Data in Brief 50 (2023) 109506

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

Whole genome sequencing data of *Raoultella ornithinolytica* PX02 isolated from San Jacinto river sediment in Baytown, Texas

Rupa Iyer*, Brian Iken

Division of Research, Innovation, and Economic Development, Tarleton State University, Stephenville 76401, Texas

ARTICLE INFO

Article history:
Received 30 March 2023
Revised 24 July 2023
Accepted 15 August 2023
Available online 19 August 2023

Dataset link: Draft genome sequence of Raoultella ornithinolytica PX02 (Original data)

Keywords:
Antibiotic resistance
incF plasmid
Raoultella ornithinolytica
San Jacinto river
Whole genome

ABSTRACT

Raoultella ornithinolytica PX02 is a Gram-negative, encapsulated bacterium, part of the *Enterobacteriaceae* family, emerging as a notable human pathogen. Here, we present the whole genome sequence of *R. ornithinolytica* PX02 isolated from San Jacinto River sediment near a Burnet Shores community in Baytown, Texas. This microorganism harbors a large 200,000 bp incF plasmid and can potentially be a significant antibiotic reservoir. The PX02 genome consists of 5,970,914 base pairs encoding approximately 5,661 functional proteins. Strain PX02 (chromosomal and plasmid) was compiled at the scaffold level and can be accessed through the National Center for Biotechnology Information database under accession NZ_NJBC00000000.1.

* Corresponding author.
E-mail address: iyer@tarleton.edu (R. Iyer).

https://doi.org/10.1016/j.dib.2023.109506

Specifications Table

Subject	Microbiology
Specific subject area	Microbial genetics
Type of data	Figures, Tables
How data was acquired	DNA sequencing: Illumina Miseq Bioinformatics: NCBI Prokaryotic Genomes Automatic Annotation Pipeline (PGAAP), the Rapid Annotation using Subsystem Technology (RAST) server, the Comprehensive Antibiotic Resistance Database (CARD)
Data format	Raw and analyzed
Description of data collection	Cellular DNA from a *Raoultella ornithinolytica* strain isolated from San Jacinto River sediment was extracted using a DNEasy blood and tissue kit and sequenced using Illumina MiSeq technology.
Data source location	Sediment from USA: Texas, Baytown, (29.778404, -95.059108)
Data accessibility	Strain data is uploaded to National Center for Biotechnology Information database under accession NZ_NJBC00000000.1. Direct link to data: https://www.ncbi.nlm.nih.gov/Traces/wgs/NJGC01?val=NJBC01

1. Value of the Data

- The whole genome data of *R. ornithinolytica* PX02 highlights a microorganism with significant potential as an antibiotic reservoir from a community along the shore of the polluted San Jacinto River in Baytown, Texas.
- A large >200,000 bp plasmid is present within this organism. Many of the hypothetical proteins are of unknown function. Still, of those identified, there are proteins encoding for horizontal transmission of the plasmid to other bacterial species and toxin/antitoxin determinants that commonly regulate pathogenicity.
- This data can be used to explore the emerging role of *R. ornithinolytica* in human infections by comparing this environmental isolate to those found in wastewater and clinical settings.

2. Objective

Raoultella sp. are hardy, ubiquitous microorganisms found in soil and marine habitats. Members of the *Enterobacteriaceae* family are closely related to *Klebsiella*, a genus they once shared, and possess many of the same genomic features [1]. This includes a robust stress response and a propensity to harbor large transmissible plasmids. These plasmids are often the source of antimicrobial resistance genes and contribute to the antibiotic resistome spread amongst enteric bacterial species [2]. One such species, *R. ornithinolytica*, has emerged as an opportunistic human pathogen in recent years. While most cases of pathogenicity in *R. ornithinolytica* still originate from within nosocomial settings, the capacity for this species to act as an antibiotic reservoir in the outside environment needs to be clarified. However, a polluted environment, such as the San Jacinto River, that sees considerable human traffic and activity is a potential hotspot for antibiotic resistance exchange. Our objective then was to sequence, annotate, and analyze the genome of *R. ornithinolytica* PX02, collected from sediment along the San Jacinto River, to provide a new environmental isolate for this species to study the presence and potential capacity of transferring antimicrobial resistance.

