriptome data, especially in relation to bacterial infection and virulence factors.

## Conclusion

In summary, we have demonstrated that the plasticity of the *B. mallei* genome interferes with its gene content and expression. Accordingly, methods of glanders serodiagnosis using only purified recombinant antigens could be potentially compromised. Multiple instances have demonstrated that certain targeted antigens, including BimA, HCP1, TssA, TssB, TssM, A03050H, or potentially valuable antigens like virulence factors, may undergo loss or significant changes in expression due to genetic variations. As a result, these alterations can potentially lead to modifications in the immune response, including antibody production, during infection. This study provides novel insights into *B. mallei* genome plasticity, including details and consequences of

genomic variation. These findings will be useful for informing improved glanders serodiagnosis and molecular typing in the future.

## Data availability statement

The whole-genome assemblies of strains ATCC 23344, Zagreb, Mukteswar and Bogor have been deposited in GenBank under accession number CP124607-CP124608, CP124605-CP124606, CP124602-CP124603, and CP124604, respectively. All the raw sequencing data are available under BioProject number PRJNA954830. Specifically, the RNA-sequencing raw data are available in the NCBI Sequence Read Archive under accession numbers SRR24153017 to SRR24153022 and whole genome sequencing raw data are available in the NCBI Sequence Read Archive under accession numbers SRR24860065 to SRR24860072.

## Author contributions

MK designed the experiments. MK, RG, and PC wrote the manuscript. PC, RG, JC, EH, SN-D, DC, JH, KM, M-OD, and MK reviewed and edited the manuscript. JC, EH, and SN-D performed WGS. PC, MK, RG, JC, and M-OD analyzed sequencing data. DC, JH, and KM conducted bacterial culture and RNA purification experiments. All authors contributed to the article and approved the submitted version.

## Funding

This research was supported by the Canadian Food Inspection Agency.

## Acknowledgments

We would like to thank the Friedrich Loeffler Institute for sharing *B. mallei* strains Zagreb, Mukteswar, and Bogor with us. We also thank Centre d'expertise et de service Génome Québec for performing the RNA sequencing.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

- Whitlock GC, Valbuena GA, Popov VL, Judy BM, Estes DM, Torres AG. *Burkholderia mallei* cellular interactions in a respiratory cell model. *J Med Microbiol.* (2009) 58:554–62. doi: 10.1099/jmm.0.007724-0
- Van Zandt KE, Greer MT, Gelhaus HC. Glanders: an overview of infection in humans. *Orphanet J Rare Dis.* (2013) 8:131. doi: 10.1186/1750-1172-8-131
- Kettle AN, Wernery U. Glanders and the risk for its introduction through the international movement of horses. *Equine Vet J.* (2016) 48:654–8. doi: 10.1111/evj.12599
- Khan I, Wieler LH, Melzer F, Elschner MC, Muhammad G, Ali S, et al. Glanders in animals: a review on epidemiology, clinical presentation, diagnosis and countermeasures. *Transbound Emerg Dis.* (2013) 60:204–21. doi: 10.1111/j.1865-1682.2012.01342.x
- Estes DM, Dow SW, Schweizer HP, Torres AG. Present and future therapeutic strategies for melioidosis and glanders. *Expert Rev Anti-Infect Ther.* (2010) 8:325–38. doi: 10.1586/eri.10.4
- Elschner MC, Melzer F, Singha H, Muhammad S, Gardner I, Neubauer H. Validation of a commercial glanders ELISA as an Alternative to the CFT in international trade of equidae. *Front Vet Sci.* (2021) 8:628389. doi: 10.3389/fvets.2021.628389
- Elschner MC, Scholz HC, Melzer F, Saqib M, Marten P, Rassbach A, et al. Use of a Western blot technique for the serodiagnosis of glanders. *BMC Vet Res.* (2011) 7:4. doi: 10.1186/1746-6148-7-4
- Sengyee S, Yoon SH, West TE, Ernst RK, Chantratita N. Lipopolysaccharides from different burkholderia species with different lipid a structures induce toll-like receptor

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fvets.2023.1217135/full#supplementary-material>

### SUPPLEMENTARY FIGURE S1

Pan-genome analysis of 112 *B. mallei* genomes. (A) Rarefaction curves for the core (persistent) genome size and pan-genome. (B) U-shaped distribution demonstrating gene cluster count among 112 *B. mallei* genomes.

