Novel Research in Microbiology Journal (2024), 8(5): 2653-2665

(Print) (ISSN 2537-0286)
(Online) (ISSN 2537-0294)
https://nrmj.journals.ekb.eg/

Research Article

DOI: 10.21608/NRMJ.2024.322471.1720

Comparative genomic analysis of enterobacterial clinical isolates reveals differences in resistome and stress signaling pathway

Sarah Sonbol[1]; Nahla A. Hussein[2*]

[1]Microbiology and Immunology Department, Faculty of Pharmacy, Heliopolis University, El-Salam City, Cairo, Egypt; [2]Molecular Biology Department, Biotechnology Research Institute, National Research Centre, Cairo, Egypt

[*]Corresponding author E-mail: nahlahussein@aucegypt.edu

Received: 7 September, 2024; Accepted: 13 October, 2024; Published online: 15 October, 2024

Abstract

The *Enterobacteriaceae* is a large family of bacteria that includes many important bacterial pathogens known for their high virulence and contributions to the global antimicrobial resistance threat. We aimed, through this study, to provide a comparative overview of the genomes, resistance determinants, and stress signaling pathway of 22 enterobacterial clinical isolates. Our results showed that *Escherichia coli* harbored the lowest percentage of core genes (44 % versus 58 % in *Salmonella enterica*, and 52 % in *Klebsiella pneumoniae*) and the highest percentage of strain- specific genes (31 % versus 18 % in *S. enterica* and 24 % in *K. pneumoniae*). All bacterial strains included in the analysis encoded genes for resistance to commonly used antibacterials, including cell envelope acting antibiotics, aminoglycosides, quinolones, and sulphonamide. The resistance genes showed different distributions among the core, dispensable and/or strain-specific genomes, ranging from an almost equal distribution in *E. coli* and *S. enterica* to higher occurrence in the dispensable/ strain specific genome in *K. pneumoniae*. Depending on the bacterial genus, cell envelope acting antibiotics occupied between 15 % - 30 % of the core genome resistome. In addition, the sequence diversity of different proteins of the signalling pathway "the regulator of colanic acid capsule synthesis" (Rcs), which is involved in sensing and responding to cell envelope threats was examined. High diversity of two domains in Rcs system proteins (*i.e.*, IgaA first cytosolic domain and RcsD pseudokinase) was observed despite high conservation of all other proteins. All these differences might point out to specific adaptation strategies of the studied bacterial isolates to antibiotic exposure and ecological niche variation.

Keywords: Resistome, Enterobacteria, Pangenome, Rcs system, Cell-envelope acting antibiotics

1. Introduction

The *Enterobacteriaceae* family includes many important pathogens such as *Escherichia* spp., *Shigella* spp., *Salmonella* spp., *Klebsiella* sp., *Yersinia* spp., etc. (Janda and Abbott, 2021). They are characterized by sophisticated arrays of virulence factors and antimicrobial resistance (AMR) genes (Arafah, 2024).

Many members of this family; particularly those of the order Enterobacterales, are among the World Health Organization (WHO) priority list of pathogens, including those cases of carbapenem, third-generation cephalosporins resistance Enterobacterales critical group, and/ or high-risk fluoroquinolone-resistant *Salmonella typhi* group (Jesudason, 2024). Other genera such as *Enterobacter* spp. and *K. pneumoniae* are two main players in the ESKAPE pathogens, recognized by their multidrug resistance phenotypes (De Oliveira *et al*., 2020). In Egypt and other countries worldwide, multi drug-resistant Gram-negative enterobacteria are increasingly isolated from the community and clinical settings causing life-threatening infections (El-Kholy *et al*., 2020).

Some studies followed an epidemiological approach to help combat infections caused by *Enterobacteriaceae*. Epidemiological studies involve tracking pathogenic strains and variation within their genotypes and phenotypes across geographical locations and seasonal time points (Logan and Weinstein, 2017).

With the advent of sequencing technologies, genome assemblies, and annotation pipelines, a plethora of genomes from clinical bacterial isolates became available. The information obtained from the detailed analysis of these genomes can give valuable insights into the current status of AMR, its geographical distribution, its dissemination potential, and many other aspects of the human and environmental microbiome (Xiao *et al*., 2024). Noteworthy, discrimination between genes that are always encoded by a certain genus (core genome), those that may be encoded (dispensable), or gens that are strain-specific is an essential aspect of clinical isolates' genomics (Sherman and Salzberg, 2020). This pangenome analysis, when combined with full AMR profiling, gives useful information about the intrinsic AMR mechanisms versus the acquired ones that carry high potential of dissemination and useful insights about comparative resistome and mobilome (*i.e*. all mobile genetic elements encoded by a certain genome) (Anani *et al*., 2020).

