

# Exploration of Evolutionary Dynamics and Genomic Architecture in *Acinetobacter baumannii*

Rhythm Sharma

Central University of Himachal Pradesh

Dinesh Lakhanpal

[dineshlakhanpal@hpcu.ac.in](mailto:dineshlakhanpal@hpcu.ac.in)

Central University of Himachal Pradesh

---

## Research Article

**Keywords:** Nosocomial infection, Carbapenem Resistance, Information theory, Kullback–Leibler divergence, Pan-drug resistance, pan-genome

**Posted Date:** August 31st, 2026

**DOI:** <https://doi.org/10.21203/rs.3.rs-9765709/v1>

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

[Read Full License](#)

**Additional Declarations:** No competing interests reported.

---

# Exploration of Evolutionary Dynamics and Genomic Architecture in *Acinetobacter baumannii*

Rhythm Sharma<sup>1</sup>, and Dinesh Lakhanpal<sup>2\*</sup>

<sup>1</sup>Centre for Computational Biology & Bioinformatics, School of Life Sciences, Central University of Himachal Pradesh, Kangra, Himachal Pradesh, India (Bharat)

<sup>2</sup>Centre for Computational Biology & Bioinformatics, School of Life Sciences, Central University of Himachal Pradesh, Kangra, Himachal Pradesh, India (Bharat)

## Abstract

*Acinetobacter baumannii* responsible to cause nosocomial infections in the hospital settings and a leading cause of 80% death across the globe. The genome of *Acinetobacter baumannii* comprises diverse genetic elements that play a critical role in carrying and disseminating antimicrobial resistance (AMR) genes. Thus, we integrated pan-genome analysis, phylogenomics and compositional analysis to elucidate the genetic landscape. Analyses revealed that *A. baumannii* has an open, highlighting its remarkable adaptability. Additionally, phylogenomic analysis showed the global population was dominated by the high-risk carbapenem resistant sequence type ST2 lineage, although numerous other sequence types co-existed. Mapping of AMR genes on different genomic segments indicated that resistance genes were primarily acquired through horizontal transfer, with genes from other Gram-negative bacteria including *Klebsiella pneumoniae* and *Shewanella frigidimarina* as donor organisms. Further Kullback–Leibler divergence–based analysis of genomic composition showed that genomic islands (GIs) act as semi-stable, compositionally constrained elements that preferentially retain AMR determinants. In other words, these islands are not transient collections of foreign DNA but stable genetic loci containing resistance genes. Clinically, our findings emphasized the need to monitor chromosomes, GIs, prophages and plasmids as reservoirs and vehicles of carbapenem resistance genes. Thus, our analysis revealed the evolutionary dynamics and genome architecture of *A. baumannii* underscores the importance of genomic surveillance and targeted interventions against this formidable pathogen.<sup>1</sup>

---

<sup>1</sup> \*Corresponding author Tel No. +91-9888609771  
email: [dineshlakhanpal@hpcu.ac.in](mailto:dineshlakhanpal@hpcu.ac.in)

## Keywords

Nosocomial infection, Carbapenem Resistance, Information theory, Kullback–Leibler divergence, Pan-drug resistance, pan-genome.

## 1. Introduction

With the growing threat of antimicrobial resistance (AMR), pathogens have evolved from being simply classified as multi-drug resistant (MDR) to now being considered extensively drug-resistant (XDR) or pan-drug resistant (PDR) [1]. In recent years, *A. baumannii*, one of the ESKAPE pathogens, has emerged as a major concern in hospital settings across the world [2], [3]. WHO 2024 report, classified *A. baumannii* as a critical priority pathogen [4]. *A. baumannii* is a Gram negative, opportunistic, and nosocomial bacterium, majorly responsible for causing hospital acquired infections like pneumonia, urinary tract infection and blood infection [5], [6]. It has developed remarkable ability to resist a broad spectrum of antibiotics including carbapenems and colistin [7], [8]. The prevalence of MDR reported for 70% to 90% in the different parts of the world [9]. The mortality rate of *A. baumannii* infections ranges from 30% to 70% worldwide [9], [10].

The potential of *A. baumannii* to acquire antibiotic resistance, form biofilms, and persist in hostile environments is largely attributed to its open pan-genome [11], [12]. On the basis of gene pool, pan-genome are classified in two categories i.e. open and closed. Open pan-genome is responsible for continuous evolution due to acquisition of new genes over time. The closed pan-genome represents a limited and fixed gene repertoire. The pan-genome consists of the conserved core genome and diverse accessory genome [13]. The core genes are essential for bacterial survival while the accessory genes provide advantage to adapt to different ecological niches. The accessory genome flexibility also plays a significant role in carrying AMR in the species [14]. Global epidemiological studies show that resistance dissemination is not random but strongly associated with specific high-risk clonal lineages, particularly sequence type ST2, one of the most dominant international clones, carrying *blaOXA-23* gene [9]. According to 2012-2014, antimicrobial testing leadership and surveillance (ATLAS), carbapenem resistant strain of *A. baumannii* from Asian countries have reported the presence of Class D  $\beta$ -lactamase gene *blaOXA-23* in countries like Australia, China, Thailand, Philippines and Malaysia etc [9], [15]. Additionally, ST2 has been linked to the spread of carbapenem and colistin resistance genes, underscoring the evolutionary and epidemiological relevance of genome expansion.

The primary mechanism driving accessory genome diversity is horizontal gene transfer (HGT). HGT serves as a major vehicle of bacterial genome evolution due to acquisition of AMR genes and virulence factors [16]. Through plasmids, genomic islands (GIs), insertion sequences, and phages, *A. baumannii* acquires resistance genes such as carbapenemase-encoding *blaOXA* variants including blaOXA-23 [17]. Chromosomal integration of foreign DNA can generate GIs enriched in resistance determinants [18]. Studies have reported the role of GIs in various pathogenic bacteria for the spread of AMR [19]. GIs, particularly AbaR-type resistance islands, often carry clusters of resistance genes in *A. baumannii* [20]. Additionally, horizontally acquired elements are major contributors to the strain-level variability and clinical persistence.

HGT does not merely alter gene content, it can also reshape genome composition. Genome composition, particularly AT and GC content, provides a foundational yet static snapshot of genetic architecture [21]. Genes acquired through HGT often differ significantly from the host chromosome, retaining distinct genomic signatures such as altered GC content and dinucleotide composition. Plasmids are the main contributors to genetic variation in bacteria due to their frequent movement between cells [22]. Although plasmids share some similarity with the host genome, their nucleotide composition often differs significantly, whereas phages are assumed to co-evolve with the chromosome and show higher similarity [23]. Also, plasmid–host similarity has been shown to correlate with GC content, GC-rich genomes accumulate fewer random mutations, while AT-rich genomes show greater sequence variation [24]. AT richness leads to the accumulation of DNA regions and mutational biases [25]. While GC content provides a broad measure of genome composition, it lacks sensitivity to detect subtle structural variation. Higher-order nucleotide dependencies, such as dinucleotide or tetranucleotide frequencies, capture more detailed sequence organization and lineage-specific signatures, making them better suited for detecting the compositional footprints of horizontal gene transfer [26]. Although such regions may gradually ameliorate toward host composition, detectable compositional deviations frequently remain. Therefore, genome architecture reflects both vertical evolutionary history and episodic horizontal acquisition events. Quantitative approaches are therefore needed to assess how horizontal acquisition influences genome-wide compositional structure.

Information theory as a quantitative measure used for assessing genomic complexity. Kullback–Leibler divergence (DKL), or relative entropy, has been widely applied to molecular biology for analyzing DNA and protein sequences [27]. In genomic applications, DKL quantifies how one probability distribution diverges from another based on probability

distributions [28]. By comparing tetranucleotide frequencies and the zero-order Markov model (mononucleotide), DKL provides a measure of information or relative entropy in DNA sequences [29]. Higher DKL values mean high relative entropy that indicates stronger nucleotide dependencies in the corresponding tetranucleotides. Stronger dependencies between adjacent nucleotides also suggest that the sequence carries more information. Conversely, lower DKL values reflect lower relative entropy means weaker dependencies, hence less information is carried by the DNA sequences. Importantly, DKL quantifies structural divergence rather than directly detecting HGT, but it provides a sensitive metric for identifying deviations in genomic organization.

Despite extensive research on pan-genome structure, mobile genetic elements, resistance gene distribution, and clonal phylogeny in *A. baumannii*, these components are often examined independently. Pan-genome studies typically focus on gene presence–absence patterns without incorporating compositional metrics. HGT investigations frequently identify mobile elements without quantifying their genome-wide structural impact. Similarly, phylogenetic analyses describe lineage relationships without linking clonal expansion to compositional divergence. Thus, there is no integrated framework that quantitatively connects pan-genome expansion, horizontal gene transfer, genome compositional divergence, and clonal lineage evolution in *A. baumannii*. Understanding how accessory genome acquisition reshapes nucleotide architecture and how these changes relate to AMR dissemination requires a unified analytical approach.

In this study, we analyzed 421 complete genomes of *A. baumannii* to investigate the evolutionary forces shaping its genomic architecture. We performed pan-genome analysis to characterize core and accessory gene dynamics and applied multilocus sequence typing (MLST) and phylogenetic reconstruction to assess lineage structure. Using an information-theoretic approach based on DKL, we quantified compositional divergence across genomic regions to evaluate structural heterogeneity potentially associated with horizontal acquisition. Carbapenem-resistance and colistin-resistance genes were mapped across chromosomes and mobile genetic elements to assess their genomic context. By integrating pan-genome dynamics, compositional analysis, resistance gene localization, and phylogenetic structure, this study provides a comprehensive framework for understanding genome plasticity and AMR evolution in *A. baumannii*.

