Data in Brief 53 (2024) 110117

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

Whole genome sequencing data of *Acinetobacter venetianus* JKSF06 collected from Houston ship channel sediment in La Porte, Texas

Rupa Iyer*, Brian Iken

Division of Research, Innovation and Economic Development, Tarleton State University, Stephenville, TX 76401, United States

ARTICLE INFO

Article history:
Received 12 April 2023
Revised 22 December 2023
Accepted 23 January 2024
Available online 1 February 2024

Dataset link: Draft Genome sequences of Acinetobacter venetianus strain JKSF06 (Original data)

Keywords:
Acinetobacter venetianus
Alkane degradation
Emulsan
Houston ship channel
Whole genome

ABSTRACT

Acinetobacter venetianus is a Gram-negative, mesophilic bacterium that thrives in aquatic environments. Here, we present the whole genome sequence of *A. venetianus* JKSF06, isolated from sediment that was collected in La Porte, Texas, near the southern terminus of the Houston Ship Channel into the Gulf of Mexico. The JKSF06 strain harbors multiple xenobiotic gene determinants targeting environmental waste that can be found here, including petroleum hydrocarbons and n-alkanes. In addition, JKSF06 can actively degrade organophosphate phophotriesters such ethyl paraoxon. In total, the genome of JKSF06 consists of 3,462,857 bp encoding for 3173 putative proteins. The complete sequence of *A. venetianus* JKSF06 can be viewed under accession LSVD00000000.1 through the National Center for Biotechnology Information (NCBI).

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

https://doi.org/10.1016/j.dib.2024.110117

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
Data format	Raw and analyzed
Description of data collection	Cellular DNA from an unknown bacterial strain isolated from Houston Ship Channel sediment was extracted using a DNEasy blood and tissue kit and sequenced using Illumina MiSeq technology.
Data source location	Sediment: Houston Ship Channel off of Morgan's Point in La Porte, Texas (29.678722, -94.982404)
Data accessibility	Strain data is uploaded to GenBank database under accession LSVD00000000.1. Raw sequences data uploaded to Sequence Read Archive under accession SRR25375834. **Direct link to data:** https://www.ncbi.nlm.nih.gov/Traces/wgs/NJGC01?val=LSVD01

1. Value of the Data

- The genome data of *A. venetianus* JKSF06 reveals a possible candidate for the bioremediation of crude oil and other petroleum hydrocarbons through its biodegradative pathways and putative emulsifier gene cluster.
- While *A. venetianus* JKSF06 does share many of the biomarkers and pathways reported in other strains of this species, there are notable differences in organization, structure, and underlying sequence of the associated genes and gene clusters. This sequence data, therefore, also benefit efforts to determine the various mechanisms involved in this microorganism's capacity for petroleum hydrocarbon biodegradation.
- *A. venetianus* JKSF06 can degrade organophosphate phosphotriesters and thiophosphates to a lesser degree. This sequence data can aid in our understanding of how this novel characteristic not previously reported in the species has evolved or has been acquired.

2. Objective

Acinetobacter venetianus is a Gram-negative bacterium belonging to the family *Moraxellaceae*. They are noted for their capacity to adhere to and degrade a broad range of petroleum hydrocarbons, even in complex mixtures such as crude oil and diesel fuel [1]. Aromatic biodegradation has also been reported in environments with heavy pollution [1]. One such polluted environment could include a portion of the San Jacinto River through which the Houston Ship Channel flows serving shipping traffic for the port of Houston, Texas, and surrounding communities. This region has seen the inundation of recalcitrant environmental waste from nearby illegal dumping sites, including pesticides and dioxin-like aromatic waste. Commercial shipping currently presents another factor in the region, as the oil and gas industry, and by association, petroleum hydrocarbons are prevalent here as well. Our objective for this project was to assess the genome of *A. venetianus* JKSF06 first to understand its potential for xenobiotic degradation of the environmental waste common to this region and second to gain insight into the genetic basis underlying the petroleum hydrocarbon biodegradation pathways common to this *Acinetobacter* species.

3. Data Description

This data comprises a total of 53 assembled contigs for *A. venetianus* JKSF06. Genome statistics have been summarized in Table 1. Annotation was performed through the National Cen-

Table 1
Acinetobacter venetianus JKSF06 genome statistics.