Another interesting approach is the study of stress signaling pathways, their components, and molecular mechanisms enabling bacterial sensing, response, and final resistance to antibiotics, among other stressful environmental stimuli (Lazar *et al*., 2021). These studies enhance our understanding of the mechanisms of bacterial monitoring and adaptation to their environmental conditions, including antibiotics, physical, and chemical stressors (Bleul *et al*., 2022). A good example of these studies are the ones conducted on two component systems such as the regulator of colanic acid capsule synthesis (Rcs phosphorelay), which are present only and conserved within the members of the *Enterobacteriaceae* (Wall *et al*., 2018). This complex phosphorelay is formed of an outer membrane stress sensor (RcsF), an essential inner membrane negative regulator (IgaA), two inner membrane proteins, the histidine kinase (HK) and phosphotransfer protein RcsC and RcsD; respectively, and finally a cytosolic response regulator (RcsB) (Meng *et al*., 2021). The Rcs system is known to sense and respond to envelope threats including cell envelope-acting antibiotics such as β-lactams and

antimicrobial peptide (*i.e*., polymyxin) (Wall *et al*., 2018). Cell-envelope acting antimicrobials are a cornerstone of antibacterial therapy due to their wide activity spectrum, relative specificity and safety (Cho *et al*., 2014). This group includes β-lactams (penicillins and cephalosporins), glycopeptides (e.g., vancomycin), and penams (such as carbapenems and related antimicrobials). Additionally, Rcs over activation has been reported to increase *E. coli* resistance to β-lactams (Hirakawa *et al*., 2003 a, b). For convenience, most of these studies have used non-pathogenic laboratory strains as model microorganisms (Hussein *et al*., 2018).

Despite tremendous efforts, the insights we have from investigating the publicly available clinical isolates' genomes remain unsatisfactory compared to the large number of genomes available. In addition, only a limited number of studies succeeded to apply the research outcomes regarding stress signaling pathways obtained from laboratory strains to clinical isolates (Brannon *et al*., 2024), thus digging an important basic research-clinical application gap. The objective of this study was to investigate in detail 22 clinical isolates of three Gram-negative bacteria of the *Enterobacteriaceae* family; mainly *E. coli*, *S. enterica*, and *K. pneumoniae*. They are the causative agents of several infectious diseases such as gastroenteritis, typhoid fever, respiratory and urinary tract infections, and many of them are multi-drug resistant (Mazzariol *et al*., 2017). To achieve this objective, a comparative analysis of the pangenomes and resistomes of the chosen clinical isolates was performed. Moreover, a detailed analysis of the Rcs system in the studied clinical isolates was provided in an attempt to understand antimicrobial resistance profile against cell envelope-acting antibiotics within the analyzed isolates.

2. Materials and methods

2.1. Criteria of selection of clinical isolate's genomes

Using Bacterial and Viral Bioinformatics Resource Center (BV-BRC) (Olson *et al*., 2023), we looked for *Enterobacteriaceae* clinical isolates obtained from human subjects with different health status conditions. We focused our search on isolates belonging to *E. coli*, *K. pneumoniae*, and *S. enetrica* (Supplementary Table 1). We chose complete assembled genomes with good quality (CheckM contamination ≤ 0.2), which were collected between 2020 and 2024. In case of *E. coli* and *Klebsiella* sp., samples isolated from stool or gastrointestinal tract (GIT) were not included to exclude normal flora. We then focused our search on 7 *E. coli* genomes (GCA_034640615.1, GCA_026968085.1, GCA_023238365.1, GCA_022919055.1, GCA_028421485.1, GCA_030053855.1, GCA_034643955.1), 8 *K. pneumoniae* genomes (GCA_037023225.1, GCA_022982575.1, GCA_023093995.1, GCA_025118435.1, GCA_025118455.1, GCA_035201625.1, GCA_035658455.1, GCA_036599745.1), and 7 *S. enterica* genomes (GCA_032465755.1, GCA_032465795.1, GCF_025725745.1, GCA_032465775.1, GCA_026013315.1, GCA_032465835.1, GCA_026013295.1).

2.2. Pangenome analysis

Pangenome analysis was performed using pangenome analysis pipeline (Roary) (Page *et al*., 2015), which was available *via* the web-based tool Panexplorer (Dereeper *et al*., 2022). Based on the obtained results, we determined the percentage of core genome, dispensable genome, and strain-specific genome compared to the reference strain for each genus (*E. coli* GCA_000005845.2, *K. pneumoniae* GCA_000240185.2, and *S. enterica* GCA_000006945.2).

2.3. Determination of antibiotic resistome

The complete resistome of each chosen clinical isolate was determined using Resistance Gene Identifier (RGI) of Comprehensive Antibiotic Resistance database (CARD) (Alcock *et al*., 2023).

They were classified manually based on a group of antibacterials. The antimicrobial resistance profile (AMR) profile of the clinical isolates was also confirmed from the NCBI database.

2.4. Multiple sequence alignment of Rcs system components

Gene sequences of the Rcs system components *rcsF*, *igaA* (*yrfF*), *rcsC*, *rcsD*, and *rcsB* were retrieved from each genome of the selected clinical isolates. The EMBL-EBI Job Dispatcher sequence analysis tools were used for gene translation and multiple sequence alignment (Madeira *et al*., 2022). Gene sequences of *rcsF*, *igaA*, *rcsC*, *rcsD*, and *rcsB* from the studied isolates were translated into their corresponding peptide sequences using a Transeq tool. Each peptide sequence was then aligned with its homologues from selected genomes using a MUSCLE web tool. The clustalW alignment was then visualized using Jalview tool (Waterhouse *et al*., 2009). Percent similarity and divergence was calculated by submitting the multiple sequence alignment files into https://www.bioinformatics.org/msareveal/index2.html. Whenever needed, the number of amino acids substitutions that may lead to significant changes in the domain/ protein properties (*i.e*., acidic to basic/ small aliphatic to charged or aromatic) at specific positions was manually counted from the multiple sequence alignment, which were divided over the total number of amino acids in the domain/ protein and multiplied by 100 as inferred by Alzan *et al*., (2016).