3. Data Description

This data set includes assembled contigs, compiled at the scaffold level, for the bacterial isolate *R. ornithinolytica* PX02. This sequence data has been quality checked and annotated by the

Table 1
Raoultella ornithinolytica PX02 genome statistics.

Assembly statistics	
platform	Illumina MiSeq (2*250) paired-end
genome size(bp)	5,970,914 (Chromosome - 5,743,580 Plasmid - 227,334)
number of contigs	78(chromosome) + 11 (plasmid)
average coverage	270.65
raw reads (paired-end)	6,464,112
# of assembled contigs	89
Annotation statistics	
GC content	55.60 (chromosome) + 50.80 (plasmid)
total genes	6,003
coding genes	5,561
RNAs	131

Table 2
Intrinsic antimicrobial resistance gene mechanisms (perfect and strict hits only).

Antimicrobial Resistance Mechanism	# of Hits	Representative Genes
Antibiotic inactivation	2	ORN-1 beta-lactamase, fosA5
Resistance-nodulation-cell division (RND) antibiotic efflux	7	adeF, baeR, H-NS, rsmA, CRP, marA,marR
Major facilitator superfamily (MFS) antibiotic efflux	4	H-NS, emrR, KpnG, KpnH
ATP-binding cassette (ABC) antibiotic efflux pump	2	msbA, lptD
Small multidrug resistance (SMR) antibiotic efflux	3	qacJ, KpnE, KpnF
Antibiotic target alternation	7	arnT, eptB, marR, UhpT, vanG, EF-Tu mutation, Haemophilus influenzae PBP3 conferring resistance to beta-lactam antibiotics
Bacterial porin with reduced permeability	1	marA

National Center for Biotechnology Information (NCBI) Prokaryotic Genomes Automatic Annotation Pipeline to confirm its listed genus and species identification. Genome statistics for this whole genome sequencing project are summarized in Table 1. In addition to NCBI, subsequent annotation was performed through the Rapid Annotation using Subsystem Technology (RAST) server. RAST categorized approximately 31% of chromosomal sequences and 13% of plasmid-based coding sequences into 25 of 27 total subsystems, as seen in Fig. 1. No sequences were found to apply to the motility or photosynthesis subsystems. Chromosomal sequences not assigned to a subsystem had 2,547 sequences assigned putative roles and 1,416 annotated as uncharacterized hypothetical proteins. Strain PX02's large plasmid (227,334 bp) is a member of the inch family limited to the *Enterobacteriaceae* family and was closely related to those found in *Klebsiella pneumoniae* (Fig. 2). Of 316 unique plasmid-based sequences, a total of 156 could be assigned a putative role, of which 39 were identified through at least 1 RAST designated subsystem. The remaining 170 sequences were listed as uncharacterized proteins. Several unassigned plasmid-based sequences were attributed to toxin/antitoxin systems commonly seen in *Klebsiella* sp. and *Pseudomonas aeruginosa*. A summary of identified intrinsic antimicrobial resistance biomarkers within the PX02 genome is summarized in Table 2. Antimicrobial resistance determinants were predominately grouped as antibiotic efflux and alteration resistance mechanisms. In addition, two determinants were grouped into antibiotic inactivation. These include a single ORN-1 Ambler Class A beta-lactamase highly conserved within the *R. ornithinolytica* species and a single instance of the fosA5 gene, which is a mobilized variant with increased resistance to fosfomycin [3,4]. Putatively impacted drug classes based on intrinsic (chromosomal) and acquired (plasmid) resistance is presented in Table 3.