Assembly statistics	
platform	Illumina MiSeq (2*250) paired-end
# of paired (forward and reverse) raw sequence reads	3,594,783
# of base pairs (bp) in raw sequence	1.4 G (billion)
# of contigs after assembly	53
Contig N50	319,182
Contig L50	4
genome size (bp)	3,462,857 bp
average coverage	397.00
Annotation statistics	
GC content %	39.1
total genes	3321
total protein-coding sequences	3244
functional protein-coding genes	3173
total RNAs	77

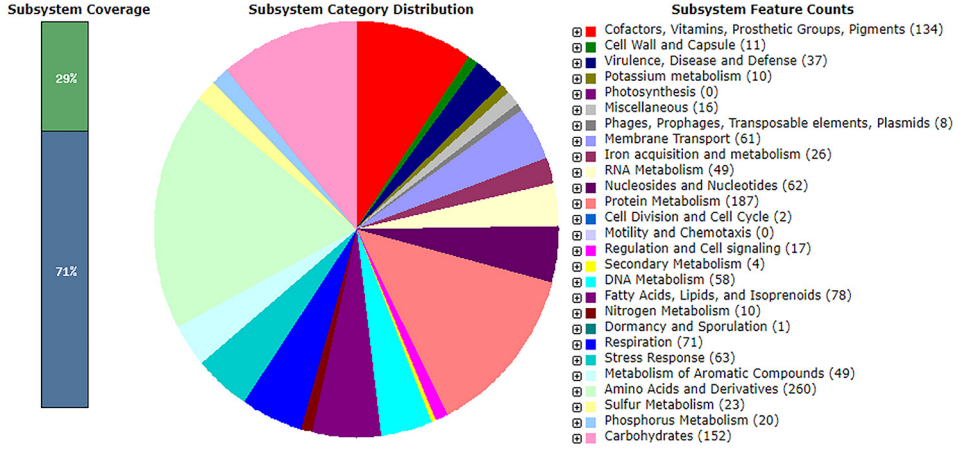

Fig. 1. Subsystem category distribution of protein-coding genes of *Acinetobacter venetianus* JKSF06 as annotated by the Rapid Annotation using Subsystem Technology (RAST) server. The bar chart shows the subsystem coverage in percentage (the blue bar corresponds to the percentage of proteins not identified). The pie chart shows the distribution of the 27 most abundant subsystem categories. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

ter for Biotechnology Information (NCBI) Prokaryotic Genomes Automatic Annotation Pipeline and the Rapid Annotation using Subsystem Technology (RAST) server. Average Nucleotide Identity (ANI) analysis identified JKSF06 as *A. venetianus* (97.52%) in comparison to type strain *A. venetianus* RAG-1 = CIP 110063. The assignment of annotated sequences covering approximately 29% of the JKSF06 genome into 27 different subsystems is shown in Fig. 1. While no motility genes were assigned to the associated subsystem during annotation, as seen in Fig. 1, a manual search of features within RAST subsystems using key terms “motility” and “chemotaxis” identified several pilus and chemotaxis-associated genes involved in twitching motility. These include *alkN* and type IV pilus biogenesis proteins. Proteins associated with hydrocarbon metabolism in *A. venetianus* species and biomarkers for organophosphate biodegradation were also revealed through BLAST analysis of the unassigned sequences. Aromatic metabolism was primarily limited to the beta-ketoadipate pathway common to many bacteria, with no significant hits for biomarker genes associated with dioxin-like compounds. All relevant n-alkane and hydrocarbon biodegradation biomarkers found in the JKSF06 strain are shown in Table 2. The *wee* gene clus-

Table 2
Long-chain alkane degradation pathway genes found in *Acinetobacter venetianus* JKSF06.

Gene designation	Name	Representative species
alkB	Alkane-1-monooxygenase	*Acinetobacter venetianus* RAG-1
alkH	Aldehyde dehydrogenase	*Acinetobacter venetianus* RAG-1
alkJ	Alcohol dehydrogenase	*Acinetobacter venetianus* RAG-1
alkK	Acyl-CoA synthetase	*Acinetobacter venetianus* RAG-1
alkN	Methyl-accepting chemotaxis protein	*Pseudomonas putida* GPo-1
alkT	Rubredoxin-NAD (+) reductase	*Acinetobacter venetianus* RAG-1
RubA	Rubredoxin	*Alcanivorax borkumensis* SK2
RubB	Rubredoxin reductase	*Alcanivorax borkumensis* SK2
AlmA	NAD(P)/FAD-dependent oxidoreductase	*Acinetobacter* sp. DSM 17874
LadA	Long-chain alkane monooxygenase	*Geobacter thermodenitrificans*
Ald1	Aldehyde dehydrogenase family protein	*Acinetobacter* sp. M-1

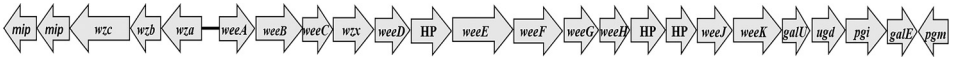

Fig. 2. Structure of the emulsan biosynthesis cluster found in *Acinetobacter venetianus* JKSF06. HP denotes the placement of a gene encoding for a hypothetical protein that does not share sequence or functionality with any known gene within this cluster.

ter responsible for synthesizing the emulsan polysaccharide is presented in Fig. 2. The structure of this cluster is most similar to *A. venetianus* RAG-1. However, the cluster appears to be missing *wzy*, which encodes for a polymerase that is believed to play a significant role in lipid polysaccharide biosynthesis in Gram-negative bacteria. In addition, at least three unknown hypothetical genes do not share similarity to other known sequences found within the *wee* gene cluster of *Acinetobacter* species.