## 2. Method

## **2.1 Data Retrieval and Annotation**

We screened the databases of microbes at Bacterial and Viral Bioinformatics Resource Center (BV-BRC) and Genbank at the National Center for Biotechnology Information database (NCBI) [30], [31]. The complete information of *A. baumannii* strains was retrieved from the databases (accessed on June 16, 2024). The high quality, complete genomes were downloaded and annotated using Prokka version 1.14.6 [32]. Initially, we separated the chromosomes and plasmids for downstream analysis.

## **2.2 Phage and Genomic Island Prediction**

Phages and GIs were predicted from chromosomal sequences of complete strains. Phigaro version 2.4.0 and PHASTER (PHAge Search Tool Enhanced Release) webserver were used to detect prophage regions [33], [34], [35]. Only 420 prophages were predicted in the strains. For GI prediction, we used AlienHunter version 1.7 using compositional bias [36].

## **2.3 Pan-genome Analysis, Multilocus Sequence Typing, and Phylogenetic Relationship**

Roary version 3.13 was applied with default settings to generate the pan-genome of complete genomes of *A. baumannii*. [37]. The sequence typing of the 421 *A. baumannii* genomes was carried out using *mlst* v2.23.0 [38], [39]. For the sequence typing, we used the Pasteur scheme. For phylogenetic analysis, we used *trimAl* version v1.5.rev0 to remove extra or poorly aligned regions [40]. The evolutionary analysis was done using *RAxML* v8.2.12 [41] with 100 bootstrap using maximum likelihood method. Tree visualization was performed using a webserver tool named iTOL v7 [42].

## **2.4 Anti-microbial Resistance Prediction**

421 genomes of *A. baumannii* were selected for AMR analysis using AMRFinderPlus version 3.12.8 [43]. Genes for carbapenem and colistin drug classes were identified in the genomes. To determine their genomic context, the positions of AMR genes were mapped onto the chromosome, GIs, phages and plasmids. The color gradient heatmap was constructed to visualize the distribution and relative abundance of carbapenem and colistin resistance genes across genome segments.

## **2.5 Horizontal Gene Transfer**

For HGT analysis, protein sequences of AMR genes located within GIs, phages and plasmids were extracted. These sequences were analyzed using HGTector version 2.0b3 with default parameters [44]. The homologous sequences were identified via DIAMOND searches against

the NCBI non-redundant (nr) database [45]. For taxonomic classification, the self group was set to *A. baumannii* (rank: species; taxon ID: 470), while the close group corresponded to the order *Moraxellales* (rank: order; taxon ID: 2887326). Homologs outside these groups were treated as distal taxa. Genes showing strong similarity to distal taxa but weak or no hits in the close group were flagged as putative horizontally transferred genes.

## 2.6 Calculation of Kullback–Leibler divergence

The DKL, also referred to as relative entropy, was used to quantify the compositional difference between the observed and expected tetranucleotide frequencies in DNA sequences.

For a given DNA sequence, the set  $\{z_1, z_2, \dots, z_{256}\}$  represents all possible tetranucleotides (4-mers). The DKL between the observed distribution  $P$  (i.e., actual 4-mer frequencies in the sequence) and the expected distribution  $Q$  (estimated from mononucleotide composition under a zero-order Markov model) was calculated using the following formula:

$$D_{KL}(P \parallel Q) = \sum_{i=1}^{256} P(z_i) \log \left( \frac{P(z_i)}{Q(z_i)} \right)$$

Here,  $P(z_i)$  represents the observed frequency of the  $i^{\text{th}}$  tetranucleotide in a given DNA sequence  $X$ . The expected frequency,  $Q(z_i)$ , is estimated using a zero-order Markov model based on the mononucleotide composition of the DNA sequence  $X$ . That is, all chromosomal sequences were used to calculate  $Q(z_i)$  when the given sequence  $X$  is a chromosome; all GIs sequences were used when  $X$  is a GI; all phage sequences were used when  $X$  is a phage; and all plasmid sequences were used when  $X$  is a plasmid.

Linear regression model was used to compare the DKL to AT content and sequence size across different types of sequences i.e chromosomes, GIs, phages and plasmids [29]. The Mann–Whitney U test was used to compare DKL distributions among chromosomes, GIs, phages and plasmids due to the skewed nature of the data.

## 3. Results

### 3.1 Pan-genome of *A. baumannii*

The pan-genome of *A. baumannii* was analyzed for the openness of its genome repertoire. Using the formula  $n = k * x^\alpha$  where  $k = 4240.29$  and  $\gamma = 0.35$ , it was determined that the pan-

genome is open, as indicated by the condition  $0 < \gamma < 1$ . A total of 35475 genes were predicted in 421 strains. These genes were classified into core genes (801;  $99\% \leq \text{strains} \leq 100\%$ ), soft core genes (819;  $95\% \leq \text{strains} < 99\%$ ), shell genes (3167;  $15\% \leq \text{strains} < 95\%$ ), and cloud genes (30688;  $0\% \leq \text{strains} < 15\%$ ). These categories accounted for 2.25%, 2.30%, 8.92%, and 86.50% of the total gene pool, respectively. Additionally, the pan-genome curve of *A. baumannii* did not reach a plateau as shown in **Figure 1**, indicating the potential for additional genes to be introduced with the inclusion of new strains.

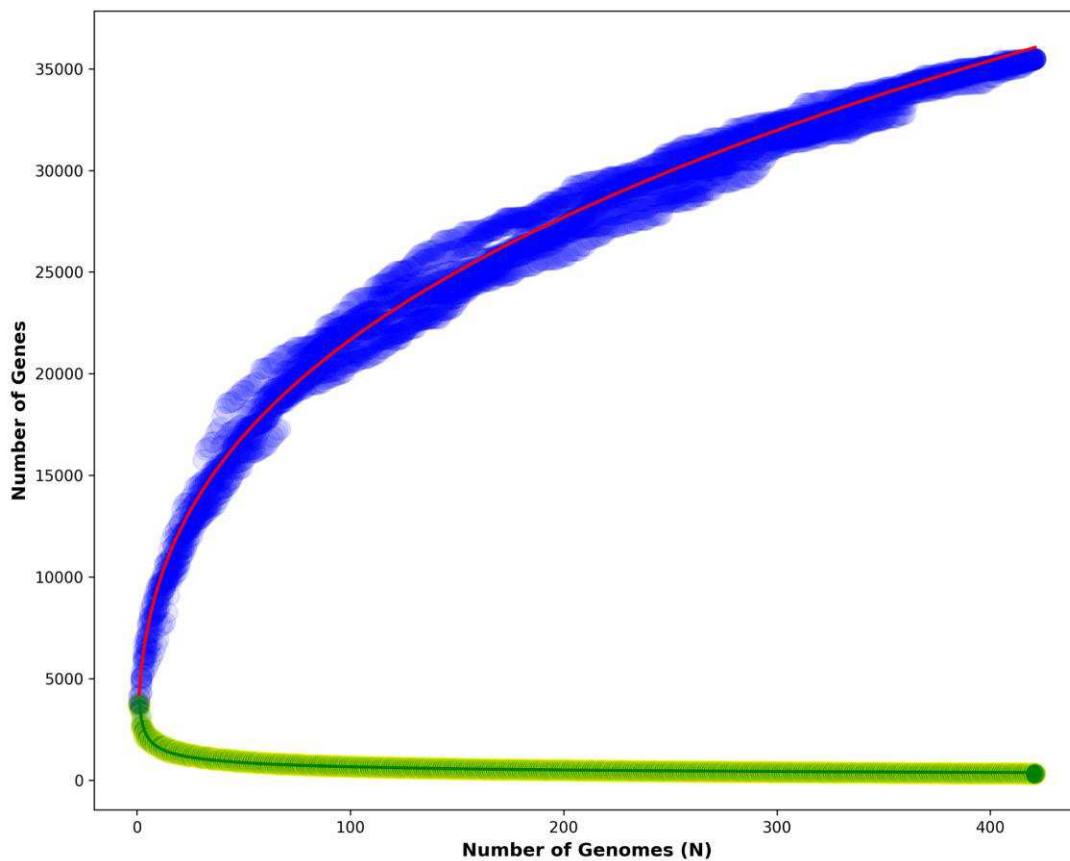


Figure 1: Pan-genome analysis of *A. baumannii*. The accumulative curve for the genomes of *A. baumannii* represented the size of pan and core genomes as the number of genomes increased. Green color represents the core genome and purple color represents the accessory genome.

