Genomics Data 14 (2017) 7–9

Contents lists available at ScienceDirect

Genomics Data

journal homepage: www.elsevier.com/locate/gdata

Data in Brief

Whole genome of *Klebsiella aerogenes* PX01 isolated from San Jacinto River sediment west of Baytown, Texas reveals the presence of multiple antibiotic resistance determinants and mobile genetic elements

Rupa Iyer[a,*], Brian Iken[a], Ashish Damania[b]

[a] Center for Life Sciences Technology, Department of Engineering Technology, University of Houston, Houston, TX, USA
[b] Department of Tropical Medicine-Pediatrics, Baylor College of Medicine, Houston, TX, USA

ABSTRACT

Klebsiella aerogenes is a Gram-negative bacterium of the family *Enterobacteriaceae* which is widely distributed in water, air and soil. It also forms part of the normal microbiota found in human and animal gastrointestinal tracts. Here we report the draft genome sequence (chromosome and 1 plasmid) of *K. aerogenes* strain PX01 compiled at the scaffold level from 97 contigs totaling 5,262,952 bp. *K. aerogenes* PX01 was isolated from sediment along the northern face of Burnet Bay west of Baytown, Texas. The nucleotide sequence of this genome was deposited into NCBI GenBank under the accession NJBB00000000.

Specifications

Organism/cell line/tissue	*Klebsiella aerogenes*
Strain	PX01
Sequencer or array type	Illumina Miseq
Data format	Analyzed
Experimental factors	Bacterial strain
Experimental features	Whole genome analysis and gene annotation of PX01
Sample source location	San Jacinto River sediment near Burnet Bay (Baytown, Texas)
GPS coordinates	29.779500, − 95.057105

1. Direct link to deposited data

https://www.ncbi.nlm.nih.gov/nuccore/NJBB00000000.

2. Experimental design, materials and methods

Polluted sediment was collected from the banks of the San Jacinto River along the north face of Burnet Bay in Baytown, Texas as part of an Environmental Sampling Research Module with the goal of tracking pesticide-degrading activity and isolating putative bacterial degraders across the Houston-metropolitan area [1]. Selective media was prepared for screening purposes: Carbon Selective Media (CSM) which has a composition of 2 mM NTA, 0.8 mM $MgSO_4{\cdot}7H_2O$, 0.17 mM $CaNO_3$, 0.018 mM $FeSO_4{\cdot}7H_2O$, 20% *v*/v Phosphate Buffer. 5 mL of CSM media was aliquoted into culture tubes with 100 µg/mL ethyl paraoxon as a screening agent. These tubes were prepared fresh each week for each new subculture set for a period of five weeks. The culture was then diluted into minimal media with glycerol added as a supplementary carbon source and plated onto an agar plate with 100 µg/mL ethyl paraoxon. A resulting tan colored bacterium was isolated from the agar plate and shipped to Genewiz (South Plainfield, NJ), where library construction and whole genome sequencing of the bacterium was performed as described below.

Samples were visually inspected upon receipt and genomic DNA was extracted from bacterial colonies using the PureLink Genomic DNA extraction kit as per manufacturer's protocols. The resulting genomic DNA was quantified using both the Nanodrop and the Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA). A total of 50–60 ng of each sample was run on a 0.6% agarose gel to check for quality. The Illumina Nextera XT library preparation, clustering, and sequencing reagents were used throughout the process following the manufacturer's recommendations (Illumina, San Diego, CA, USA). DNA libraries were analyzed on the Agilent TapeStation (Agilent Technologies, Palo Alto, CA, USA) and quantified using the Qubit 2.0 Fluorometer.

* Corresponding author.
E-mail address: riyer@uh.edu (R. Iyer).

http://dx.doi.org/10.1016/j.gdata.2017.07.012
Received 23 June 2017; Received in revised form 13 July 2017; Accepted 18 July 2017
Available online 23 July 2017

Table 1
K. aerogenes PX01 genome statistics.

