PNAS

Supporting Information for-
Antimicrobial resistance heterogeneity among multidrug-resistant Gram-negative pathogens: Phenotypic, Genotypic and Proteomic analysis

Tanshi Mehrotra[1#], Dipasri Konar[2#], Agila Kumari Pragasam[1#], Shakti Kumar[1], Pradipta Jana[1], Prabhakar Babele[1], Deepjyoti Paul[1], Ayushi Purohit[1], Subhash Tanwar[3], Susmita Bakshi[1], Santanu Das[1], Jyoti Verma[1], Daizee Talukdar[1], Lekshmi Narendrakumar[1], Akanksha Kothidar[1], Sonali Porey Karmakar[1], Susmita Chaudhuri[3], Sujoy Pal[4], Kajal Jain[4], Chittur V Srikanth[5], M. Jeeva Sankar[6], Krishnamohan Atmakuri[1], Ramesh Agarwal[6], Rajni Gaind[7], Mamatha Ballal[8], Nagamani Kammili[9], Rupak K Bhadra[10], Thandavarayan Ramamurthy[1,11], G. Balakrish Nair[1,12@] Bhabatosh Das[1@]

1-Infection and Immunology Division, Translational Health Science and Technology Institute, Faridabad, India

2-Jan Swasthya Sahyog, Ganiyari, Bilaspur, India

3-Multidisciplinary Clinical and Translational Research, Translational Health Science and Technology Institute, Faridabad, India

4-Department of Gastroenterology and Human Nutrition, All India Institute of Medical Sciences, New Delhi, India

5-Laboratory of Gut Infection and Inflammation Biology, Regional Centre for Biotechnology, Faridabad, India

6-Department of Pediatrics, All India Institute of Medical Sciences, New Delhi, India

7-Department of Microbiology, Vardhman Mahavir Medical College and Safdarjung Hospital, New Delhi, India

8-Department of Microbiology, Kasturba Medical College, Manipal Academy of Higher Education, Manipal, Karnataka, India

9-Department of Microbiology, Gandhi Medical College and Hospital, Secunderabad, India

10-Infectious Diseases and Immunology Division, CSIR-Indian Institute of Chemical Biology, Kolkata-700 032, India

11-Division of Bacteriology, ICMR-National Institute of Cholera and Enteric Diseases, Kolkata

12- Rajiv Gandhi Centre for Biotechnology, Thiruvananthapuram, Kerala 695014, India

@ **Corresponding author:**

***Bhabatosh Das*:** *Functional Genomics Laboratory, Centre for Antimicrobial Resistance Research, Infection and Immunology Division, Translational Health Science and Technology Institute, NCR Biotech Science Cluster, 3rd Milestone, Faridabad – Gurgaon Expressway, PO box 04, Faridabad – 121001, Haryana, India.* Phone: +91-129-2876471; Fax: 0129-2876500; *Email: bhabatosh@thsti.res.in*

G. Balakrish Nair: *Rajiv Gandhi Centre for Biotechnology, Thiruvananthapuram, Kerala 695014, India. E-mail: gbnair_2000@yahoo.com Mob: +919999829379*

1. Supplementary Materials and Methods:

1.1 Patient recruitment and sample collection

Samples were collected from 5 different participating sites these are All India Institute of Medical Sciences (New Delhi), Safdarjung Hospital (New Delhi), Jan Swasthya Sahyog (Chattisgarh), Gandhi Medical College (Telangana) and Manipal Academy of Higher Education (Karnataka) mapping Northern, Central and Southern regions of India. **(Dataset S1)**.

1.2 Phenotypic AST and DNA extraction

Antimicrobial susceptibility testing included seventeen antibiotics these are penicillin (10IU), ampicillin (10µg), carbenicillin (100µg), imipenem (10µg), streptomycin (10µg), spectinomycin (100µg), chloramphenicol (30µg), tetracycline (5µg), kanamycin (30µg), neomycin (30µg), erythromycin (15µg), doxycycline (30µg), SXT (23.75µg+ 1.25µg), polymyxin B (300IU), nalidixic acid (30µg), ciprofloxacin (5µg) and rifampin (5µg/25µg). DNA was extracted using the GenElute bacterial genomic kit (Sigma Aldrich). DNA quality was ensured by 260/280 and 260/230 ratios through Nanodrop and AGE (Agarose Gel Electrophoresis). Only pure DNA with no RNA contamination was selected for downstream processing.

1.3 Genome analysis- Whole Genome Sequencing, Raw reads processing, Assembly, Quality assessment, and Annotation

For sequencing 1ng input DNA was used for paired-end library preparation with read chemistry of 2 × 300 bp. After sequencing quality was checked for the raw reads generated using trimmomatic software version-0.39 (1), samples having QC value < 20 and read length <36 bp were trimmed followed by checkm.

1.4 MLST based typing

MLST profiles for *E. coli* were analyzed using the Achtman scheme that includes *adk*, *fumC*, *gyrB*, *icd*, *mdh*, *purA* and *recA*. For *K. pneumoniae* and *P. aeruginosa,* the PubMLST scheme has been followed, which used *gapA*, *infB*, *mdh*, *pgi*, *phoE*, *rpoB,* and *tonB* genes; and *acsA*, *aroE*, *guaA*, *mutL*, *nuoD*, *ppsA,* and *trpE* respectively. While for *A. baumannii, cpn60*, *fusA*, *gltA*, *pyrG*, *recA*, *rplB,* and *rpoB* genes were used as per the Pasteur scheme (2).

1.5 Global phylogenomic framework

After the the downloaded genomes were first screened for the quality check by checkM tool, further, these large data sets were trimmed down using dereplicator tool version v0.1.0 (https://github.com/rrwick/Assembly-Dereplicator). For *K. pneumoniae, P. aeruginosa,* and *A. baumannii, a* reference genome-based mapping approach using Snippy v.4.6.0 (https://github.com/tseemann/snippy, Seeman 2015), has been followed to capture the SNPs (3), which was then used to generate the phylogenetic tree (4). Whereas, for *E.coli,* core genome based approach was followed by generating core genome alignment using panaroo (5), followed by SNPs using snp-sites (3), and phylogenetic tree construction using GTR-GAMMA based maximum likelihood (ML) tree by RAxML (4). The total number of genomes included for these four pathogens varies with respect to the available resources at NCBI and as well based on the inclusion criteria as stated above.

1.6 Screening of antimicrobial-resistant determinants, mobile genetic elements, IS elements, and virulence factors

For *E. coli* all the genomes have been identified as respective phylogroups based on an *in-silico* phylotyping tool using Clermontyping (6).

1.7 LC-MS/MS based proteome profiling and data analysis

Approximately 1ml culture (OD_{600} >0.8) of *K. pneumoniae, A. baumannii, E. coli,* and *P. aeruginosa* were harvested and lysed with 100µl 8M urea, 2% CHAPS, 50mM Tris, pH 8.0, and 1% protease inhibitor cocktail followed by sonication at 40% amplitude. Protein concentration in the supernatant was determined using BCA according to the manufacturer protocol and subjected to in-solution digestion using trypsin. The resulting mixture was desalted after that 10µg peptide samples were first loaded on a trap column and then separated on a reverse phase column using a 90 min gradient of buffer A (water and 0.1% formic acid) and buffer B (acetonitrile and 0.1% formic acid): initial 5% B for 5 min, followed by a linear gradient to 35% B in 70 min and 50% B in 80 min followed by peptide elution window for 7 min at 90% B. Protein Pilot was used for library generation: a proteome database was constructed using protein sequences from UniProt reference proteomes of *K. pneumonia, A. baumannii, E. coli,* and *P. aeruginosa*. A reverse database search strategy was adopted to estimate the FDR (1%) for peptide and protein identification. The proteins and associated peptides were filtered by PeakView and MarkerView for relative quantitative analysis. For each sample 3 technical replicates were run and analysis of relative quantitation

was performed by t-test to determine if the mean for each group was significantly different. The proteins showing log2 (fold change) and 1.5 fold (increase or decrease) with respect to control ($p \leqslant .05$) were considered to exhibit altered abundance.

2.0 Supplementary Figures

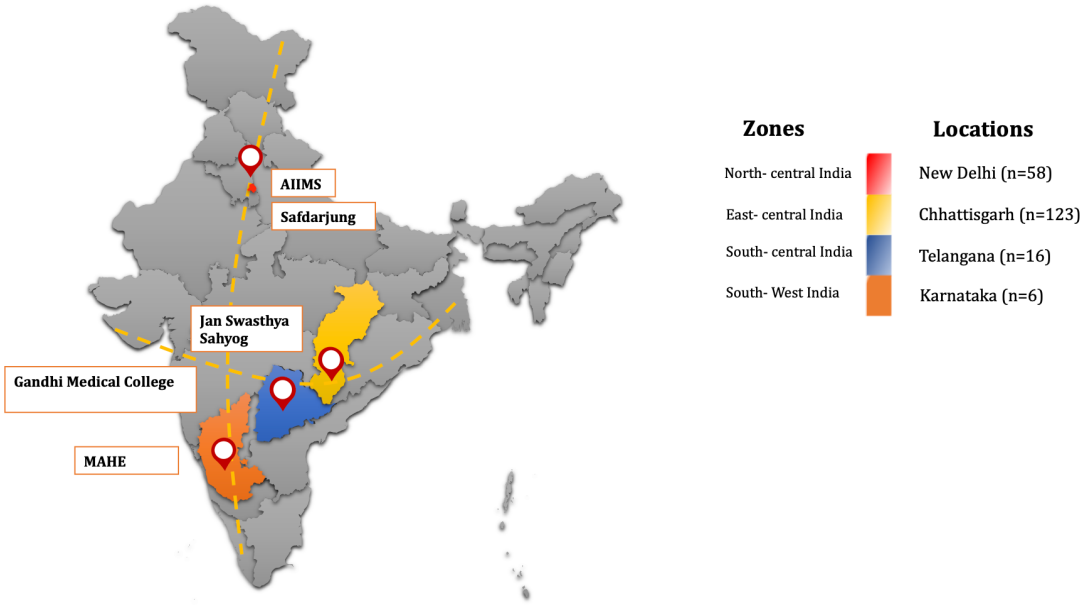

Fig. S1- Showing different study sites in India along with zones, locations, and number of samples collected from each site.