### 3.2 Phylogenetic relatedness

The core-gene based phylogenetic tree revealed a distinct distribution of strains according to their antibiotic susceptibility profiles. A total of 421 strains were classified into three groups

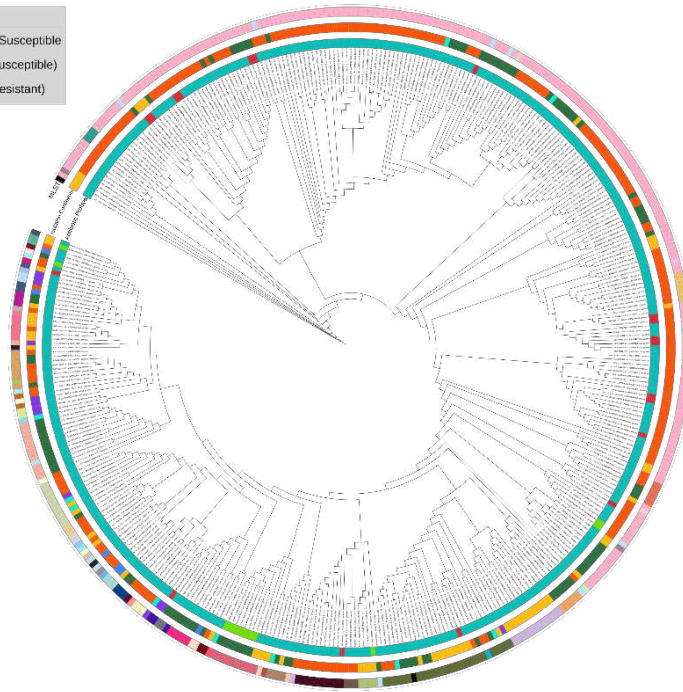
named group 1 (CCS), group 2 (CR) and group 3 (CCR), based on their resistance to carbapenem and colistin antibiotics, as shown in **Figure 2**. The majority of strains in this group 2 belonged to sequence type ST2 and were predominantly distributed across Asia. Asian strains were assigned to ST622, ST25, and ST10, all of which exhibited carbapenem resistance. In North America, ST2 strains were prevalent, and ST2 was also commonly detected in Europe. Additionally, all ST49 strains from North America were carbapenem resistant. European strains were again dominated by ST2, with all ST636 strains from this region also displaying carbapenem resistance. In contrast, strains from South America and Oceania within Group 2 were more heterogeneous in ST composition. South American isolates belonged to ST79 and ST162, whereas strains from Oceania were assigned to ST2, ST25, and ST1. Group 1 showed limited clonal diversity but wide geographic distribution, whereas Group 3 was phylogenetically and geographically dispersed, comprising multiple STs across several continents.

**Antibiotic Profiles**

- Group 1 Carbapenem Susceptible and Colistin Susceptible
- Group 2 (Carbapenem Resistant and Colistin Susceptible)
- Group 3 (Carbapenem Resistant and Colistin Resistant)

**Isolation Continents**

- Europe
- Asia
- North America
- Africa
- South America
- Oceania



**Multilocus Sequence Typing**

- 2
- 1
- 267
- 2
- 23
- 10
- 464
- 1
- 437
- 1549
- NA
- 422
- 79
- 156
- 25
- 52
- 138
- 1543
- 20
- 1544
- 78
- 77
- 345
- 1512
- 575
- 622
- 49
- 203
- 1472
- 103
- 107
- 570
- 108
- 374
- 239
- 63
- 85
- 438
- 45
- 1554
- 221
- 142
- 132
- 499
- 636
- 158
- 16
- 46
- 126
- 162
- 109
- 318
- 15
- 164
- 33
- 193
- 400
- 58
- 867
- 54
- 32
- 268
- 111
- 40
- 1579
- 1572
- 1580
- 724
- 282
- 406
- 1142
- 1336
- 195

Figure 2: Circular phylogenetic tree of 421 *A. baumannii* strains. The first ring represents the antibiotic profile (inner side). The second ring represents multilocus sequence typing (MLST) and the third ring represents the different continents (outer side).

### 3.3 Antimicrobial resistant gene across different genome segment

We identified various genes associated with carbapenem and colistin resistance in 421 genomes of *A. baumannii* as shown in **Figure 3**. After mapping the positions of these genes across different genome segments, we found a total of 965 AMR genes in the chromosome, 101 in GIs, 11 in phages, and 1 in plasmids. Within the chromosome, out of 965 genes, 872 were associated with carbapenem resistance, 75 associated with carbapenem/sulbactam-durlobactam and 18 with colistin resistance. In GIs, all predicted genes were associated with carbapenem resistance. Among the identified resistance genes, *blaNDM-2* (n=1), *blaNDM-6* (n=1), and *blaIMP-19* (n=1) and *blaOXA-78* (n=1) were exclusively identified within GIs. In addition, *blaOXA-23* (n=35), *blaNDM-1* (n=30), *blaOXA-58* (n=7), *blaOXA* (n=5), and *blaOXA-235* (n=1) were also detected in GIs. In phages, genes named *blaOXA-23* (n=6), *blaOXA* (n=3), and *blaOXA-72* (n=1) encoded for carbapenem resistance while *mcr4.3* (n=1) gene responsible for colistin resistance. The single plasmid-associated gene named *blaOXA-23* (n=1) was present in *A. baumannii* A20AB02 to confer resistance to carbapenem. Additionally, a gene named *blaOXA-23* was distributed in all the genome segments i.e. chromosomes, plasmids, phages, and GIs.

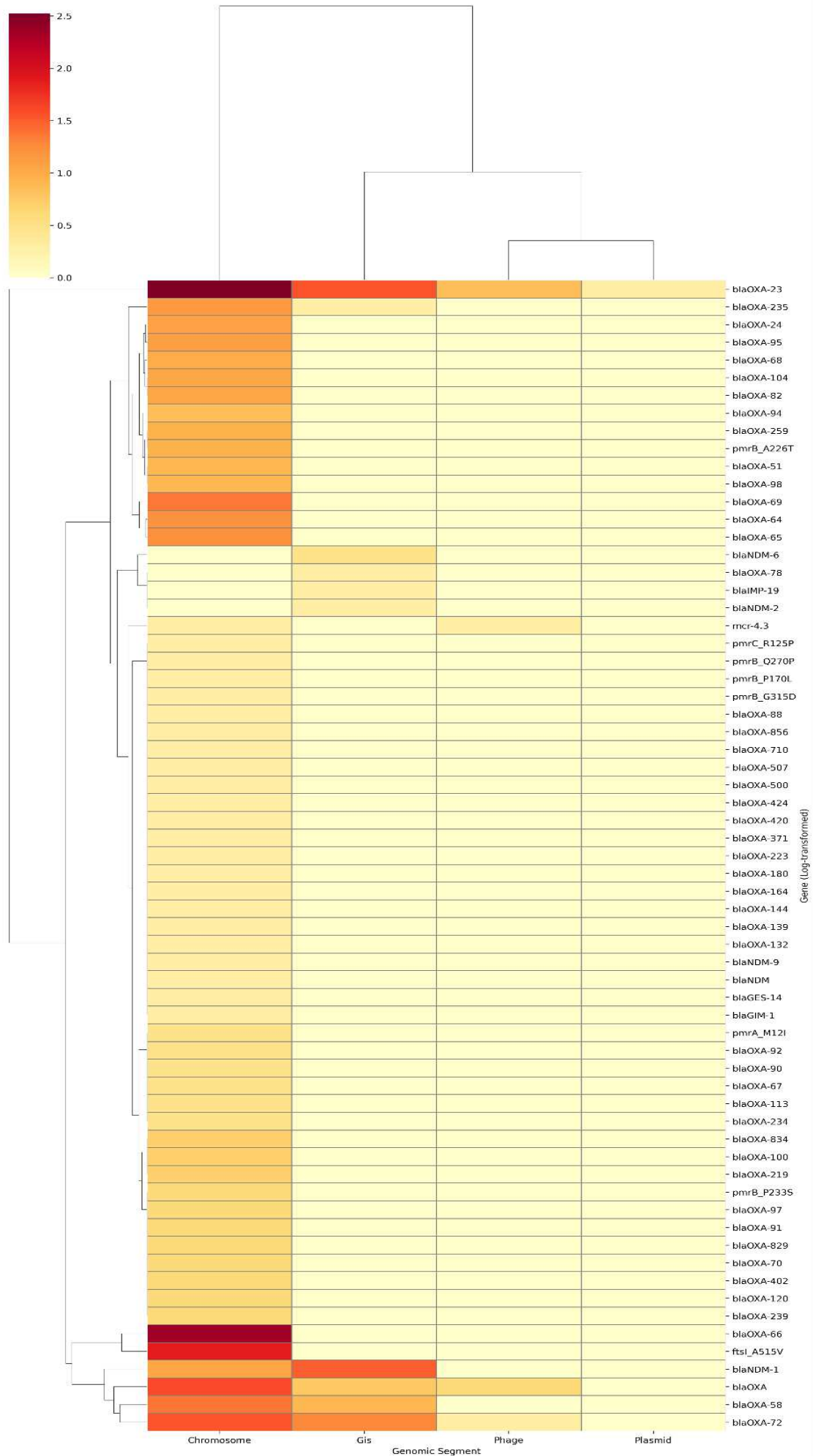


Figure 3: Distribution of carbapenem and colistin resistant genes in different genomic segments i.e. chromosome, GIs, phages and plasmids. The color gradient shows the log count of each gene.

### 3.4 Horizontal gene transfer in *A. baumannii*

HGT plays a crucial role in bacterial evolution. Through lateral gene transfer, bacteria gain new phenotypic traits that enhance their ability to colonize diverse ecological niches. A total of 15 representative proteins involved in carbapenem and colistin resistance were analysed for HGT from mobile genetic elements, including GIs, phages, and plasmids. Genes encoding these proteins belong to metallo- $\beta$ -lactamases like *blaNDM-1*, *blaNDM-2*, *blaNDM-6*, *blaIMP-19*, and *mcr-4.3*. The potential donors of the transferred genes were mainly from the species of *Klebsiella pneumoniae* for *blaNDM-1*, *blaNDM-2*, *blaNDM-6*, *blaIMP-19* and *Shewanella frigidimarina* for *mcr-4.3*.