### SUPPLEMENTARY FIGURE S2

SNP-based Phylogenetic tree of *B. mallei* isolates and related metadata. The phylogenetic tree was obtained by IQtree with 1000 ultrafast bootstraps. Lineages (L1-8) were determined by Fastbp. Geographic location was collected from BioSample and Assembly database from NCBI: Bahrain (BH), Brazil (BR), China (CN), Hungary (HU), India (IN), Indonesia (ID), Iran (IR), Myanmar (MM), Pakistan (PK), Russia (RU), Turkey (TR), United Kingdom (UK), United States (US), and Yugoslavia (YU).

### SUPPLEMENTARY FIGURE S3

Differences in gene expression and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways between strain ATCC 23344 and Zagreb. (A) PCoA plot of gene expression between strain ATCC 23344 and Zagreb based on Bray-Curtis dissimilarity. Three biological replicates for each strain. (B) KEGG pathways correlating to virulence DEGs between strain ATCC 23344 and Zagreb. The red color bars represent pathways related to bacterial pathogeny and virulence.

### SUPPLEMENTARY TABLE S1

Summary of genome assembly information.

### SUPPLEMENTARY TABLE S2

Number of reads for each sample in RNA-seq experiment.

### SUPPLEMENTARY FILE S1

Genomes and assemblies used in this study.

### SUPPLEMENTARY FILE S2

Summary of DEGs generated from RNA-seq analysis.

### SUPPLEMENTARY FILE S3

Virulence factors predicted by PathoFact.

### SUPPLEMENTARY FILE S4

Operons on CH1 predicted by OperonSEQer.

### SUPPLEMENTARY FILE S5

Operons on CH2 predicted by OperonSEQer.