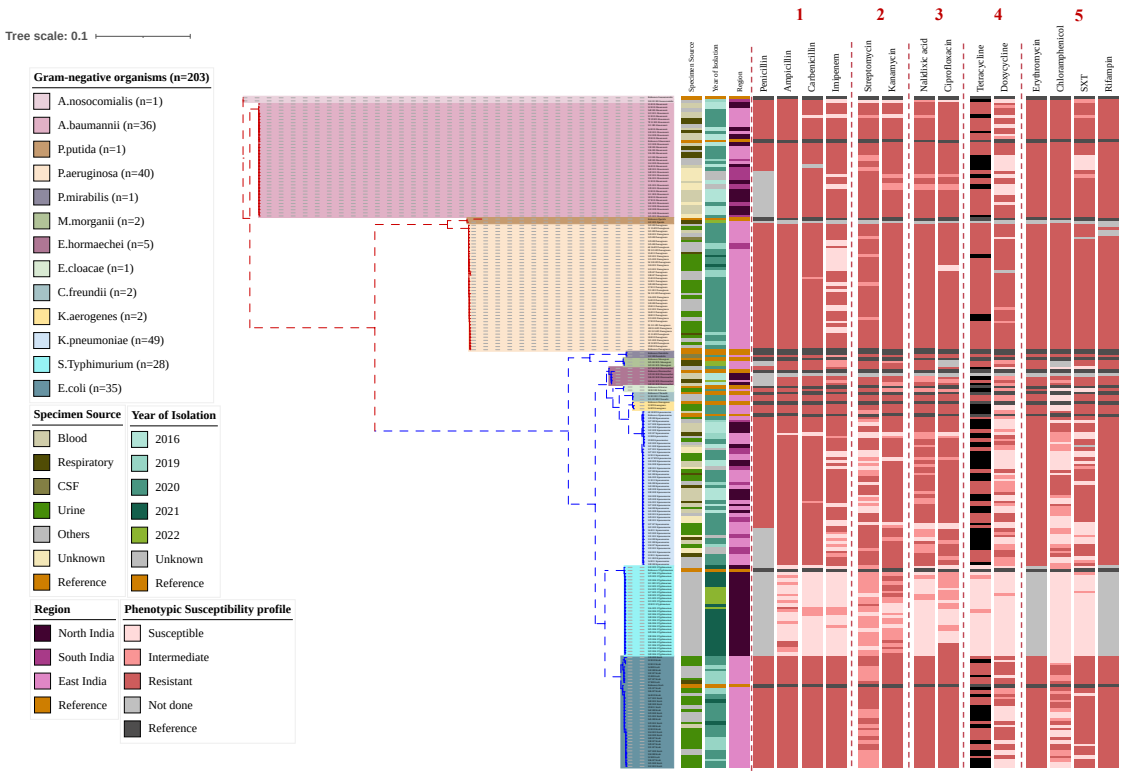

Fig. S2- Phenotypic antimicrobial susceptibility profile overlaid on the maximum likelihood phylogeny based on 16S, 23S, and 5S rRNA sequences. Tree scale indicates number of substitutions per genome per site. Specimen type, year of isolation and region is mentioned along with antimicrobials as colour strips. Antimicrobials are divided into five major classes 1- Beta-lactams, 2- Aminoglycosides, 3-Fluoroquinolones, 4-Tetracyclines, 5-Others.

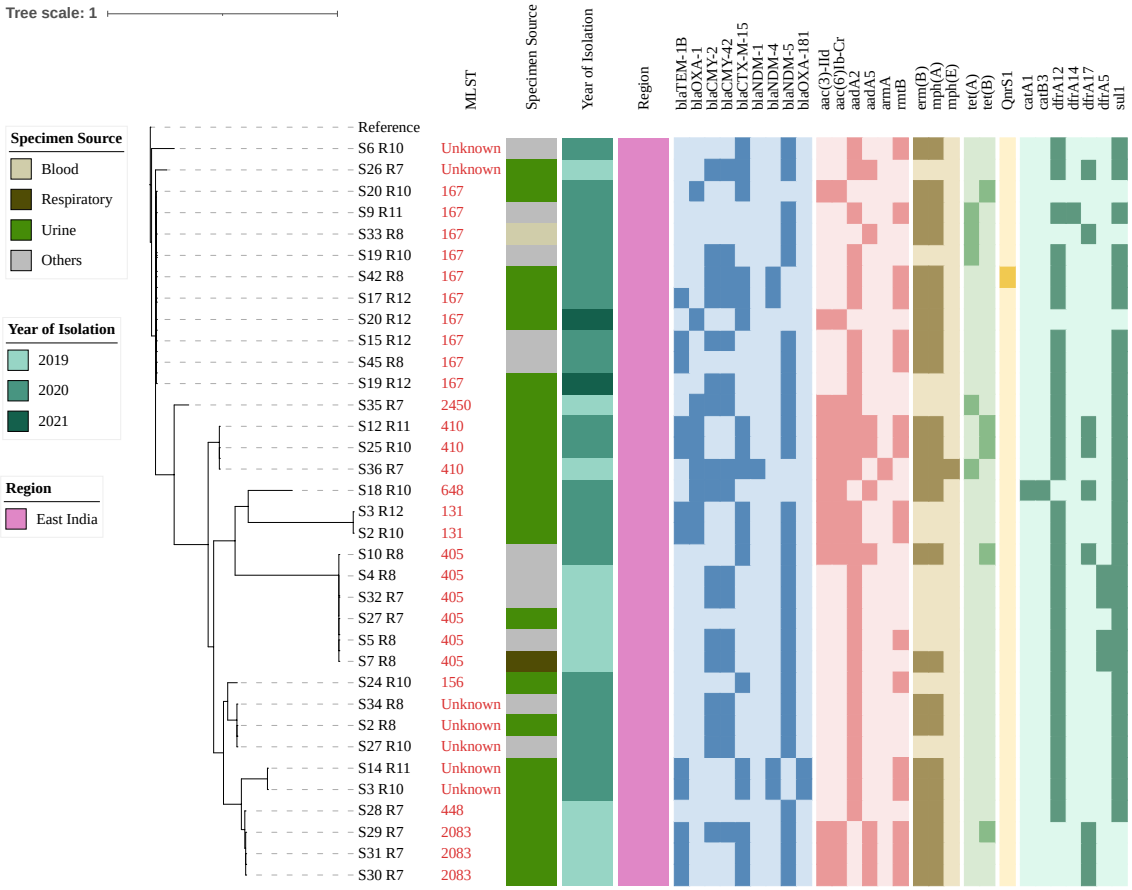

Fig. S3- Maximum likelihood phylogenetic tree based on the nucleotide sequences of the 7 housekeeping genes for *E.coli* (*adk*, *fumC*, *gyrB*, *icd*, *mdh*, *purA* and *recA*). Metadata displaying the MLST STs, specimen type, and year of isolation labelled as a colour strip. The presence and absence of antimicrobial-resistant genes against different classes of antimicrobial agents categorized as beta-lactams, aminoglycosides, macrolides, tetracycline, and others are labelled as heat maps. Tree scale indicates number of substitutions per genome per site.

