AMERICAN SOCIETY FOR MICROBIOLOGY

**Microbiology**® Resource Announcements

# Complete Genome Sequences of Gram-Negative Opportunistic Pathogens Isolated in Hospitals in Almaty, Kazakhstan

Ilya S. Korotetskiy,[a] Ardak B. Jumagaziyeva,[a] Bahkytzhan Kerimzhanova,[a] Oleg N. Reva,[b] Sergey V. Shilov,[a] Tatyana Kuznetsova,[a] Natalya Zubenko,[a] Nadezhda Korotetskaya,[a] Aimana Bekmukhamedova,[c] Ganiya Satylgankyzy,[c] Nailya G. Klivleyeva[d]

aScientific Center for Anti-Infectious Drugs, Almaty, Kazakhstan
bCentre for Bioinformatics and Computational Biology, Department of Biochemistry, Genetics, and Microbiology, University of Pretoria, Pretoria, South Africa
cDepartment of Vascular Surgery of the Syzganov National Scientific Center of Surgery, Almaty, Kazakhstan
dThe Research and Production Center for Microbiology and Virology, Almaty, Kazakhstan

**ABSTRACT** The problem of nosocomial infections is growing due to the introduction of new treatment regimens involving immunosuppressive drugs. The genomes of seven Gram-negative clinical isolates of *Escherichia*, *Klebsiella*, and *Pseudomonas* were sequenced and analyzed in this study to serve as model microorganisms to study drug-induced antibiotic resistance reversion.

*E*scherichia coli, *Klebsiella pneumoniae*, and *Pseudomonas aeruginosa* are the most common agents of nosocomial infections. The emergence of nosocomial pathogens often involves the acquisition of a virulence plasmid, horizontal gene transfer, and adaptive mutations. The dynamics of these processes require constant monitoring.

Seven strains of Gram-negative bacteria were isolated at the Department of Vascular Surgery of the Syzganov National Scientific Center of Surgery in Almaty, Kazakhstan. Isolates were obtained by direct plating from biological material onto selective and differential diagnostic media (Table 1). The aim of the study was to identify and perform genotyping of the potential agents of nosocomial infections. For more details on the isolates, see BioProject accession number PRJNA754843. This study was approved by the Committee of Institutional Animal Care and Use at the Scientific Center for Anti-Infectious Drugs (SCAID), Almaty.

For DNA extraction, cultures were grown on nutrient agar (HiMedia) for 24 h at 37°C. DNA was extracted using the PureLink genomic DNA minikit (Invitrogen, USA). DNA was sheared using the Megaruptor 3 shearing kit. A library was prepared using the PacBio SMRTbell Express template prep kit v2.0. SMRTbell templates were annealed using the Sequel binding and internal control kit v3.0. The Sequel sequencing kit v3.0 and a single-molecule real-time (SMRT) cell 1M v3 tray were used for sequencing. For each SMRT cell (Pacific Biosciences), 600 min movies were captured by Macrogen (Seoul, South Korea) using the PacBio Sequel I sequencing platform. Peaks smaller than 8 kb were removed using the BluePippin system. The numbers of generated reads and $N_{50}$ values for each sample are shown in Table 1. Further processing of the DNA reads was performed using software tools as described below, with default parameter settings if not indicated otherwise. The DNA reads were quality controlled and checked for remaining adapters using LongQC v1.2.0c (1) and assembled using Canu v2.0 (2). Plasmid contigs were identified using Platon v1.6 (3). The contigs were scaffolded and joined using MeDuSa at http://combo.dbe.unifi.it/medusa (4) by comparison with the most closely related reference genomes identified in GenBank by BLASTN (Table 1). The original DNA reads were mapped to the scaffolds using pbmm2 (SMRT Link v10.10.119588) for error correction, and consensus sequences were generated from