2.5. Graphs, figures, and data analyses

Graphs were generated using MS Excel and Venn diagrams using https://bioinformatics.psb.ugent.be/webtools/Venn/. The software's for multiple sequence alignment generation and visualization were mentioned in the previous section.

3. Results

3.1. Pangenome analysis of *E. coli, S. enterica,* and *K. pneumoniae* clinical isolates

A number of 22 clinical isolates were chosen from the BV-BRC database, applying filters of completeness and quality of sequences/ assemblies (CheckM contamination ≤ 0.2). Collection was conducted between 2020 and 2024 and three species were isolated from patients mainly; *E. coli*, *S. enterica*, and *K. pneumoniae*. This led us to investigate 7 clinical isolates of *E. coli*, 7 of *S. enterica*, and 8 of *K. pneumoniae*. Next, the genomes of each genus were analysed using the pangenome analysis pipeline (Roary) available via the web-based tool Panexplorer. The analysis results revealed that *E. coli* harbored the lowest percentage of core genes (44 % versus 58 % in *S. enterica*, and 52 % in *K. pneumoniae*), while the highest percentage of strain-specific genes was 31 % versus 18 % in *S. enterica*, and 24 % in *K. pneumoniae*) (Fig. 1A). Since all three genera belong to the family *Enterobacteriaceae* and order Enterobacterales, we investigated whether they shared common genes in their core and dispensable genomes or not. Accordingly, the list of genes included in the core and dispensable genomes of all isolates were compared and results displayed that they shared 1631 genes in the core genomes and 143 in the dispensable genomes (Fig. 1B).

3.2. Antimicrobial resistance profiles of *E. coli, S. enterica,* and *K. pneumoniae*

Members of the *Enterobacteriaceae* family are known to encode a large repertoire of AMR genes. These genes affect the clinical choice of appropriate antibiotic(s) for therapy or prophylaxis. Additionally, they represent a real threat of AMR dissemination as some of their resistant determinants are acquired via horizontal gene transfer, carrying the potential of inter- and intra-species spread. Based on this, we searched for the AMR genes in genome of the selected clinical isolates using the resistance genes identifier (RGI) of the comprehensive antibiotic resistance database (CARD). Stringent parameters were applied during our search by choosing the resistance genes identifier (RGI) criteria of identifying only “perfect, strict and complete” genes.

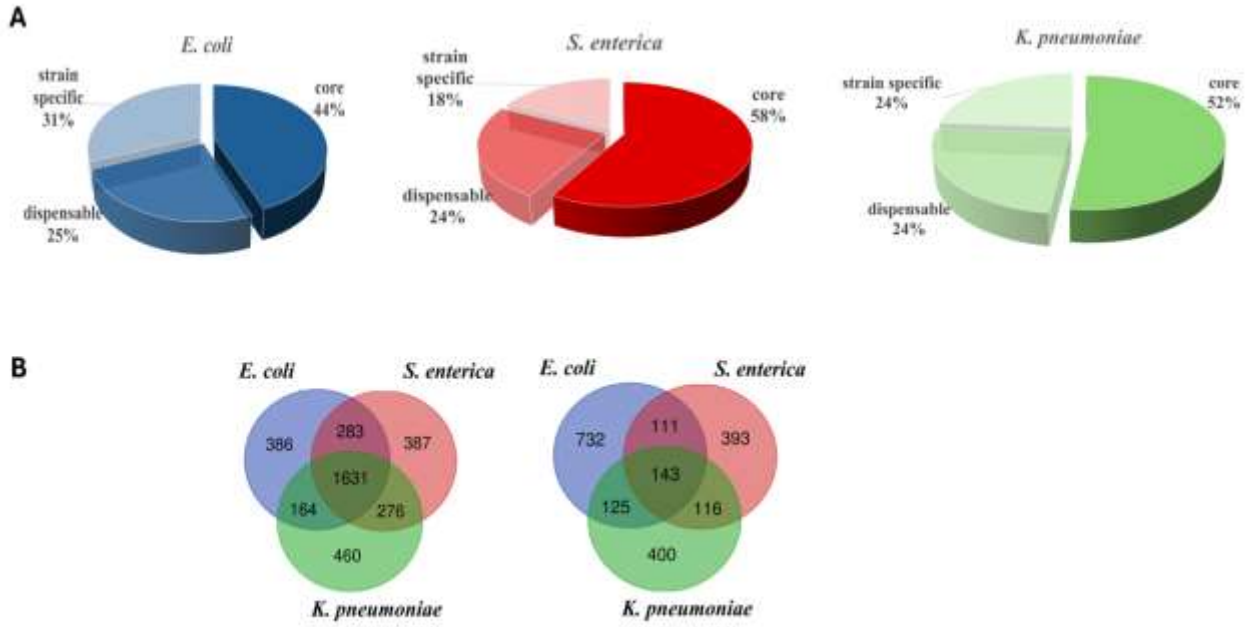

Fig. 1A, B: Pangenome analysis of clinical isolates of *E. coli*, *S. enterica*, and *K. pneumoniae*. A) Percentage of core, dispensable, and strain- specific genome in each of the three included genera, B) Number of shared genes in the core and dispensable/ strain- specific genome of each genus

