BMC Genomics

RESEARCH

Open Access

Large-scale genomic analysis of *Elizabethkingia anophelis*

Pavel Andriyanov[1*], Pavel Zhurilov[1], Alena Menshikova[1†], Anastasia Tutrina[1], Ivan Yashin[1] and Daria Kashina[1†]

Abstract

The recent emergence of *Elizabethkingia anophelis* as a human pathogen is a major concern for global public health. This organism has the potential to cause severe infections and has inherent antimicrobial resistance. The potential for widespread outbreaks and rapid global spread highlights the critical importance of understanding the biology and transmission dynamics of this infectious agent. We performed a large-scale analysis of available 540 *E. anophelis*, including one novel strain isolated from raw milk and sequenced in this study. Pan-genome analysis revealed an open and diverse pan-genome in this species, characterized by the presence of many accessory genes. This suggests that the species has a high level of adaptability and can thrive in a variety of environments. Phylogenetic analysis has also revealed a complex population structure, with limited source-lineage correlation. We identified diverse antimicrobial resistance factors, including core-genome and accessory ones often associated with mobile genetic elements within specific lineages. Mobilome analysis revealed a dynamic landscape primarily composed of genetic islands, integrative and conjugative elements, prophage elements, and small portion of plasmids emphasizing a complex mechanism of horizontal gene transfer. Our study underscores the adaptability of *E. anophelis*, characterized by a diverse range of antimicrobial resistance genes, putative virulence factors, and genes enhancing fitness. This adaptability is also supported by the organism's ability to acquire genetic material through horizontal gene transfer, primarily facilitated by mobile genetic elements such as integrative and conjugative elements (ICEs). The potential for rapid evolution of this emerging pathogen poses a significant challenge to public health efforts.

Keywords *Elizabethkingia anophelis*, Emerging pathogen, Genomics, Antimicrobial resistance, Virulence

Background

Emerging infectious diseases (EIDs) pose significant threats to global public health, encompassing newly identified pathogens or those increasing in prevalence within susceptible populations [1, 2]. Bacteria, alongside viruses, play a critical role in this global challenge [3–6].

Elizabethkingia anophelis exemplifies a particularly concerning emerging bacterial pathogen. *E. anophelis* is an aerobic, non-fermentative, non-motile, Gram-negative bacilli that belong to the *Weeksellaceae* family, *Flavobacteriales* order, and *Bacteroidota* phylum. Recently recognized for its potential to cause severe and life-threatening infections, *E. anophelis* poses a significant threat to both immunocompromised and immunocompetent individuals, with reported mortality rates reaching up to 70% [7–9]. Originally isolated from the midgut of the *Anopheles gambiae* mosquito, a known malaria vector [10], *E. anophelis* is now recognized as a prevalent member of the mosquito gut microbiota [11, 12]. While mosquitoes can transmit certain diseases, their exact role in *E. anophelis* transmission remains unclear, with no documented transmission events. It is hypothesized that *E. anophelis* may be transmitted through alternative routes, such as contact with contaminated environments

†Alena Menshikova and Daria Kashina contributed equally to this work.

*Correspondence:
Pavel Andriyanov
andriyanovpvl@gmail.com
[1] Federal Research Center for Virology and Microbiology, Branch in Nizhny Novgorod, Nizhny Novgorod, Russia

or person-to-person spread. Many epidemiological investigations suggest water as a primary reservoir for this pathogen [13–18], consistent with its ubiquitous nature across diverse environments [14, 19, 20].

The first documented case of *E. anophelis* infection occurred in 2011 in the Central African Republic, where the bacterium was isolated from the cerebrospinal fluid of an 8-day-old girl [21]. Before 2016, nosocomial outbreaks were reported in Singapore [22], Hong Kong [23], and England [24]. During the period from 2015 to 2016, a major *Elizabethkingia* outbreak was first observed in the USA, with the onset of illness primarily occurring in community settings [15]. The outbreak strain exhibited the disruption of the DNA repair *mutY* gene caused by the insertion of an integrative and conjugative element (ICE), resulting in the emergence of a hypermutator variant. During the outbreak, there were a total of 66 confirmed cases of infection, with 20 fatalities. Following the 2015–2016 USA outbreak, *E. anophelis* has become a global concern. Sporadic cases of both hospital-acquired and community-acquired infections have been reported worldwide [14, 25–27]. Notably, France witnessed the first in Europe outbreak in 2020–2021, with 20 reported cases and 9 fatalities [16]. More recently, cases have been identified in Nepal, Japan, and India (2022–2023) [28–30], with the most recent one reported in the Netherlands in February 2024 and in Hanoi, Vietnam, in March 2024 [18, 31]. Nearly all reported cases of *E. anophelis* infection manifest as severe illnesses, with meningitis, pneumonia, and bloodstream infections being the most frequent presentations.

Another notable feature of this bacterium is its natural resistance to a wide range of antimicrobials. *E. anophelis* is characterized by intrinsic resistance to a broad spectrum of clinically important antibiotics, primarily mediated by chromosomally encoded determinants. Notably, this resistance encompasses almost all beta-lactam antibiotics due to the presence of three distinct beta-lactamase genes: bla_{CME}, an Ambler class A serine extended-spectrum beta-lactamase (ESBL), and bla_{B} and bla_{GOB}, which encode Ambler class B metallo-beta-lactamases (MBLs) [32–34]. Additionally, *E. anophelis* genomes harbor various other chromosomally encoded genes that confer resistance to a diverse range of antimicrobials. These genes include those encoding efflux pumps, enzyme-degrading enzymes, and enzyme-modifying enzymes, further complicating therapy options [25, 31, 35, 36]. Moreover, *E. anophelis* strains exhibiting multidrug resistance phenotypes have also been shown to form biofilms [37]. Biofilms are structured bacterial communities enclosed within an extracellular matrix composed of a mixture of chemicals that makes bacteria tolerant to antimicrobial agents and host responses [38, 39].

The severity of clinical manifestations, coupled with the global emergence of multidrug-resistant strains, underscores the urgent need for comprehensive research to better understand *E. anophelis* and develop effective therapeutic strategies. This concern is further amplified by recent findings from Shaohua Hu et al., who employed next-generation sequencing (NGS) technology for an extensive analysis of the global distribution of *E. anophelis*, with a particular focus on Asian clinical isolates. Their study highlighted the potential for large-scale outbreaks and the rapid global spread of this pathogen [7].

In the present study, we report an expanded large-scale genomic analysis of *E. anophelis*, encompassing 540 genomes. The analysis included one novel strain isolated and sequenced in our laboratory, along with 339 publicly available genomes and 200 *de novo* assembled genomes from NCBI SRA. We employed a multifaceted approach, including pan-genome analysis, phylogenetic analysis, virulome and resistome analysis, as well as mobilome analysis. Additionally, phenotypic characterization of our isolate was performed to understand its biological properties.

Results

Bacterial isolation, identification, and microbiological characteristics

E. anophelis MR2-16/1 was isolated from an unpasteurized bovine milk sample on tryptic soy agar (TSA) medium supplemented with antibiotics. On TSA, MR2-16/1 colonies exhibited a slight yellow pigmentation. The obtained pure culture displayed positive oxidase and catalase reactions. Gram-stained smears revealed typical gram-negative bacilli. The assembled 16 S rRNA gene fragment was 1,364 base pairs long, demonstrating 99.93% identity with the *E. anophelis* $R26^T$ 16 S rRNA gene sequence.

Colony morphology on blood agar is illustrated in Fig. 1A. On blood agar, MR2-16/1 formed circular, large-sized (1–3 mm), opaque colonies with an entire margin and convex elevation. Colonies appeared smooth and glistening, with no observed pigmentation on blood agar media. Beta-hemolysis was observed after 24 h of incubation at 37 °C on blood agar supplemented with 5% O-human blood, but not with 5% bovine blood (Fig. 1B, C). The hemolytic activity exhibited a delayed pattern, with clear lysis zones visible only beneath extensive colonies characterized by higher biomass. Hemolysis was not evident around individual colonies after 24 h of incubation.

Strain MR2-16/1 exhibited a non-motile, capsule-negative phenotype (Table S1). It also lacked lecithinase,

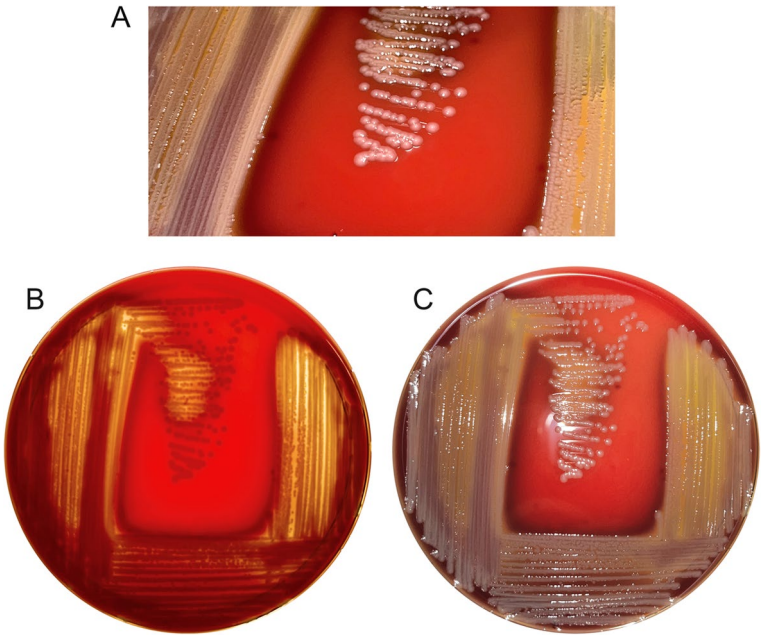

Fig. 1 Colony morphology of *E. anophelis* MR2-16/1 on blood agar with 5% O-human blood and hemolysis after 24 h incubation: **A** The colony morphology of *E. anophelis* MR2-16/1 after 24 h incubation. The colonies are circular-shaped, large-sized (1–3 mm), opaque with an entire margin and convex elevation; **B** The colonies in passing light: the "lagging" hemolysis is noticeable, with clear lysis zones beneath extensive biomass; **C** Photography in directional light

protease, lipase, and gelatinase activity, and displayed no adhesion to human erythrocytes. Positive reactions were recorded for urease, ornithine decarboxylase, lysine decarboxylase, esculin hydrolysis, and β-galactosidase tests. Negative results were obtained for arginine decarboxylase, melibiose fermentation, and trehalose fermentation.

Comparative analysis of phenotypic features between our strain and the R26 type strain revealed differences in lysine decarboxylase, ornithine decarboxylase, and urease activities (Table S1).

Antimicrobial susceptibility profile of *E. anophelis* MR2-16/1

The MR2-16/1 strain displayed an extensive antimicrobial resistance (AMR) profile, exhibiting resistance to 17 out of 24 tested antibiotics. The results of the testing are listed in Table 1.

The strain was susceptible to the following antibiotics: cefoperazone-sulbactam, tigecycline, piperacillin-tazobactam, and trimethoprim-sulfamethoxazole. In vitro, susceptibility to piperacillin-tazobactam was widely reported among various *Elizabethkingia* spp. clinical isolates [37, 40–42]. However, in a study published in 2023 by Mei-Chen Tan and colleagues, it was found that none of the methods used to determine the susceptibility to piperacillin-tazobactam correlated satisfactorily with the minimum bactericidal concentrations and clinical outcomes in *Elizabethkingia* isolates [42]. Therefore, it appears that our MR2-16/1 strain also had an intrinsic resistance to this combination.

The Intermediate (I - susceptible, increased exposure) category was interpreted for the following antibiotics: cefepime, cefoperazone, and ciprofloxacin.

The MR2-16/1 strain was resistant to amikacin, ampicillin, ampicillin-sulbactam, aztreonam, cefazolin, cefuroxime, chloramphenicol, colistin, ertapenem, gentamicin, meropenem, netilmicin, piperacillin, tetracycline, tobramycin. Such an AMR profile meets the criteria for MDR due to resistance across multiple antibiotic classes.

Phylogenomics of *E. anophelis*

We used the filtered core genome alignment from Panaroo to examine phylogenetic relationships among 540 *E. anophelis* genomes. The phylogenetic tree was built from SNP data of 2,922 core genes, with a total alignment length of 169,247 bp. Clustering with fastbaps produced 30, 87, 158, and 214 clusters at levels 1, 2, 3, and 4, respectively. We chose level 1 resolution for further analysis since genomes belongs to the same species.