**TABLE 1** Deposited complete genome sequences of Gram-negative isolates

| Strain name | Sample type | Media | Total no. of reads | $N_{50}$ (bp) | Resistance[a] | MLST[b] | Serotype | Replicon(s) | Length (bp) | GC content (%) | Reference genome GenBank accession no. | GenBank accession no. | SRA accession no. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Escherichia coli* strain SCAID WND1-2021 (1/128) | Swab from wound | CHROMagar Orientation, Endo agar | 136,334 | 12,221 | AMX, CTR, CZ, OX, E, AZM, AMP, CB,[c] AK[c] | ST43 | H4:O25 | Chromosome / Plasmid | 5,449,567 / 186,443 | 50.56 / 52.21 | CP041581 / LS992193 | CP082831 / CP082832 | SRX11971100 |
| *E. coli* SCAID WND2-2021 (3/145) | Swab from wound | CHROMagar Orientation, Endo agar | 170,744 | 11,412 | AMX, OX, E, AMP, AZM[c] | ST3 | H18:O17 | Chromosome / Plasmid 1 / Plasmid 2 / Plasmid 3 | 5,134,206 / 139,267 / 106,249 / 32,040 | 50.77 / 51.5 / 49.56 / 41.4 | CP046396 / CP044193 / CP003290 / CP021735 | CP082827 / CP082828 / CP082829 / CP082830 | SRX11971101 |
| *Klebsiella pneumoniae* SCAID PHRX1-2021 (13/97) | Swab from pharynx | CHROMagar Orientation, Endo agar | 72,020 | 11,598 | AMX, OX, E, AZM, AMP | ST23 | Kp1 | Chromosome / Plasmid | 5,498,275 / 217,781 | 57.43 / 50.16 | CP026021 / MF993442 | CP082805 / CP082806 | SRX11971103 |
| *K. pneumoniae* SCAID PHRX2-2021 (20/245) | Swab from pharynx | CHROMagar Orientation, Endo agar | 132,267 | 10,935 | AMX, OX, E, AMP | ST380 | Kp1 | Chromosome / Plasmid 1 / Plasmid 2 | 5,319,600 / 162,135 / 95,203 | 57.54 / 50.25 / 49.97 | CP076322 / MZ156797 / FO834904 | CP082796 / CP082797 / CP082798 | SRX11971105 |
| *Pseudomonas aeruginosa* SCAID TST-2021 (7/157) | Swab from tracheostomy tube | CHROMagar Orientation, cetrimide agar | 61,010 | 10,499 | AMX, CTR,[c] CZ, OX, E, AZM,[c] IPM, AMP | ST308 | | Chromosome | 7,173,620 | 65.81 | NZ_CP027172 | CP082823 | SRX11971107 |
| *P. aeruginosa* SCAID WND1-2021 (9/195) | Swab from wound | CHROMagar Orientation, cetrimide agar | 142,460 | 11,343 | AMX, CTR, CZ, OX, E, AMP, IPM[c] | ST244 | | Chromosome | 7,093,992 | 65.9 | CP032257 | CP082822 | SRX11971108 |
| *P. aeruginosa* SCAID PLC1-2021 (16/222) | Pleural cavity | CHROMagar Orientation, cetrimide agar | 79,229 | 11,145 | AMX, CZ, OX, E, AZM, AMP | ST308 | | Chromosome | 7,124,329 | 65.85 | CP027172 | CP082821 | SRX11971109 |

[a]AK, amikacin; AMP, ampicillin; AMX, amoxicillin; AZM, azithromycin; CB, carbenicillin; CTR, ceftriaxone; CZ, cefazolin; E, erythromycin; IPM, imipenem; OX, oxacillin.

[b]MLST, multilocus sequence typing.

[c]Intermediate resistance. The resistance to antibiotics was determined experimentally. The susceptibility was determined by the disk diffusion method in Mueller-Hinton agar (HiMedia, India). The results of the threshold inhibition zones were evaluated according to the CLSI.

the alignments using the gcpp Arrow algorithm (SMRT Link v10.10.119588). The consensus sequences were annotated using the RAST server (https://rast.nmpdr.org/) with the RASTtk algorithm (5) and the "Fix frameshifts" setting. The chromosomal sequences were rotated to start with *dnaA* on the positive strand, and the plasmid sequences were shifted for 50 kb to perform circularization, final error fixation, and genome completion by another round of mapping of the initial PacBio reads using pbmm2. The final consensus sequences of the complete genomes were generated from the alignments and deposited at NCBI (Table 1). The GenBank annotation robot PGAP was used for annotation of the deposited genomes. Multilocus sequence typing (MLST) was performed using the BIGSdb (https://bigsdb.pasteur.fr/) and CBS (https://www.cbs.dtu.dk/services/MLST) databases (6, 7). The *E. coli* serotypes were predicted by genotype using SerotypeFinder v2.0.1 (http://cge.cbs.dtu.dk/services/SerotypeFinder/) (8).