### 3.5 DKL difference in Chromosomes, GIs, Phages and Plasmids

We examined whether information capacity, measured as DKL, differed among chromosomes, GIs, phages, and plasmids. As shown in **Figure 4**, GI exhibited the highest DKL values overall with median DKL values 0.082. DKL was significantly lower in Phages with median value 0.066 compared to GIs. Chromosomes showed a further reduction in DKL value with median 0.046 relative to GIs, while plasmids had the lowest DKL values with median 0.045 among all segment types, being significantly lower than chromosomes. Additionally, the largest contrast in DKL was observed between GIs and plasmids, indicating the greatest divergence in tetranucleotide composition relative to random expectations. Overall, all the results are statistically significant. GIs were the most compositionally biased DNA sequences, whereas plasmids exhibited the most random (least biased) nucleotide composition.

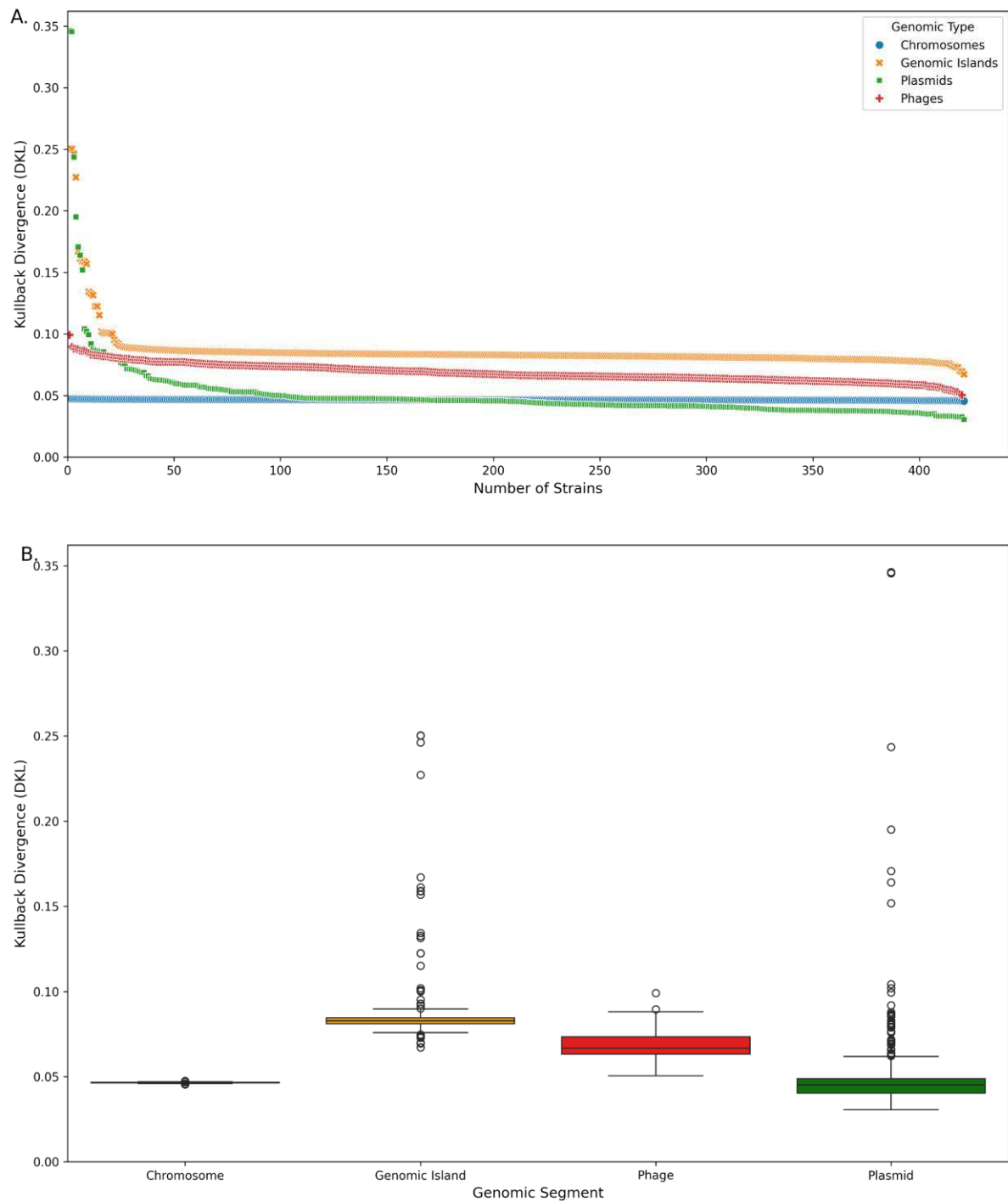


Figure 4: Kullback distribution of complete bacterial genomes segregated into chromosomes, GIs, phages and plasmids. (A) Scatter plot showing DKL distribution. (B) Box plots showing the distribution of DKL across genomic segments. Median DKL values are indicated (GI = 0.082, Phage = 0.066, Chromosome = 0.046, Plasmids = 0.045).

### 3.4 Kullback Divergence vs AT content

We examined the relationship between AT content and DKL across chromosomes, GIs, plasmids, and phages using log-transformed DKL values. DKL was treated as the response variable and AT content as the predictor in all regression analyses as shown in **Figure 5**. DKL exhibited segment-specific associations with AT content. GIs showed a moderate negative correlation, with AT content explaining most of the variance in DKL with  $R^2 = 0.59$ , whereas plasmids exhibited weak correlations with  $R^2 = 0.17$ . No correlation was identified in chromosomes and phages with  $R^2 = 0.09$  and,  $R^2 \approx 0.002$  respectively. All the given results for GIs, phages and chromosomes were predicted as statistically significant with p value  $> 0.05$ . The explanation for  $R^2$  values and other details are given in Table 1 and Table 2.

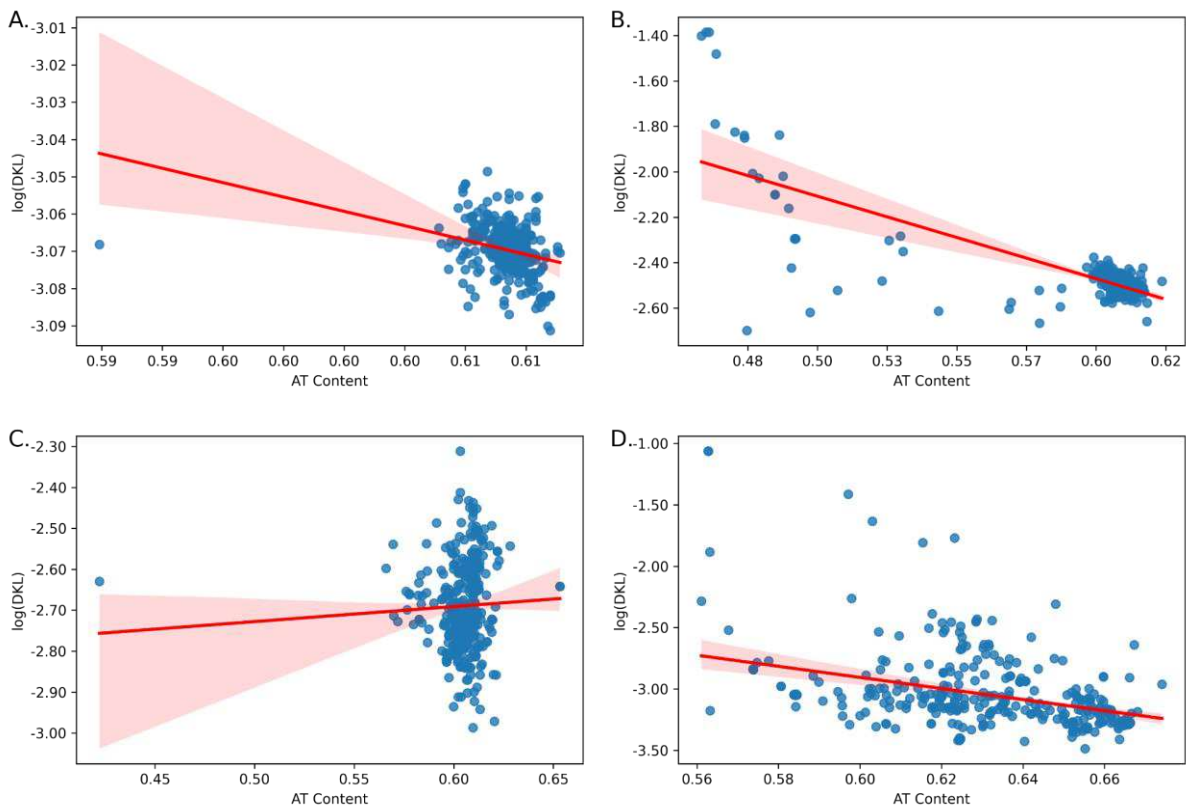


Figure 5: Log-transformed DKL is plotted against AT content using linear regression model. A. Plot represents the correlation between Chromosome AT content. B. Plot represents the correlation between GIs and AT content. C. Plot represents the correlation between phages and

AT content. D. Plot represents the correlation between plasmids and AT content. Among all the highest correlation was observed between DKL in GIs and AT content.