In general, *E. anophelis* demonstrated a complex phylogeny (Fig. 2). Instances of strains with divergent metadata (source type and country) sharing common ancestors suggested a shared origin. Within the 30

Table 1 Minimal inhibitory concentrations (MICs) of E. Anophelis MR2-16/1 strain determined for 24 antibiotics

Antimicrobial Class	Antibiotic	MIC	Interpretation	Guidance*
Beta-lactams	Ampicillin	> 128	R	EU_PK/PD
	Ampicillin-sulbactam	64/32	R	EU_PK/PD
	Aztreonam	> 16	R	EU_PK/PD
	Cefazolin	> 16	R	EU_PK/PD
	Cefepime	8	I	EU_PK/PD
	Cefuroxime	> 64	R	EU_PK/PD
	Cefoperazone	32	I	CLSI_Entero_2020
	Cefoperazone-Sulbactam	4/2	S	EU_PK/PD
	Cefotaxime	> 8	R	EU_PK/PD
	Ceftazidime	> 16	R	EU_PK/PD
	Piperacillin	32	R	EU_PK/PD
	Piperacillin-tazobactam	8/4	S	EU_PK/PD
	Ertapenem	> 2	R	EU_PK/PD
	Meropenem	> 16	R	EU_PK/PD
Aminoglycosides	Amikacin	> 64	R	EU_PK/PD
	Gentamicin	> 32	R	EU_PK/PD
	Tobramycin	> 8	R	EU_PK/PD
	Netilmicin	> 16	R	EU_Entero
Tetracyclines	Tetracycline	32	R	CLSI_Entero_2020
	Tigecycline	0.5	S	EU_PK/PD
Fluoroquinolones	Ciprofloxacin	0.5	I	EU_PK/PD
Phenicols	Chloramphenicol	> 32	R	EU_PK/PD
Polymyxins	Colistin	> 16	R	EU_Entero
Antimetabolites	Trimethoprim-sulfamethoxazole	1/19	S	EU_PK/PD

EU_PK/PD* - EUCAST 2023 Pharmacokinetic/Pharmacodynamic

EU_Entero* - EUCAST 2023 *Enterobacteriaceae*

CLSI_Entero_2020* - CLSI M100 2020 *Enterobacteriaceae*

clusters, the most substantial ones were Baps clusters 14 ($n=76$ strains), 3 ($n=62$), 5 ($n=57$), 13 ($n=37$), 18 ($n=34$), 9 ($n=32$), and 19 ($n=31$). The resulting phylogenetic analysis revealed a division into two main branches, with clusters 14, 25, and 26 constituting a distinct lineage, and the remaining clusters forming a second lineage (Fig. 2).

Cluster 14 consisted of the most diverse strains, originating from various sources such as clinical, environmental, veterinary, and food. The strains were collected predominantly in USA (40 strains) and China (16 strains), but also in other countries from Europe, Asia, and Australia. Cluster 3 contained mostly USA strains isolated and sequenced during the *Elizabethkingia* outbreak in the United States between 2015 and 2016 [15], with 8 additional strains from human clinical samples originating from China, 4 from environmental samples in Brazil, and 3 from clinical samples from Australia. Cluster 5 was also represented mainly by USA clinical strains, as well as strains from China, Taiwan, India, and Sweden, with one environmental strain from Australia. Cluster 13 mainly consisted of clinical strains originating from China and a few from the United States, while cluster 18 included strains from China, the USA, Australia, Taiwan and Canada. Cluster 9 was primarily comprised of strains originating in Asia (China, Taiwan, Vietnam and Japan), with a small number of strains from Australia and the United States included in this cluster. Cluster 19 consisted only of strains from China, with a very small number of strains from the United States and France included in this cluster.

Some other clusters, both medium-sized and smaller, were composed of strains with mixed origin or originating solely from a single country. Cluster 8 consisted of strains isolated from Taiwan, while clusters 1, 2, and 4 were composed of strains originating from China. A single-member cluster 30 was isolated in Canada. In addition, some clusters displayed a mixed origin, with strains originating from multiple countries.

The MR2-16/1 strain has been assigned to cluster 27 and demonstrates the closest relationship to the clinical strain Ytc isolated from Turkey in 2022. Strains from

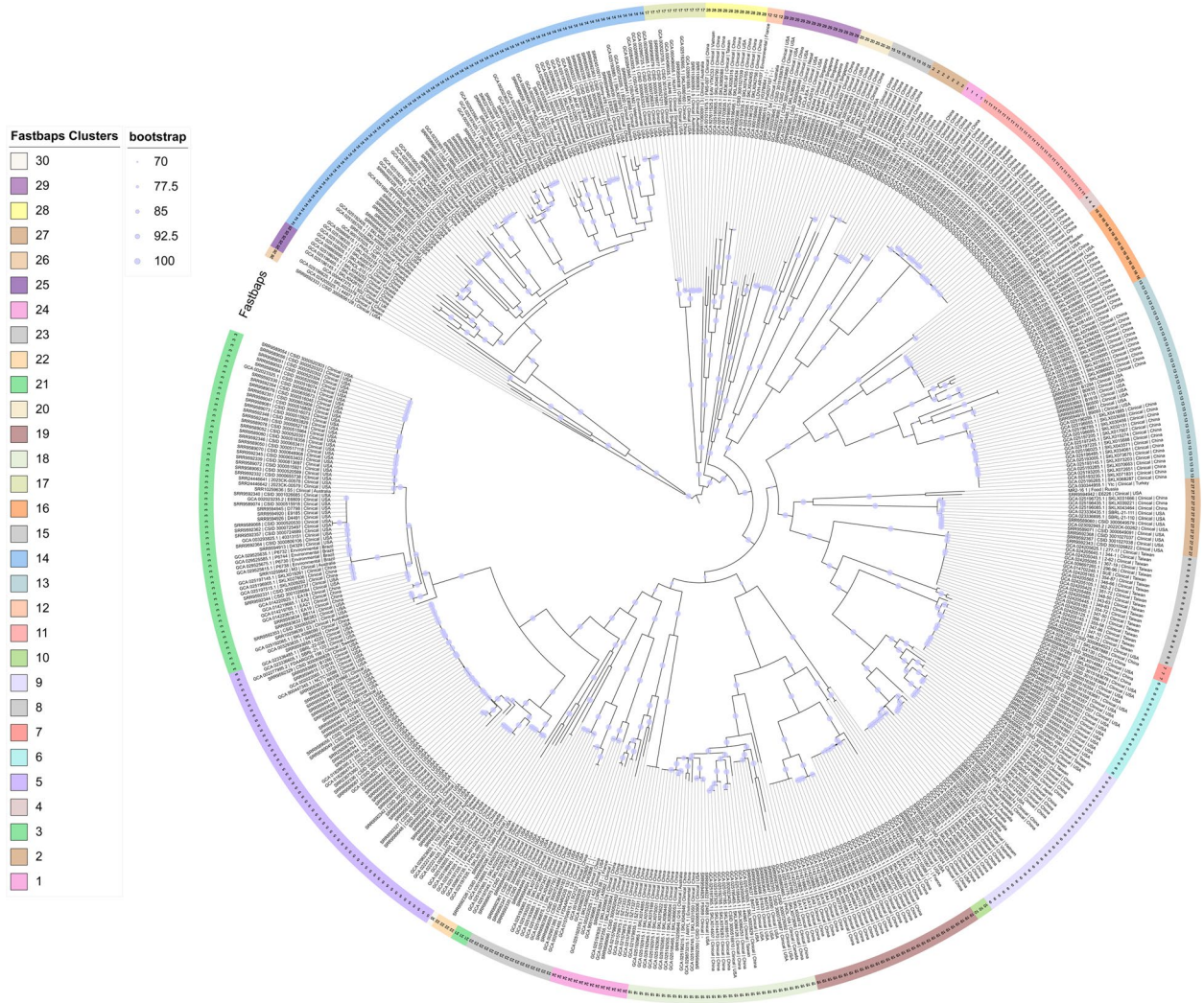

Fig. 2 Core genes SNP-based phylogenetic reconstruction of 540 *E. anophelis* strains. Clusters are represented by color ring

clinical cases have been identified within cluster 22 that originate from the United States, China, and Turkey.

Pan-genome analysis

The 540 genomes displayed a mean of 3,722 CDS (range: 3,462-4,328). Panaroo predicted a total of 15,822 gene clusters across the 540 *E. anophelis* genomes (Fig. 3). The total pan-genome was categorized into 4 groups with Panaroo: the core genome (present in 99-100% of strains), comprising 2,922 clusters; the soft-core (95–99%), encompassing 129 clusters; the shell (15–95%), totaling 1,167 clusters; and the cloud (0–15%), comprising a substantial 11,604 clusters. Only 18.46% of clusters were shared by ≥534 strains (99–100%), indicative of extensive genetic divergence.

Pan-genome categories histogram for 540 *E. anophelis* genomes, illustrating the number and proportion of core, soft-core, shell, and cloud genes. The percentage values were rounded to simplify presentation.

The identified core ($n=2,922$ genes), accessory ($n=10,291$), and unique genes ($n=2,609$) underwent functional annotation utilizing the COGclassifier tool (Table S2, Fig. 4). Of the 15,822 protein clusters, only 6,580 sequences (41.58%) were classified into functional categories based on COG: "Core" 2,161/2,922 (73.95%), "Accessory" 3,574/10,291 (34.73%), and "Unique" 845/2,609 (32.39%).

Within the spectrum of COG functional groups, "Function unknown (S)" and "General function prediction only (R)" constituted 236 (3.58%) and 537 (8.16%) annotated genes, respectively. To ensure a meaningful functional analysis, these groups were excluded from subsequent investigations. Among all analyzed proteins, the following categories were unoccupied (=0), as they are more

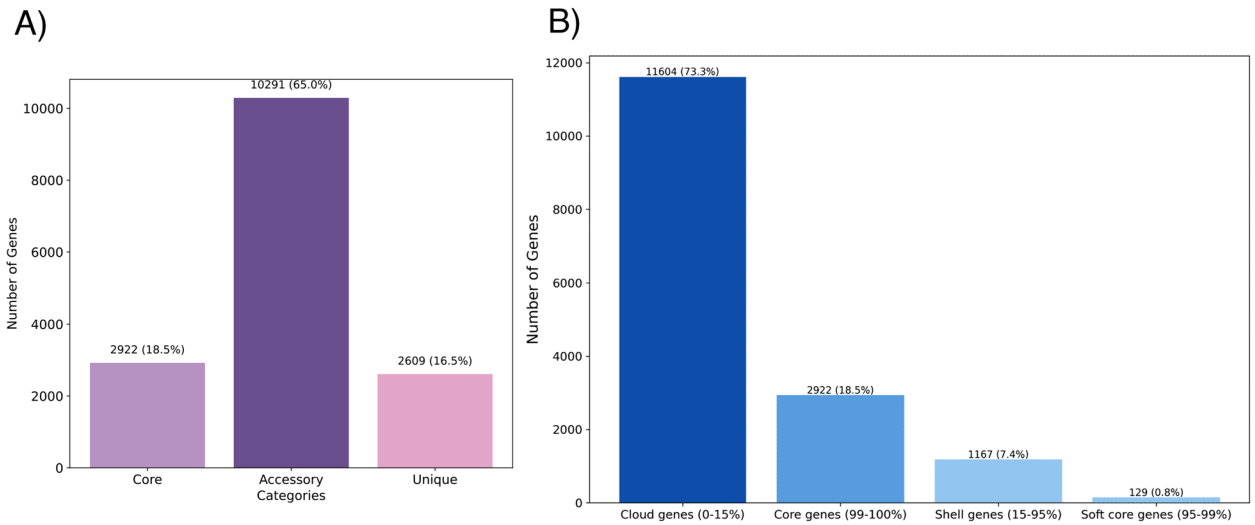

Fig. 3 Distribution of gene clusters among pan-genome categories for 540 *E. anophelis* genomes: **A** Core, accessory, and unique categories; **B** Core, soft core, shell, and cloud categories

characteristic of eukaryotes: "RNA processing and modification", "Chromatin structure and dynamics", "Nuclear structure", "Cytoskeleton", and "Extracellular structures". The major functional category was "Replication, Recombination, and Repair (L)" with 661/6,580 clusters (10.04%), while the smaller category was "Cell Motility (N)" with 19/6,580 clusters (0.28%) (Fig. 4, A).