- 4 activation and react with melioidosis patient sera. *Infect Immun.* (2019) 87:e00692-19. doi: 10.1128/IAI.00692-19
9. Stone JK, Mayo M, Grasso SA, Ginther JL, Warrington SD, Allender CJ, et al. Detection of *Burkholderia pseudomallei* O-antigen serotypes in near-neighbor species. *BMC Microbiol.* (2012) 12:250. doi: 10.1186/1471-2180-12-250
10. Elschner MC, Laroucau K, Singha H, Tripathi BN, Saqib M, Gardner I, et al. Evaluation of the comparative accuracy of the complement fixation test, Western blot and five enzyme-linked immunosorbent assays for serodiagnosis of glanders. *PLoS One.* (2019) 14:e0214963. doi: 10.1371/journal.pone.0214963
11. Singha H, Malik P, Goyal SK, Khurana SK, Mukhopadhyay C, Eshwara VK, et al. Optimization and validation of indirect ELISA using truncated TssB protein for the serodiagnosis of glanders amongst equines. *ScientificWorldJournal.* (2014) 2014:469407:1-6. doi: 10.1155/2014/469407
12. Godoy D, Randle G, Simpson AJ, Aanensen DM, Pitt TL, Kinoshita R, et al. Multilocus sequence typing and evolutionary relationships among the causative agents of melioidosis and glanders, *Burkholderia pseudomallei* and *Burkholderia mallei*. *J Clin Microbiol.* (2003) 41:2068-79. doi: 10.1128/JCM.41.5.2068-2079.2003
13. Losada L, Ronning CM, Deshazer D, Woods D, Fedorova N, Kim HS, et al. Continuing evolution of *Burkholderia mallei* through genome reduction and large-scale rearrangements. *Genome Biol Evol.* (2010) 2:102-16. doi: 10.1093/gbe/evq003
14. Nierman WC, Deshazer D, Kim HS, Tettelin H, Nelson KE, Feldblyum T, et al. Structural flexibility in the *Burkholderia mallei* genome. *Proc Natl Acad Sci U S A.* (2004) 101:14246-51. doi: 10.1073/pnas.0403306101
15. Appelt S, Rohleder AM, Jacob D, Von Buttlar H, Georgi E, Mueller K, et al. Genetic diversity and spatial distribution of *Burkholderia mallei* by core genome-based multilocus sequence typing analysis. *PLoS One.* (2022) 17:e0270499. doi: 10.1371/journal.pone.0270499
16. Bochkareva OO, Moroz EV, Davydov I, Gelfand MS. Genome rearrangements and selection in multi-chromosome bacteria *Burkholderia* spp. *BMC Genomics.* (2018) 19:965. doi: 10.1186/s12864-018-5245-1
17. Romero CM, Deshazer D, Feldblyum T, Ravel J, Woods D, Kim HS, et al. Genome sequence alterations detected upon passage of *Burkholderia mallei* ATCC 23344 in culture and in mammalian hosts. *BMC Genomics.* (2006) 7:228. doi: 10.1186/1471-2164-7-228
18. Cooper VS, Vohr SH, Wrocklage SC, Hatcher PJ. Why genes evolve faster on secondary chromosomes in bacteria. *PLoS Comput Biol.* (2010) 6:e1000732. doi: 10.1371/journal.pcbi.1000732
19. Laroucau K, Aaziz R, Vorimore F, Varghese K, Deshayes T, Bertin C, et al. A genetic variant of *Burkholderia mallei* detected in Kuwait: Consequences for the PCR diagnosis of glanders. *Transbound Emerg Dis.* (2021) 68:960-3. doi: 10.1111/tbed.13777
20. Sela U, Euler CW, Correa Da Rosa J, Fischetti VA. Strains of bacterial species induce a greatly varied acute adaptive immune response: The contribution of the accessory genome. *PLoS Pathog.* (2018) 14:e1006726. doi: 10.1371/journal.ppat.1006726
21. Feron R, Waterhouse RM. Assessing species coverage and assembly quality of rapidly accumulating sequenced genomes. *Gigascience.* (2022) 11:giac006. doi: 10.1093/gigascience/giac006
22. Lawniczak MKN, Durbin R, Flicek P, Lindblad-Toh K, Wei X, Archibald JM, et al. Standards recommendations for the Earth BioGenome Project. *Proc Natl Acad Sci U S A.* (2022) 119:e2115639118. doi: 10.1073/pnas.2115639118
23. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics.* (2015) 31:3210-2. doi: 10.1093/bioinformatics/btv351
24. Lopez J, Copps J, Wilhelmsen C, Moore R, Kubay J, St-Jacques M, et al. Characterization of experimental equine glanders. *Microbes Infect.* (2003) 5:1125-31. doi: 10.1016/j.micinf.2003.07.004
25. Kang M, Chmara J, Naushad S, Huang H. Complete genome sequence of a canadian strain of *Raoultella planticola* with metal and antimicrobial resistance genes. *Microbiol Resour Announc.* (2021) 10:e0041521. doi: 10.1128/MRA.00415-21
26. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics.* (2018) 34:i884-90. doi: 10.1093/bioinformatics/bty560
27. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol.* (2017) 13:e1005595. doi: 10.1371/journal.pcbi.1005595
28. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol.* (2019) 37:540-6. doi: 10.1038/s41587-019-0072-8
29. Hu J, Fan J, Sun Z, Liu S. NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics.* (2020) 36:2253-5. doi: 10.1093/bioinformatics/btz891
30. Wick RR, Holt KE. Polypolish: Short-read polishing of long-read bacterial genome assemblies. *PLoS Comput Biol.* (2022) 18:e1009802. doi: 10.1371/journal.pcbi.1009802
31. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics.* (2018) 34:3094-100. doi: 10.1093/bioinformatics/bty191
