Genomics Data 14 (2017) 40–43

Contents lists available at ScienceDirect

Genomics Data

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

Data in Brief

Genome of *Pseudomonas nitroreducens* DF05 from dioxin contaminated sediment downstream of the San Jacinto River waste pits reveals a broad array of aromatic degradation gene determinants

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

[a] Center for Life Sciences Technology, Engineering Technology, University of Houston, 300 Technology Building, Houston, TX 77204, USA
[b] Department of Pediatrics-Tropical Medicine, Baylor College of Medicine, Houston, TX, USA

ABSTRACT

P. nitroreducens DF05 is a Gram negative, motile, aerobic, rod-shaped and psychrotrophic bacterium that was isolated from contaminated San Jacinto River sediment near River Terrace Park in Channelview, Texas. The draft genome of strain DF05 consists of a total of 192 contigs assembled at the scaffold level totaling 6,487,527 bp and encoding for 5862 functional proteins, 1116 of which are annotated as hypothetical proteins. The bacterial chromosome has a GC content of 65.15% and contains 22 rRNA and 70 tRNA loci. In addition, approximately 142 proteins localized on the bacterial chromosome are associated with metabolism of aromatic compounds. A single plasmid approximately 95 kb in size was also detected carrying copies of RNA genes and multiple phage assembly proteins.

Specifications	
Organism/cell line/tissue	*Pseudomonas nitroreducens*
Strain	DF05
Sequencer or array type	Illumina Miseq
Data format	Analyzed
Experimental factors	Bacterial strain
Experimental features	Whole genome analysis and gene annotation of DF05
Sample source location	San Jacinto River sediment near River Terrace Park (Channelview, Texas)
GPS coordinates	29.781723, −95.103015

1. Direct link to deposited data

https://www.ncbi.nlm.nih.gov/nuccore/NJBA00000000.

2. Experimental design, materials and methods

Sediment was collected from the banks of the San Jacinto River in Terrace River Park southwest of submerged dioxin-laden waste pits that are the source of widespread contamination across the San Jacinto River system [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 dibenzofuran. 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 dibenzofuran. Resultant colonies, yellow-brown in coloration, were then 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

* Corresponding author.
E-mail addresses: riyer@uh.edu (R. Iyer), bniken@central.uh.edu (B. Iken), ashish.damania@bcm.edu (A. Damania).

http://dx.doi.org/10.1016/j.gdata.2017.07.011
Received 26 June 2017; Received in revised form 13 July 2017; Accepted 19 July 2017
Available online 17 August 2017

Table 1
P. nitroreducens DF05 genome statistics.

Assembly statistics	
Platform	Illumina MiSeq (2 ∗ 250) paired end
Total raw reads	2,918,773
Total filtered reads	2,763,660
Genome size (bp)	6,487,527
Number of contigs	186 (chromosome) + 9 (plasmid)
Largest contig	918,767
Contigs (≥ 200 bp)	195
Contigs (≥ 500 bp)	41
N50	729,875 (Based on contigs of size ≥ 500)
L50	4 (Based on contigs of size ≥ 500)
Average coverage	158.24 ×
Annotation statistics	
GC content	65.15 (chromosome) + 57.76 (plasmid)
Total genes	6023
Coding genes	5862
rRNAs	22
tRNAs	70

Technologies, Palo Alto, CA, USA) and quantified using the Qubit 2.0 Fluorometer.

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). Whole genome sequencing of bacterial strains was performed through Illumina MiSeq paired-end sequencing (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]. The Mash program [5] was used for species identification using k = 21 and sketch size of 1000 against the Mash Refseq (release 70) database. Fasta files for the five top bacterial hits sorted by distance were downloaded from RefSeq database and used to calculate the Mash distance. The resulting distance file from the previous step was imported into R using the readr package [6]. Finally, the Ggdendrogram [7] package was used to create a dendrogram plot through hclust function output using UPGMA method (Fig. 1). The Quast program [8] was used to calculate assembly statistics using scaffold mode Preliminary reference based annotation using PATRIC [9] web resources was carried out to identify conserved pathways. Final *de novo* annotation was performed with the NCBI Prokaryotic Genomes Automatic Annotation Pipeline (http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html) and RAST server [10,11].