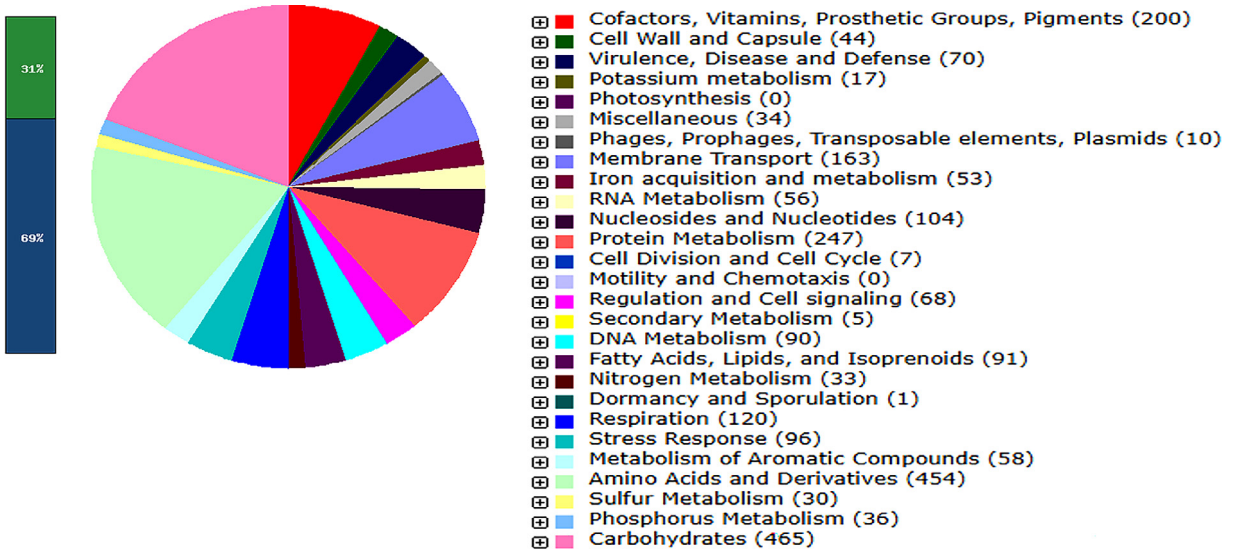

Fig. 1. Subsystem category distribution of chromosomal coding genes of *Raoultella ornithinolytica* PX02 as annotated by the Rapid Annotation using Subsystem Technology (RAST) server. The bar chart shows the subsystem coverage in percentage (the blue bar corresponds to the percentage of proteins not identified). The pie chart shows the distribution of the 27 most abundant subsystem categories.