Assembly statistics	
Platform	Illumina MiSeq (2 * 250) paired end
Total raw reads	3,622,205
Total filtered reads	3,476,896
Genome size(bp)	5,224,354 (chromosome) 38,598 (plasmid)
Number of contigs	97 (chromosome) 6 (plasmid)
Average coverage	240.65 ×
Annotation statistics	
GC content	55.09% (chromosome) 51.34% (plasmid)
Total genes	5265
Coding genes	5.123
rRNAs	41
tRNAs	89

The DNA libraries were quantified by real time PCR (Applied Biosystems, Carlsbad, CA, USA), and multiplexed in equal molar mass. DNA libraries were multiplexed and loaded on an Illumina MiSeq instrument according to manufacturer's instructions (Illumina, San Diego, CA, USA). Sequencing was performed using a 2 × 250 paired-end (PE) configuration. Image analysis and base calling were conducted by the MiSeq Control Software (MCS) on the MiSeq instrument (Table 1). Sequence reads were checked for quality using Fastqc [2] and filtered using BBTools [3] with minimum Phred score of 20. Paired-end reads were assembled into contigs with the Spades 3.10.1 program [4]. Preliminary reference based annotation using PATRIC [5] web resources was carried out to identify conserved pathways. Final *de novo* annotation performed through the NCBI Prokaryotic Genomes Automatic Annotation Pipeline (http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html) and the Rapid Annotation System Technology (RAST) server identified the unknown bacterium as *Klebsiella aerogenes* [6,7].

2.1. Data description

Klebsiella species are Gram-negative encapsulated bacteria commonly found in mammalian gastrointestinal tracts. While typically benign, some species do possess the capacity to act as opportunistic human pathogens. *Klebsiella aerogenes* (formerly *Enterobacter aerogenes*) is a motile, non-spore forming, bacterium that has emerged as a multidrug-resistant (MDR) threat and is often included as part of the ESKAPE (*Enterococcus faecium*, *Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and Enterobacter sp.) group of human pathogens, bacteria that are the source of most clinical infections worldwide [8]. Hospital outbreaks due to *K. aerogenes* is often related to the acquisition of novel antibiotic resistance determinants through mobile genetic elements as well as constitutive β-lactamase overexpression [9].

The genome of *Klebsiella aerogenes PX01* includes a circular bacterial chromosome with a GC content of 65.09%, consists of 5,224,354 bp and a single plasmid of 38,598 bp. Combined, the total genome of PX01 contains approximately 5265 gene sequences, 5123 coding sequences including 41 rRNAs and 89 tRNAs. An overview of genome subsystem features (Fig. 1) shows that the greatest number of identified genes were foremost allocated to subsystems for general cell survival and metabolism including carbohydrate and protein metabolism as well as amino acid and vitamin synthesis and degradation. Secondary subsystems with over 100 + genes identified through RAST are principally survival and stress response oriented and include both cell wall and capsule production as well as antibiotic resistance and virulence. CARD