3. Data description

Sequence analysis identifies this isolate, designated as DF05, as a novel strain of *Pseudomonas nitroreducens*, a member of the *Pseudomonadaceae* family from the *Gammaprotebacteria* class, strongly related to *P. fluorescens*, *P. aeruginosa* and *P. putida* species. This Gram-negative soil bacterium is ubiquitously distributed throughout the environment and primarily known for its denitrifying capabilities and propensity for petroleum hydrocarbons and/or crude oil [12,13]. According to RAST (Rapid Annotation using Subsystem Technology) and MASH genome analysis, the closest relative to *P. nitroreducens* DF05 is *P. nitroreducens* TX1, an isolate collected from rice field drainage in China with the capacity to degrade nonionic surfactants and alkylphenols (Fig. 1) [14]. RAST analysis *P. nitroreducens* DF05 (Fig. 2) identified a total of 142 genes associated with the metabolism of aromatics including 61 genes involved in the metabolism of central aromatic metabolites, 71 genes involved in peripheral aromatic degradation pathways and 10 genes involved in amino acid catabolism and gentisate/salicylate degradation pathways. Representative subsystems include *n*-heterocyclic aromatics found in crude oil and tar, *n*-phenylalkanoic acid, benzoate, hydroxybenzoate and chloroaromatics as well as both the catechol and protocatechuate branches of the β-ketoadipate pathway (Table 2). In addition, *P. nitroreducens* DF05 also possesses specialized genes involved in alkylphenol and halogenated aromatic degradation including alcohol dehydrogenase, aldehyde dehydrogenase, phenol hydroxylase, 2,3 biphenyl dioxygenase as well as multiple dehalogenase enzymes. Taken together, this data suggest that this microorganism may have significant value in the bioremediation of aromatic xenobiotics.

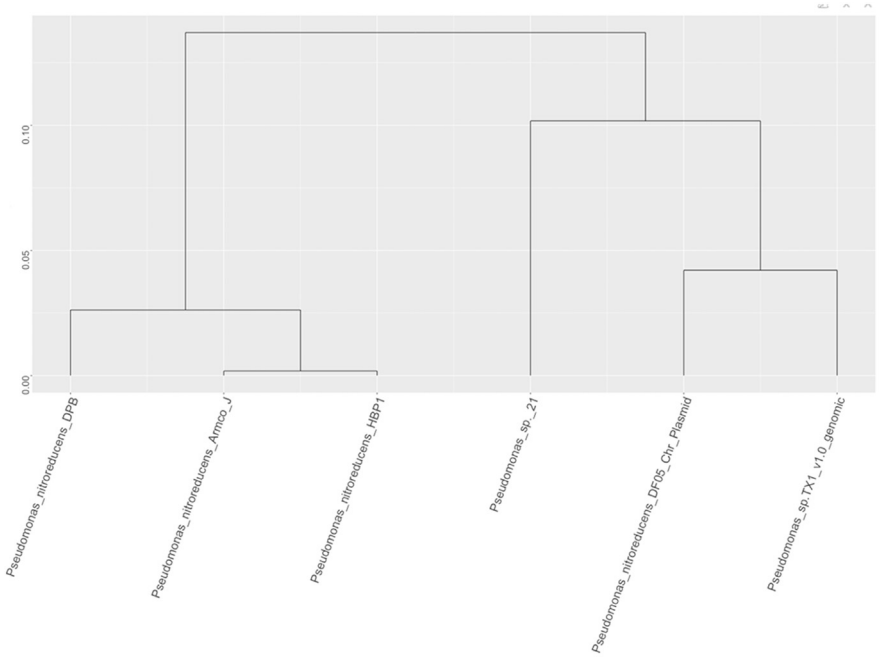

Fig. 1. Dendogram of *P. nitroreducens* DF05 and the five closest neighboring genomes.