Within the *E. anophelis* pan-genome, the core genome predominated in the following categories: J - Translation, ribosomal structure and biogenesis; E - Amino acid transport and metabolism; C - Energy production and conversion; F - Nucleotide transport and metabolism; G - Carbohydrate transport and metabolism; H - Coenzyme transport and metabolism (Fig. 4, B).

In contrast, accessory genes were found to be prevalent in the following: V - Defense mechanisms; L - Replication, recombination and repair; X - Mobilome: prophages, transposons; K – Transcription; T - Signal transduction mechanisms; M - Cell wall/membrane/envelope biogenesis; D - Cell cycle control, cell division, chromosome partitioning; P - Inorganic ion transport and metabolism; U - Intracellular trafficking, secretion, and vesicular transport (Fig. 4, B).

"Accessory" and "Unique" genes were highly represented in metabolic categories such as Q - Secondary metabolites biosynthesis, transport and catabolism – 45.2% and 8.2%, respectively; I - Lipid transport and metabolism – 35.9% and 14.3%; G - Carbohydrate transport and metabolism – 36.7% and 8.9%; H - Coenzyme transport and metabolism – 37.7% and 7.8%; C - Energy production and conversion – 31.9% and 12.4%. Such dominance of "Accessory" as well as the high prevalence of "Unique" genes in some categories indicates extensive genetic diversity.

Despite the fact that some categories of COGs are traditionally considered to be conservative (core-dominant), including L – Replication, Recombination, and Repair; K – Transcription, and D – Cell Cycle Control, Cell Division, Chromosome Partitioning, and a few others, in our case, the latter were significantly enriched with "accessory" or "unique" clusters. However, after manually reviewing the annotated sequences, we observed that the majority of proteins belong to mobile genetic elements, such as integrases, resolvases, regulators, prophage proteins, etc.

Post-processing of the pan-genome analysis results, as presented in Fig. 5 and Table S3, revealed a significant p-value (2.90e-48) for the "core" term, indicating a strong positive association between core branch lengths and gene exchange (gain/loss) events. In contrast, the p-value for the "tip" term was not significant (0.799) (Fig. 5A). The cumulative pan-genome plot also demonstrates an increase in the number of cumulative gene gains and losses corresponding with the increase in branch length from the root to the leaves (Fig. 5B). This pattern supports the existence of a temporal component in the gene gain and loss process, suggesting that *E. anophelis* has an open pan-genome. Although the pan-genome accumulation curve may be subject to error and might not fully represent the pan-genome, it has been included as a traditional representation (Fig. 5C). Additionally, we calculated the Heaps' law γ parameter for our dataset, a classic metric for assessing pan-genome "openness." The Heaps' γ parameter

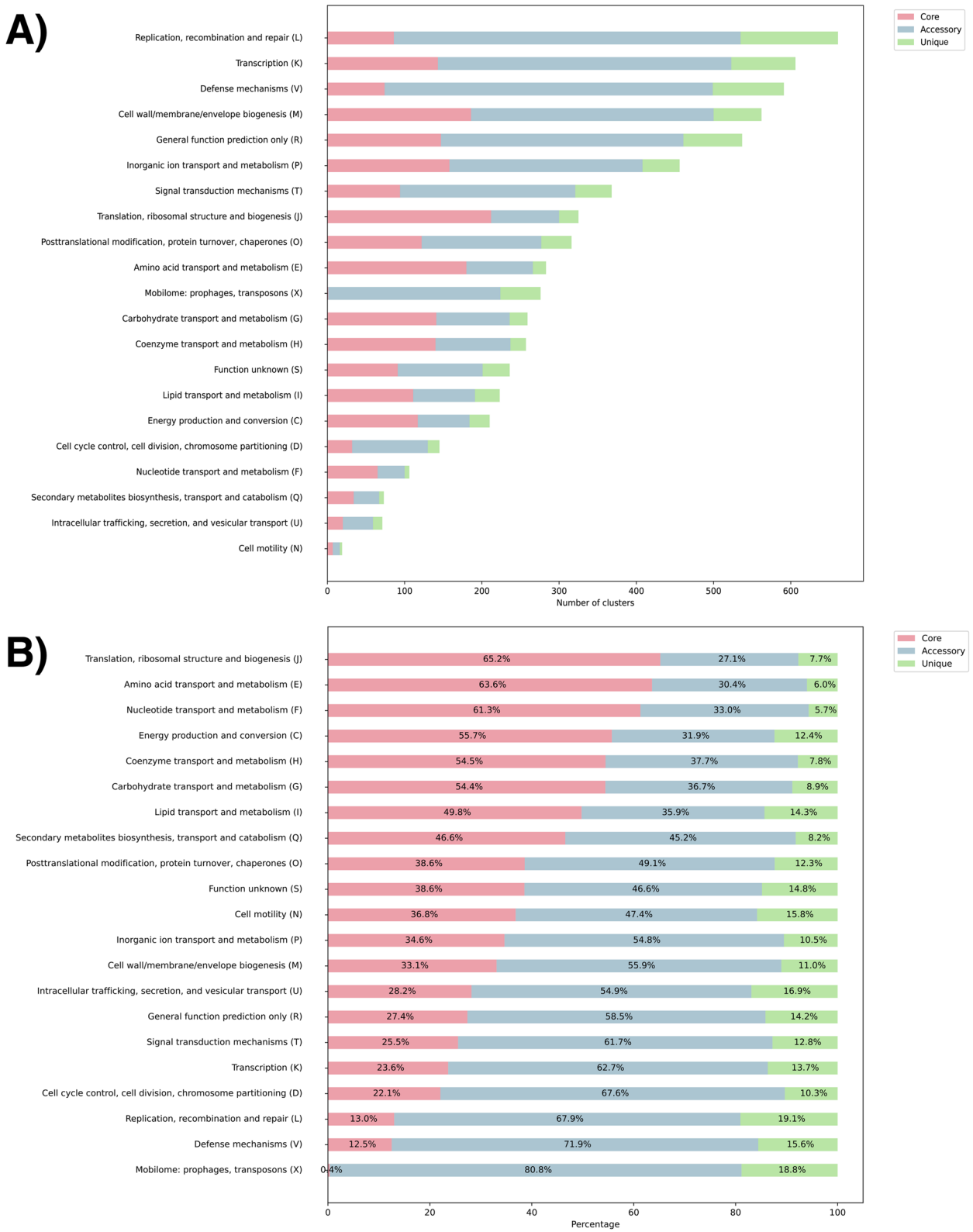

Fig. 4 Cluster of Orthologous Genes (COG) functional categories distribution among predicted protein clusters in 540 *E. anophelis* genomes: **A** Stacked bar chart displaying the protein clusters count in each pan-genome part; **B** Stacked bar chart illustrating the percentage of each pan-genome part within the respective COG category. The total gene cluster count sorts the A chart; the B chart is sorted by "Core"

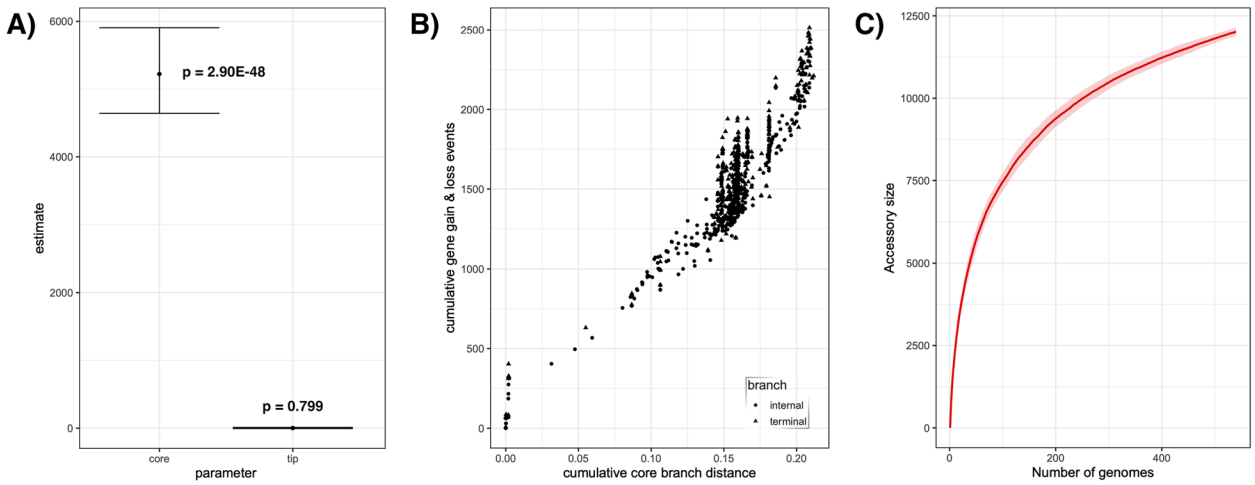

Fig. 5 Panstripe-generated plots: **A** Estimation plot for the "core" and "tip" parameters; **B** Plot showing the cumulative number of gene gain/loss events versus cumulative branch length; **C** Pan-genome accumulation curve

was greater than zero ($\gamma = 0.2164$), further indicating an open pan-genome. In summary, our results, based on both classical methods and statistical analysis, reveal that *E. anophelis* possesses an open pan-genome with a diverse array of accessory genes.

Virulence factors of *E. anophelis*

A protein homology search using local Blastp in 540 strains revealed the presence of 5 to 13 putative virulence factor (VF) protein homologs across various classes, categorized according to VFDB classification (Table S4). These categories included stress survival, immune modulation, adherence, nutritional/metabolic factors, exotoxins, and type IV secretion system components (TIVSS).

The putative virulence factors exhibited varying frequencies of occurrence among the studied population. Only 5 out of 16 detected genes were found to be present in the vast majority of the population (> 99.5%): *htpB* (Hsp60 - heat shock protein, $n = 538$), *tufA* (elongation factor Tu, $n = 539$), *katA* (catalase, $n = 540$), *wbtL* (glucose-1-phosphate thymidyltransferase, $n = 540$), and *cbu*_0270 (IVB SSIV effector, $n = 540$). The isocitrate lyase (*icl*) homolog was identified in 90% ($n = 486$) of the genomes. The capsular polysaccharide synthesis gene, *cap8A* homolog, and the GDP-mannose 4,6-dehydratase homolog *gmd* were present in more than 40% of the genomes ($n = 222$ and $n = 249$, respectively). Other factors, such as *msrA*/*BpilB* (trifunctional thioredoxin/methionine sulfoxide reductase A/B protein) and urease genes (*ureA*, *ureB*, and *ureG*), were detected in less than 3% of the genomes. The urease genes were specifically found in strains belonging to the monophyletic cluster 14 clade, as well as in all strains within cluster 15.

Antimicrobial resistance factors of *E. anophelis*

A comprehensive analysis identified 68 known resistance factors (RF), including their variants and homologs, which were categorized into two main functional groups: antibiotic inactivation enzymes and efflux pumps (Table S5). Each strain possessed a diverse range of RF, varying in number from 8 to 17.

Comparative genomic analysis of antimicrobial resistance genes (AMR) revealed that all genomes examined in this study contained the following six AMR genes (Table S5): *aadS* (6-aminoglycoside adenyltransferase AadS), *catB* (chloramphenicol type B O-acetyltransferase), *ranA*/*B* (ABC multidrug efflux transporter), bla_{CME} (extended-spectrum class A beta-lactamase of the CME family), and bla_{GOB} (subclass B3 metallo-beta-lactamase of the GOB family). The presence of these factors in all the studied genomes suggests that they are core-genome elements. Some genomes harbored two copies of bla_{GOB}, specifically SZ17329 (GCA_021579875.1) and SZ17325 (GCA_021579855.1) strains. Similarly, 17 strains exhibited 2 variants of chloramphenicol acetyltransferase homologs (Cat-B/Cat-B3).

We observe 14 AMR factors in 56 genomes with a high level of sequence similarity (greater than 90%), which are believed to have been spread among members of the *Bacteroidota* phylum. To our knowledge, these antimicrobial resistance determinants are not specific to the *Elizabethkingia* genus (Table S5).