Collectively, resistomes of the analyzed strains comprised resistance genes to aminoglycosides, fluroquinolones, cell envelope-acting antibiotics (*i.e*., β-lactams, glycopeptides, and carbapenams), sulfonamides, macrolides, tetracycline, and rifampin. As inferred from the AMR genes, resistance to these antibacterials relied on known mechanisms, including drug modification/ inactivation (for aminoglycosides, β- lactams, and macrolides), change of target binding site (e.g. rifampin, β- lactams, glycopeptides, and carbapenams), efflux pumps (e.g. macrolides and tetracycline), and use of alternative metabolic pathways (*i.e*., sulfonamides). Summing up the number of AMR genes in all clinical isolates per genus, we observed that 7 *E. coli* isolates harbored the most numerous AMR genes (99 genes), followed by 8 *K. pneumoniae* isolates (74 genes), and by 7 *S. enterica* isolates (62 genes). Results showed that there was a remarkable variation among the number of AMR genes encoded by each isolate within the same genus. For example, one *K. pneumoniae* isolate (GCA_025118455.1) encoded only 31 genes (42 %) of the resistome detected in all *K. pneumoniae* isolates, whereas another encoded 59 genes (79 %) of the same resistome (GCA_037023225.1). Similar patterns with narrower range were observed in *E. coli* and *S. enterica*.

On compiling the resistome of all isolates within the same genus, genes conferring resistance to more than three classes of antibiotics represented more than 10 % of the resistome genes (18.81 %, 10.78 % and 10.54 % in *E. coli*, *K. pneumonia*, and *S. enterica* isolates, respectively). As expected, most of them

encoded efflux pumps (*i.e*., *AcrAB-TolC, mdsA, KpnE, KpnF*…etc.) or general transcriptional regulators (*i.e.*, marA, soxR…etc.).

In addition, comparing different genera, there was a variation of the percentage of AMR to specific classes of antibiotics. For example, AMR genes against cell-envelope- acting antibiotics formed 18 %, 17 %, and 5 % of *E. coli*, *K. pneumonia*, and *S. enterica* resistome, respectively. This variation motivated us to investigate whether the AMR genes identified here, in particular genes conferring resistance to cell envelope-acting antibiotics, occurred in the core or accessory/ dispensable genome. Relying on the pangenome analysis obtained from the previous section, we found that in both *E. coli* and *S. enterica*, AMR genes were almost equally partitioned between core and dispensable/ strain-specific genomes, in contrast to *K. pneumoniae* where AMR genes existed in the dispensable/ strain- specific genome. Regarding genes encoding resistance to cell envelope-acting antibiotics, *S. enterica* isolates had almost equal distribution in both core and dispensable/strain-specific genes, in contrast to *E. coli* and *K. pneumonia*, where, surprisingly most of these genes were localized in the dispensable/ strain- specific genome.

This discrepancy in distribution of resistance determinants against β-lactams, carbapenams, and glycopeptides, raised the question if detecting these antibiotics in the surrounding milieu varied between different included genera. As Rcs system was the main signaling pathway responsible of sensing and responding to cell envelope threats, we decided to investigate the pangenome distribution of Rcs system components, the degree of conservation of its components, and their primary structures across the tested clinical isolates.

3.3. Conservation of Rcs system in *E. coli, S. enterica,* and *K. pneumoniae*

To this point, we thought to investigate if there are differences in the Rcs system among the tested genera. As previously outlined, the Rcs system is a two component signaling pathway conserved in *Enterobacteriaceae*. Rcs system senses envelope threats, including cell-envelope active antibiotics, and its overexpression was previously shown to increase resistance to β-lactams, as outlined in the introduction section. Rcs system is formed of 5 components. Thus, we aimed to investigate if differences in the sequences of Rcs system components correlate with the differences in genomic frequency and distribution of cell envelope-acting antibiotics resistance genes that we observed in the previous section. For all *E. coli*, *K. pneumonia*, and *S. enterica* isolates, Rcs system components (*i.e*., *rcsF, rcsC, rcsD, rcsB, rcsA, igaA*, and *yrfF*) were encoded by the core genome (Supplementary Table 2). We aligned the sequences of each component separately and examined the differences in alignment sequences. We focused on amino acid substitutions that were expected to alter physicochemical properties of the domain/ protein (e.g. substitution of basic amino acids by acidic ones or vice versa), or its conformation (e.g. substitution of small aliphatic amino acids by proline or by charged amino acids or vice versa).

Broadly, two features of the multiple sequence alignment were observed. First, RcsF, IgaA, RcsD, RcsC, and RcsB exhibited very low divergence within isolates of the same genus. Second, as expected, *S. enterica* Rcs components showed more similarity to *E. coli* homologues than to *K. pneumoniae* (Table 1). Alignment of RcsF, the outer membrane stress sensor of the Rcs system, showed a high degree of conservation (Table 1). Most substitutions (with few exceptions) included amino acids from the same chemical group and are thus unlikely to change the function or activity of RcsF as a stress sensor. Next, the repressor of Rcs system; *IgaA*, was examined. *IgaA* is an essential gene in *E. coli* and *S. enterica* that functions as a negative regulator of the Rcs system. It is formed of two N-terminal cytosolic domains and a single large C-terminal periplasmic domain that was embedded in the inner membrane by five trans-membrane domains.