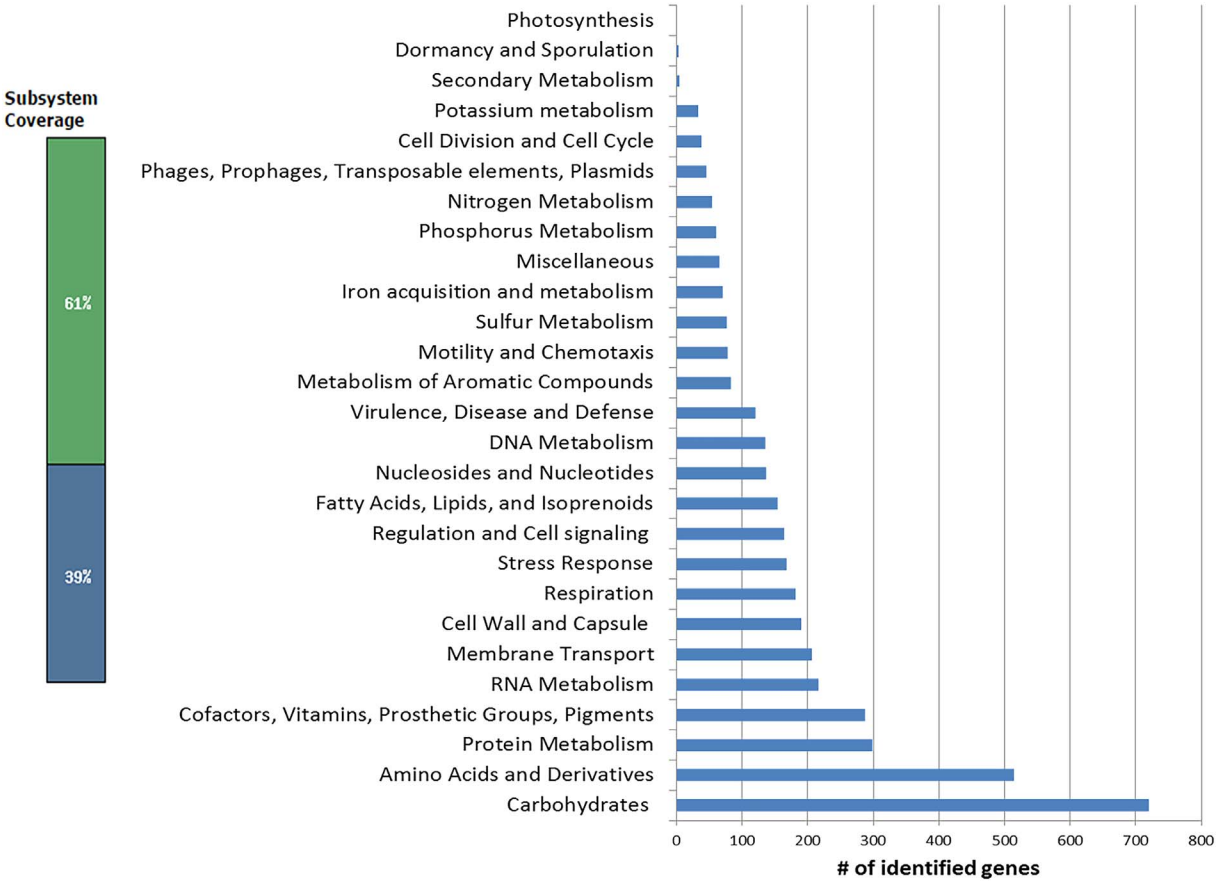

Fig. 1. Subsystem category distribution of major protein coding genes of *Enterobacter aerogenes* strain PX01 as annotated by the RAST annotation server. The bar chart on the left shows the subsystem coverage in percentage (blue bar corresponds to percentage of proteins included). The bar chart to the right shows the distribution of the 27 most abundant subsystem categories.

Table 2
CARD antibiotic resistance gene results.

Category	Hits	Notable genes involved in antibiotic resistance
Determinant of sulfonamide resistance	1	leuO
Determinant of fluoroquinolone resistance	1	mfd
Antibiotic target protection protein	1	mfd
Determinant of fosfomycin resistance	2	FosA5, UhpT
Determinant of isoniazid resistance	1	katG
Antibiotic resistant gene variant or mutant	2	katG, UhpT
Determinant of aminoglycoside resistance	1	kdpE
Determinant of beta-lactam resistance	1	CMY-108
Antibiotic inactivation enzyme	2	CMY-108, FosA5
Protein modulating permeability to antibiotic	2	marA, ramA
Gene altering cell wall charge	4	arnA, PmrE, PmrC, PmrF
Determinant of polymyxin resistance	4	arnA, PmrE, PmrC, PmrF
Gene conferring antibiotic resistance *via* molecular bypass	1	bacA
Determinant of resistance to peptide antibiotics	1	bacA
Protein(s) and two-component regulatory system modulating antibiotic efflux	13	adeL, adeL, CRP, marA, H-NS, kdpE, ramA, cpxA, baeR, baeS, emrR, leuO, robA
Efflux pump complex or subunit conferring antibiotic resistance	43	mdtC, robA, emrA, mdtH, mdtK, acrF, oqxA, mdtA, oqxB, mdtD, acrB, acrB, rosB, rosA, emrB, *K. pneumoniae* acrA, acrD, mdtM, patA, macB, mexB, mdtB, tolC, acrE, mdtG, mdtL, cpxA, mdfA, yojI, msbA, emrD, baeS, baeR, norB, CRP, hmrM, emrR, marA, adeL, adeL, ramA, macA, H-NS