Additionally, it is believed that the resistance of *E. anophelis* to fluoroquinolones is mediated by mutations in GyrA (S83I and S83R) and GyrB (D431N or D431H) [43]. These mutations were absent in our strain, therefore the intermediate susceptibility level to

ciprofloxacin might be explained by other mechanisms, such as efflux [44, 45].

Mobilome analysis and investigation of MGEs distribution across population

The large-scale screening of MGEs against the mobileOG-db Beatrix 1.6 v1 database identified 12,531 different MGEs in 540 *E. anophelis* genomes (Fig. 6). Among the five groups of MGEs suggested by mobileOG-DB, the most prevalent groups were "Phage" and "Transfer", with 7,135 (56.9%) and 3,838 (30.6%) ORFs, respectively. The "Phage" category, as the name implies, represents a group of proteins that are associated with the life cycle of a prophages. The "Transfer" category includes proteins that contribute to the inter-organism transfer of MGE, such as type IV secretion system machinery. The "Replication/Recombination/Repair" category accounted for 586 (4.67%) ORFs, comprising replication initiation proteins and regulators, as well as several repair systems and homologous recombination protein systems. "Integration/Excision" was represented by 585 (4.66%) ORFs, which includes proteins mediating MGE recombination such as transposases and tyrosine recombinases. The "Stability/Transfer/Defense" category was represented by 387 (3.08%) ORFs, which contains elements related to restriction-modification systems, CRISPR, and anti-CRISPR proteins as well as components of several anti-phage defense systems.

A total of 371 putative ICEs were identified among 540 *E. anophelis* genomes, with 54 found in complete genomes and 317 in draft assemblies. The number of ICEs per genome ranged from 1 to 3 (Tables S4 and S5). Based on the presence of integrase genes, the identified ICEs were classified as mobilizable (integrase-positive) or non-mobilizable (integrase-negative). Notably, 32 out of 54 ICEs (59.3%) in complete genomes and 104 out of 317 (33.1%) in draft genomes were categorized as mobilizable, suggesting a potential for HGT. 84 ICEs (22.6%) were found to harbor antimicrobial resistance (AMR) genes. Among complete genomes, 10 ICEs contained AMR genes (8 of which were mobilizable), while 74 were identified in draft assemblies (55 of which were mobilizable). The number of AMR genes per ICE ranged from 1 to 8 (Table S6, S7, and Figure S1).

Analysis of Genomic Islands (GIs) revealed a widespread presence of these elements in *E. anophelis* genomes. A total of 105 GIs were identified across 540 genomes (60 in complete and 45 in draft assemblies), with the number of GIs per genome ranging from 1 to 6 (Tables S4 and S5). 79 GIs (75.2%) harbored integrase genes, suggesting their potential for horizontal gene transfer. Furthermore, 46 GIs contained antimicrobial resistance (AMR) genes, with 22 of these (47.8%) also possessing integrase genes.

The analysis of AMR-positive GIs and ICEs distribution within the population revealed that most of these elements were associated with specific monophyletic lineages of clinical origin (mostly nosocomial infection cases) (Table S6, S7, and Figure S1). Strains harboring these elements were isolated from a wide range of countries, including China, the USA, Taiwan, Singapore, Vietnam, Nepal, Australia, Russia, and Sweden. In some unrelated lineages unusual for *E. anophelis* AMR genes with 90–100% of identity were observed: bla_{TEM}, bla_{OXA}, *lnu*(*H*), aminoglycoside modifying enzymes, *erm*(F), etc. Notably, some phylogenetic clusters of clinical origin lacked AMR-positive GIs and ICEs.

Using the Plasmer tool, we identified that 43 strains (7.96%) harbored plasmid sequences in total number of 72 (Table S8). Number of detected plasmid sequences ranged 1–6 per genome, and the length of fragments ranged 1025-58,106 bp. Notably, the length of plasmid sequences with AMR genes ranged 1025–1658 bp. Taxonomic classification of 72 identified plasmids revealed that 42 of them belongs to *Bacteroidota* phylum members (*Bacteroides* spp., *Empedobacter* spp., *Riemerella anatipestifer*, *E. anophelis*, *M. odoratimimus*). However, we found that 4 plasmid sequences belonged to *Pseudomonadales* order: 3 – *Acinetobacter baumannii*, and 1 – *Pseudomonas* spp. These plasmids were associated with clinical strain EM361-97 isolated in Taiwan in 2010. Notably, AMR factors were predicted only in plasmids originated from *Bacteroides* spp., *Empedobacter* spp., *R. anatipestifer*, and *M. odoratimimus*, but not from *E. anophelis*. (*aadS*, bla_{TEM}, *mef(C)*, *mph(G)*, *tet(X2)*, *tet(36)*, and *sul2*). Additionally, 4 plasmids had general "Bacteria (taxid 2)" origin, and 23 plasmids had unclassified origin. Analysis of plasmid distribution among the population revealed that plasmids carrying AMR genes exhibit an Asian origin (China and Taiwan) (Figure S1). These plasmids exhibit a clustering pattern, with three monophyletic groups consisting of clinical strains from China, as well as single strains from both China and Taiwan, forming outgroups with respect to the closest monophyletic groups. Plasmids without AMR were mostly dispersed across *E. anophelis* population with rare formation of 3 member clusters. Notably, 6 plasmids were harbored by 2 strains with environmental origin: 3 by CSID_3000656206 strain isolated from frog, 2 by DSM-23,781 isolated from mosquito (according with DSM database), and 1 by FP12_13 strain isolated from protist. It was also observed that some strains carried both plasmids GIs or ICEs simultaneously.

Among the 540 *E. anophelis* genomes analyzed, 451 prophage regions were identified in 248 genomes,

accounting for 37.57% of the total (Table S9). The number of prophages per genome varied from 1 to 7. Among all detected prophages only 9 exhibited intact structures. The genome length ranged from 3,749 bp to 62,844 bp, and the GC percentage (GC%) among all detected prophage genomes spanned from 30.53 to 42.2%. AMR genes search among prophage genomes revealed 3 prophages with AMR carriage in the following strains: 2002C02-176 — intact (Clinical, Taiwan), 2008N07-201 — defective (Clinical, Taiwan), F3201 — defective (Clinical, Taiwan). Notably, prophages in the first two strains were found inserted in ICE sequences with tandem ICE-phage formation (Figure S2, S3). Most AMR genes in these tandems were embedded in the prophage genomes (Table S9).

Large-scale analysis of insertion sequences (IS elements) across 540 *E. anophelis* genomes identified 4,491 elements belonging to 16 known families.

Notably, 15 "new" IS elements were discovered (Table S10). IS3 and IS21 were the most abundant families, with 1,267 and 874 elements respectively, followed by IS200/IS605 ($n=479$), IS110 ($n=378$), and IS481 ($n=347$). Number of IS per genome ranged 1–61. The distribution of these elements across genomes is visualized in a boxplot (Figure S4). The counts of IS elements exhibit a right-skewed pattern: most samples have low counts (median = 7.0), while 30 genomes show disproportionately high numbers of IS elements (outliers with ≥ 18). The latter had mixed geographical origin and were represented by 30 clinical, 1 veterinary, and 1 environmental strains. Notably, two monophyletic clinical strains from Taiwan, 2002C02-176 and 2008N07-201, exhibited exceptionally high numbers of IS elements (59 and 61).

Discussion

E. anophelis has recently emerged as a human pathogen, posing a serious threat to global human health due to its severe infection processes and intrinsic antimicrobial resistance [46]. In the present study, we conducted a large-scale genome analysis among 540 genomes of *E. anophelis*. We employed pan-genome analysis, phylogenetic analysis, virulome, and resistome analysis, as well as mobilome analysis.

Our MR2-16/1 strain, derived from raw bovine milk, is the second instance of *E. anophelis* isolation from this source [47]. MR2-16/1 exhibited resistance to commonly available antibiotics, with detected AMR factors aligning with the observed resistance pattern, except for tetracycline and colistin. In this instance, we are unable to explain this phenotypic resistance profile. *E. anophelis* seems to possess unknown resistance factors or mechanisms, which could include specific proteins (enzymes, etc.), individual genomic characteristics in the form of mutations in promoters or in specific target genes, or other complex mechanisms. To further contextualize our strain and investigate *E. anophelis* on a population level, we conducted a comprehensive genomic analysis using nearly all available *E. anophelis* data in the NCBI.

Our pan-genomic analysis revealed extensive genomic heterogeneity within *E. anophelis*. The core genome comprised 18.46%, while accessory and unique gene clusters constituted a substantial 81.54%, showing a relatively large fraction among a wide variety of COG categories, including relatively conserved groups like those related to central metabolism (e.g., amino acid and carbohydrate metabolism). It indicates a high level of adaptability. These observations is also confirmed by post-processing pan-genome analysis, where significant association was identified between core branch lengths and gene gain/loss events. Thus, *E. anophelis* has an open pan-genome with a diverse array of accessory genes.

Previous pan-genomic studies of *E. anophelis* reported findings with slightly variable number of core-genome clusters compared to our data [15, 48, 49]. Differences in core genome size could be attributed to various factors, including the approach, sampling strategy, and the tools or parameters used. We suggest that the choice of tools and parameters may be one of the key determinants. Nonetheless, the mentioned studies also emphasized the extensive genetic diversity within the *E. anophelis* population. Although we are aware of the limitations of our pan-genome approach: the 93.7% (506/540) of genomes used were sequenced from isolates of clinical origin. This is related to the fact that *E. anophelis* has been mostly studied as an infectious agent with a multidrug-resistant (MDR) profile. Additionally, 80% of genomes used were sequenced from strains isolated from both China ($n=223$) and USA ($n=209$). All this limits our understanding and interpretation of the results related to the biology of this microorganism. Therefore, our results are rather applicable to strains of clinical origin. To further our understanding of *E. anophelis*, its ecology must be investigated in depth. It is important to study strains that have been isolated from different sources and compare them to clinical ones.

Phylogenetic analysis based on core-genome SNP data showed a complex structure among 540 *E. anophelis* strains [50]. We observed 30 phylogenetic clusters among the 540 strains analyzed. While most clusters displayed a mix of sources and geographic origins, some clusters were remarkably homogeneous, consisting entirely of strains with a shared origin. This complex phylogenetic structure suggests high diversity among strains, consistent with our pan-genomic findings, and highlights a lack of strict correlation between isolation source and phylogeny for most lineages. However, in some instances, a

putative correlation was observed. Initially, *E. anophelis* was isolated from the midgut of the *Anopheles gambiae* mosquito in 2011 ($R26^T$ strain). Interestingly, some strains originating from *A. gambiae* and *A. sinensis* ($R26^T$, Ag1, AR6-8, MSU001, AR4-6) formed a distinct monophyletic group within cluster 16. This suggests that this lineage may possess unique features that contribute to its survival in mosquito hosts. Similar phylogenetic observations were reported by Breurec S. in 2016 [48]. However, the Ngousso strain, isolated from *A. coluzzii*, clustered with clinical strains within cluster 14. Most studies focus on clinical strains and cases of infection, and more data is needed to elucidate the genomic features and biodiversity of *E. anophelis* in various environments. Such research could assist in understanding the basis of pathogenicity, as well as other complex modes of interaction with different hosts.

Pathogenesis of *E. anophelis* infections remains largely unknown, especially certain virulence factors of this microbe (VF). Using protein homology search, we identified putative VFs present in almost all strains. These factors included: *katA* (catalase), *wbtL* (glucose-1-phosphate thymidylyltransferase), *htpB* (heat shock protein), *tufA* (elongation factor Tu), and CDU_0270 (trans-2-enoyl-CoA reductase). Interestingly, these virulence factors are typically associated with functions that benefit intracellular pathogens in their original hosts. For instance, *tufA* and *htpB* are linked to adhesion and invasion of non-phagocytic cells, while *katA* and *wbtL* are associated with persistence within monocytes and macrophages as well as for their activation inhibition [51–54]. We hypothesized that in *E. anophelis*, homologs of these factors could potentially contribute to a facultatively intracellular lifestyle. Such a lifestyle is typical for pathogens like *Legionella pneumophila* and *Listeria monocytogenes* [55–57]. These bacteria can persist in the environment but can also establish themselves within bacteriovorus protists such as amoebae [58, 59]. Notably, the NCBI SRA database contains genomic libraries of *E. anophelis* isolated from protists (SRR22102837 and SRR22102838). Moreover, the ability of *E. anophelis* to inhibit macrophage function, as demonstrated by Mayura et al. [60], aligns with strategies employed by facultative intracellular pathogens [57, 61]. Additionally, *E. meningosepticum* (formerly *Chryseobacterium meningosepticum*) was observed to invade murine respiratory tract epithelial cells, potentially through macropinocytosis accompanied by the formation of membrane ruffles [62]. Nonetheless, our hypothesis requires an experimental basis to be checked. Future investigations are needed to elucidate the mechanisms employed by *E. anophelis* to interact with host cells and establish a potential intracellular niche.