Table 1. Percentage similarity among different homologues of Rcs system components compared to *E. coli* sequence

Protein	Domain	*E. coli*	*S. enterica*	*K. pneumoniae*	% Variation
RcsF		100 %	91 %	83 %	
IgaA	Full length	100 %	84 %	72 %	
	First cytosolic	100 %	80 %	64 %	12 %
	Second cytosolic	100 %	92 %	87 %	3 %
	Periplasmic	100 %	83 %	77 %	8 %
RcsD	Full length	100 %	84 %	72 %	
	Periplasmic	100 %	85 %	71 %	8 %
	Cytosolic	100 %	84 %	72 %	10 %
RcsC	Full length	100 %	89 %	80 %	
	Periplasmic	100 %	86 %	75 %	6 %
	Cytosolic	100 %	90 %	83 %	5 %
RcsB		100 %	99 %	96 %	

IgaA first cytosolic domain was reported to play a main repressor role, while IgaA periplasmic domain interacts with RcsF and RcsD. *S. enterica* and *K. pneumoniae* showed 84 % and 72 % similarity to *E. coli* IgaA, respectively (Table 1). When each of IgaA soluble domains was aligned (cytosolic or periplasmic) separately, a high degree of conservation of the periplasmic and second cytosolic domain was observed; with relatively fewer substitutions that were expected to change the conformation or the physicochemical character of the protein (Table 1). In contrast, IgaA first cytosolic domains contained many non-conservative substitutions. *K. pneumoniae* IgaA first cytosolic domain was only 64 % similar to *E. coli* IgaA, with important substitutions expected to change the domain properties (around 12 %) (Table 1).

Changes from the consensus were more numerous in *K. pneumoniae* isolates than *S. enterica*, suggesting a possible difference in regulation and/or basal level of activation of the Rcs system.

Because the interaction of RcsF and IgaA represents the first step towards envelope stress sensing and based on our results of high conservation of RcsF and variability in IgaA first cytosolic domain, we thought to investigate the other partners of IgaA: RcsD. RcsD is the phospho transfer protein of the Rcs system mediating the transfer between histidine kinase RcsC and the response regulator, *i.e*., RcsB. Meanwhile, RcsD periplasmic domain was shown to interact with IgaA periplasmic domain. Corroborating with previous alignment results, RcsD periplasmic domain of *K. pneumonia* showed less similarity to *E. coli* than S. *enterica* (Table 1). Although many substitutions were present, only a few could be expected to alter the activity/ function of RcsD (approximately 8 % of the substitutions).

RcsD cytosolic domain was also conserved with only 9 % significant substitutions. However, the sequence fragment extending from amino acid residues 467 to 711, showed a high degree of variability. Interestingly, this region, according to the uniport database, included the pseudokinase domain of RcsD. Meanwhile, the multiple sequence alignment of RcsC showed comparable results in terms of sequence conservation and similarity % to *E. coli* (Table 1). RcsB, on the other hand, displayed very little variability and was almost completely conserved.

Current results showed that although the Rcs system was conserved in all the 22 isolates, important sequence diversity occurred in some domains; particularly IgaA first cytosolic and RcsD pseudokinase domains. This variability might cause a different regulatory level of the Rcs system that altered the resistance to the cell envelope-acting antibiotics. This resistance was predicted to be intrinsic as all Rcs components were present in the core genome.

4. Discussion

Antimicrobial resistance represents one of the global health threats humanity is facing. The qualitative and quantitative increase of multidrug resistant microorganisms urges the scientific community to implement strategies and find innovative solutions to this problem. The primary step is to have a clear picture of the AMR problem, particularly in the clinical settings. This picture will include identifying resistant microorganisms from patients, genome sequencing and assembly, identifying the resistome, assessing its dissemination potential, and finally understanding the molecular pathways regulating bacterial sensing and their response to the presence of antibiotics (Strateva and Peykov, 2024). The currently available databases are rich in sequences and assemblies of genomes obtained from clinical isolates. However, investigation of their full AMR profiles and the molecular mechanisms contributing and/ or regulating AMR is still insufficient.

In this study, 22 clinical isolates of three enterobacterial genera were selected; mainly *E. coli*, *K. pneumonia*, and *S. enterica* isolated from the clinical settings between 2020-2024. Several criteria were applied to the current sequence selection as described earlier in terms of sequence quality and genome completeness. First, the percentage of core, accessory, and strain-specific genes was determined. The usefulness of pangenome analysis in the clinical setting was outlined previously as an important step to characterize the resistome and mobilome of the different strains (Anani *et al*., 2020). In addition, combining pangenome analysis and immunoinformatic approach led to identification of several candidates antigenic proteins from *Klebsiella* sp. for vaccine development (Wang *et al*., 2022).

For pangeome analysis, the web based tool panexplorer was used by means of the Roary tool as described earlier. The Roary tool determines core and accessory genomes by converting coding DNA sequences to protein sequences and clustering them

before comparing them using BLASP. The outputs are clusters assigned to either core or accessory or strain-specific genomes (Page *et al*., 2015).

Compared to *S. enterica* and *K. pneumoniae*, *E. coli* isolates showed the smallest set of core genes and the highest percentage of dispensable and strain-specific genes. In addition, it carried collectively larger number of AMR genes. This may be attributed to high prevalence of *E. coli* as a commensal or as a pathogen, enabling it to acquire more resistant traits as adaptive mechanisms to repeated exposure to environmental stressors, including antibiotics (Nji *et al*., 2021). Our study revealed the existence of 3124 core genes and 1750 dispensable genes in *E. coli*, in contrast to a previous study showing an *E. coli* genome size of 5020 (+/- 446 genes), including a core genome of 2344 +/- 43 genes (Rasko *et al*., 2008).