32. Okonechnikov K, Conesa A, Garcia-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics.* (2016) 32:292-4. doi: 10.1093/bioinformatics/btv566
33. Tatusova T, Dicuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, et al. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* (2016) 44:6614-24. doi: 10.1093/nar/gkw569
34. O'donnell S, Fischer G. MUM&Co: accurate detection of all SV types through whole-genome alignment. *Bioinformatics.* (2020) 36:3242-3. doi: 10.1093/bioinformatics/btaa115
35. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol.* (2020) 37:1530-4. doi: 10.1093/molbev/msaa015
36. Tonkin-Hill G, Lees JA, Bentley SD, Frost SDW, Corander J. Fast hierarchical Bayesian analysis of population structure. *Nucleic Acids Res.* (2019) 47:5539-49. doi: 10.1093/nar/gkz361
37. Mortazavi M, Ren Y, Saini S, Antaki D, St Pierre CL, Williams A, et al. SNPs, short tandem repeats, and structural variants are responsible for differential gene expression across C57BL/6 and C57BL/10 substrains. *Cell Genom.* (2022) 2:100102. doi: 10.1016/j.xgen.2022.100102
38. Gautreau G, Bazin A, Gachet M, Planel R, Burlot L, Dubois M, et al. PpanGGOLiN: Depicting microbial diversity via a partitioned pangenome graph. *PLoS Comput Biol.* (2020) 16:e1007732. doi: 10.1371/journal.pcbi.1007732
39. Yu G. Using ggtree to Visualize Data on Tree-Like Structures. *Curr Protoc Bioinformatics.* (2020) 69:e96. doi: 10.1002/cpbi.96
40. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics.* (2017) 33:2938-40. doi: 10.1093/bioinformatics/btx364
41. De Nies L, Lopes S, Busi SB, Galata V, Heintz-Buschart A, Laczny CC, et al. PathoFact: a pipeline for the prediction of virulence factors and antimicrobial resistance genes in metagenomic data. *Microbiome.* (2021) 9:49. doi: 10.1186/s40168-020-00993-9
42. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol.* (2016) 34:525-7. doi: 10.1038/nbt.3519
43. Pimentel H, Bray NL, Puente S, Melsted P, Pachter L. Differential analysis of RNA-seq incorporating quantification uncertainty. *Nat Methods.* (2017) 14:687-90. doi: 10.1038/nmeth.4324
44. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol.* (2016) 428:726-31. doi: 10.1016/j.jmb.2015.11.006
45. Kanehisa M, Sato Y, Kawashima M. KEGG mapping tools for uncovering hidden features in biological data. *Protein Sci.* (2022) 31:47-53. doi: 10.1002/pro.4172
46. Lawrence JG. Shared strategies in gene organization among prokaryotes and eukaryotes. *Cells.* (2002) 110:407-13. doi: 10.1016/S0092-8674(02)00900-5
47. Krishnakumar R, Ruffing AM. OperonSEQR: A set of machine-learning algorithms with threshold voting for detection of operon pairs using short-read RNA-sequencing data. *PLoS Comput Biol.* (2022) 18:e1009731. doi: 10.1371/journal.pcbi.1009731
48. Hamala T, Wafula EK, Guiltinan MJ, Ralph PE, Depamphilis CW, Tiffin P. Genomic structural variants constrain and facilitate adaptation in natural populations of *Theobroma cacao*, the chocolate tree. *Proc Natl Acad Sci U S A.* (2021) 118:e2102914118. doi: 10.1073/pnas.2102914118
49. Wagner GE, Berner A, Lipp M, Kohler C, Assig K, Lichtenegger S, et al. Protein microarray-guided development of a highly sensitive and specific dipstick assay for glanders serodiagnostics. *J Clin Microbiol.* (2023) 61:e0123422. doi: 10.1128/jcm.01234-22
50. Pal V, Kumar S, Malik P, Rai GP. Evaluation of recombinant proteins of *Burkholderia mallei* for serodiagnosis of glanders. *Clin Vaccine Immunol.* (2012) 19:1193-8. doi: 10.1128/CI.00137-12
51. Kumar S, Malik P, Verma SK, Pal V, Gautam V, Mukhopadhyay C, et al. Use of a recombinant burkholderia intracellular motility a protein for immunodiagnosis of glanders. *Clin Vaccine Immunol.* (2011) 18:1456-61. doi: 10.1128/CI.05185-11
52. Shanks J, Burtnick MN, Brett PJ, Waag DM, Spurgers KB, Ribot WJ, et al. *Burkholderia mallei* tssM encodes a putative deubiquitinase that is secreted and expressed inside infected RAW 264.7 murine macrophages. *Infect Immun.* (2009) 77:1636-48. doi: 10.1128/IAI.01339-08
53. Brockhurst MA, Harrison E, Hall JJP, Richards T, McNally A, Maclean C. The ecology and evolution of pangenomes. *Curr Biol.* (2019) 29:R1094-103. doi: 10.1016/j.cub.2019.08.012
54. Brangsch H, Saqib M, Sial AUR, Melzer F, Linde J, Elschner MC. Sequencing-based genotyping of Pakistani *Burkholderia mallei* strains: a useful way for investigating glanders outbreaks. *Pathogens.* (2022) 11:614. doi: 10.3390/pathogens11060614
55. Brangsch H, Singha H, Laroucau K, Elschner M. Sequence-based detection and typing procedures for *Burkholderia mallei*: Assessment and prospects. *Front Vet Sci.* (2022) 9:1056996. doi: 10.3389/fvets.2022.1056996
56. Wagner JK, Setayeshgar S, Sharon LA, Reilly JP, Brun YV. A nutrient uptake role for bacterial cell envelope extensions. *Proc Natl Acad Sci U S A.* (2006) 103:11772-7. doi: 10.1073/pnas.0602047103