We identified 68 antimicrobial resistance factors (RF) including their variants and homologs in 540 *E. anophelis* strains. Among them 6 core-genome RFs were identified: *aadS* (6-aminoglycoside adenyltransferase AadS), *catB* (chloramphenicol type B O-acetyltransferase), *ranA*/*B* (ABC multidrug efflux transporter), bla_{CME} (extended-spectrum class A beta-lactamase of the CME family), and bla_{GOB} (subclass B3 metallo-beta-lactamase of the GOB family). Notably, some strains harbored two copies of the bla_{GOB} gene. This duplication could potentially confer a significant advantage during antimicrobial therapy [63]. Other RF homologs was identified, however, it is important to note that our "homology search" method has certain limitations. The relaxed search parameters we used enabled us to identify a large number of putative homologs, but the precise function these genes remains unknown. Therefore, interpretations of the results should be made with caution. Nevertheless, the homologous sequences identified in this study may be used in experimental investigations to verify their role as RFs.

Different strains exhibited varied RF profiles. In particular, RF profiles were found to be associated with different phylogenetic lineages. Moreover some RFs had a 90–100% of identity with well-known factors. Observed association and atypical RF genes suggests a possible link between genetic relatedness and the acquisition of MGEs with resistance determinants. To test this hypothesis we performed mobilome investigation and checked whether observed MGEs associated with AMR and phylogeny.

Our analysis revealed that the mobilome of *E. anophelis* is primarily composed of genetic islands (GIs), including integrative and conjugative elements (ICEs), prophage genomes, and relatively small number of plasmids. Integrons, commonly found MGEs in other bacteria, were not detected in this study. This suggests a unique mechanism for horizontal gene transfer in this pathogen. We found that 22.6% of ICEs and 43.8% of GIs carried various RFs. The majority of AMR-positive GIs and ICEs were associated with certain phylogenetic lineages of nosocomial origin. The presence of exclusive AMR genes in these lineages — unusual for *E. anophelis* — such as bla_{TEM}, bla_{OXA}, aminoglycoside-modifying enzymes, etc., as well as certain biocide resistance genes (e.g., *qacH*, *arsB*), suggests possible HGT events between *E. anophelis* and resident nosocomial microflora. Moreover, strains harboring these elements were isolated from a wide range of countries and formed unrelated clusters, suggesting the ubiquity of such events. GIs and ICEs appear to be the natural MGEs of *E. anophelis* and, under antibiotic and biocide selective pressure, can rapidly acquire, exchange, and disseminate AMR factors, thereby contributing to the adaptability of the host. We also observed instances of tandem formations between prophages and ICEs, which

aligns with previous reports in *E. anophelis* strains from Taiwan [64]. These findings also indicate the rapid evolution of the ICEs under selective pressure. Taken together this observations highlights a link between MGEs and the dissemination of AMR genes in specific lineages of *E. anophelis*.

We also identified 72 plasmid sequences among all strains. Plasmids are well known as one of the most common HGT vectors in prokaryotes [65]. However, they appear to have a more restricted presence in *E. anophelis* compared to ICEs, with only 7.96% of strains being plasmid-positive. Plasmids were found predominantly in clinical isolates, but also in 3 environmental isolates. This suggests that plasmids may play certain roles in environmental isolates as well. However, due to the biased dataset (with the majority of strains of clinical origin), we are unable to accurately estimate the true distribution of these MGEs in the "natural" *E. anophelis* population.

The difference in the clustering patterns of plasmids with and without AMR genes suggests that AMR-positive plasmids have a more localized distribution and are more likely to be selected in a clinical environment due to antibiotic pressure. Additionally, we identified one strain with plasmid sequences originating from members of the *Pseudomonadales* order, specifically *Acinetobacter baumannii* and *Pseudomonas* spp. These latter species are well known as some of the most frequently observed ESCAPE group members in hospital environments [66]. These findings indicate a potential HGT event between *E. anophelis* and a member of the nosocomial microflora. The continued circulation of *E. anophelis* in the hospital environment may lead to increased HGT with established residents, the emergence of new plasmids and other MGEs, and consequently, the intensification of reciprocal AMR gene exchange and spread. The observed abundance of IS elements in *E. anophelis* may also contribute to host adaptation through mutation or gene regulation [67]. Through the formation of composite transposons IS also could promote the diversification and evolution of other MGEs (GIs, ICEs, prophages, plasmids, etc.) [68].

All this suggests that *E. anophelis* has a complex mobilome, with GIs and ICEs as the main players. The presence of antimicrobial resistance genes in certain lineages suggests an evolutionary trend for this pathogen and its adaptation, particularly in nosocomial settings. The presence of lineage-specific resistance factors may present challenges for future treatment approaches, particularly if specific lineages become more prevalent through selection processes. The widespread use of antibiotics in healthcare environments creates a strong selective pressure that may favor the emergence and spread of resistance genes, especially those located on mobile genetic elements (MGEs) that can be easily transmitted between bacteria. This highlights the potential challenges presented by lineage-specific resistance factors in future treatment strategies. It is especially true if certain lineages become increasingly prevalent during antibiotic selection in healthcare settings, as demonstrated by the case of Changhua Christian Hospital in Taiwan [64].

Taken together our large-scale genomic analysis of *E. anophelis* sheds light on this emerging pathogen's diversity and potential threats. We found extensive genetic heterogeneity in the open pan-genome of this bacterium. Phylogenetic analysis revealed a complex population structure, with limited source-lineage correlation. Observed putative virulence factors suggest a high variability of pathogenesis mode with common factors conferring possible inclination towards a facultatively intracellular lifestyle. We also identified a high diversity of antimicrobial resistance factors, often associated with MGEs within specific lineages. Notably, the *E. anophelis* mobilome is dynamic, primarily consisting of GIs and ICEs. However prophage elements, plasmids, and high diversity of IS elements were also observed. Our findings suggest that *E. anophelis* is a highly adaptable pathogen, possessing a wide range of intrinsic AMR genes and virulence factors as well as genes that contribute to its fitness. This adaptability is further enhanced by its ability to acquire genetic material via horizontal gene transfer through mobile genetic elements like integrative and conjugative elements (ICEs). The potential for the rapid evolution of this emerging pathogen poses a significant challenge.

Materials and methods

Bacterial isolation and microbiological characteristics

E. anophelis strain MR2-16/1 was isolated from an unpasteurized bovine milk sample collected during a study investigating the prevalence of multidrug-resistant bacteria in raw bovine milk. The milk sample originated from a retail seller located in the Nizhny Novgorod district, European Russia, collected between June and August 2022.

To concentrate bacterial mass, 50 ml of the milk sample was centrifuged at 8000 rpm for 15 min. The sediment was then diluted in 2.5 ml of sterile phosphate-buffered saline (PBS) solution (pH 7.2; Himedia, Mumbai, India) and mixed with 2.5 ml of sterile tryptic soy broth (TSB) supplemented with antibiotics. The antibiotics were used in the following final concentrations: 50 μg/ml ampicillin and 25 μg/ml gentamicin. To inhibit fungal growth, 50 μg/ml fluconazole was added to the medium. The resulting broth was incubated at 37 °C for 48 h.

A loopful of the broth was plated onto tryptic soy agar (TSA) medium (Himedia, Mumbai, India) supplemented with the same concentration of antibiotics listed above.

The plates were incubated at 37 °C for 18–24 h. Visible grown colonies were further checked for purity using Gram staining and light microscopy. The OXItest (Erba Lachema, Brno, Czechia) was utilized to detect oxidase activity, and the 3% hydrogen peroxide test was conducted to identify catalase activity. The stock culture was stored at -80 °C in tryptic soy broth (TSB) (Himedia, Mumbai, India) supplemented with 15% glycerol before use.

Motility testing using Motility test agar (HiMedia, India) was performed on *E. anophelis* MR2-16/1. After incubation at 37 °C for 24 h, motility was assessed based on growth patterns compared to an uninoculated control.

Egg yolk agar plates were used to assess the enzyme activity of *E. anophelis* MR2-16/1: lecithinase (precipitate), lipase (iridescence), and protease (clearing zone). Incubation at 37 °C was 48–72 h for lecithinase and protease, and up to 7 days for lipase. *Staphylococcus aureus* 209P and *Escherichia coli* XL1-Blue served as positive and negative controls, respectively [69].

The gelatinase activity of *E. anophelis* MR2-16/1 was determined using the Micro-Gelatinase-NICF kit (RCP, Russia) following the manufacturer's protocol. After 48 h incubation at 37 °C, tubes were chilled at 4 °C for 2 h. A clear liquid fraction after chilling indicated a positive test. A non-inoculated control tube served as the negative control.

The biochemical profile of the *E. anophelis* MR2-16/1 strain was characterized using the ENTEROtest 24 N (Lachema, Czech) biochemical test system adhering to the manufacturer's instructions. The kit encompassed 24 biochemical assays.

Hemolytic activity of *E. anophelis* MR2-16/1 was determined on sheep and human blood agar plates (HiMedia, India) after incubation at 37 °C for 24 and 48 h. Alpha-hemolysis (green zone), beta-hemolysis (clear lysis), and gamma-hemolysis (no change) were differentiated based on lysis patterns [70].

Adherence to human erythrocytes was assessed following the protocol by Lenchenko et al. [71]. Briefly, washed human O-type blood cells were mixed with a bacterial suspension (2.0 McFarland) at a 1:1 ratio and incubated at 37 °C, 200 rpm for 24 h. After washing, smears were stained and examined microscopically to calculate the average adhesion index (AAI), adhesion coefficient (AC), and microorganism adhesion index (AI).

E. anophelis MR2-16/1 capsule production was assessed using the Capsule Stain-Kit (HiMedia, India) following the manufacturer's protocol. An overnight culture grown on Nutrient agar at 37 °C was used. *Klebsiella pneumoniae* ML-9 served as the positive control.

Biofilm formation was assessed using a microtiter plate assay [72]. Briefly, *E. anophelis* MR2-16/1 suspension (0.5 McFarland) in MHB-glucose was diluted 20-fold, and 200 µl aliquots were added to a 96-well plate. After incubation at 37 °C for 24 h, wells were washed, biofilms were fixed and stained with safranin, and absorbance was measured at 620 nm. Non-inoculated wells served as negative controls. Biofilm formation was indicated by OD values exceeding the negative control.

Isolate identification

Bacterial identification was performed using nearly full-length 16 S rRNA gene sequencing. DNA was extracted from a single colony grown on tryptic soy agar media through thermal lysis of the bacterial culture suspension (heated at 95 °C for 15 min). The universal primers 27 F (5'-AGAGTTTGATCMTGGCTCAG-3') and 1492R (5'-TACGGYTACCTTGTTACGACTT-3') were used for PCR amplification [73]. PCR products of approximately 1500 bp were visualized in a 1% agarose gel containing ethidium bromide. Amplicons were then extracted from the gel and purified using a GeneJET Gel Extraction Kit (Dia-m, Moscow, Russia). Sanger sequencing was performed at the GENOME Center for Collective Use (Moscow, Russia). Raw reads were manually checked and a consensus sequence was assembled from both reads using UGENE v39.0 software [74]. The BLAST search was conducted against the EzTaxon server database (Database ver. 2021.07.07) [75]. The recommended criteria for accurate species identification suggested in the "Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology" were employed: consensus length > 1300 nt; <0.5% ambiguity [76].

Antimicrobial susceptibility testing (AST)

Antimicrobial susceptibility testing was conducted using the standard microdilution method for determining minimum inhibitory concentrations (MICs). MIKROLATEST MIC (Erba Lachema) kits G-I and G-II were employed. All procedures followed the kit manufacturers' instructions precisely. Mueller Hinton Broth No.2 Control Cations (HiMedia, India) served as the suspension medium.