The percentage of cell envelope-acting antibiotics resistance genes present in the core genomes of the tested isolates represented 15 %, 18 %, and 30 % in *E. coli*, *S. enterica*, and *K. pneumoniae*, respectively. As core genes are shared among all strains of a certain species (Sherman and Salzberg, 2020), this could suggest the presence of intrinsic resistant mechanisms to this class of antibacterials in the enterobacterial species under investigation. Interestingly, *S. enterica* clinical isolates had approximately half of cell envelope-acting antibiotics resistance genes encoded by the core genome, although they had an overall smaller resistome. This could be attributed to the largest core genome of *S. enterica* genera (as depicted in Fig. 1A) or to a difference in sensing and responding to cell envelope threats, leading to differences in response to cell envelope-acting antibiotics (Rodríguez *et al*., 2023).

There was also a noticeable difference between the numbers of resistance genes in each isolate; however, whether this difference could be attributed to specific traits of the isolates or to particular patient characteristics is not clear yet. For example, *E. coli* isolate GCA_023238365.1 had larger number of resistance genes compared to GCA_028421485.1, although both were isolated from the same geographical location (Supplementary Table 1). However, the first was isolated from blood while the second from urine of a urinary tract infected patient, thus raising the possibility of different exposure to antibiotics or different extracellular milieu, leading to a different AMR profile. Difference in the AMR profile according to the infection site was observed in a previous study reported by Regassa *et al*., (2023).

Next, sequence analysis of the Rcs system component was conducted to assess the diversity/ conservation between the tested isolates. The importance of Rcs system as one of three signaling pathways in enterobacteria sensing cell envelope challenges was described previously in details (Wall *et al*., 2018). Our reasons to study this system in the 22 isolates stemmed from the discrepancy in percentage and genome partitioning of cell envelope-acting antibiotics resistant genes, and previous findings unrevealing an incomplete Rcs system and/ or truncated variants of its proteins in some enterobacterial species (Stella *et al*., 2018). Current analysis revealed that the 22 isolates had all the 5 components of the Rcs system present in their core genomes. Although highly conserved, two main domains showed remarkable variability; mainly IgaA first cytosolic and RcsD pseudohistidne domains.

IgaA first cytosolic domain was previously shown to play a major role of repressing the Rcs system, in contrast to the second cytosolic loop (Hussein *et al*., 2018). Recently, Rodríguez *et al*., (2023) outlined an interaction between 2 β- strands of the first and second cytosolic domains, which was predicted to be essential for phenotypic complementation in *Salmonella* spp. Their analysis also suggested differences in the Rcs system; particularly IgaA related to the life style of Enterobacteriales.

In this study, our analysis revealed that *K. pneumoniae* showed a highly variable IgaA first cytosolic domain; in particular the region spanned between H192 to P249, which was shown previously to be important for the interaction among the two

cytosolic domains and Rcs system repression (Rodríguez *et al*., 2023). The same region; however, was conserved in *E. coli* and *S. enterica*. IgaA first cytosolic domain was previously shown to play a major role of repressing the Rcs system, in contrast to the second cytosolic loop (Hussein *et al*., 2018). Accordingly, we reasoned that in *K. pneumoniae*, the Rcs system might exhibit different levels of basal activity and/ or repression compared to *S. enterica* and *E. coli*, due to the difference in IgaA first cytosolic domain sequence. The observed high resemblances between *E. coli* and *S. enterica* could not be attributed to their evolutionary closeness only as both RcsB and RcsF in the three genera (including *Klebsiella* sp.) were highly conserved regardless of their evolutionary distance. Thus, the differences between RcsD and IgaA in *K. pneumoniae* from one side and *S. enterica* and *E. coli* from the other side could be attributed to other factors than the evolutionary distance, including AMR genes presence/ absence, partitioning between the core genomes, and the life style and ecological niche (Rodríguez *et al*., 2023).

Interestingly, the core genes against the cell envelope acting antibiotics represented 30 % of the AMR genes in *Klebsiella* sp., 18 % in *S. enterica* and 15 % in *E. coli*, raising an important possibility that an altered repression of the Rcs system can be compensated by an increase of the core resistance gene. Whether this possibility is just a tempting hypothesis or a plausible explanation remains to be tested. In parallel, *E. coli* and *S. enterica* tended to have similar and more conserved Rcs system, raising the possibility of an inter-correlation between signaling pathways and core genome resistome.

Conclusion

The results obtained in this study provided a comparative overview of the pangenome analysis of 22 clinical isolates of *E. coli*, *S. enterica*, and *K. pneumoniae*, outlining differences in the sets of core and dispensable/ strain-specific genes. Partitioning of AMR genes between core and dispensable genomes and the total number of genes against specific classes of antibiotics varied widely among the three genera. In parallel, important sequence diversity was observed between the two key domains in Rcs signaling pathway. Altogether, these differences may indicate distinct adaptation strategies of the investigated clinical isolates to antibiotic exposure and ecological niche changes.

Acknowledgement

The authors would like to acknowledge Dr\ Heba Alzan for critical reading of the manuscript and providing very helpful comments and Dr\ Naiera Helmy for encouragement, scientific and logistic help.

Conflict of interests

The authors declare that no competing or conflicts of interest exist.

Funding

This work was supported by Grant number 43867 from the Sciences and Technology Development Fund (STDF).

Ethical approval

None applicable.

Author's Contributions

S.S: Methodology, Analysis, Draft Writing and Revision. NH: Fund acquisition, Conceptualization, Methodology, Analysis, Draft Writing and Revision.