**Table 1: Interpretation of  $R^2$  values used to describe the strength of associations.**

| S. No. | $R^2$ range   | Interpretation                           |
|--------|---------------|------------------------------------------|
| 1.     | $< 0.15$      | Negligible (very weak explanatory power) |
| 2.     | $0.15 - 0.50$ | Weak                                     |
| 3.     | $0.50 - 0.75$ | Moderate                                 |
| 4.     | $> 0.75$      | Strong                                   |

### 3.5 Kullback Divergence vs Size

DKL was examined in relation to genome size across chromosomes, genomic islands, phages, and plasmids using regression analysis. The DKL was used as the predictor variable, with corresponding segment size as the response, to assess the correlation between segment size and compositional divergence. Based on predefined  $R^2$  thresholds, GIs exhibited a weak negative correlation between size and DKL with  $R^2 = 0.18$ . In contrast to GIs, other segments like chromosomes, phages and plasmids showed a negligible association with  $R^2 = 0.10$ ,  $R^2 = 0.01$  or less. Except for plasmids, other results were statistically significant as given in Table 2.

### 3.6 Size vs AT content

In prokaryotic organisms, chromosome size and AT content have been shown to be negatively correlated, with segment size treated as the response variable and AT content as the predictor. In this study, we carried out a regression analysis between size and AT content for chromosomes, GI, phages, and plasmids. We found a strong negative correlation for GIs between size and AT content with  $R^2=0.82$  followed by plasmids with  $R^2=0.092$ . With  $R^2=0.092$ ,  $R^2=0.02$  or less, the size of plasmids, chromosomes and phages showed negligible association with AT content. All the results were statistically significant as given in Table 2.

### 3.7 AT content vs Kullback Divergence and Size

We assessed the combined effect of DKL, sequence size, and size squared on AT content using multiple regression across different genomic segments i.e. chromosomes, plasmids, phages,

and GIs. Here AT content was a response variable while DKL, sequence size, and size squared as the predictor variable. Among these GIs showed a strong negative correlation with  $R^2=0.80$ , explaining a 80% of variance in AT content. Plasmids also exhibited a weak association with  $R^2=0.20$ . On the other hand, with  $R^2=0.02$  or less for phages and chromosomes showed negligible associations and were not statistically significant given in Table 2.

### 3.8 Kullback Divergence Between Different Antibiotic Resistant Groups

DKL was measured for different AMR groups as shown in **Figure 6**. We divided the data into three groups i.e. group 1 comprised strains that were both carbapenem susceptible and colistin susceptible (CCS). Group 2 comprised strains that were characterized by carbapenem-resistant but colistin-susceptible (CR). Additionally, Group 3 included strains that were both carbapenem resistant and colistin resistant (CCR). A significant difference was observed between CCS and CR strains, indicating that the variable of interest differs between these groups with  $p = 0.01$ . With  $p < 0.05$ , no significant differences were detected between CCS and CCR strains or CR and CCR strains.

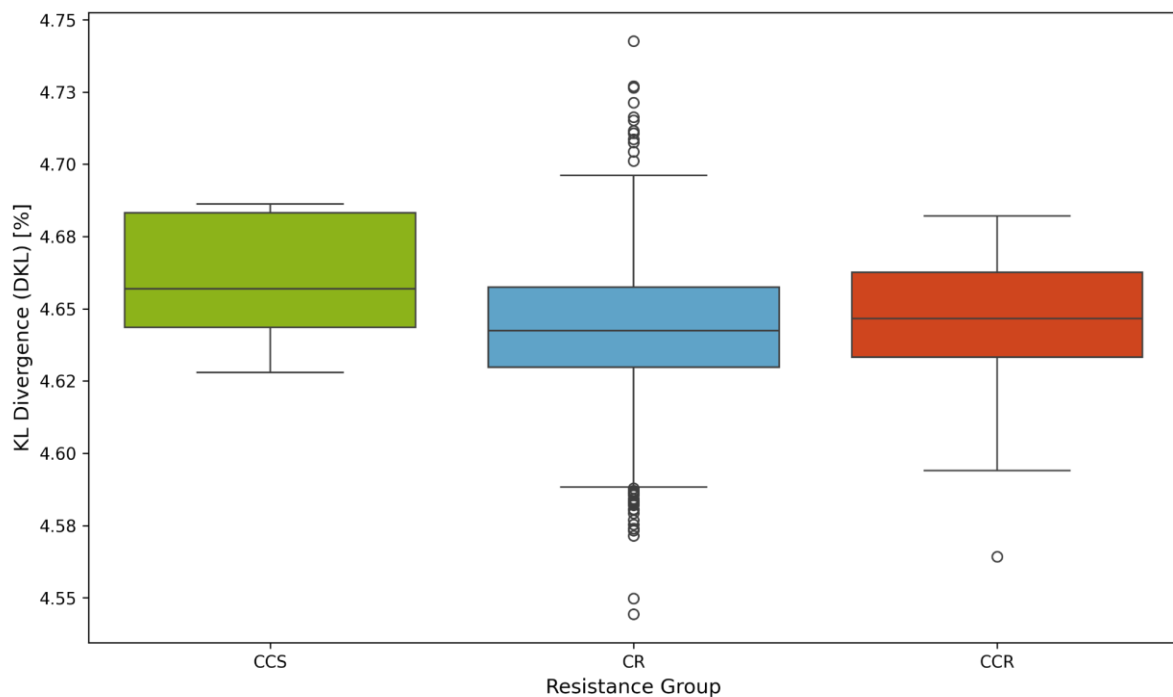


Figure 6: Box plots showing comparison of DKL values among different resistance groups. group 1 comprised strains that were susceptible to both carbapenem and colistin. Group 2

comprised strains that were characterized by carbapenem-resistant but colistin-susceptible. Additionally Group 3 included strains that were resistant to both carbapenem and colistin.

**Table 2: Table of regression analysis for different genomic segments showing their R<sup>2</sup> and p value.**

| S.No. | Relation                                                                                            | Genomic Segment | R <sup>2</sup> | p value   |
|-------|-----------------------------------------------------------------------------------------------------|-----------------|----------------|-----------|
| 1.    | Kullback divergence and AT content for<br>$\log(Y_{DKL}) = a + b \cdot X_{AT} + \varepsilon$        | Chromosome      | 0.09           | p < 0.001 |
|       |                                                                                                     | Genomic Islands | 0.59           | p < 0.001 |
|       |                                                                                                     | Phage           | 0.00           | p = 0.355 |
|       |                                                                                                     | Plasmid         | 0.16           | p < 0.001 |
| 2.    | Size and kullback divergence<br>$Y_{Size} = a + b \cdot X_{DKL} + \varepsilon$                      | Chromosome      | 0.10           | p < 0.001 |
|       |                                                                                                     | Genomic Islands | 0.18           | p < 0.001 |
|       |                                                                                                     | Phage           | 0.01           | p = 0.02  |
|       |                                                                                                     | Plasmid         | 0.00           | p = 0.568 |
| 3.    | Size and AT content<br>$Y_{Size} = a + b \cdot X_{AT} + \varepsilon$                                | Chromosome      | 0.02           | p = 0.001 |
|       |                                                                                                     | Genomic Islands | 0.82           | p < 0.001 |
|       |                                                                                                     | Phage           | 0.02           | p = 0.002 |
|       |                                                                                                     | Plasmid         | 0.09           | p < 0.001 |
| 4.    | AT content vs kullback divergence and size<br>$Y_{AT} = a + b \cdot X_{DKL} + c \cdot X_{Size} + d$ | Chromosome      | -2.47          | p < 0.001 |
|       |                                                                                                     | Genomic Islands | 0.80           | p < 0.001 |
|       |                                                                                                     | Phage           | 0.02           | p > 0.05  |
|       |                                                                                                     | Plasmid         | 0.20           | p < 0.001 |

## 4. Discussion

Due to its continuously evolving genome, *A. baumannii* has emerged as a critical nosocomial pathogen, leading to high morbidity and mortality worldwide [46]. According to a global cohort study conducted in 47 countries, approximately 96% strains of *A. baumannii* were reported to have carbapenem resistant genes [47]. Its adaptability under antimicrobial pressure and diverse environments arises from its genomic architecture, balancing a conserved core genome with a dynamic accessory genome [48].

Our genomic analysis of 421 *A. baumannii* strains revealed an exceptionally diverse and dynamic gene pool. Using the Heap's law model ( $\gamma=0.35$ ) we confirmed that *A. baumannii* has an open pan-genome, meaning that each newly sequenced strain contributes novel genes. Such openness ( $0<\gamma<1$ ) in *A. baumannii* is characteristic of species with high genome plasticity and frequent horizontal gene acquisition [49]. In our dataset, only ~2.3% of all genes are core genes and 2.3% soft core, whereas a vast majority ~86.5% are rare cloud genes found in strains. This remarkably small core-genome fraction implies that essentially all housekeeping functions are encoded by a relatively limited set of conserved genes, while the accessory genome is enormous [50]. According to Bhargavi *et al.* an open pan-genome of *A. baumannii* reflects its lifestyle as an opportunistic pathogen, helps in rapid adaptation to diverse environmental conditions due to acquisition of new virulence or resistance traits [51].