57. Laroucau K, De Assis L, Santana V, Girault G, Martin B, Miranda Da Silveira PP, et al. First molecular characterisation of a Brazilian *Burkholderia mallei* strain isolated from a mule in 2016. *Infect Genet Evol.* (2018) 57:117–20. doi: 10.1016/j.meegid.2017.11.014
58. Girault G, Wattiau P, Saqib M, Martin B, Vorimore F, Singha H, et al. High-resolution melting PCR analysis for rapid genotyping of *Burkholderia mallei*. *Infect Genet Evol.* (2018) 63:1–4. doi: 10.1016/j.meegid.2018.05.004
59. Morrow JD, Cooper VS. Evolutionary effects of translocations in bacterial genomes. *Genome Biol Evol.* (2012) 4:1256–62. doi: 10.1093/gbe/evs099
60. Kim HS, Schell MA, Yu Y, Ulrich RL, Sarria SH, Nierman WC, et al. Bacterial genome adaptation to niches: divergence of the potential virulence genes in three *Burkholderia* species of different survival strategies. *BMC Genomics.* (2005) 6:174. doi: 10.1186/1471-2164-6-174
61. Diaz Caballero J, Clark ST, Wang PW, Donaldson SL, Coburn B, Tullis DE, et al. A genome-wide association analysis reveals a potential role for recombination in the evolution of antimicrobial resistance in *Burkholderia multivorans*. *PLoS Pathog.* (2018) 14:e1007453. doi: 10.1371/journal.ppat.1007453
62. Hassan AA, Dos Santos SC, Cooper VS, Sa-Correia I. Comparative evolutionary patterns of *Burkholderia cenocepacia* and *B. multivorans* during chronic co-infection of a cystic fibrosis patient lung. *Front Microbiol.* (2020) 11:574626. doi: 10.3389/fmicb.2020.574626
63. Lee AH, Flibotte S, Sinha S, Paiero A, Ehrlich RL, Balashov S, et al. Phenotypic diversity and genotypic flexibility of *Burkholderia cenocepacia* during long-term chronic infection of cystic fibrosis lungs. *Genome Res.* (2017) 27:650–62. doi: 10.1101/gr.213363.116
64. Lieberman TD, Michel JB, Aingaran M, Potter-Bynoe G, Roux D, Davis MR Jr, et al. Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes. *Nat Genet.* (2011) 43:1275–80. doi: 10.1038/ng.997
65. Diard M, Hardt WD. Evolution of bacterial virulence. *FEMS Microbiol Rev.* (2017) 41:679–97. doi: 10.1093/femsre/fux023
66. Bernhards RC, Cote CK, Amemiya K, Waag DM, Klimko CP, Worsham PL, et al. Characterization of *in vitro* phenotypes of *Burkholderia pseudomallei* and *Burkholderia mallei* strains potentially associated with persistent infection in mice. *Arch Microbiol.* (2017) 199:277–301. doi: 10.1007/s00203-016-1303-8
67. Mott TM, Vijayakumar S, Sbrana E, Endsley JJ, Torres AG. Characterization of the *Burkholderia mallei* tonb mutant and its potential as a backbone strain for vaccine development. *PLoS Negl Trop Dis.* (2015) 9:e0003863. doi: 10.1371/journal.pntd.0003863
68. Tan KS, Chen Y, Lim YC, Tan GY, Liu Y, Lim YT, et al. Suppression of host innate immune response by *Burkholderia pseudomallei* through the virulence factor TssM. *J Immunol.* (2010) 184:5160–71. doi: 10.4049/jimmunol.0902663
69. Burtneck MN, Shaffer TL, Ross BN, Muruato LA, Sbrana E, Deshazer D, et al. Development of subunit vaccines that provide high-level protection and sterilizing immunity against acute inhalational melioidosis. *Infect Immun.* (2018) 86:e00724-17. doi: 10.1128/IAI.00724-17
70. Grote A, Earl AM. Within-host evolution of bacterial pathogens during persistent infection of humans. *Curr Opin Microbiol.* (2022) 70:102197. doi: 10.1016/j.mib.2022.102197
71. Kraft C, Stack A, Josenhans C, Niehus E, Dietrich G, Correa P, et al. Genomic changes during chronic *Helicobacter pylori* infection. *J Bacteriol.* (2006) 188:249–54. doi: 10.1128/JB.188.1.249-254.2006
72. Mikonranta L, Mappes J, Laakso J, Ketola T. Within-host evolution decreases virulence in an opportunistic bacterial pathogen. *BMC Evol Biol.* (2015) 15:165. doi: 10.1186/s12862-015-0447-5
73. Dohre SK, Kamthan A, Singh S, Alam SI, Kumar S. Identification of a new diagnostic antigen for glanders using immunoproteome analysis. *Comp Immunol Microbiol Infect Dis.* (2017) 53:26–32. doi: 10.1016/j.cimid.2017.06.007
74. Kai K. Bacterial quorum sensing in symbiotic and pathogenic relationships with hosts. *Biosci Biotechnol Biochem.* (2018) 82:363–71. doi: 10.1080/09168451.2018.1433992
75. Olson ND, Lund SP, Colman RE, Foster JT, Sahl JW, Schupp JM, et al. Best practices for evaluating single nucleotide variant calling methods for microbial genomics. *Front Genet.* (2015) 6:235. doi: 10.3389/fgene.2015.00235