The G-I kit included the following antibiotics: ampicillin, ampicillin-sulbactam, cefazolin, cefuroxime, aztreonam, gentamicin, amikacin, colistin, trimethoprim-sulfamethoxazole, ciprofloxacin, chloramphenicol, and tetracycline.

The G-II kit included the following antibiotics: piperacillin, piperacillin-tazobactam, cefotaxime, ceftazidime, cefoperazone, cefoperazone-sulbactam, cefepime, meropenem, ertapenem, tigecycline, netilmicin, and tobramycin.

The interpretation of the results largely adhered to European Committee on Antimicrobial Susceptibility

Testing 2023 (EUCAST) 2023 v13.0 guidelines. Since specific breakpoints for MIC interpretation for *Elizabethkingia* spp. are not yet available, the PK/PD (non-species-related) breakpoints were employed, as the EUCAST recommended. The Clinical and Laboratory Standards Institute (CLSI) M100 2020 manual for *Enterobacteriaceae* was utilized for cefoperazone and tetracycline MIC interpretation. *E. coli* ATCC 25,922 served as the quality control strain.

Whole-genome sequencing, assembly, and annotation

Bacterial biomass was obtained from an overnight broth culture (TSB) for the isolation of genomic DNA. DNA was extracted using the QIAamp DNA Kit (Qiagen, Düsseldorf, Germany) following the manufacturer's instructions. DNA quality and concentration were assessed employing 1% agarose gel electrophoresis and the Qubit dsDNA BR Assay Kit (Fisher Scientific, Waltham, MA, USA) on the Qubit 3.0 fluorometer (Fisher Scientific), respectively. Sequencing was carried out by Geneanalytics LLC (Moscow, Russia) using an Illumina HiSeq 1500 machine (Illumina, San Diego, CA, USA) generating 150-bp paired-end reads.

The obtained read libraries were checked using FastQC v0.11.9 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Reads were trimmed and filtered using fastp v0.23.2 [77]. Processed forward and reverse libraries contained 14,286,960 reads in sum.

The Shovill v0.9.0 assembler (https://github.com/tseemann/shovill) was employed for genome assembly with a SPAdes option [78] and minimal contig length of 200 bases [79]. To assess the quality of the final assembly, QUAST v5.2.0 and CheckM2 v1.0.2 were used [80, 81]. The genome of the *Elizabethkingia anophelis* R26^T type strain served as the reference genome (GenBank accession: GCA_002023665.2). The assembly was annotated using Prokka v1.14.6 with default parameters (--kingdom Bacteria --genus *Elizabethkingia* --species *anophelis*) [82].

Public genomic data acquisition and processing

All publicly available genome assemblies of *E. anophelis* (accessed on August 10, 2023; $n=477$) were retrieved from the NCBI GenBank database using the "exclude atypical genomes" option. The collection of genome assemblies was further filtered, selecting the most complete version for each isolate from among any available assemblies.

Additionally, 203 non-duplicated publicly available libraries of paired-end reads generated using the Illumina platform were obtained from the Sequence Read Archive (SRA). These libraries underwent quality control (QC) using FastQC v0.11.9 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and MultiQC v1.17 [83] as well as were trimmed and filtered using fastp v0.23.2 [77]: reads with a Phred score < 28 were discarded. Assembly quality was assessed using QUAST v5.2.0 and CheckM2 v1.0.2.

Two paired-end libraries were excluded based on GC% analysis indicating library contamination (SRA acc: SRR11060090, SRR24509966). Libraries that passed QC were assembled using the Shovill v0.9.0 (https://github.com/tseemann/shovill) tool and the SPAdes assembler option, with a minimum contig length of 200 bases.

Both genome collection datasets (from GenBank and the manually assembled SRA library) were subjected to a quality check using the QUAST v5.2.0 and the CheckM2 v1.0.2 tools. Genomes that did not meet the following criteria were discarded: >250 contiguous sequences, an average sequence completeness of < 99%, a contamination level of > 2%, and an N50 length of < 50,000 bp [80, 81].

The filtered collection comprised 540 genome assemblies of *E. anophelis* (339 obtained from GenBank, 200 manually assembled from SRA, and MR2-16/1 genome assembly) (Table S11). All 540 genome assemblies were annotated using the Prokka v1.14.6 tool with the default parameters [82].

FastANI v1.3 was employed to verify the species affiliation of all resulting and annotated assemblies ($n=540$). The reference genome employed was the complete *E. anophelis* R26^T chromosome (GenBank accession: GCA_002023665.2). The ANI percentage ranged from 97.32 to 99.99%, indicating that the assembled genomes belonged to the *E. anophelis* species.

Pan-genome analysis, core genome-based phylogenomic analysis, and pan-genome postprocessing of *E. anophelis*

To evaluate the *E. anophelis* pan-genome, the Panaroo v1.5.0 tool was employed [84]. "Moderate" clean mode and a sequence identity threshold of 80% were used. Paralogous sequences were merged to minimize possible inflation in gene cluster number. Multifasta core genome alignment was generated using the MAFFT v7.490 tool [85].

As a default, Panaroo categorizes the pan-genome into the following groups: core genes (genes present in 99–100% of strains), soft-core genes (95–99%), shell genes (15–95%), and cloud genes (0–15%). For subsequent analysis, we utilized these groups and introduced an additional category for unique genes (genes exclusive to individual strains).

All extracted gene groups underwent functional annotation using COGclassifier v1.0.5 (https://github.com/moshi4/COGclassifier), employing default search options. The resulting COG function annotations for

each Panaroo genome fraction were partitioned into three distinct groups: the "Core" comprising genes present in 99–100% of strains, the "Accessory" encompassing merged soft-core, shell, and cloud genes excluding unique genes, and the "Unique" grouping genes not categorized in either of the preceding groups. Annotation data were visualized based on the Cluster of Orthologous Genes (COG) category count [86] using Python packages matplotlib [87], and pandas [88].

The maximum-likelihood phylogeny was reconstructed based on the core genome alignment filtered with Panaroo to investigate the phylogenetic relationships between the MR2-16/1 strain and other *E. anophelis* strains as well as investigate population structure. Initially, the filtered core genome alignment obtained from Panaroo was processed with the SNP-sites tool v2.5.1 [89] to concatenate single nucleotide polymorphisms (SNPs) using the only ATGC filter option. IQ-Tree v2.2.3 [90] was employed to reconstruct the core gene SNP phylogeny using a symmetric model with unequal rates but equal base frequencies (SYM) [91] evolution model and FreeRate ($R=6$) [92, 93] with 1000 ultrafast bootstrap iterations and the ascertainment bias correction option (ASC) [94]. The ModelFinder tool was used to select an appropriate evolution model [95], and the UFBoot2 tool was employed to calculate an ultrafast bootstrap approximation [96]. The fastbaps tool was used for phylogenetic clusterization with "baps" prior option [97]. It uses the Dirichlet process mixture model (DPM) for clustering multilocus genotype data. Maximum-likelihood (ML) tree generated with IQ-Tree was used as a tree primer for fastbaps with SNP-filtered alignment of the core genome. Thus, combinations of the ML and Bayesian approaches were used to infer the phylogeny of *E. anophelis*. We set 4 levels of fastbaps analysis to deeply resolve phylogenetic structure. Level 1 was further used to depict clusters. The resulting phylogenetic tree was midpoint rooted. iTOL v6.8 was utilized to visualize the SNP-based core genome phylogenetic tree with meta-information [98].

Pan-genome post processing was performed with Panstripe tool v0.3.0 [99]. Panstripe was used in conjunction with the IQ-Tree-generated midpoint rooted phylogeny and the presence/absence gene matrix. The lengths of each branch in the phylogenetic tree were compared to the numbers of gene gains and losses, using a Gaussian distribution model. Panstripe evaluates two main terms: "core" and "tip". The significant p-value ($p<0.001$) for the "core" term indicates whether the branch lengths in the phylogeny are associated with gene gain/ loss events. The significant p-value ($p<0.001$) for the "tip" term indicates associations with genes that are found at the tips of the phylogeny (it can be associated with annotation errors, sampling density, and highly mobile elements that are not observed in multiple genomes).

The genome and pan-genome size parameters can also be utilized to assess the "openness" or "closeness" of the pan-genome, as per Heaps' law [100]. By employing this law, it is possible to calculate the γ parameter. The magnitude of this parameter determines the "openness" ($\gamma>0$) or "closeness" ($\gamma<0$). The Heaps' law was used to calculate the γ parameter for the *E. anophelis* pan-genome, using 1,000 permutations (number of random genome orderings).

Analysis of *E. anophelis* virulence factors and antimicrobial resistance genes

Blastp search against VFDB setA was used to detect homologs of known virulence factors v2023.09.29 (Virulence Factor Database) [101, 102]. Fasta files containing translated CDSs (amino acid sequences) obtained from 540 genome assemblies were used as input. The following blastp search options were applied: E-value ≤ 1e-20, percentage of identity ≥ 60%, and the maximum number of target sequences = 1. In general, when selecting the values for these parameters, we followed the recommendations of Pearson W.R. for determining homology [103].

NCBI Antimicrobial Resistance Gene Finder (AMRFinderPlus) v3.12.8 with --plus option was used to detect known AMR factors as well as their homologs [104]. Additionally, biocide and heavy metal resistance factors were considered. Fasta files containing translated CDSs (amino acid sequences) obtained from 540 genome assemblies were used as input. The following options were utilized: identity percentage ≥ 55%, coverage ≥ 60%. Such relaxed parameters were chosen to detect putative homologs of known AMR factors because the majority of AMR factors in *E. anophelis* remain unknown.

Analysis of Mobile Genetic Elements (MGE)

For large-scale Mobile Genetic Elements (MGEs) screening, we employed the mobileOG-db Beatrix 1.6 v1 database, utilizing a bash script available on the GitHub repository [105]. The script is based on Diamond v2.1.8 [106] for a search against the mobileOG database. The specified options included a maximum E-value of 1e-20, a percent of identical matches ≥ 90%, and a percent of query coverage to the sample ≥ 90%. The generated CSV files were employed to assess the distribution of identified genetic elements across categories, including "phage," "transfer," "integration/excision," "replication/recombination/repair," and "stability/transfer/defense." Elements falling within these categories may play diverse roles in the life cycle of recognized MGEs, encompassing plasmids, phages, integrative elements, transposable

elements, and conjugative elements. For a more detailed analysis, the MGE search was performed using the special databases.

Integrative and conjugative elements (ICEs) were identified with the IslandCompare online tool using the IslandPath-incorporated methodology [107]. IslandCompare is designed to process collections of microbial genome sequences and present information about genomic islands (including ICEs). To minimize potential bias introduced by variations in genome assembly completeness, complete and draft genomes were analyzed separately for genomic island (GI) prediction. For draft genomes, contigs were assembled into a single, non-reference-guided consensus sequence to maximize the available sequence data for GI identification. Subsequently, Gbk files generated using Prokka software were analyzed to identify a preliminary set of potential GIs. The extracted potential GI sequences were structurally annotated using Prodigal v2.6.3 (https://github.com/hyattpd/Prodigal). The obtained CDSs were functionally annotated using the bakta_proteins script, which can be found in the Bakta repository [108]. The predicted protein sequences from each GI in TSV format, were manually inspected for the presence of key integrase and mobilization system components (ICE proteins) essential for GI mobility. These included integrase, relaxase, VirD4 coupling protein (T4CP), VirB4/TraG, and various transport proteins (TraA, B, C, D, etc.). ICEs lacking integrase genes were categorized as non-mobile or degenerate elements. Additionally, the CDSs were searched for antimicrobial resistance genes using AMRFinderPlus v3.12.8 with the following parameters: identity percentage of ≥60%, and coverage of ≥60% [104]. Notably, GIs lacked essential integrase and mobilization system components (ICE proteins) but with AMR genes presence were also included for further analysis.

Insertion sequence (IS elements) and related CDS were detected using ISEScan v1.7.2.3 [109]. Structurally annotated IS-related CDS were subjected to functional annotation using the bakta_proteins option with Bakta database v5.0 [108]. All CDS were searched against AMRFinderPlus v3.12.8 to detect known AMR factors as well as their homologs associated with IS (identity percentage ≥55%, coverage ≥60%) [104].