The core-gene based phylogeny of *A. baumannii* delineated three major clades with distinct clonal and geographic profiles. In our analysis, Group 2 is dominated by ST2 as the predominant lineage across Asia, Europe, and North America, and it was consistently associated with carbapenem resistance. Sharma *et al.* found that 80% - 90% of carbapenem-resistant *A. baumannii* worldwide belonged to ST2 across the globe [9], [52]. We also observed continent-specific lineages, with Asian strains comprising ST622, ST25, and ST10, all of which were carbapenemase-positive. In contrast, strains from Oceania showed a more heterogeneous composition, including ST2, ST25, and the historically widespread ST1. Another study by Gopikrishnan *et al.* noted that ST2 remains the dominant clone in Asia, while local clones such as ST622 (India) and ST724 (Taiwan) are also emerging. Additionally, we observed ST636 in several European isolates, consistent with that pattern. In North America, we identified ST49 and ST636 with carbapenem resistance. ST49 is not a classical global clone, but its presence in our data suggested local expansion of a resistant lineage. In South America,

strains belonged to ST79 and ST162. Studies have identified ST79 as a lineage with broad geographic distribution across multiple countries, although it is particularly associated with Latin America [52]. ST162 has been reported as an emerging lineage with a geographically diverse distribution, particularly in South America, where it co-circulates with other regional clones such as ST79 and ST15 [53]. Group 1 had very limited ST diversity but was still widely distributed. Group 3 is phylogenetically and geographically mixed, containing numerous STs. Colistin resistance in *A. baumannii* remains relatively infrequent compared to carbapenem resistance [9]. Studies have reported a diverse range of sequence types distributed globally across Asia, Europe, and the Americas, encompassing both well-established epidemic clones and region-specific lineages. This underscores the epidemiology distribution of carbapenem-resistant *A. baumannii* strains globally, one “high-risk” lineage ST2 predominates, while other lineages circulate to a lesser extent.

Mapping of AMR genes showed that carbapenemase and colistin-resistance genes are widely distributed but unevenly partitioned, with the chromosome GIs, prophages, and plasmids [29]. 965 resistance genes in the chromosome, followed by genomic islands (101 genes), phages (11 genes), and plasmids, which carried only one carbapenem resistance gene. In the chromosome, nearly all AMR genes (872/965) were carbapenemases or related  $\beta$ -lactamases, whereas only 18 were colistin-related and 75 were dual carbapenem/sulbactam-durlobactam. This predominance of chromosomal  $\beta$ -lactamases showed *A. baumannii* carries multiple innate carbapenemase genes *blaOXA* variants in its core genome leading to an intrinsic resistance [54]. Additionally, GIs were strikingly enriched for resistance genes as we identified 101 AMR genes in GIs, all conferring carbapenem resistance. This is consistent with GIs functioning as resistance islands or “AbaR” elements in *A. baumannii* [20]. Importantly, we found several carbapenemase genes were found exclusively present on GIs including *blaNDM-2*, *blaNDM-6*, *blaIMP-19* and *blaOXA-78*. We also found dozens of other carbapenemase genes on GIs i.e. *blaOXA-23* and *blaNDM-1* indicating that resistance islands have accumulated a broad arsenal of  $\beta$ -lactamases. By contrast, phage carried only 11 AMR genes. These included *blaOXA-23*, *blaOXA-72*, *blaOXA* for carbapenem resistance and *mcr-4.3* gene for colistin resistance. Karampatakis et.al. also reported the presence of *blaOXA-23* and related genes in phages [49]. We have also identified a single *mcr-4.3* gene within a phage. Previous study has shown that *mcr-4* variants are typically plasmid-mediated and are believed to originate from *Shewanella* species reservoirs [55]. This suggests that, in addition to plasmids, phages may also contribute to the mobilization and dissemination of colistin resistance in *A. baumannii*. By comparison,

we found only one gene named *blaOXA-23* on a plasmid. Wang *et al.* showed that *blaOXA-23* gene can insert in multiple sites and even co-exist on chromosome and plasmid within the same cell [56]. It is consistent with prior observations that many *A. baumannii* carbapenemases were carried on chromosomes as well as on plasmids.

Together, these findings showed that AMR genes in *A. baumannii* are present not only within the chromosome but also across diverse mobile elements [57]. The presence of these genes in diverse bacterial species, including *Escherichia coli*, *Klebsiella pneumoniae*, and *Pseudomonas aeruginosa*, strongly supports their dissemination through HGT rather than vertical inheritance [58]. Consistent with this, our HGT analysis identified *Klebsiella pneumoniae* as the donor of several carbapenemases, including *blaNDM-1*, *blaNDM-2*, *blaNDM-6*, and *blaIMP-19*, whereas *Shewanella frigidimarina* was implicated as the source of *mcr-4.3*. *Klebsiella* is a notorious reservoir for *blaNDM* and other metallo- $\beta$ -lactamases as the prototypical *blaNDM-1* was first identified in *K. pneumoniae* from Indian strains [59]. *blaNDM-1* exhibits higher binding affinity toward most cephalosporins, while showing comparatively lower affinity for carbapenems. Previous studies indicate that *blaNDM* and *blaIMP* genes are widely disseminated across Gram-negative bacteria, including *Escherichia coli*, *Pseudomonas aeruginosa*, and *Acinetobacter* spp [60]. Similarly, *Shewanella baltica* has long been recognised as the source of the *mcr-4* family, which confers colistin resistance [55].

Finally, we examined the compositional bias or information content of different genomic segments using DKL. We identified that information content of *A. baumannii* varies significantly among different segments of the genome. GIs showed the greatest DNA composition bias among all underscores their central role as hubs of HGT and evolutionary diversity in *A. baumannii* [61]. The high DKL observed in GIs is consistent with their acquisition from diverse donor lineages, resulting in nucleotide compositions that differ substantially from the host chromosome [18]. This suggests that over the period of time, GIs have undergone complete compositional amelioration into the host chromosome. In contrast, plasmids exhibited lower DKL, indicating a more compositionally heterogeneous, more random or partially ameliorated nucleotide structure consistent with other studies [29]. Although plasmids play a crucial role as vectors of HGT, their reduced information content likely reflects broad host range and less stability [24]. Nonetheless, the transition of plasmids into stably inherited genetic elements remains understudied.

We also predicted the negative correlation between DKL and AT content observed in GIs further supports this interpretation. Regression analysis revealed the moderate negative

correlation for GIs between chromosomes, phages and plasmids with  $R^2 = 0.59$ . Plasmids and chromosomes showed weak correlation between DKL and AT content, This suggests that information was more linked to change in the AT content of *A. baumannii*. GIs with lower AT content and high DKL, consistent with the studies that GC-enriched regions are more stable or less prone to mutational changes [62]. Because AT-rich sequences tend to be more permissive and compositionally degenerate, increases in AT content can reduce detectable divergence from the genomic background, thereby lowering DKL. Bohlin and colleagues also showed that high DKL was generally associated with low AT content, indicating that more compositionally structured sequences tend to be relatively GC-enriched [29].

Additionally, GIs with  $R^2 = 0.18$  showed weak inconsistent correlations with DKL and genome size. The weak correlation between size and DKL suggests that increases in size do not necessarily correspond to increases in information. Instead, high DKL in GIs appears to be driven primarily by nucleotide composition, donor lineage, and degree of compositional amelioration rather than by size [63]. On the other hand, lack of a correlation across chromosomes, phages and plasmids implies that information content is shaped by sequence organization and evolutionary history rather than genome expansion. Additionally, various studies have demonstrated that chromosome sizes are inversely proportional to in prokaryotic organisms.

Generally, AT content is widely reported to correlate negatively with genome size. Notably, GIs displayed a strong negative correlation between size and AT content ( $R^2 = 0.82$ ). In contrast, this relationship was not observed in chromosomes, phage and plasmids. This indicates that the relation between genome size and AT is independent of AT and DKL. Again, only GIs showed strong but negative correlation of AT content with size and DKL with  $R^2 = 0.80$ . On the other hand, plasmids also showed little negative correlation of AT content with size and DKL. This pattern suggests that GIs' higher information content (higher DKL) tend to be relatively smaller and more GC-enriched. Overall, high information potential reflects strong compositional bias, highlighting their central role in genome plasticity and rapid adaptive evolution.compositional amelioration into the host chromosome. In contrast, plasmids exhibited lower DKL, indicating a more compositionally heterogeneous, more random or partially ameliorated nucleotide structure consistent with other studies [29]. Although plasmids play a crucial role as vectors of HGT, their reduced information content likely reflects broad host range and less stability [24]. Nonetheless, the transition of plasmids into stably inherited genetic elements remains understudied.

## 5. Conclusion

Our study integrates pan-genomics, phylogenomics and genome composition analysis to offer a detailed picture of *A. baumannii* genetics. We confirm that *A. baumannii* has an open, highly variable pan-genome, underscoring its adaptability. Its global population structure is dominated by ST2, spreading resistance worldwide, yet many diverse lineages coexist. AMR mapping analyses showed that resistance genes were frequently acquired horizontally rather than vertical transfer. These genes were acquired from two bacterial species named *Klebsiella pneumoniae* and *Shewanella frigidimarina* as potential donors. Additionally, DKL-based genome architecture analysis revealed that GIs are not merely transient collections of foreign DNA but semi-stable, compositionally constrained elements that preferentially retain AMR determinants. From a clinical perspective, these findings emphasize the importance of monitoring chromosomes, GIs, phages and plasmids as reservoirs and vehicles of critical resistance determinants in *A. baumannii*. Ultimately, understanding these aspects is essential for designing more effective therapeutic interventions and limiting the spread of this formidable pathogen.