analysis (Table 2) of the PX01 strain reveals a microorganism that is well suited for opportunistic pathogenicity through a combination of multidrug efflux, β-lactamase activity and antibiotic inactivation enzymes, many of which are also associated with clinical strains of both *K. pneumoniae* and *E. coli* [10]. In addition, this strain harbors a minimum of 46 prophage related sequences located on the bacterial chromosome suggesting a possible vehicle for lateral transfer to and from this bacterial strain.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

Whole genome sequencing of *Enterobacter aerogenes* PX01 was funded through NIST (G110008 58106).

References

[1] R. Iyer, K. Smith, B. Kudrle, A. Leon, Detection and location of OP-degrading activity: a model to integrate education and research, New Biotechnol. 32 (4) (2015) 403–411.
[2] B. Bioinformatics, FastQC a Quality Control Tool for High Throughput Sequence Data. Babraham Institute, Cambridge, United Kingdom. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/, (2011).
[3] B. Bushnell, BBMap Short Read Aligner. http://sourceforge.net/projects/bbmap.
[4] A.R. Wattam, D. Abraham, O. Dalay, T.L. Disz, T. Driscoll, J.L. Gabbard, J.J. Gillespie, R. Gough, D. Hix, R. Kenyon, D. Machi, C. Mao, E.K. Nordberg, R. Olson, R. Overbeek, G.D. Pusch, M. Shukla, J. Schulman, R.L. Stevens, D.E. Sullivan, V. Vonstein, A. Warren, R. Will, M.J.C. Wilson, Y.H. Seung, C. Zhang, Y. Zhang, B.W. Sobral, PATRIC, the bacterial bioinformatics database and analysis resource, Nucleic Acids Res. 42 (2014) D581–D591.
[5] A. Bankevich, S. Nurk, D. Antipov, A.A. Gurevich, M. Dvorkin, A.S. Kulikov, V.M. Lesin, S.I. Nikolenko, S. Pham, A.D. Prjibelski, A.V. Pyshkin, A.V. Sirotkin, N. Vyahhi, G. Tesler, M.A. Alekseyev, P.A. Pevzner, SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing, J. Comput. Biol. 19 (2012) 455–477.
[6] T. Tatusova, M. DiCuccio, A. Badretdin, V. Chetvernin, S. Ciufo, W. Li, Prokaryotic genome annotation pipeline, The NCBI Handbook, second ed., National Center for Biotechnology Information, Bethesda, MD, 2013.
[7] R.K. Aziz, D. Bartels, A.A. Best, M. DeJongh, T. Disz, R.A. Edwards, K. Formsma, S. Gerdes, E.M. Glass, M. Kubal, F. Meyer, The RAST server: rapid annotations using subsystems technology, BMC Genomics 9 (2008) 75.
[8] J.N. Pendleton, S.P. Gorman, B.F. Gilmore, Clinical relevance of the ESKAPE pathogens, Expert Rev. Anti-Infect. Ther. 11 (3) (2013) 297–308.
[9] A. Davin-Regli, J.M. Pagès, *Enterobacter aerogenes* and *Enterobacter cloacae*; versatile bacterial pathogens confronting antibiotic treatment, Front. Microbiol. 6 (2015) 392.
[10] A.G. McArthur, N. Waglechner, F. Nizam, A. Yan, M.A. Azad, A.J. Baylay, K. Bhullar, M.J. Canova, G. De Pascale, L. Ejim, L. Kalan, The comprehensive antibiotic resistance database, Antimicrob. Agents Chemother. 57 (7) (2013) 3348–3357.