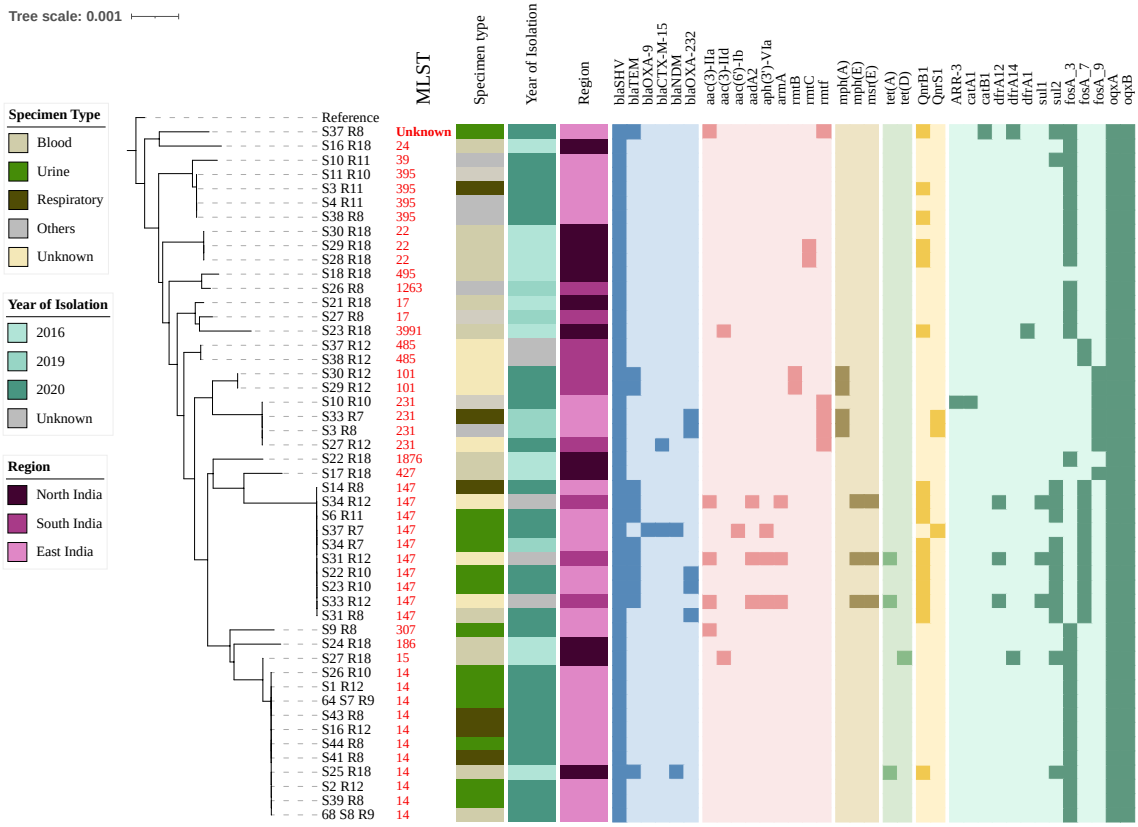

Fig. S4- Maximum likelihood phylogenetic tree based on the nucleotide sequences of the 7 housekeeping genes for *K. pneumoniae* (*gapA*, *infB*, *mdh*, *pgi*, *phoE*, *rpoB,* and *tonB*). Metadata displaying the MLST STs, specimen type, and year of isolation labelled as a colour strip. The presence and absence of antimicrobial-resistant genes against different classes of antimicrobial agents categorized as beta-lactams, aminoglycosides, macrolides, tetracycline, and others are labelled as heat maps. Tree scale indicates number of substitutions per genome per site.

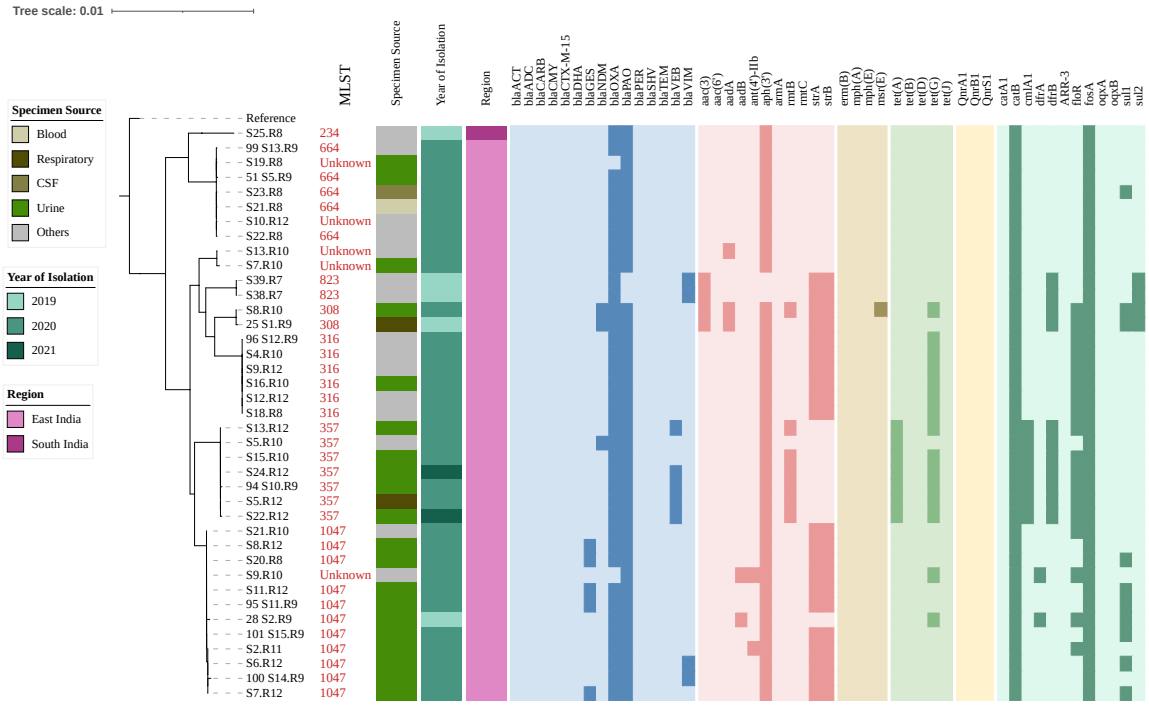

Fig. S5- Maximum likelihood phylogenetic tree based on the nucleotide sequences of the 7 housekeeping genes for *P. aeruginosa* (*acsA*, *aroE*, *guaA*, *mutL*, *nuoD*, *ppsA,* and *trpE*). Metadata displaying the MLST STs, specimen type, and year of isolation labelled as a colour strip. The presence and absence of antimicrobial-resistant genes against different classes of antimicrobial agents categorized as beta-lactams, aminoglycosides, macrolides, tetracycline, and others are labelled as heat maps. Tree scale indicates number of substitutions per genome per site.

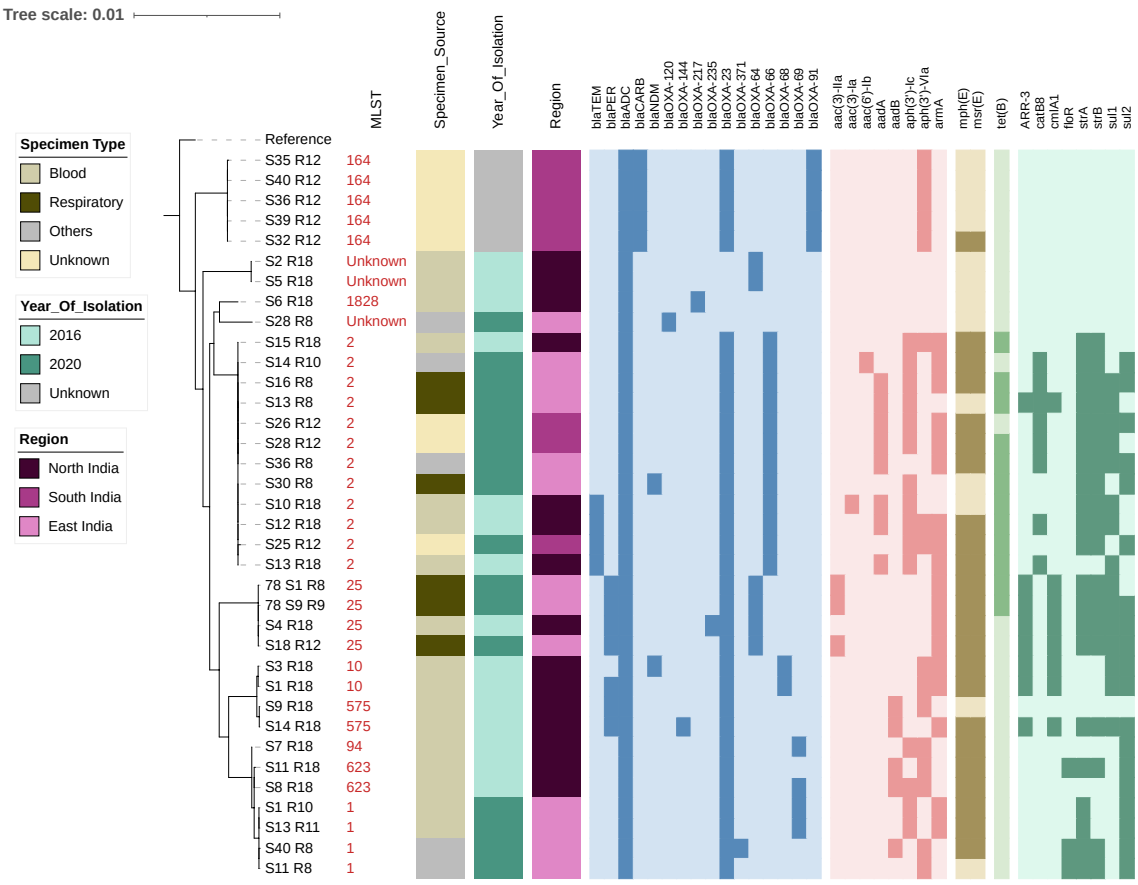

Fig. S6- Maximum likelihood phylogenetic tree based on the nucleotide sequences of the 7 housekeeping genes for *A. baumannii* (*cpn60*, *fusA*, *gltA*, *pyrG*, *recA*, *rplB,* and *rpoB*). Metadata displaying the MLST STs, specimen type, and year of isolation labelled as a colour strip. The nine *bla*$_{\text{OXA-51}}$ variants such as *bla*$_{\text{OXA-64}}$, *bla*$_{\text{OXA-66}}$, *bla*$_{\text{OXA-68}}$, *bla*$_{\text{OXA-69}}$, *bla*$_{\text{OXA-91}}$, *bla*$_{\text{OXA-120}}$, *bla*$_{\text{OXA-217}}$, *bla*$_{\text{OXA-144}}$, *bla*$_{\text{OXA-371}}$ are included for analysis. The presence and absence of antimicrobial-resistant genes against different classes of antimicrobial agents categorized as beta-lactams, aminoglycosides, macrolides, tetracycline, and others are labelled as heat maps. Tree scale indicates number of substitutions per genome per site.