5. References

Alcock, B.P.; Huynh, W.; Chalil; R.; Smith, K.W.; Raphenya, A.R.; Wlodarski; M.A. et al. (2023). CARD 2023: expanded curation : support for machine learning : and resistome prediction at the Comprehensive Antibiotic Resistance Database. Nucleic Acids Research. 51(1) : 690-699. https://doi.org/10.1093/nar/gkac920.

Alzan, H.F.; Lau, A.O.T.; Knowles, D.P.; Herndon, D.R.; Ueti, M.W.; Scoles, G.A. et al. (2016).

Expression of 6-Cys gene superfamily defines *Babesia bovis* sexual stage development within *Rhipicephalus microplus*. PLoS ONE. 11(9): e0163791. https://doi.org/10.1371/journal.pone.0163791.

Anani, H.; Zgheib, R.; Hasni, I.; Raoult, D. and Fournier, P.E. (2020). Interest of bacterial pangenome analyses in clinical microbiology. Microbial Pathogenesis. 149: 104275. https://doi.org/10.1016/j.micpath.2020.104275.

Arafah, A.A. (2024). Antibiotic resistance trends among *Enterobacteriaceae* in Saudi Arabia: A Systematic Review. Cureus. 16(3): e56614. https://doi.org/doi:10.7759/cureus.56614.

Bleul, L.; Francois, P. and Wol, C. (2022). Two-component systems of *S. aureus*: Signaling and sensing mechanisms. Genes. 13(1): 34. https://doi.org/10.3390/genes13010034.

Brannon, J.R.; Reasoner, S.A.; Bermudez, T.A.; Comer, S.L.; Wiebe, M.A.; Dunigan, T.L. et al. (2024). Mapping niche-specific two-component system requirements in uropathogenic *Escherichia coli*. Microbiology Spectrum. 12(4): e0223623. https://doi.org/10.1128/spectrum.02236-23.

Cho, S.H.; Szewczyk, J.; Pesavento, C.; Zietek; M.; Banzhaf, M.; Roszczenko, P. et al. (2014). Detecting envelope stress by monitoring β-barrel assembly. Cell. 159(7) : 1652-1664. https://doi.org/10.1016/j.cell.2014.11.045.

De Oliveira, D.M.P.; Forde, B.M.; Kidd, T.J.; Harris, P.N.A.; Schembri, M.A.; Beatson, S.A. et al. (2020). Antimicrobial Resistance in ESKAPE Pathogens. Clinical Microbiology Review. 33(3): e00181-19. https://doi.org/10.1128/CMR.

Dereeper, A.; Summo, M. and Meyer, D.F. (2022). PanExplorer: a web-based tool for exploratory analysis and visualization of bacterial pan-genomes. Bioinformatics. 38(18): 4412-4414. https://doi.org/10.1093/bioinformatics/btac504.

El-Kholy, A.A.; Girgis, S.A.; Shetta, M.A.F.; Abdel-Hamid, D.H. and Elmanakhly, A.R. (2020). Molecular characterization of multidrug-resistant Gram-negative pathogens in three tertiary hospitals in Cairo, Egypt. European Journal of Clinical Microbiology and Infectious Diseases. 39: 987-992. https://doi.org/10.1007/s10096-020-03812-z.

Hirakawa, H.; Nishino, K.; Hirata, T. and Yamaguchi, A. (2003a). Comprehensive studies of drug resistance mediated by overexpression of response regulators of two-component signal transduction systems in *Escherichia coli*. Journal of Bacteriology; 185(6) : 1851-1856. https://doi.org/10.1128/JB.185.6.1851-1856.2003.

Hirakawa, H.; Nishino, K.; Yamada, J.; Hirata, T. and Yamaguchi, A. (2003b). β-Lactam resistance modulated by the overexpression of response regulators of two-component signal transduction systems in *Escherichia coli*. Journal of Antimicrobial Chemotherapy. 52(4): 576-582. https://doi.org/10.1093/jac/dkg406.

Hussein, N.; Cho, S.H.; Laloux, G.; Siam, R. and Collet; J.F. (2018). Distinct domains of *Escherichia coli* IgaA connect envelope stress sensing and down-regulation of the Rcs phosphorelay across subcellular compartments. PLoS Genetics. 14(5): e1007398. https://doi.org/10.1371/journal.pgen.1007398.

Janda, J.M. and Abbott, S.L. (2021). The changing face of the family *Enterobacteriaceae* (Order: Enterobacterales): New members; taxonomic issues; geographic expansion; and new diseases and disease syndromes. Clinical Microbiology Reviews. 34(2): e00174-20. https://doi.org/10.1128/CMR.00174-20.

Jesudason, T. (2024). WHO publishes updated list of bacterial priority pathogens. The Lancet Microbe. 5(9): e100940. https://doi.org/10.1016/j.lanmic.2024.07.003.

Lazar, V.; Holban, A.M.; Curutiu, C. and Chifiriuc, M.C. (2021). Modulation of Quorum Sensing and Biofilms in Less Investigated Gram-

Negative ESKAPE Pathogens. Frontiers in Microbiology. 12: 676510. https://doi.org/10.3389/fmicb.2021.676510.

Logan, L.K. and Weinstein, R.A. (2017). The epidemiology of Carbapenem-resistant *Enterobacteriaceae*: The impact and evolution of a global menace. Journal of Infectious Diseases. 215(1): 28–36. https://doi.org/10.1093/infdis/jiw282.