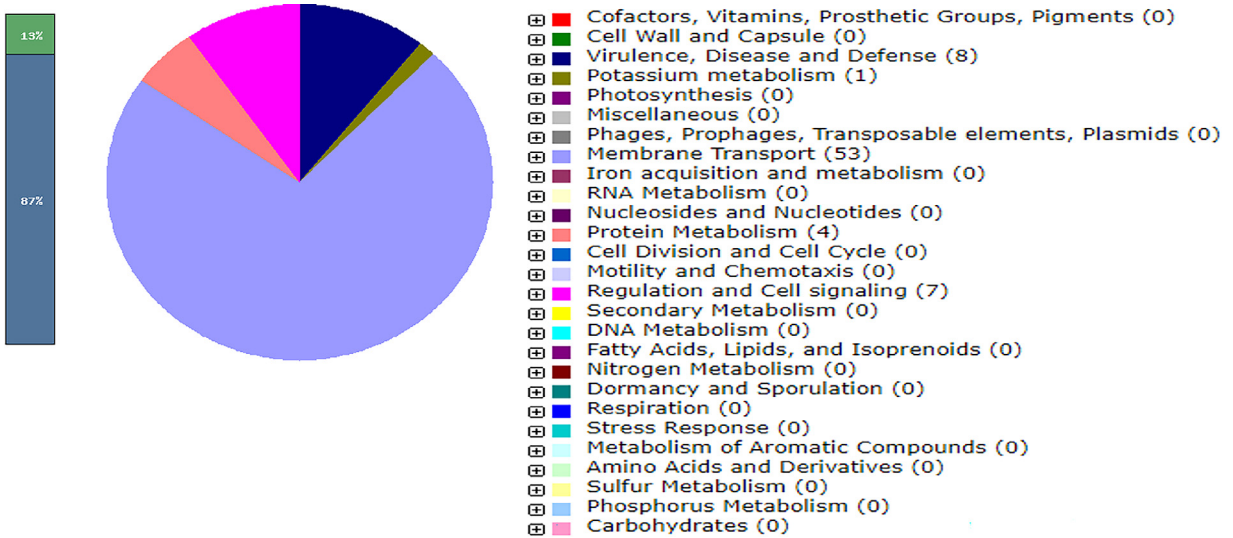

Fig. 2. Subsystem category distribution of plasmid coding genes of *Raoultella ornithinolytica* PX02 as annotated by the Rapid Annotation using Subsystem Technology (RAST) server. The bar chart shows the subsystem coverage in percentage (the blue bar corresponds to the percentage of proteins not identified). The pie chart shows the distribution of the 27 most abundant subsystem categories.

Table 3
Antibiotic drug classes putatively impacted in the *Raoultella ornithinolytica* PX02 genome.

Drug Class	# of Hits (Perfect)	# of Hits (Strict)	# of Hits (Loose)
Carbapenam	1	3	39
Cephalosporin	1	8	52
Fluoroquinolone		11	121
Tetracycline		7	130
Aminoglycoside		6	62
Aminocoumarin		2	26
Macrolide		6	91
Cephamycin		3	39
Penam		7	60
Nitroimidazole		1	10
Diaminopyrimidine		1	13
Phenicol		3	54
Rifamycin		5	23
Phosphonic acid		2	15
Elfamycin		1	4
Monobactam		1	27
Glycopeptide		1	19
Glycylcycline		2	18
Lincosamide			8
Streptogramin A			4
Streptogramin B			2
Oxazolidinone			4
Pleuromutilin			4
Bicyclomycin-like			7
Fusidane			3
Mupirocin-like			3
Nitrofuran			2
Sulfonamide			4
Isonazid-like			6
Pyrazine			2
Salicylic acid			1
Antibacterial free fatty acids			5
Disinfecting agents and antiseptics		5	57
Peptide antibiotic		7	51
Nucleoside antibiotic			13

4. Experimental Design, Materials and Methods

4.1. Reagents and Reagent Preparation

Ethyl paraoxon was purchased from Sigma Aldrich, diluted to a stock concentration of 100 mg/mL in acetonitrile, and stored at room temperature. All base reagents used to prepare carbon selective medium (CSM), composed of 2 mM nitrilotriacetic acid, 0.8 mM $MgSO_4 \cdot 7H_2O$, 0.17 mM $CaNO_3$, 0.018 mM $FeSO_4 \cdot 7H_2O$, and 20% v/v phosphate buffer, were also obtained from Sigma Aldrich and stored at room temperature.

4.2. Sample Collection

A hole approximately 12 inches in depth was dug along the shore of the San Jacinto River along the shoreline near Ilfrey Dr. in Baytown, Texas (GPS coordinates: 29.778404, -95.059108). A total of 1 gram of soil was deposited into a labeled amber centrifuge tube for transport.

4.3. Sample Screening

5 mL of carbon selective media was inoculated with ethyl paraoxon at a final 100 µg/mL concentration. Approximately 1 gram of collected soil was added to CSM medium and allowed to grow overnight at 37°C and 200 rpm. 200 µL of the resulting overnight culture was added to fresh CSM + ethyl paraoxon and incubated at 30°C and 120 rpm for one week. Successive sub-cultivations were performed over five weeks before plating onto minimal agar plates spread with ethyl paraoxon as a screening agent.

Colonies were assessed visually for signs of pesticide degradation through the discoloration of the agar resulting from the breakdown of ethyl paraoxon and forming a halo-like effect around the colony. One colony demonstrating this effect was labeled PX02, grown overnight in Luria-Bertani broth and made into glycerol stock for future preparations.

4.4. Genomic DNA Preparation

A 1.5 mL overnight culture was grown from a working preparation of the PX02 glycerol stock from the Iyer laboratory. Cellular DNA was extracted from the overnight culture using a Qiagen DNEasy blood and tissue kit according to the manufacturer's instructions for microbial cells.