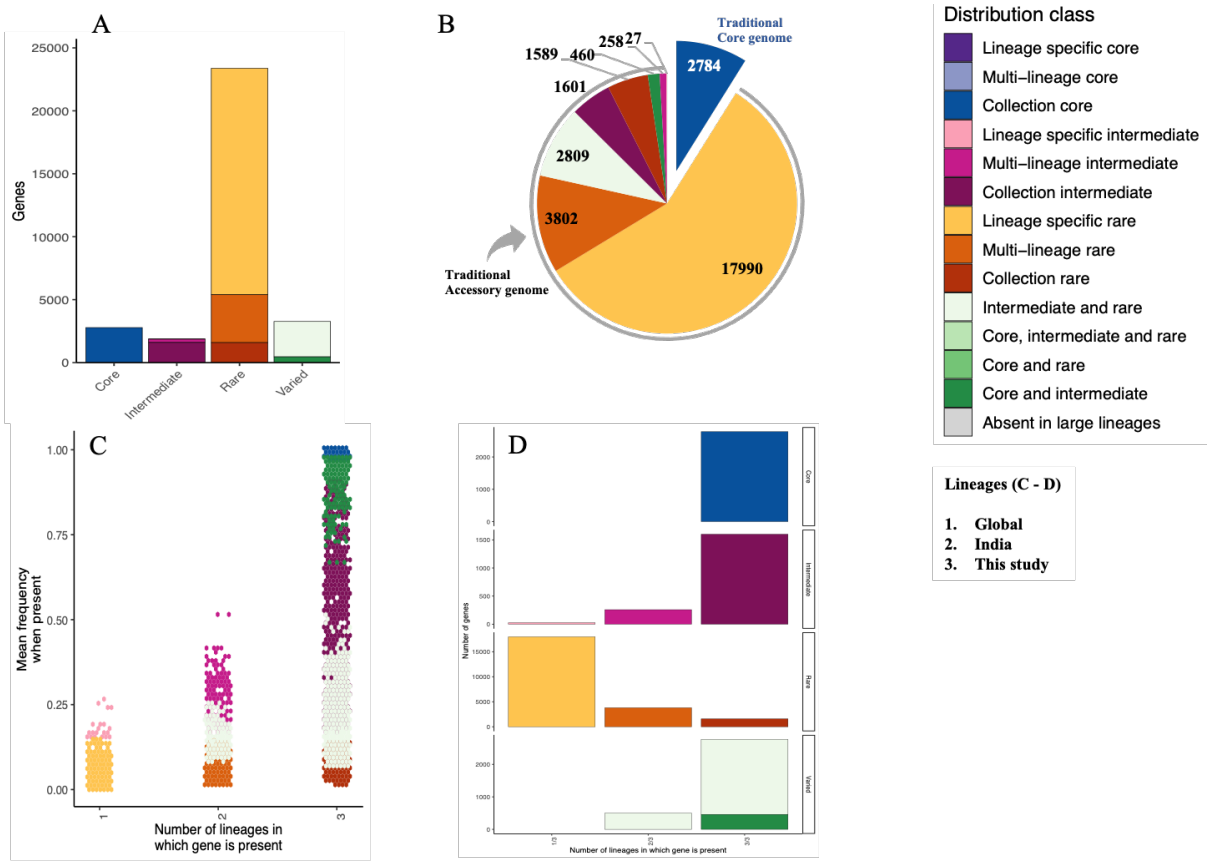

Fig. S7- Pan genome analysis of *E. coli* genomes *(n= 995)* showing (a) Number of gene clusters present in each distribution class, (b) Pie chart showing number of genes present in each distribution class classified as traditional core or accessory genome, (c) Mean distribution frequency for different lineages, (d) Number of genes distributed in different lineages. Lineages from C-D are 1) Global genomes, 2) Genomes deposited from India, 3) Genomes analysed in this study.

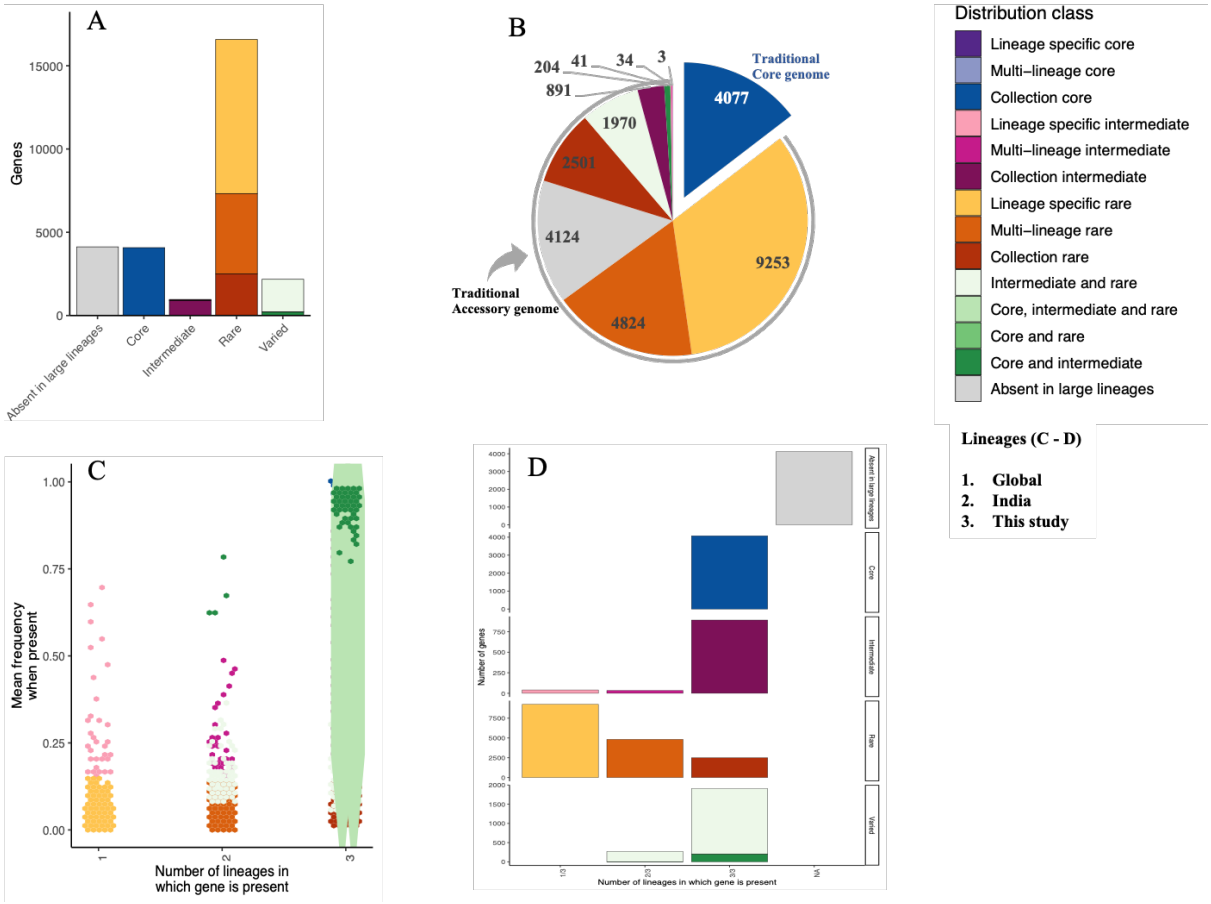

Fig. S8- Pan genome analysis of *K. pneumoniae* genomes *(n= 776)* showing (a) Number of gene clusters present in each distribution class, (b) Pie chart showing number of genes present in each distribution class classified as traditional core or accessory genome, (c) Mean distribution frequency for different lineages, (d) Number of genes distributed in different lineages. Lineages from C-D are 1) Global genomes, 2) Genomes deposited from India, 3) Genomes analysed in this study.