4. Experimental Design, Materials, and Methods

4.1. Reagents and reagent preparation

All reagents were purchased from Sigma-Aldrich. The carbon selective medium (CSM) described below was prepared at 100 mL total volume using the following recipe: 95 mL of water, 200 µL of 1M NTA, and 100 µL of 20% (w/v) $MgSO_4$ $7H_2O$ was mixed, autoclaved and allowed to sit until at room temperature. Then 100 µL of 4% (w/v) $CaNO_3$, 100 µL of 0.5% (w/v) $FeSO_4$ $6H_2O$ and 4.5 mL of phosphate buffer was added to complete the solution.

4.2. Sample collection

Soil/sediment samples were collected from a hole dug 0.3 m along the San Jacinto River shoreline near Morgan's Point in La Porte, Texas (29.678722, -94.982404).

4.3. Sample screening

Approximately 1 g of collected soil was added to carbon selective medium (CSM) supplemented with ethyl paraoxon at a 100 µg/mL concentration. The resulting overnight culture was then subcultivated into 5 mL of fresh CSM and left to incubate at 30 °C and 120 rpm. This process was repeated once a week for five consecutive weeks. Aliquots of 100 µL from the last sub-cultivation were spread onto ethyl paraoxon inoculated agar plates. Growing colonies were

then assessed for ethyl paraoxon degradation based on discoloration of the surrounding agar due to the presence of the primary metabolite, *p*-nitrophenol. A colony with the desired effect was labeled JKSF06 and made into a glycerol stock.

4.4. Genomic DNA preparation

To check for purity prior to extraction, a standard non-selective LB agar plate was streaked from an aliquot of glycerol stock and allowed to grow for 72 h. In the absence of any visual contamination, total cellular DNA from an overnight culture grown from the same glycerol stock was extracted with a Qiagen DNEasy blood and tissue kit and associated protocol. Isolated DNA was shipped to Genewiz/Azenta (South Plainfield, NJ) for whole genome sequencing.

4.5. Whole genome sequencing

Illumina paired-end sequencing was performed by Genewiz/Azenta (South Plainfield, NJ).

4.6. Genome assembly, annotation, and sequence analysis

Raw sequence reads were quality checked [2], filtered [3], and assembled into 53 contigs with SPAdes [4]. Assembled contigs were uploaded to the National Center for Biotechnology Information (NCBI) Prokaryotic Genomes Automatic Annotation Pipeline [5] for annotation and species clarification using average nucleotide identity (ANI) analysis [6,7]. The Rapid Annotation using Subsystem Technology (RAST) server was used as a supplemental source for genomic annotation [8]. BLAST analysis was then used to search the uploaded JKSF06 genome for the presence of hydrocarbon and alkane degradation biomarkers [9–11].

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 sequences of Acinetobacter venetianus strain JKSF06 (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.

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).

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.

References

[1] M. Fondi, I. Maida, E. Perrin, V. Orlandini, L. La Torre, E. Bosi, A. Negroni, G. Zanaroli, F. Fava, F. Decorosi, L. Giovannetti, Genomic and phenotypic characterization of the species *Acinetobacter venetianus*, Sci. Rep. 6 (1) (2016) 21985.
[2] B. Bioinformatics, FastQC A Quality Control Tool for High throughput Sequence Data, Babraham Institute, Cambridge, United Kingdom, 2011 http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
[3] B. Bushnell. BBMap short read aligner. (2016), http://sourceforge.net/projects/bbmap.
[4] 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.
[5] 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.
[6] S. Ciufo, S. Kannan, S. Sharma, A. Badretdin, K. Clark, S. Turner, S. Brover, C.L. Schoch, A. Kimchi, M. DiCuccio, Using average nucleotide identity to improve taxonomic assignments in prokaryotic genomes at the NCBI, Int. J. Syst. Evol. Microbiol. 68 (7) (2018) 2386–2392.
[7] S. Kannan, S. Sharma, S. Ciufo, K. Clark, S. Turner, P.A. Kitts, C.L. Schoch, M. DiCuccio, A. Kimchi, Collection and curation of prokaryotic genome assemblies from type strains at NCBI, Int. J. Syst. Evol. Microbiol. 73 (1) (2023) 005707.
[8] 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.
[9] S.F. Altschul, T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller D.J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res. 25 (17) (1997) 3389–3402.
[10] S.F. Altschul, J.C. Wootton, E.M. Gertz, R. Agarwala, A. Morgulis, A.A. Schäffer, Y.K. Yu, Protein database searches using compositionally adjusted substitution matrices, FEBS J. 272 (20) (2005) 5101–5109.
[11] A. Marchler-Bauer, Y. Bo, L. Han, J. He, C.J. Lanczycki, S. Lu, F. Chitsaz, M.K. Derbyshire, R.C. Geer, N.R. Gonzales, M. Gwadz, CDD/SPARCLE: functional classification of proteins via subfamily domain architectures, Nucleic Acids Res. 45 (D1) (2017) D200–D203.