4.5. Whole Genome Sequencing and Assembly

Sample DNA was shipped to Genewiz/Azenta (South Plainfield, NJ) to conduct Illumina MiSeq paired-end sequencing per the manufacturer's protocol.

4.6. Genome Annotation and Sequence Analysis

Raw sequence reads obtained from Genewiz/Azenta were checked in Fastqc [5] under default settings. Poor reads were filtered out through bbduk, a part of BBTools [6], using the following settings, minlen =50, qtrim = rl, trimq = 20, ktrim = r, k=25, mink = 11, ref=adapters.fa and hdist =1. Filtered reads were assembled with the Spades 3.10 program under default settings [7]. Annotation was performed by uploading assembled contigs to Genbank, which then transferred the data to the National Center for Biotechnology Information (NCBI) Prokaryotic Genomes Automatic Annotation Pipeline (PGAP) [8]. The latest annotation of strain PX02 was run on PGAP version 6.3 using the standard GeneMarkS-2+ annotation method. A second annotation was conducted using the Rapid Annotation using Subsystem Technology (RAST) server [9]. RAST annotation was carried out with default settings enabled. Analysis of antibiotic resistance gene determinants was conducted similarly through the Resistance Gene Identifier (RGI main) analysis tool part of the Comprehensive Antibiotic Resistance Database (CARD) with the perfect, strict, and loose selection criterion enabled and all other selections set to default [3,4]. A PERFECT hit means that the predicted gene perfectly matches a curated resistance gene at the amino acid level. PERFECT hits have a known positive impact on the organism's Minimum Inhibitory Concentration (MIC) against a particular antibiotic or class of antibiotics. STRICT hits are not exact matches to a curated sequence, but are within the detection model cut-off for that sequence as defined by CARD curators. STRICT hits are likely to provide some additional functionality in resistance. LOOSE hits are sequences below the cutoff value and their impact on resistance is likely limited [3,4].

Ethics Statement

This work does not involve human subjects or animal subjects. The authors declare that this manuscript is original work and has not been published elsewhere.

Data Availability

Draft genome sequence of Raoultella ornithinolytica PX02 (Original data) (National Center for Biotechnology Information).

CRediT Author Statement

Rupa Iyer: Conceptualization, Methodology, Data curation, Supervision; **Brian Iken:** Methodology, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare that the research was conducted without commercial or financial relationships that could be interpreted as a conflict of interest.

Acknowledgments

This work was supported by the National Institute of Standards and Technology (NIST) (G110008 58106) and by the National Science Foundation (NSF) (award no. 1726968).

References

[1] R. Hajjar, G. Ambaraghassi, H. Sebajang, F. Schwenter, S.H. Su, *Raoultella ornithinolytica*: emergence and resistance, Infect. Drug Resist. 13 (2020) 1091–1104.
[2] A. Carattoli, Plasmids and the spread of resistance, Int. J. Med. Microbiol. 303 (6-7) (2013) 298–304.
[3] 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.
[4] B.P. Alcock, A.R. Raphenya, T.T. Lau, K.K. Tsang, M. Bouchard, A. Edalatmand, W. Huynh, A.L.V. Nguyen, A.A. Cheng, S. Liu, S.Y. Min, CARD 2020: antibiotic resistance surveillance with the comprehensive antibiotic resistance database, Nucleic Acids Res. 48 (D1) (2020) D517–D525.
[5] G. de Sena Brandine, A.D. Smith. Falco: high-speed FastQC emulation for quality control of sequencing data. F1000Res. 8 (2019) 1874.
[6] B. Bushnell. BBMap short read aligner. University of California, Berkeley, California. http://sourceforge.net/projects/bbmap, 2016 (accessed 15 April 2017).
[7] 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.
[8] T. Tatusova, M. DiCuccio, A. Badretdin, V. Chetvernin, S. Ciufo, W. Li, Prokaryotic genome annotation pipeline, The NCBI Handbook, National Center for Biotechnology Information, Bethesda, MD, 2013 second ed.https://www.ncbi.nlm.nih.gov/books/NBK174280/.
[9] 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 Genom. 9 (2008) 75.