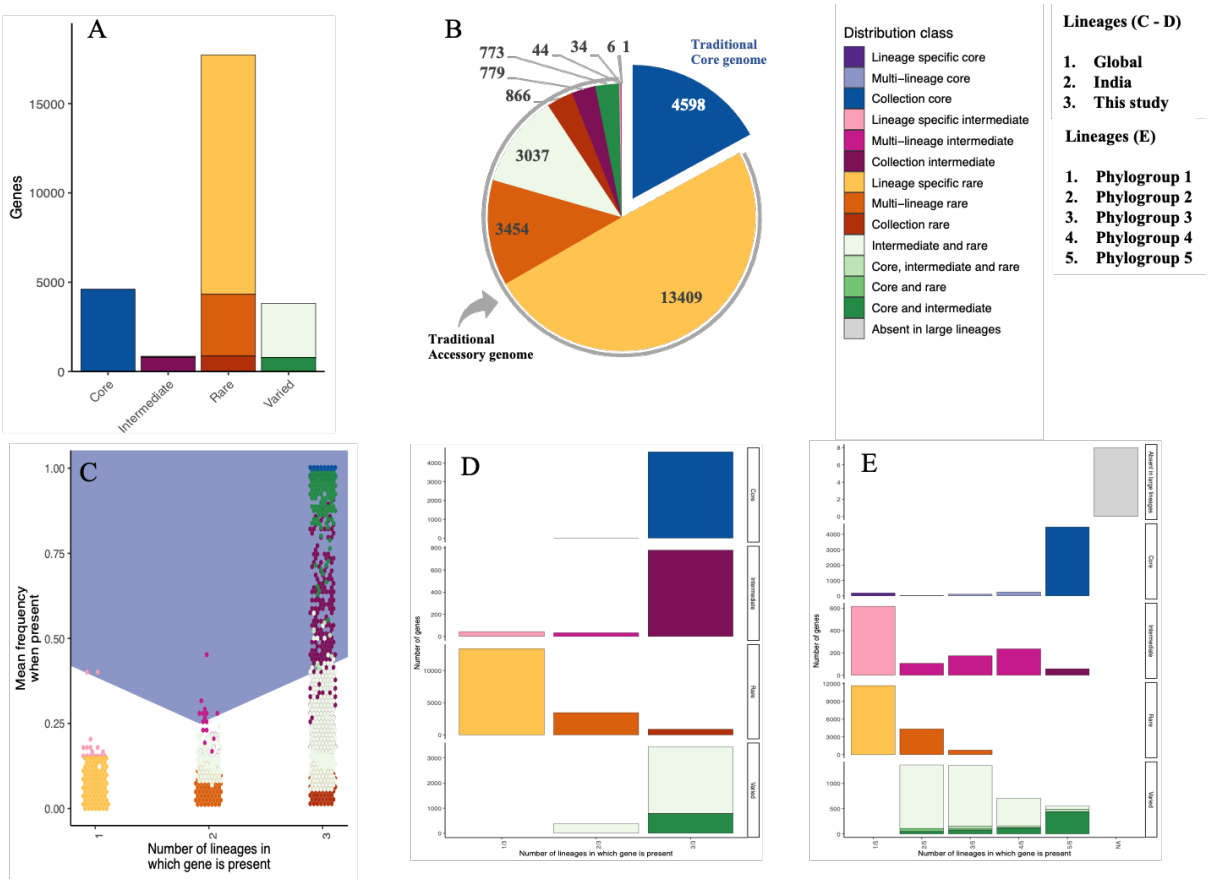

Fig. S9- Pan genome analysis of *P. aeruginosa* genomes *(n= 323)* showing (a) Number of gene clusters present in each distribution class, (b) Pie chart showing number of genes present in each distribution class classified as traditional core or accessory genome, (c) Mean distribution frequency for different lineages, (d) Number of genes distributed in different lineages, (e) Number of genes present in lineages categorised as Phylogroups 1-5. Lineages from C-D are 1) Global genomes, 2) Genomes deposited from India, 3) Genomes analysed in this study.

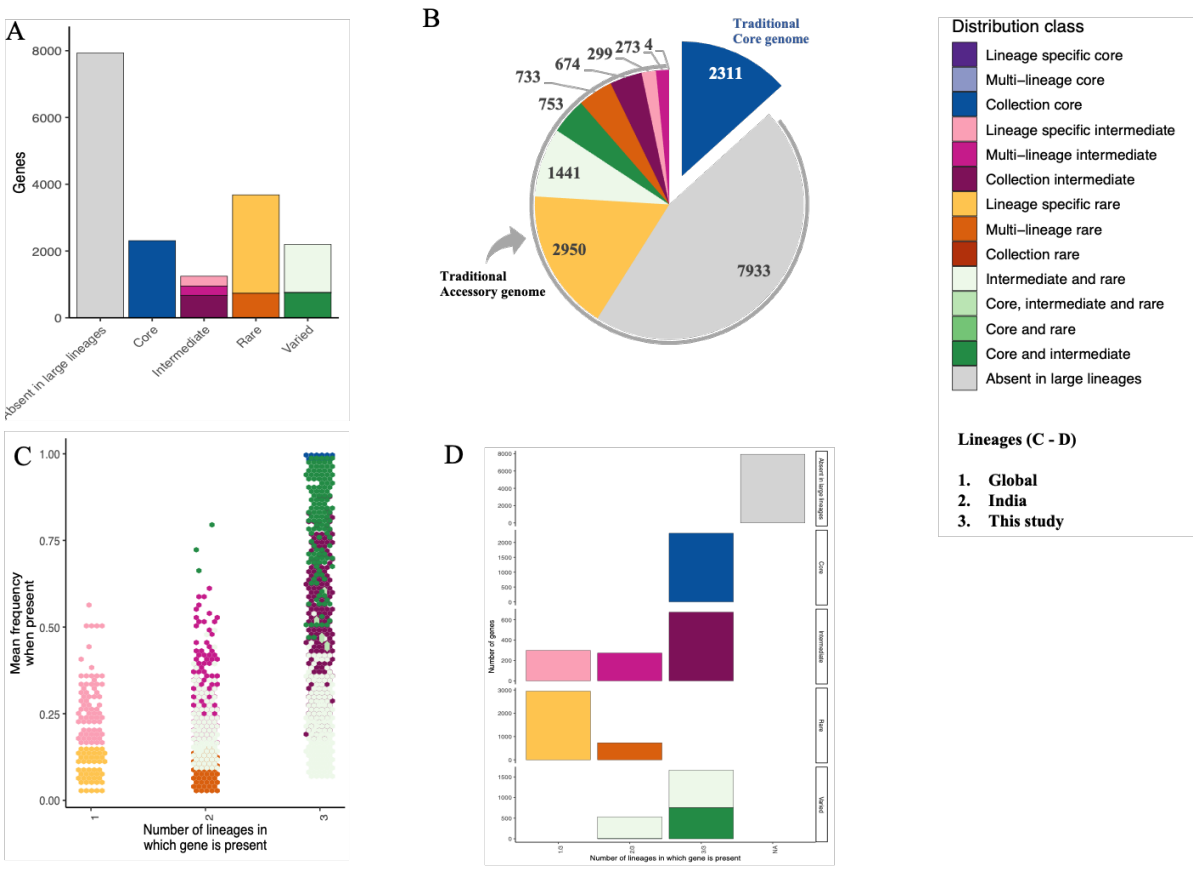

Fig. S10- Pan genome analysis of *A. baumannii* genomes *(n= 582)* showing (a) Number of gene clusters present in each distribution class, (b) Pie chart showing number of genes present in each distribution class classified as traditional core or accessory genome, (c) Mean distribution frequency for different lineages, (d) Number of genes distributed in different lineages. Lineages from C-D are 1) Global genomes, 2) Genomes deposited from India, 3) Genomes analysed in this study.