## 6. Acknowledgement

Authors acknowledge the computational facility provided by Central University of Himachal Pradesh, at Center for Computational Biology and Bioinformatics.

## 7. Funding Declaration

No funding was received for this work.

## 8. References

- [1] “Global burden of bacterial antimicrobial resistance 1990–2021: a systematic analysis with forecasts to 2050 - The Lancet.” Accessed: Jul. 28, 2025. [Online]. Available: [https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(24\)01867-1/fulltext](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(24)01867-1/fulltext)
- [2] C. Ma and S. McClean, “Mapping Global Prevalence of *Acinetobacter baumannii* and Recent Vaccine Development to Tackle It,” *Vaccines*, vol. 9, no. 6, Jun. 2021, doi: 10.3390/vaccines9060570.
- [3] M. Castanheira, R. E. Mendes, and A. C. Gales, “Global Epidemiology and Mechanisms of Resistance of *Acinetobacter baumannii*-calcoaceticus Complex,” *Clin. Infect. Dis.*, vol. 76, no. Supplement\_2, pp. S166–S178, May 2023, doi: 10.1093/cid/ciad109.

- [4] WHO *Bacterial Priority Pathogens List 2024: Bacterial Pathogens of Public Health Importance, to Guide Research, Development, and Strategies to Prevent and Control Antimicrobial Resistance*, 1st ed. Geneva: World Health Organization, 2024.
- [5] A. Howard, M. O'Donoghue, A. Feeney, and R. D. Sleator, "Acinetobacter baumannii: An emerging opportunistic pathogen," *Virulence*, vol. 3, no. 3, pp. 243–250, May 2012, doi: 10.4161/viru.19700.
- [6] I. Cavallo *et al.*, "Acinetobacter baumannii in the critically ill: complex infections get complicated," *Front. Microbiol.*, vol. 14, Jun. 2023, doi: 10.3389/fmicb.2023.1196774.
- [7] N. Thadtapong, S. Chaturongakul, S. Soodvilai, and P. Dubbs, "Colistin and Carbapenem-Resistant Acinetobacter baumannii Aci46 in Thailand: Genome Analysis and Antibiotic Resistance Profiling," *Antibiot. Basel Switz.*, vol. 10, no. 9, p. 1054, Aug. 2021, doi: 10.3390/antibiotics10091054.
- [8] A. Papazachariou, R.-N. Tziolos, S. Karakonstantis, P. Ioannou, G. Samonis, and D. P. Kofteridis, "Treatment Strategies of Colistin Resistance Acinetobacter baumannii Infections," *Antibiotics*, vol. 13, no. 5, Art. no. 5, May 2024, doi: 10.3390/antibiotics13050423.
- [9] R. Sharma and D. Lakhanpal, "Acinetobacter baumannii: A comprehensive review of global epidemiology, clinical implications, host interactions, mechanisms of antimicrobial resistance and mitigation strategies," *Microb. Pathog.*, vol. 204, p. 107605, Jul. 2025, doi: 10.1016/j.micpath.2025.107605.
- [10] "Risk factors for mortality in Acinetobacter baumannii bloodstream infections and development of a predictive mortality model - PubMed." Accessed: Jul. 28, 2025. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/39029658/>
- [11] C. A. Matthews, N. S. Watson-Haigh, R. A. Burton, and A. E. Sheppard, "A gentle introduction to pangenomics," *Brief. Bioinform.*, vol. 25, no. 6, p. bbae588, Nov. 2024, doi: 10.1093/bib/bbae588.
- [12] "Comparative Genomics of Herpesviridae Family to Look for Potential Signatures of Human Infecting Strains - Sharma - 2016 - International Journal of Genomics - Wiley Online Library." Accessed: Feb. 19, 2026. [Online]. Available: <https://onlinelibrary.wiley.com/doi/full/10.1155/2016/9543274>
- [13] V. Sharma, A. Sood, P. Ray, and A. Angrup, "Comparative genomics reveals the evolution of antimicrobial resistance in *Bacteroides nordii*," *Microb. Pathog.*, vol. 173, p. 105811, Dec. 2022, doi: 10.1016/j.micpath.2022.105811.
- [14] J. de Korne-Elenbaas, S. M. Bruisten, A. P. van Dam, M. C. J. Maiden, and O. B. Harrison, "The Neisseria gonorrhoeae Accessory Genome and Its Association with the Core Genome and Antimicrobial Resistance," *Microbiol. Spectr.*, vol. 10, no. 3, pp. e02654-21, May 2022, doi: 10.1128/spectrum.02654-21.

- [15] C. Liu *et al.*, “Epidemiological and genetic characteristics of clinical carbapenem-resistant *Acinetobacter baumannii* strains collected countrywide from hospital intensive care units (ICUs) in China,” *Emerg. Microbes Infect.*, vol. 11, no. 1, pp. 1730–1741, Dec. 2022, doi: 10.1080/22221751.2022.2093134.
- [16] G. Liu, L. E. Thomsen, and J. E. Olsen, “Antimicrobial-induced horizontal transfer of antimicrobial resistance genes in bacteria: a mini-review,” *J. Antimicrob. Chemother.*, vol. 77, no. 3, pp. 556–567, Mar. 2022, doi: 10.1093/jac/dkab450.
- [17] H. R. Noel, J. R. Petrey, and L. D. Palmer, “Mobile genetic elements in *Acinetobacter* antibiotic-resistance acquisition and dissemination,” *Ann. N. Y. Acad. Sci.*, vol. 1518, no. 1, pp. 166–182, Dec. 2022, doi: 10.1111/nyas.14918.
- [18] E. F. Boyd, S. Almagro-Moreno, and M. A. Parent, “Genomic islands are dynamic, ancient integrative elements in bacterial evolution,” *Trends Microbiol.*, vol. 17, no. 2, pp. 47–53, Feb. 2009, doi: 10.1016/j.tim.2008.11.003.
- [19] X.-Q. Wang, L. Xu, M.-J. Du, H.-N. Wang, and C.-W. Lei, “Genomic islands associated with the dissemination of multidrug resistance in clinically important pathogenic bacteria,” *J. Glob. Antimicrob. Resist.*, vol. 48, pp. 1–11, May 2026, doi: 10.1016/j.jgar.2026.02.007.
- [20] D. Bi *et al.*, “Large-Scale Identification of AbaR-Type Genomic Islands in *Acinetobacter baumannii* Reveals Diverse Insertion Sites and Clonal Lineage-Specific Antimicrobial Resistance Gene Profiles,” *Antimicrob. Agents Chemother.*, vol. 63, no. 4, pp. e02526-18, Apr. 2019, doi: 10.1128/AAC.02526-18.
- [21] G. C. diCenzo and T. M. Finan, “The Divided Bacterial Genome: Structure, Function, and Evolution,” *Microbiol. Mol. Biol. Rev. MMBR*, vol. 81, no. 3, pp. e00019-17, Aug. 2017, doi: 10.1128/MMBR.00019-17.
- [22] S. Halary, J. W. Leigh, B. Cheaib, P. Lopez, and E. Baptiste, “Network analyses structure genetic diversity in independent genetic worlds,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 107, no. 1, pp. 127–132, Jan. 2010, doi: 10.1073/pnas.0908978107.
- [23] O. N. Reva and B. Tümmler, “Global features of sequences of bacterial chromosomes, plasmids and phages revealed by analysis of oligonucleotide usage patterns,” *BMC Bioinformatics*, vol. 5, p. 90, Jul. 2004, doi: 10.1186/1471-2105-5-90.
- [24] S. Köstlbacher, A. Collingro, T. Halter, D. Domman, and M. Horn, “Coevolving Plasmids Drive Gene Flow and Genome Plasticity in Host-Associated Intracellular Bacteria,” *Curr. Biol.*, vol. 31, no. 2, pp. 346-357.e3, Jan. 2021, doi: 10.1016/j.cub.2020.10.030.
- [25] M. Rajewska, K. Wegrzyn, and I. Konieczny, “AT-rich region and repeated sequences – the essential elements of replication origins of bacterial

replicons,” *FEMS Microbiol. Rev.*, vol. 36, no. 2, pp. 408–434, Mar. 2012, doi: 10.1111/j.1574-6976.2011.00300.x.

[26] J. Bohlin, M. W. Van Passel, L. Snipen, A. B. Kristoffersen, D. Ussery, and S. P. Hardy, “Relative entropy differences in bacterial chromosomes, plasmids, phages and genomic islands,” *BMC Genomics*, vol. 13, no. 1, p. 66, 2012, doi: 10.1186/1471-2164-13-66.

[27] S. Kullback and R. A. Leibler, “On Information and Sufficiency,” *Ann. Math. Stat.*, vol. 22, no. 1, pp. 79–86, Mar. 1951, doi: 10.1214/aoms/1177729694.

[28] V. Bonnici and V. Manca, “Informational laws of genome structures,” *Sci. Rep.*, vol. 6, p. 28840, Jun. 2016, doi: 10.1038/srep28840.

[29] J. Bohlin, M. W. van Passel, L. Snipen, A. B. Kristoffersen, D. Ussery, and S. P. Hardy, “Relative entropy differences in bacterial chromosomes, plasmids, phages and genomic islands,” *BMC Genomics*, vol. 13, p. 66, Feb. 2012, doi: 10.1186/1471-2164-13-66.