Madeira, F.; Pearce, M.; Tivey, A.R.N.; Basutkar, P.; Lee, J.; Edbali, O. et al. (2022). Search and sequence analysis tools services from EMBL-EBI in 2022. Nucleic Acids Research. 50 (1): 276-279. https://doi.org/10.1093/nar/gkac240.

Mazzariol, A.; Bazaj, A. and Cornaglia, G. (2017). Multi-drug-resistant Gram-negative bacteria causing urinary tract infections: a review. Journal of Chemotherapy. 29(1): 2-9. https://doi.org/10.1080/1120009X.2017.1380395.

Meng, J.; Young, G. and Chen, J. (2021). The Rcs System in *Enterobacteriaceae*: Envelope Stress Responses and Virulence Regulation. Frontiers in Microbiology. 12: 627104-627110. https://doi.org/10.3389/fmicb.2021.627104.

Nji, E.; Kazibwe, J.; Hambridge, T.; Joko; C.A.; Larbi; A.A.; Damptey; L.A.O. et al. (2021). High prevalence of antibiotic resistance in commensal *Escherichia coli* from healthy human sources in community settings. Scientific Reports. 11(1): 3372-3377. https://doi.org/10.1038/s41598-021-82693-4.

Olson, R.D.; Assaf, R.; Brettin, T.; Conrad, N.; Cucinell, C.; Davis, J.J. et al. (2023). Introducing the Bacterial and Viral Bioinformatics Resource Center (BV-BRC): a resource combining PATRIC; IRD and ViPR. Nucleic Acids Research. 51(D1): D678-D689. https://doi.org/10.1093/nar/gkac1003.

Page, A.J.; Cummins, C.A.; Hunt, M.; Wong, V.K.; Reuter, S.; Holden, M.T.G. et al. (2015). Roary: Rapid large-scale prokaryote pan genome analysis. Bioinformatics. 31(22): 3691-3703. https://doi.org/10.1093/bioinformatics/btv421.

Rasko, D.A.; Rosovitz, M.J.; Myers, G.S.A.; Mongodin, E.F.; Fricke, W.F.; Gajer, P. et al. (2008). The pangenome structure of *Escherichia coli*: Comparative genomic analysis of *E. coli* commensal and pathogenic isolates. Journal of Bacteriology. 190(20): 619-628. https://doi.org/10.1128/JB.00619-08.

Regassa, B.T.; Tosisa, W.; Eshetu, D.; Beyene, D.; Abdeta, A.; Assefa, A. et al. (2023). Antimicrobial resistance profiles of bacterial isolates from clinical specimens referred to Ethiopian Public Health Institute: analysis of 5-year data. BMC Infectious Diseases. 23(1) : 798-803. https://doi.org/10.1186/s12879-023-08803-x.

Rodríguez, L.; Peñalver, M.; Casino, P. and García-Del Portillo, F. (2023). Evolutionary analysis and structure modelling of the Rcs-repressor IgaA unveil a functional role of two cytoplasmic small β-barrel (SBB) domains. Heliyon. 9(6): e16661. http://doi.org/10.1016/j.heliyon.2023.e16661.

Sherman, R.M. and Salzberg, S.L. (2020). Pan-genomics in the human genome era. Nature Reviews Genetics. 21(4): 243-254. http://doi.org/doi: 10.1038/s41576-020-0210-7.

Stella, N.A.; Brothers, K.M.; Callaghan, J.D.; Passerini, A.M.; Sigindere, C.; Hill, P.J. et al. (2018). An IgaA/UmoB Family Protein from *Serratia marcescens* Regulates Motility; Capsular Polysaccharide Biosynthesis; and Secondary Metabolite Production. Applied and Environmental Microbiology. 84(6) : 1-14. https://doi.org/10.1128/AEM.02575-17.

Strateva, T. and Peykov, S. (2024). First detection of a cefiderocol-resistant and extensively drug-resistant *Acinetobacter baumannii* clinical isolate in Bulgaria. Acta Microbiologica Immunologica Hungarica. 71(1): 25-36. https://doi.org/10.1556/030.2024.02201.

Wall, E.; Majdalani, N. and Gottesman, S. (2018). The complex Rcs regulatory cascade. Annual Review of Microbiology. 72 : 111–139. http://doi.org/doi: 10.1146/annurev-micro-090817-062640.

Wang, Z.; Guo, G.; Li, Q.; Li, P.; Li, M.; Zhou, L. et al. (2022). Combining Immunoinformatics with Pangenome Analysis To Design a Multiepitope Subunit Vaccine against *Klebsiella pneumoniae* K1; K2; K47; and K64. Microbiology Spectrum. 10(4). e01148-22. https://doi.org/10.1128/spectrum.01148-22,

Waterhouse, A.M.; Procter, J.B.; Martin, D.M.A.; Clamp, M. and Barton, G.J. (2009). Jalview Version 2-A multiple sequence alignment editor and analysis workbench. Bioinformatics. 25(9) : 1189-1191. https://doi.org/10.1093/bioinformatics/btp033.

Xiao, Y.H.; Luo, Z.X.; Wu, H.W.; Xu, D.R. and Zhao, R. (2024). Metagenomic next-generation sequencing for the identification of infections caused by Gram-negative pathogens and the prediction of antimicrobial resistance. Laboratory Medicine. 55(1) : 71-79. https://doi.org/10.1093/labmed/lmad039**.**