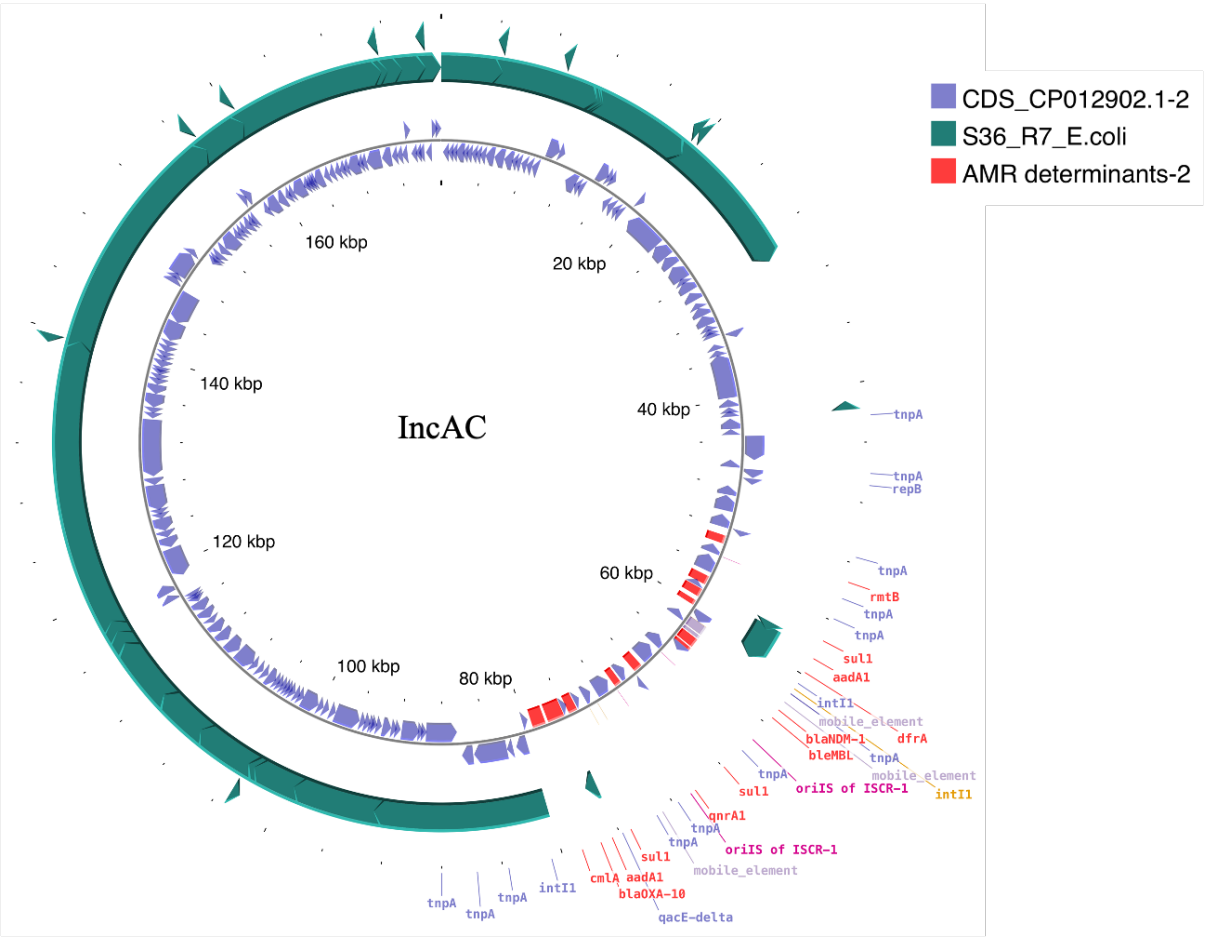

Fig. S11- Showing *IncA/C* plasmid (180 Kbp in size, S36_R7) carrying *blaNDM-1* variant in *E.coli.* AMR genes are shown in red colour. Some regions are missed due to the draft genome. Arrows are indicating the direction of the open reading frame. Figure was created using Proksee tool.

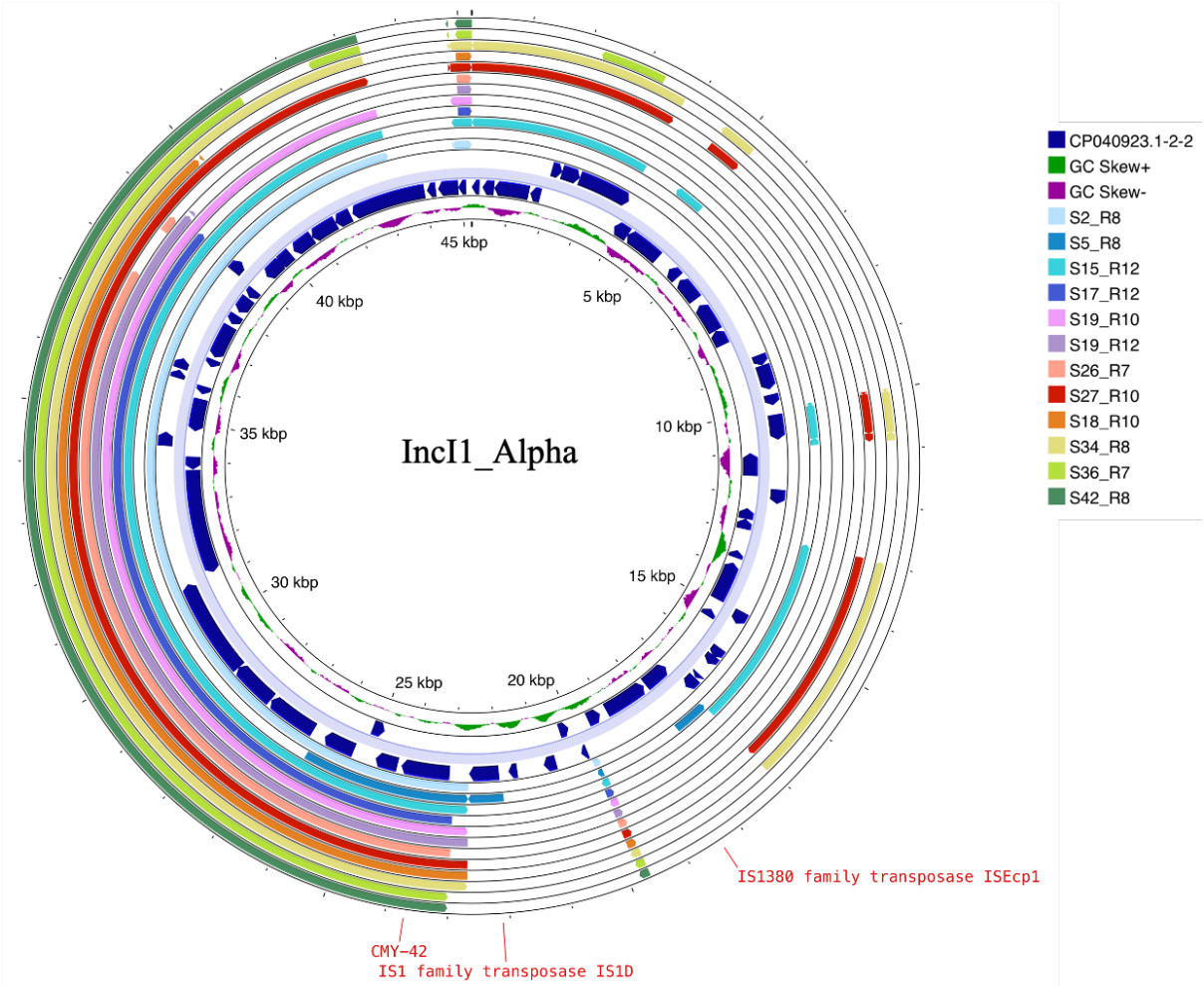

Fig. S12- Representing *IncI1-Alpha* plasmid in *E.coli* carrying *blaCMY-42* gene. Adjacent IS1 family transposases were missed due to less coverage. Arrows are indicating the direction of open reading frame. Figure was created using Proksee tool.

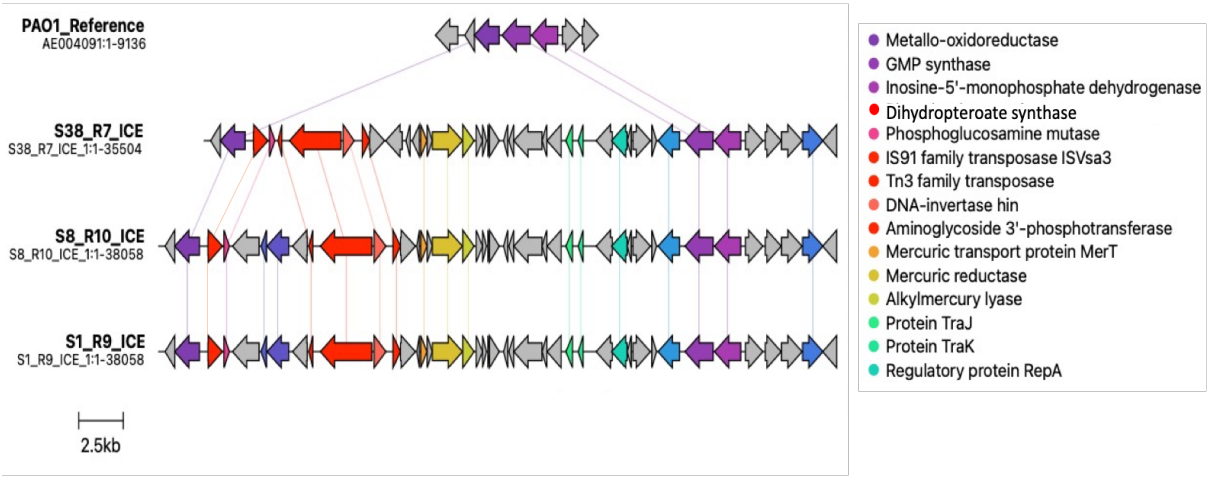

Fig. S13- Arrow representation of an integrative conjugative element (ICE) containing Tn3 family transposons carrying *aph* and *dfr* was found inserted between the genes encoding for GMP synthase and Inosine-5'-monophosphate dehydrogenase present on the chromosome. Cluster alignments are drawn to scale based using the Clinker tool. Arrows are indicating the direction of the open reading frame.