[30] R. D. Olson *et al.*, “Introducing the Bacterial and Viral Bioinformatics Resource Center (BV-BRC): a resource combining PATRIC, IRD and ViPR,” *Nucleic Acids Res.*, vol. 51, no. D1, pp. D678–D689, Jan. 2023, doi: 10.1093/nar/gkac1003.

[31] “National Center for Biotechnology Information.” Accessed: Aug. 28, 2024. [Online]. Available: <https://www.ncbi.nlm.nih.gov/>

[32] T. Seemann, “Prokka: rapid prokaryotic genome annotation,” *Bioinformatics*, vol. 30, no. 14, pp. 2068–2069, Jul. 2014, doi: 10.1093/bioinformatics/btu153.

[33] “Phigaro: high-throughput prophage sequence annotation | Bioinformatics | Oxford Academic.” Accessed: Jul. 28, 2025. [Online]. Available: <https://academic.oup.com/bioinformatics/article/36/12/3882/5822875>

[34] “PHAST: a fast phage search tool - PubMed.” Accessed: Jul. 28, 2025. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/21672955/>

[35] “PHASTER: a better, faster version of the PHAST phage search tool - PubMed.” Accessed: Jul. 28, 2025. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/27141966/>

[36] “Interpolated variable order motifs for identification of horizontally acquired DNA: revisiting the Salmonella pathogenicity islands - PubMed.” Accessed: Jul. 28, 2025. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/16837528/>

[37] A. J. Page *et al.*, “Roary: rapid large-scale prokaryote pan genome analysis,” *Bioinformatics*, vol. 31, no. 22, pp. 3691–3693, Nov. 2015, doi: 10.1093/bioinformatics/btv421.

- [38] K. A. Jolley, J. E. Bray, and M. C. J. Maiden, "Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications," *Wellcome Open Res.*, vol. 3, p. 124, Sep. 2018, doi: 10.12688/wellcomeopenres.14826.1.
- [39] T. Seemann, *tseemann/mlst*. (Nov. 15, 2024). Shell. Accessed: Nov. 29, 2024. [Online]. Available: <https://github.com/tseemann/mlst>
- [40] S. Capella-Gutiérrez, J. M. Silla-Martínez, and T. Gabaldón, "trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses," *Bioinformatics*, vol. 25, no. 15, pp. 1972–1973, Aug. 2009, doi: 10.1093/bioinformatics/btp348.
- [41] A. Stamatakis, "RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies," *Bioinformatics*, vol. 30, no. 9, pp. 1312–1313, May 2014, doi: 10.1093/bioinformatics/btu033.
- [42] "Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation | Bioinformatics | Oxford Academic." Accessed: Jul. 28, 2025. [Online]. Available: <https://academic.oup.com/bioinformatics/article/23/1/127/188940>
- [43] M. Feldgarden *et al.*, "AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence," *Sci. Rep.*, vol. 11, no. 1, p. 12728, Jun. 2021, doi: 10.1038/s41598-021-91456-0.
- [44] Q. Zhu, M. Kosoy, and K. Dittmar, "HGTector: an automated method facilitating genome-wide discovery of putative horizontal gene transfers," *BMC Genomics*, vol. 15, no. 1, p. 717, Aug. 2014, doi: 10.1186/1471-2164-15-717.
- [45] B. Buchfink, K. Reuter, and H.-G. Drost, "Sensitive protein alignments at tree-of-life scale using DIAMOND," *Nat. Methods*, vol. 18, no. 4, pp. 366–368, Apr. 2021, doi: 10.1038/s41592-021-01101-x.
- [46] "Acinetobacter baumannii: A comprehensive review of global epidemiology, clinical implications, host interactions, mechanisms of antimicrobial resistance and mitigation strategies - PubMed." Accessed: Jul. 28, 2025. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/40250495/>
- [47] C. Müller *et al.*, "A global view on carbapenem-resistant *Acinetobacter baumannii*," *mBio*, vol. 14, no. 6, pp. e02260-23, Oct. 2023, doi: 10.1128/mbio.02260-23.
- [48] "Insights into *Acinetobacter baumannii*: A Review of Microbiological, Virulence, and Resistance Traits in a Threatening Nosocomial Pathogen." Accessed: Sep. 09, 2025. [Online]. Available: <https://www.mdpi.com/2079-6382/9/3/119>

- [49] T. Karampatakis, K. Tsergouli, and P. Behzadi, "Pan-Genome Plasticity and Virulence Factors: A Natural Treasure Trove for *Acinetobacter baumannii*," *Antibiotics*, vol. 13, no. 3, p. 257, Mar. 2024, doi: 10.3390/antibiotics13030257.
- [50] D. Guillén-Navarro, S. A. Ochoa, D. De La Rosa-Zamboni, S. Giono-Cerezo, J. Xicohtencatl-Cortes, and A. Cruz-Córdova, "Comparative genomics of carbapenem-resistant *Acinetobacter baumannii* isolated from pediatric patients in a tertiary care hospital," *Microbiol. Spectr.*, vol. 13, no. 11, pp. e01676-25, Oct. 2025, doi: 10.1128/spectrum.01676-25.
- [51] "Comparative genomics reveals avian *Acinetobacter baumannii* as antimicrobial resistance reservoirs and bovine strains with elevated pathogenicity | Scientific Reports." Accessed: Mar. 27, 2026. [Online]. Available: <https://www.nature.com/articles/s41598-025-33117-0>
- [52] C. Müller *et al.*, "A global view on carbapenem-resistant *Acinetobacter baumannii*," *mBio*, vol. 14, no. 6, pp. e02260-23, doi: 10.1128/mbio.02260-23.
- [53] F. K. S. F. de Azevedo *et al.*, "Molecular epidemiology of multidrug-resistant *Acinetobacter baumannii* infection in two hospitals in Central Brazil: the role of ST730 and ST162 in clinical outcomes," *J. Med. Microbiol.*, vol. 68, no. 1, pp. 31–40, Jan. 2019, doi: 10.1099/jmm.0.000853.
- [54] M. H. Wong, B. K. Chan, E. W. Chan, and S. Chen, "Over-Expression of ISAbal-Linked Intrinsic and Exogenously Acquired OXA Type Carbapenem-Hydrolyzing-Class D-β-Lactamase-Encoding Genes Is Key Mechanism Underlying Carbapenem Resistance in *Acinetobacter baumannii*," *Front. Microbiol.*, vol. 10, p. 2809, Dec. 2019, doi: 10.3389/fmicb.2019.02809.
- [55] N. P. Marathe, F. Salvà-Serra, P. S. Nimje, and E. R. B. Moore, "Novel Plasmid Carrying Mobile Colistin Resistance Gene mcr-4.3 and Mercury Resistance Genes in *Shewanella baltica*: Insights into Mobilization of mcr-4.3 in *Shewanella* Species," *Microbiol. Spectr.*, vol. 10, no. 6, p. e0203722, Dec. 2022, doi: 10.1128/spectrum.02037-22.
- [56] Z. Wang, H. Li, J. Zhang, and H. Wang, "Co-Occurrence of blaOXA-23 in the Chromosome and Plasmid: Increased Fitness in Carbapenem-Resistant *Acinetobacter baumannii*," *Antibiotics*, vol. 10, no. 10, p. 1196, Oct. 2021, doi: 10.3390/antibiotics10101196.
- [57] B. Grodner *et al.*, "Spatial mapping of mobile genetic elements and their bacterial hosts in complex microbiomes," *Nat. Microbiol.*, vol. 9, no. 9, pp. 2262–2277, Sep. 2024, doi: 10.1038/s41564-024-01735-5.
- [58] B. J. Arnold, I.-T. Huang, and W. P. Hanage, "Horizontal gene transfer and adaptive evolution in bacteria," *Nat. Rev. Microbiol.*, vol. 20, no. 4, pp. 206–218, Apr. 2022, doi: 10.1038/s41579-021-00650-4.
- [59] D. Yong *et al.*, "Characterization of a New Metallo-β-Lactamase Gene, blaNDM-1, and a Novel Erythromycin Esterase Gene Carried on a Unique Genetic Structure in *Klebsiella pneumoniae* Sequence Type 14 from India,"

*Antimicrob. Agents Chemother.*, vol. 53, no. 12, pp. 5046–5054, Dec. 2009, doi: 10.1128/AAC.00774-09.

[60] P. Nordmann, T. Naas, and L. Poirel, “Global spread of Carbapenemase-producing Enterobacteriaceae,” *Emerg. Infect. Dis.*, vol. 17, no. 10, pp. 1791–1798, Oct. 2011, doi: 10.3201/eid1710.110655.

[61] P. Banerjee, O. Eulenstein, and I. Friedberg, “Discovering genomic islands in unannotated bacterial genomes using sequence embedding”, Accessed: Feb. 19, 2026. [Online]. Available: <https://dx.doi.org/10.1093/bioadv/vbae089>

[62] A. Almpanis, M. Swain, D. Gatherer, and N. McEwan, “Correlation between bacterial G+C content, genome size and the G+C content of associated plasmids and bacteriophages,” *Microb. Genomics*, vol. 4, no. 4, p. e000168, Apr. 2018, doi: 10.1099/mgen.0.000168.

[63] “Positive correlations between genomic %AT and genome size within strains of bacterial species - Bohlin - 2014 - Environmental Microbiology Reports - Wiley Online Library.” Accessed: Feb. 19, 2026. [Online]. Available: <https://enviromicro-journals.onlinelibrary.wiley.com/doi/10.1111/1758-2229.12145>