Plasmids identification was performed with Plasmer tool with default parameters [110]. Detected putative plasmids were structurally annotated with Prodigal v2.6.3. Translated ORFs were used for search with AMRFinderPlus v3.12.8 (identity percentage ≥55%, coverage ≥60%).

Prophage genomes and elements were detected using the PHASTER web tool [111]. Detected prophage genomes were classified following PHASTER classification: intact (score > 90), questionable (score 70–90), and incomplete (score < 70). All detected genomes were structurally annotated with Prodigal v2.6.3. The translated ORFs were employed for search with AMRFinderPlus v3.12.8 (identity percentage ≥55%, coverage ≥60%). Prophage visualization was performed using the Artemis Comparison Tool (ACT) [112].

Integron searches were performed via the IntegronFinder v2.0 tool with default search options [113, 114].

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-024-10921-y.

Supplementary Material 1.
Supplementary Material 2.
Supplementary Material 3.
Supplementary Material 4.
Supplementary Material 5.

Authors' contributions

Conceptualization: PA; Methodology: PA, PZ; Validation: PA, PZ, DK, AM, AT; Formal analysis: PA; Investigation: PA, PZ, DK, AM, AT; Resources: IY; Data Curation: PA; Writing - Original Draft: PA; Writing - Review & Editing: PA; Visualization: PA; Supervision: PA; Project administration: PA; Funding acquisition: IY.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Data availability

This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession JBCEWF000000000. The version described in this paper is version JBCEWF010000000.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 1 May 2024 Accepted: 18 October 2024
Published online: 29 October 2024

References

1. A brief guide to emerging infectious diseases. and zoonoses. https://iris.who.int/handle/10665/204722. Accessed 25 Apr 2024.
2. Morse SS. Factors in the emergence of infectious diseases. Emerg Infect Dis. 1995;1:7–15.
3. Vouga M, Greub G. Emerging bacterial pathogens: the past and beyond. Clin Microbiol Infect. 2016;22:12–21.
4. Nwannunu CE. Emerging Bacterial Infections. In: Tyring SK, Moore SA, Moore AY, Lupi O, (eds). Overcoming Antimicrobial Resistance of the Skin. Updates in Clinical Dermatology. Cham: Springer; 2021. https://doi.org/10.1007/978-3-030-68321-4_2.

5. Morens DM, Folkers GK, Fauci AS. The challenge of emerging and re-emerging infectious diseases. Nat 2004. 2004;430:6996.
6. Baker RE, Mahmud AS, Miller IF, Rajeev M, Rasambainarivo F, Rice BL, et al. Infectious disease in an era of global change. Nat Rev Microbiol 2021. 2021;20(4):193–205.
7. Hu S, Xu H, Meng X, Bai X, Xu J, Ji J, et al. Population genomics of emerging Elizabethkingia anophelis pathogens reveals potential outbreak and rapid global dissemination. Emerg Microbes Infect. 2022;11:2590–9.
8. Lau SKP, Chow WN, Foo CH, Curreem SOT, Lo GCS, Teng JLL, et al. Elizabethkingia anophelis bacteremia is associated with clinically significant infections and high mortality. Sci Rep. 2016;6:1–10.
9. Navon L, Clegg WJ, Morgan J, Austin C, McQuiston JR, Blaney DD, et al. Notes from the field: investigation of Elizabethkingia anophelis Cluster — Illinois, 2014–2016. MMWR Morb Mortal Wkly Rep. 2019;65:1380–1.
10. Kämpfer P, Matthews H, Glaeser SP, Martin K, Lodders N, Faye I. Elizabethkingia anophelis sp. nov., isolated from the midgut of the mosquito Anopheles gambiae. Int J Syst Evol Microbiol. 2011;61:2670–5.
11. Silva E, Matsena Zingoni B, Koekemoer Z, Dahan-Moss LL. YL. Microbiota identified from preserved Anopheles. Malar J. 2021;20:230.
12. Fofana A, Gendrin M, Romoli O, Yarbanga GAB, Ouédraogo GA, Yerbanga RS, et al. Analyzing gut microbiota composition in individual Anopheles mosquitoes after experimental treatment. iScience. 2021;24:103416.
13. Lee YL, Liu KM, Chang HL, Lin JS, Kung FY, Ho CM, et al. A dominant strain of Elizabethkingia anophelis emerged from a hospital water system to cause a three-year outbreak in a respiratory care center. J Hosp Infect. 2021;108:43–51.
14. Yung CF, Maiwald M, Loo LH, Soong HY, Tan CB, Lim PK, et al. Elizabethkingia anophelis and Association with tap water and handwashing, Singapore. Emerg Infect Dis. 2018;24:1730.
15. Perrin A, Larsonneur E, Nicholson AC, Edwards DJ, Gundlach KM, Whitney AM, et al. Evolutionary dynamics and genomic features of the Elizabethkingia anophelis 2015 to 2016 Wisconsin outbreak strain. Nat Commun. 2017;8:1.
16. Guerpillon B, Fangous MS, Le Breton E, Artus M, le Gall F, Khatchatourian I, et al. Elizabethkingia anophelis outbreak in France. Infect Dis Now. 2022;52:299–303.
17. Kyritsi MA, Mouchtouri VA, Pournaras S, Hadjichristodoulou C. First reported isolation of an emerging opportunistic pathogen (Elizabethkingia anophelis) from hospital water systems in Greece. J Water Health. 2018;16:164–70.
18. Brandsema BR, Fleurke GJ, Rosema S, Schins EM, Helfferich J, Bathoorn E. Neonatal Elizabethkingia anophelis meningitis originating from the water reservoir of an automated infant milk dispenser, the Netherlands, February 2024. Euro Surveill. 2024;29:14.
19. Lin JN, Lai CH, Yang CH, Huang YH. Elizabethkingia infections in humans: from genomics to clinics. Microorganisms. 2019;7:295.
20. Hem S, Jarocki VM, Baker DJ, Charles IG, Drigo B, Aucote S, et al. Genomic analysis of Elizabethkingia species from aquatic environments: evidence for potential clinical transmission. Curr Res Microb Sci. 2022;3:100083.
21. Frank T, Gody JC, Nguyen LBL, Berthet N, Le Fleche-Mateos A, Bata P, et al. First case of Elizabethkingia anophelis meningitis in the Central African Republic. Lancet. 2013;381:1876.
22. Teo J, Tan SYY, Tay M, Ding Y, Kjelleberg S, Givskov M, et al. First case of E anophelis outbreak in an intensive-care unit. Lancet. 2013;382:855–6.
23. Lau SKP, Wu AKL, Teng JLL, Tse H, Curreem SOT, Tsui SKW, et al. Evidence for Elizabethkingia Anophelis transmission from mother to infant, Hong Kong. Emerg Infect Dis. 2015;21:232–41.
24. Moore LSP, Owens DS, Jepson A, Turton JF, Ashworth S, Donaldson H, et al. Waterborne Elizabethkingia meningoseptica in adult critical care - 22, number 1—January 2016 - emerging infectious diseases journal - CDC. Emerg Infect Dis. 2016;22:9–17.
25. Bulagonda EP, Manivannan B, Mahalingam N, Lama M, Chanakya PP, Khamari B, et al. Comparative genomic analysis of a naturally competent Elizabethkingia anophelis isolated from an eye infection. Sci Rep. 2018;8: 8.
26. Nielsen HL, Tarpgaard IH, Fuglsang-Damgaard D, Thomsen PK, Brisse S, Dalager-Pedersen M. Rare Elizabethkingia anophelis meningitis case in a Danish male. JMM Case Rep. 2018;5:8.
27. Snesrud E, McGann P, Walsh E, Ong A, Maybank R, Kwak Y, et al. Clinical and genomic features of the first cases of Elizabethkingia anophelis infection in New York, including the first case in a healthy infant without previous nosocomial exposure. J Pediatr Infect Dis Soc. 2019;8:269–71.
28. Chaudhary S, Rijal A, Rajbhandari P, Acharya AB. The first reported case of Elizabethkingia anophelis from Nepal. Cureus. 2023. https://doi.org/10.7759/CUREUS.45346.
29. Ichiki K, Ooka T, Shinkawa T, Inoue S, Hayashida M, Nakamura D, et al. Genomic and phylogenetic characterization of Elizabethkingia anophelis strains: the first two cases of life-threatening infection in Japan. J Infect Chemother. 2023;29:376–83.
30. Sr. NS, Singh A, Gupta P, Agarwal A, Sr. DrNS, Singh A et al. Elizabethkingia anophelis infections: a case series from a tertiary care hospital in Uttar Pradesh. Cureus. 2022;14:e32057.
31. Commans F, Hayer J, Do BN, Tran TTT, Le TTH, Bui TT, et al. Whole-genome sequence and resistance determinants of four Elizabethkingia anophelis clinical isolates collected in Hanoi, Vietnam. Sci Rep. 2024;14:7241.
32. Bellais S, Poirel L, Naas T, Girlich D, Nordmann P. Genetic-biochemical analysis and distribution of the ambler class A β- lactamase CME-2, responsible for extended-spectrum cephalosporin resistance in Chryseobacterium (Flavobacterium) meningosepticum. Antimicrob Agents Chemother. 2000;44:1–9.
33. Yasmin M, Rojas LJ, Marshall SH, Hujer AM, Cmolik A, Marshall E, et al. Characterization of a novel pathogen in immunocompromised patients: Elizabethkingia anophelis—exploring the scope of resistance to contemporary antimicrobial agents and β-lactamase inhibitors. Open Forum Infect Dis. 2023;10:ofad014.
34. González LJ, Vila AJ. Carbapenem resistance in Elizabethkingia meningoseptica is mediated by metallo-β-lactamase BlaB. Antimicrob Agents Chemother. 2012;56:1686–92.
35. Wang M, Gao H, Lin N, Zhang Y, Huang N, Walker ED, et al. The antibiotic resistance and pathogenicity of a multidrug-resistant Elizabethkingia anophelis isolate. Microbiologyopen. 2019;8:11.
36. Johnson WL, Gupta SK, Maharjan S, Morgenstein RM, Nicholson AC, McQuiston JR, et al. A genetic locus in elizabethkingia anophelis associated with elevated vancomycin resistance and multiple antibiotic reduced susceptibility. Antibiotics. 2024;13:61.
37. Hu S, Lv Y, Xu H, Zheng B, Xiao Y. Biofilm formation and antibiotic sensitivity in Elizabethkingia anophelis. Front Cell Infect Microbiol. 2022;12:953780.
38. Flemming HC, Wingender J, Szewzyk U, Steinberg P, Rice SA, Kjelleberg S. Biofilms: an emergent form of bacterial life. Nat Rev Microbiol 2016. 2016;14:9.
39. Ciofu O, Moser C, Jensen PØ, Høiby N. Tolerance and resistance of microbial biofilms. Nat Rev Microbiol 2022. 2022;20:10.
40. Singh S, Sahu C, Singh Patel S, Ghoshal U. Clinical profile, susceptibility patterns, speciation and follow up of infections by Elizabethkingia species: study on a rare nosocomial pathogen from an intensive care unit of north India. New Microbes New Infect. 2020;38:100798.
41. Chan JC, Chong CY, Thoon KC, Tee NWS, Maiwald M, Lam JCM, et al. Invasive paediatric elizabethkingia meningoseptica infections are best treated with a combination of piperacillin/tazobactam and trimethoprim/sulfamethoxazole or fluoroquinolone. J Med Microbiol. 2019;68:1167–72.
42. Tan MC, Huang YC, Chen PJ, Huang WC, Hsu SY, Wang HY, et al. In vitro and in vivo evidence discourages routine testing and reporting of piperacillin/tazobactam susceptibility of Elizabethkingia species. J Antimicrob Chemother. 2024;79:200–2.
43. Lin IF, Lai CH, Lin SY, Lee CC, Lee NY, Liu PY, et al. Mutant prevention concentrations of ciprofloxacin and levofloxacin and target gene mutations of fluoroquinolones in Elizabethkingia anophelis. Antimicrob Agents Chemother. 2022;66:e0030122.
44. Hooper DC, Jacoby GA. Topoisomerase inhibitors: fluoroquinolone mechanisms of action and resistance. Cold Spring Harb Perspect Med. 2016;6:a025320.
45. Jian MJ, Cheng YH, Chung HY, Cheng YH, Yang HY, Hsu CS, et al. Fluoroquinolone resistance in carbapenem-resistant Elizabethkingia anophelis: phenotypic and genotypic characteristics of clinical isolates