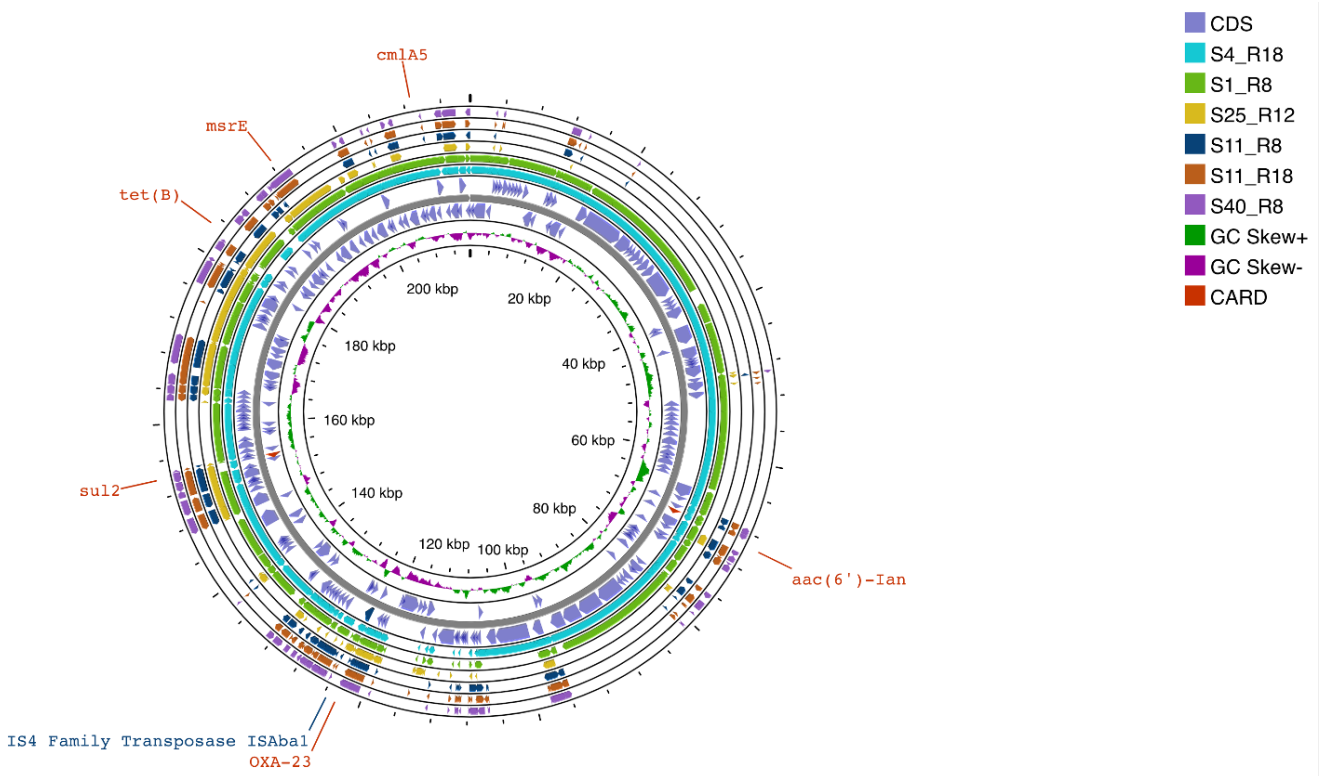

Fig. S14- Showing the genetic arrangement of *blaOxa-23* flanked by ISAba1 transposase in *A. baumannii*. AMR genes are shown in red color detected by the CARD tool. Arrows are indicating the direction of open reading frame. Figure was created using Proksee tool.

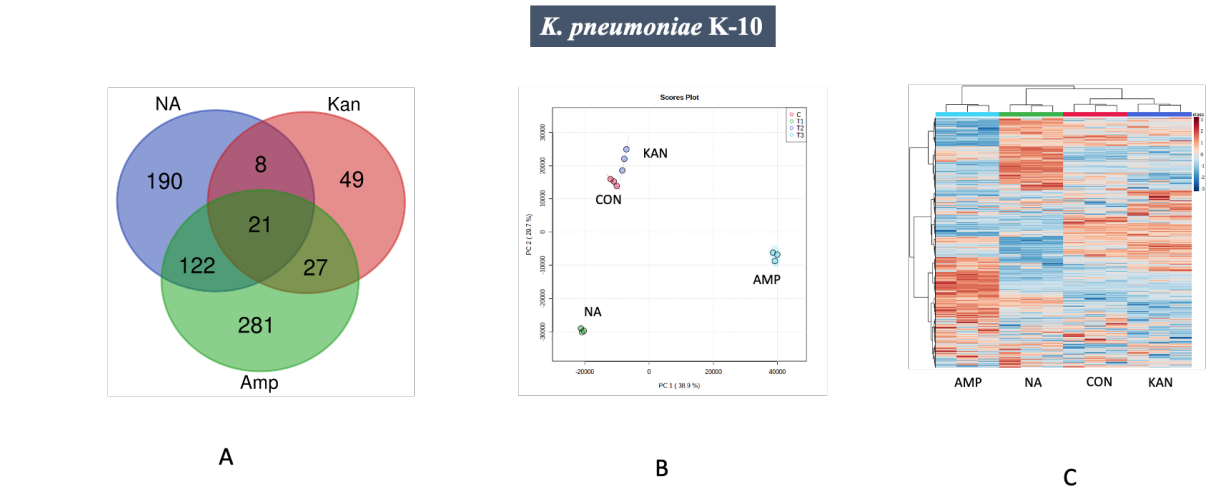

Fig. S15- Label-Free DIA-SWATH analysis identifies the pattern characteristics of MDR *K. pneumoniae* strain (K-10). Different KP isolates, including Control, NA-resistant, Kan-resistant, and Amp-resistant isolates, were collected and subjected to LC-MS/MS analysis for label-free proteome following protein extraction and trypsin digestion. (A) The number of differential proteins detected in each isolate. (B) Principal component analysis and (C) Unsupervised hierarchical clustering of the protein profiling of KP strains.

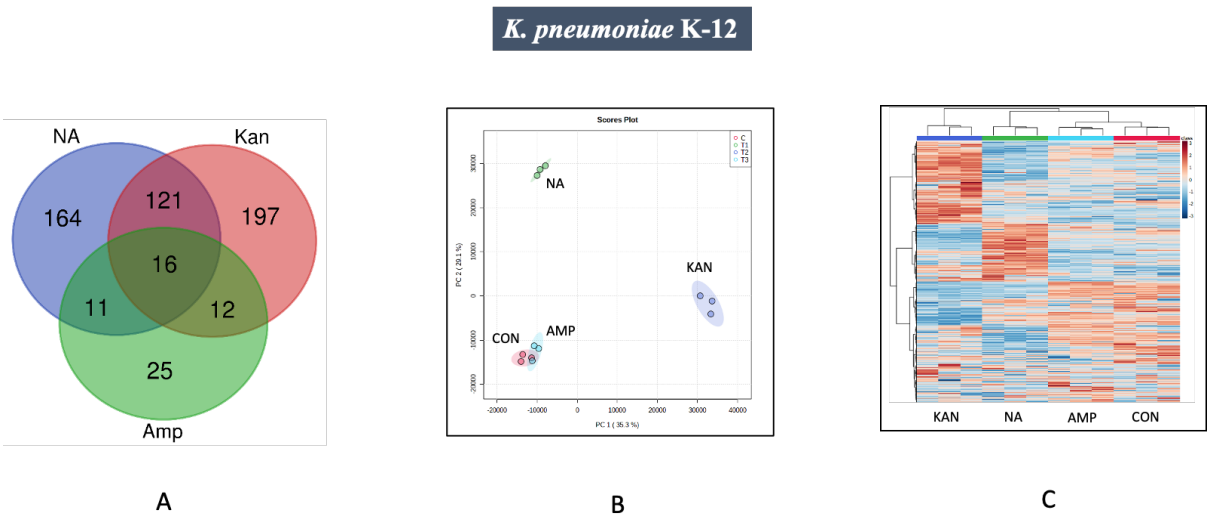

Fig. S16- Label-Free DIA-SWATH analysis identifies the pattern characteristics of MDR *K. pneumoniae* strain (K-12). Different KP isolates, including Control, NA-resistant, Kan-resistant, and Amp-resistant isolates, were collected and subjected to LC-MS/MS analysis for label-free proteome following protein extraction and trypsin digestion. (A) The number of differential proteins detected in each isolate. (B) Principal component analysis and (C) Unsupervised hierarchical clustering of the protein profiling of KP strains.

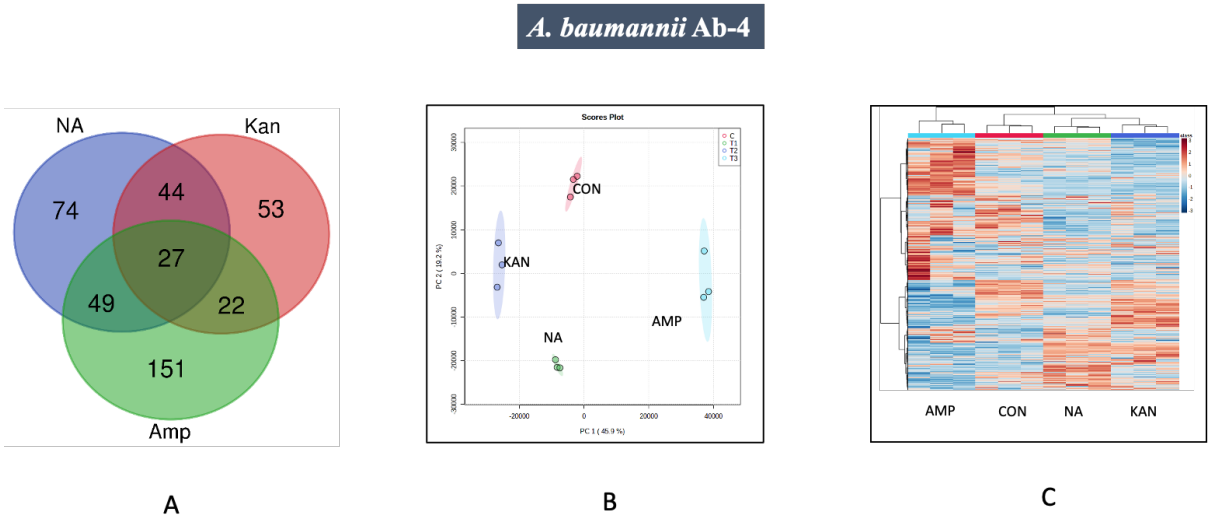

Fig. S17- Label-Free DIA-SWATH analysis identifies the pattern characteristics of MDR *A. baumannii* strain (Ab-4). Different AB isolates, including Control, NA-resistant, Kan-resistant, and Amp-resistant isolates, were collected and subjected to LC-MS/MS analysis for label-free proteome following protein extraction and trypsin digestion. (A) The number of differential proteins detected in each isolate. (B) Principal component analysis and (C) Unsupervised hierarchical clustering of the protein profiling of AB strains.