The *E. coli* isolates belonged to MLST ST43 and ST3, which are widely distributed uropathogens (9). *K. pneumoniae* ST23 and ST380 are abundant hypervirulent and multidrug-resistant variants which emerged due to the acquisition of pLVPK-type virulence plasmids (10–12). Two *P. aeruginosa* isolates belong to ST308, which is a common causative agent of nosocomial infections (13).

**Data availability.** The genome sequences are available from NCBI under BioProject accession number PRJNA754843 and the accession numbers shown in Table 1.

## REFERENCES

1. Fukasawa Y, Ermini L, Wang H, Carty K, Cheung MS. 2020. LongQC: a quality control tool for third generation sequencing long read data. G3 (Bethesda) 10:1193–1196. https://doi.org/10.1534/g3.119.400864.
2. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res 27:722–736. https://doi.org/10.1101/gr.215087.116.
3. Schwengers O, Barth P, Falgenhauer L, Hain T, Chakraborty T, Goesmann A. 2020. Platon: identification and characterization of bacterial plasmid contigs in short-read draft assemblies exploiting protein sequence-based replicon distribution scores. Microb Genom 6:mgen000398. https://doi.org/10.1099/mgen.0.000398.
4. Bosi E, Donati B, Galardini M, Brunetti S, Sagot MF, Lió P, Crescenzi P, Fani R, Fondi M. 2015. MeDuSa: a multi-draft based scaffolder. Bioinformatics 31:2443–2451. https://doi.org/10.1093/bioinformatics/btv171.
5. Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Thomason JA, III, Stevens R, Vonstein V, Wattam AR, Xia F. 2015. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci Rep 5:8365. https://doi.org/10.1038/srep08365.
6. Jolley KA, Maiden MC. 2010. BIGSdb: scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics 11:595. https://doi.org/10.1186/1471-2105-11-595.
7. Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, Jelsbak L, Sicheritz-Pontén T, Ussery DW, Aarestrup FM, Lund O. 2012. Multilocus sequence typing of total-genome-sequenced bacteria. J Clin Microbiol 50:1355–1361. https://doi.org/10.1128/JCM.06094-11.
8. Joensen KG, Tetzschner AM, Iguchi A, Aarestrup FM, Scheutz F. 2015. Rapid and easy in silico serotyping of *Escherichia coli* using whole genome sequencing (WGS) data. J Clin Microbiol 53:2410–2426. https://doi.org/10.1128/JCM.00008-15.
9. Wawire SA, Reva ON, O'Brien TJ, Figueroa W, Dinda V, Shivoga WA, Welch M. 2021. Virulence and antimicrobial resistance genes are enriched in the plasmidome of clinical *Escherichia coli* isolates compared with wastewater isolates from western Kenya. Infect Genet Evol 91:104784. https://doi.org/10.1016/j.meegid.2021.104784.
10. Chen YT, Chang HY, Lai YC, Pan CC, Tsai SF, Peng HL. 2004. Sequencing and analysis of the large virulence plasmid pLVPK of *Klebsiella pneumoniae* CG43. Gene 337:189–198. https://doi.org/10.1016/j.gene.2004.05.008.
11. Hentzien M, Rosman J, Decré D, Brenkle K, Mendes-Martins L, Mateu P. 2017. Seven hypervirulent ST380 *Klebsiella pneumoniae* septic localizations. Med Mal Infect 47:171–173. https://doi.org/10.1016/j.medmal.2016.10.002.
12. Shankar C, Jacob JJ, Vasudevan K, Biswas R, Manesh A, Sethuvel DPM, Varughese S, Biswas I, Veeraraghavan B. 2020. Emergence of multidrug resistant hypervirulent ST23 *Klebsiella pneumoniae*: multidrug resistant plasmid acquisition drives evolution. Front Cell Infect Microbiol 10:575289. https://doi.org/10.3389/fcimb.2020.575289.
13. Chew KL, Octavia S, Ng OT, Marimuthu K, Venkatachalam I, Cheng B, Lin RTP, Teo JWP. 2019. Challenge of drug resistance in *Pseudomonas aeruginosa*: clonal spread of NDM-1-positive ST-308 within a tertiary hospital. J Antimicrob Chemother 74:2220–2224. https://doi.org/10.1093/jac/dkz169.