with topoisomerase mutations and comparative genomic analysis. J Antimicrob Chemother. 2019;74:1503–10.
46. Mallinckrodt L, Huis in 't Veld, Rosema R, Voss S, Bathoorn A. E. Review on infection control strategies to minimize outbreaks of the emerging pathogen Elizabethkingia anophelis. Antimicrob Resist Infect Control. 2023;12:97.
47. Andriyanov PA, Zhurilov PA, Kashina DD, Tutrina AI, Liskova EA, Razheva IV, et al. Antimicrobial resistance and comparative genomic analysis of Elizabethkingia anophelis subsp. Endophytica isolated from raw milk. Antibiotics. 2022;11:648.
48. Breurec S, Criscuolo A, Diancourt L, Rendueles O, Vandenbogaert M, Passet V, et al. Genomic epidemiology and global diversity of the emerging bacterial pathogen Elizabethkingia anophelis. Sci Rep. 2016;6:30379.
49. Teo J, Tan SYY, Liu Y, Tay M, Ding Y, Li Y, et al. Comparative genomic analysis of Malaria Mosquito Vector-Associated novel pathogen Elizabethkingia anophelis. Genome Biol Evol. 2014;6:1158–65.
50. Oren A, Garrity GM. List of new names and new combinations previously effectively, but not validly, published. Int J Syst Evol Microbiol. 2020;70:2960–6.
51. Bandyopadhyay P, Byrne B, Chan Y, Swanson MS, Steinman HM. Legionella pneumophila catalase-peroxidases are required for proper trafficking and growth in primary macrophages. Infect Immun. 2003;71:4526–35.
52. Okan NA, Kasper DL. The atypical lipopolysaccharide of Francisella. Carbohydr Res. 2013;378:79–83.
53. Garduño RA, Chong A, Nasrallah GK, Allan DS. The legionella pneumophila chaperonin - an unusual multifunctional protein in unusual locations. Front Microbiol. 2011;2(JUNE):9837.
54. Harvey KL, Jarocki VM, Charles IG, Djordjevic SP. The diverse functional roles of elongation factor tu (Ef-tu) in microbial pathogenesis. Front Microbiol. 2019;10:485697.
55. Iliadi V, Staykova J, Iliadis S, Konstantinidou I, Sivykh P, Romanidou G, et al. Legionella pneumophila: the Journey from the environment to the blood. J Clin Med. 2022;11:11.
56. Lecuit M. Listeria monocytogenes, a model in infection biology. Cell Microbiol. 2020;22:e13186.
57. Mitchell G, Chen C, Portnoy DA. Strategies used by bacteria to grow in macrophages. Microbiol Spectr. 2016;4:3.
58. Richards AM, Von Dwingelo JE, Price CT, Kwaik YA. Cellular microbiology and molecular ecology of Legionella–amoeba interaction. Virulence. 2013;4:307.
59. Pushkareva VI, Podlipaeva JI, Goodkov AV, Ermolaeva SA. Experimental Listeria-Tetrahymena-Amoeba food chain functioning depends on bacterial virulence traits. BMC Ecol. 2019;19:1–10.
60. Mayura IPB, Gotoh K, Nishimura H, Nakai E, Mima T, Yamamoto Y, et al. Elizabethkingia anophelis, an emerging pathogen, inhibits RAW 264.7 macrophage function. Microbiol Immunol. 2021;65:317–24.
61. Yang JL, Li D, Zhan XY. Concept about the virulence factor of legionella. Microorganisms. 2022;11:74.
62. Invasion of murine. respiratory tract epithelial cells by Chryseobacterium meningosepticum and identification of genes present specifically in an invasive strain - PubMed. https://pubmed.ncbi.nlm.nih.gov/16608126/. Accessed 30 Aug 2024.
63. Sobkowiak A, Scherff N, Schuler F, Bletz S, Mellmann A, Schwierzeck V, et al. Plasmid-encoded gene duplications of extended-spectrum β-lactamases in clinical bacterial isolates. Front Cell Infect Microbiol. 2024;14:14.
64. Lee YL, Liu KM, Chang HL, Liao YC, Lin JS, Kung FY, et al. The evolutionary trend and genomic features of an emerging lineage of Elizabethkingia anophelis strains in Taiwan. Microbiol Spectr. 2022;10:e0168221.
65. Bottery MJ. Ecological dynamics of plasmid transfer and persistence in microbial communities. Curr Opin Microbiol. 2022;68:102152.
66. Miller WR, Arias CA. ESKAPE pathogens: antimicrobial resistance, epidemiology, clinical impact and therapeutics. Nat Rev Microbiol. 2024;22(10):598–616.
67. Tempel S, Bedo J, Talla E. From a large-scale genomic analysis of insertion sequences to insights into their regulatory roles in prokaryotes. BMC Genomics. 2022;23:1–19.
68. Shintani M, Vestergaard G, Milaković M, Kublik S, Smalla K, Schloter M, et al. Integrons, transposons and IS elements promote diversification of multidrug resistance plasmids and adaptation of their hosts to antibiotic pollutants from pharmaceutical companies. Environ Microbiol. 2023;25:3035–51.
69. Balows A. Manual of clinical microbiology 8th edition: P. R. Murray, E. J. Baron, J. H. Jorgenson, M. A. Pfaller, and R. H. Yolken, eds., ASM Press, 2003, 2113 pages, 2 vol, 2003 + subject & author indices, ISBN: 1-555810255-4, US$ 189.95. Diagn Microbiol Infect Dis. 2003;47:625–6.
70. Buxton R. Blood Agar Plates and Hemolysis Protocols. 2016. https://asm.org/protocols/blood-agar-plates-and-hemolysis-protocols.
71. Lenchenko E, Blumenkrants D, Sachivkina N, Shadrova N, Ibragimova A. Morphological and adhesive properties of Klebsiella pneumoniae biofilms. Vet World. 2020;13:197–200.
72. Kırmusaoğlu S. The Methods for Detection of Biofilm and Screening Antibiofilm Activity of Agents. Antimicrobials, Antibiotic Resistance, Antibiofilm Strategies and Activity Methods. 2019. https://doi.org/10.5772/INTECHOPEN.84411.
73. Weisburg WG, Barns SM, Pelletier DA, Lane DJ. 16S ribosomal DNA amplification for phylogenetic study. J Bacteriol. 1991;173:697–703.
74. Okonechnikov K, Golosova O, Fursov M, Varlamov A, Vaskin Y, Efremov I, et al. Unipro UGENE: a unified bioinformatics toolkit. Bioinformatics. 2012;28:1166–7.
75. Yoon SH, Ha SM, Kwon S, Lim J, Kim Y, Seo H, et al. Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. Int J Syst Evol Microbiol. 2017;67:1613–7.
76. Stackebrandt E, Frederiksen W, Garrity GM, Grimont PAD, Kämpfer P, Maiden MCJ, et al. Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. Int J Syst Evol Microbiol. 2002;52:1043–7.
77. Chen S, Zhou Y, Chen Y, Gu J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34:i884-890.
78. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes de novo assembler. Curr Protoc Bioinf. 2020;70:e102.
79. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS et al. SPAdes: a New Genome Assembly Algorithm and its applications to Single-Cell Sequencing. J Comput Biol. 2012;19:455–77. https://home.liebertpub.com/cmb.
80. Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A. Versatile genome assembly evaluation with QUAST-LG. Bioinformatics. 2018;34:i142-150.
81. Chklovski A, Parks DH, Woodcroft BJ, Tyson GW. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. Nat Methods 2023. 2023;20:8.
82. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30:2068–9.
83. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016;32:3047–8.
84. Tonkin-Hill G, MacAlasdair N, Ruis C, Weimann A, Horesh G, Lees JA, et al. Producing polished prokaryotic pangenomes with the Panaroo pipeline. Genome Biol. 2020;21:1–21.
85. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772.
86. Tatusov RL, Koonin EV, Lipman DJ. A genomic perspective on protein families. Science. 1997;278:631–7.
87. Hunter JD. Matplotlib: a 2D graphics environment. Comput Sci Eng. 2007;9:90–5.
88. McKinney W. Data structures for statistical computing in python. Proceedings of the 9th Python in Science Conference. 2010:56–61. https://www.scirp.org/reference/referencespapers?referenceid=3550158.
89. Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T, Keane JA, et al. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. Microb Genom. 2016;2:e000056.
90. 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. https://academic.oup.com/mbe/article/37/5/1530/5721363.
91. Zharkikh A. Estimation of evolutionary distances between nucleotide sequences. J Mol Evol. 1994;39:315–29.

92. Yang Z. A space-time process model for the evolution of DNA sequences. Genetics. 1995;139:993–1005.
93. Soubrier J, Steel M, Lee MSY, Der Sarkissian C, Guindon S, Ho SYW, et al. The influence of rate heterogeneity among sites on the time dependence of molecular rates. Mol Biol Evol. 2012;29:3345–58.
94. Lewis PO. A likelihood approach to estimating phylogeny from discrete morphological character data. Syst Biol. 2001;50:913–25.
95. Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14:6.
96. Hoang DT, Chernomor O, Von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the Ultrafast bootstrap approximation. Mol Biol Evol. 2018;35:518–22.
97. 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.
98. Letunic I, Bork P. Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021;49:W293-6.
99. Tonkin-Hill G, Gladstone RA, Pöntinen AK, Arredondo-Alonso S, Bentley SD, Corander J. Robust analysis of prokaryotic pangenome gene gain and loss rates with Panstripe. Genome Res. 2023;33(1):129–40.
100. Tettelin H, Riley D, Cattuto C, Medini D. Comparative genomics: the bacterial pan-genome. Curr Opin Microbiol. 2008;11:472–7.
101. Liu B, Zheng D, Jin Q, Chen L, Yang J. VFDB 2019: a comparative pathogenomic platform with an interactive web interface. Nucleic Acids Res. 2019;47:D687-692.
102. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421.
103. Pearson WR. An introduction to sequence similarity (homology) searching. Curr Protoc Bioinf. 2013. https://doi.org/10.1002/0471250953.bi0301s42.
104. Feldgarden M, Brover V, Gonzalez-Escalona N, Frye JG, Haendiges J, Haft DH, et al. AMRFinderPlus and the reference gene catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. Sci Rep. 2021;11:1–9.
105. Brown CL, Mullet J, Hindi F, Stoll JE, Gupta S, Choi M, Keenum I, Vikesland P, Pruden A, Zhang L. mobileOG-db: a Manually Curated Database of Protein Families Mediating the Life Cycle of Bacterial Mobile Genetic Elements. Appl Environ Microbiol. 2022;88(18):e0099122.
106. Buchfink B, Reuter K, Drost HG. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat Methods 2021. 2021;18:4.
107. Bertelli C, Gray KL, Woods N, Lim AC, Tilley KE, Winsor GL, et al. Enabling genomic island prediction and comparison in multiple genomes to investigate bacterial evolution and outbreaks. Microb Genom. 2022;8:000818.
108. Schwengers O, Jelonek L, Dieckmann MA, Beyvers S, Blom J, Goesmann A, Bakta. Rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. Microb Genom. 2021;7:000685.
109. Xie Z, Tang H. ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. Bioinformatics. 2017;33:3340–7.
110. Zhu Q, Gao S, Xiao B, He Z, Hu S. Plasmer: an accurate and sensitive bacterial plasmid prediction tool based on machine learning of shared k-mers and genomic features. Microbiol Spectr. 2023;11:11.
111. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, et al. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res. 2016;44:W16-21.
112. Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG, Parkhill J. ACT: the Artemis comparison tool. Bioinformatics. 2005;21:3422–3.
113. Néron B, Littner E, Haudiquet M, Perrin A, Cury J, Rocha EPC. IntegronFinder 2.0: identification and analysis of integrons across bacteria, with a focus on antibiotic resistance in Klebsiella. Microorganisms. 2022;10:700.
114. Cury J, Jové T, Touchon M, Néron B, Rocha EP. Identification and analysis of integrons and cassette arrays in bacterial genomes. Nucleic Acids Res. 2016;44:4539–50.

Publisher's Note