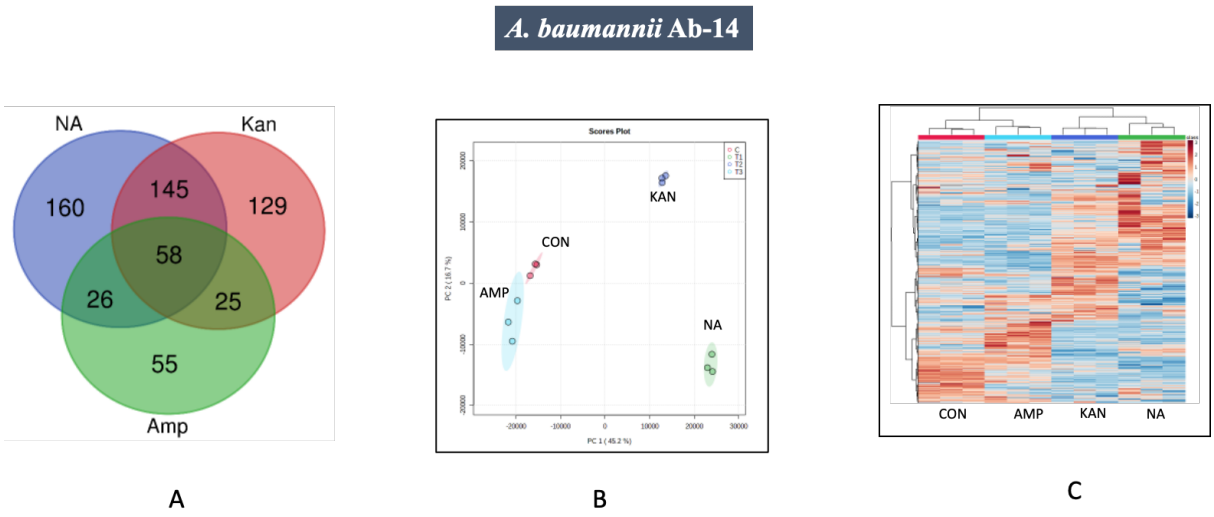

Fig. S18- Label-Free DIA-SWATH analysis identifies the pattern characteristics of MDR *A. baumannii* strain (Ab-14). Different AB isolates, including Control, NA-resistant, Kan-resistant, and Amp-resistant isolates, were collected and subjected to LC-MS/MS analysis for label-free proteome following protein extraction and trypsin digestion. (A) The number of differential proteins detected in each isolate. (B) Principal component analysis and (C) Unsupervised hierarchical clustering of the protein profiling of AB strains.

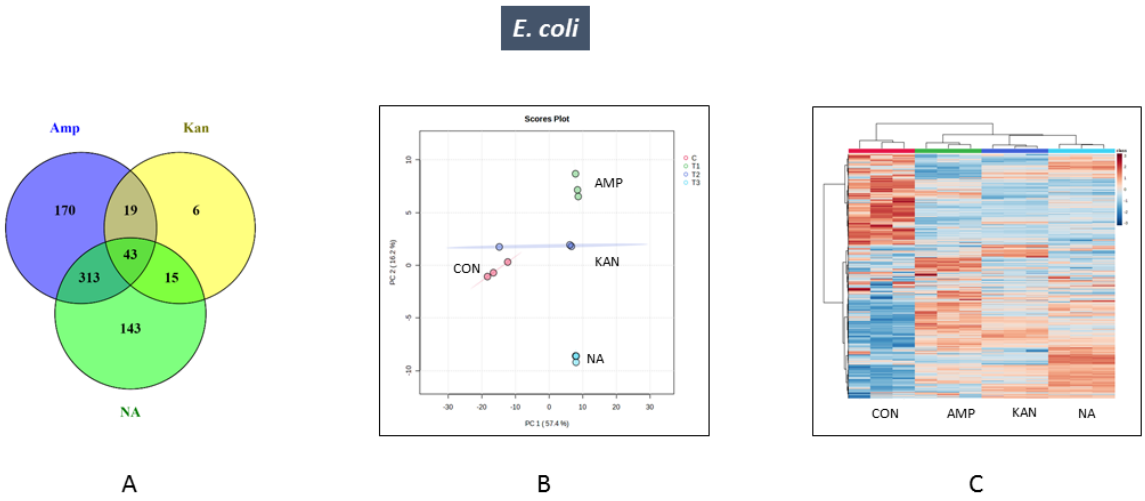

Fig. S19- Label-Free DIA-SWATH analysis identifies the pattern characteristics of MDR *E. coli* strain. Different EC isolates, including Control, NA-resistant, Kan-resistant, and Amp-resistant isolates, were collected and subjected to LC-MS/MS analysis for label-free proteome following protein extraction and trypsin digestion. (A) The number of differential proteins detected in each isolate (B) Principal component analysis and (C) Unsupervised hierarchical clustering of the protein profiling of *E. coli* strains

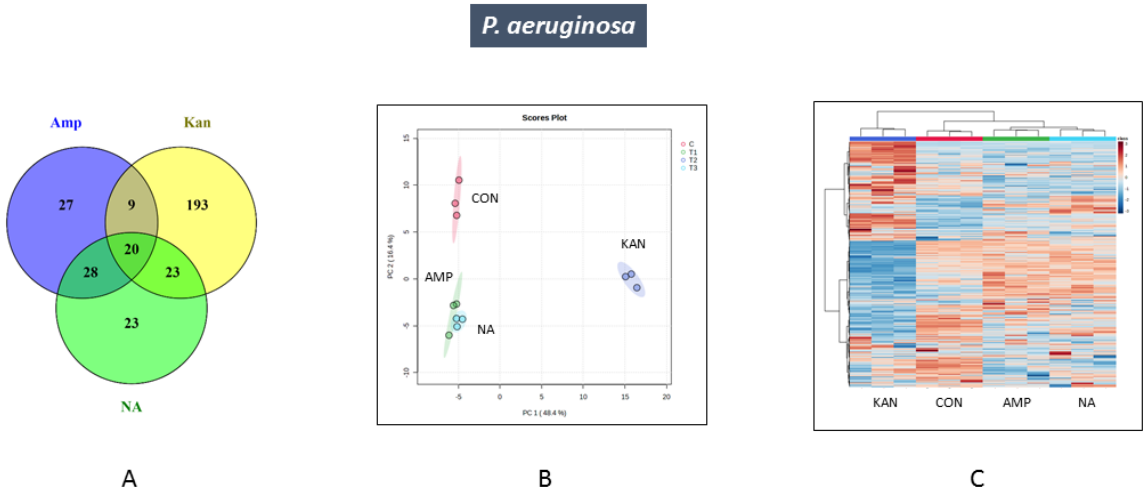

Fig. S20- Label-Free DIA-SWATH analysis identifies the pattern characteristics of MDR *P. aeruginosa* strain. Different PA isolates, including Control, NA-resistant, Kan-resistant, and Amp-resistant isolates, were collected and subjected to LC-MS/MS analysis for label-free proteome following protein extraction and trypsin digestion. (A) The number of differential proteins detected in each isolate (B) Principal component analysis and (C) Unsupervised hierarchical clustering of the protein profiling of MDR *P. aeruginosa* strains

References

1. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014 Aug 1;30(15):2114-20.
2. Jolley KA, Bray JE, Maiden MC. Open-access bacterial population genomics: BIGSdb software, the PubMLST. org website and their applications. Wellcome open research. 2018;3.
3. Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T, Keane JA, Harris SR. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. biorxiv. 2016 Jan 1:038190.
4. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014 May 1;30(9):1312-3
5. Tonkin-Hill G, MacAlasdair N, Ruis C, Weimann A, Horesh G, Lees JA, Gladstone RA, Lo S, Beaudoin C, Floto RA, Frost SD. Producing polished prokaryotic pangenomes with the Panaroo pipeline. Genome biology. 2020 Dec;21(1):1-21.
6. Beghain J, Bridier-Nahmias A, Le Nagard H, Denamur E, Clermont O. 2018. ClermonTyping: an easy-to-use and accurate in silico method for *Escherichia* genus strain phylotyping. Microb Genom 4:e000192