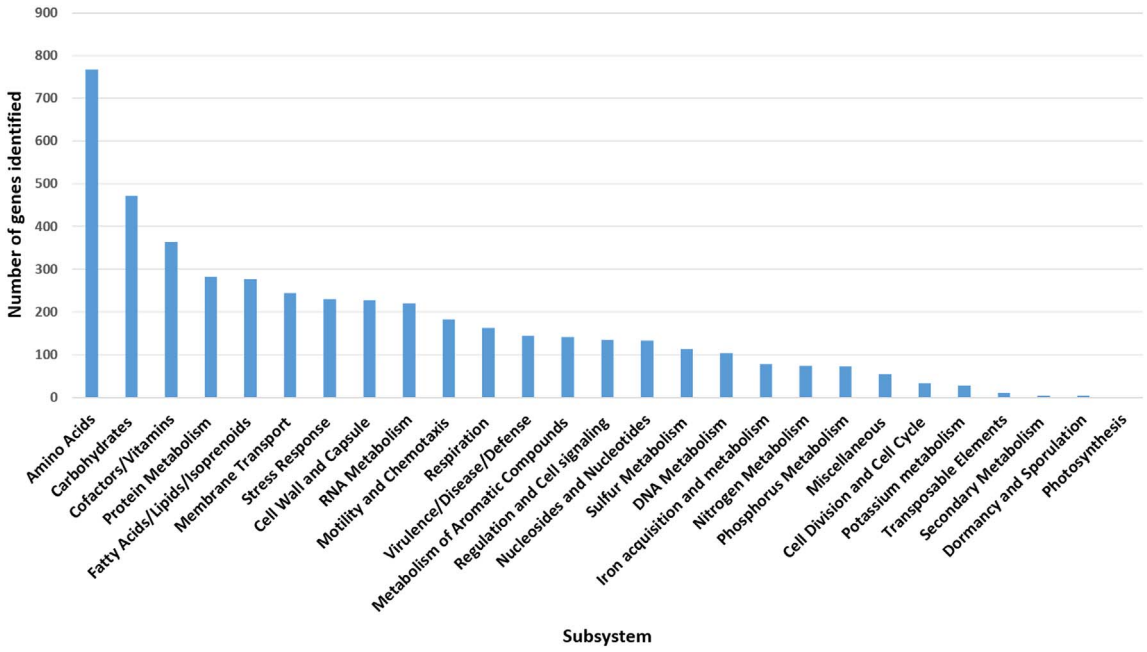

Fig. 2. Subsystem category distribution of major protein coding genes of *P. nitroreducens* DF05 as annotated by the RAST annotation server. The bar chart shows the distribution of the 27 most abundant subsystem categories.

Table 2
Genes identified by RAST in the metabolism of aromatic compounds.

Category	Representative genes
Aromatic amino catabolism	Aldehyde dehydrogenase (EC 1.2.1.3), PaaZ
Aromatic amino catabolism	4-Hydroxyphenylacetate 3-monooxygenase, reductase component (EC 1.6.8.-)
Aromatic amino catabolism	Nitrilotriacetate monooxygenase component B (EC 1.14.13.-)
Benzoate degradation	Benzoate 1,2-dioxygenase alpha subunit (EC 1.14.12.10)
Benzoate degradation	Benzoate 1,2-dioxygenase beta subunit (EC 1.14.12.10)
Benzoate degradation	Benzoate 1,2-dioxygenase, ferredoxin reductase component
Benzoate degradation	Benzoate transport protein
Benzoate degradation	benABC operon transcriptional activator BenR
Benzoate degradation	1,2-Dihydroxycyclohexa-3,5-diene-1-carboxylate dehydrogenase (EC 1.3.1.25)
Catechol branch of beta-ketoadipate pathway	3-Oxoadipate CoA-transferase subunit A (EC 2.8.3.6)
Catechol branch of beta-ketoadipate pathway	3-Oxoadipate CoA-transferase subunit B (EC 2.8.3.6)
Catechol branch of beta-ketoadipate pathway	Succinyl-CoA:3-ketoacid-coenzyme A transferase subunit A (EC 2.8.3.5)
Catechol branch of beta-ketoadipate pathway	Succinyl-CoA:3-ketoacid-coenzyme A transferase subunit B (EC 2.8.3.5)
Catechol branch of beta-ketoadipate pathway	Muconate cycloisomerase (EC 5.5.1.1)
Catechol branch of beta-ketoadipate pathway	Beta-ketoadipate enol-lactone hydrolase (EC 3.1.1.24)
Catechol branch of beta-ketoadipate pathway	Catechol 1,2-dioxygenase (EC 1.13.11.1)
Catechol branch of beta-ketoadipate pathway	Muconolactone isomerase (EC 5.3.3.4)
Central meta-cleavage pathway of aromatic compound degradation	2-Hydroxy-6-oxo-6-phenylhexa-2,4-dienoate hydrolase (EC 3.7.1.-)
Central meta-cleavage pathway of aromatic compound degradation	5-Carboxymethyl-2-hydroxymuconate delta-isomerase (EC 5.3.3.10)
Central meta-cleavage pathway of aromatic compound degradation	4-Oxalocrotonate tautomerase (EC 5.3.2.-)
Central meta-cleavage pathway of aromatic compound degradation	2-Polyprenylphenol hydroxylase and related flavodoxin oxidoreductases
Chloroaromatic degradation pathway	Beta-ketoadipate enol-lactone hydrolase (EC 3.1.1.24)
Chloroaromatic degradation pathway	Beta-ketoadipyl CoA thiolase (EC 2.3.1.-)
Gentisate degradation	putative 4-hydroxybenzoyl-CoA thioesterase
Gentisate degradation	4-Hydroxybenzoate transporter
Gentisate degradation	Maleylacetoacetate isomerase (EC 5.2.1.2)
Gentisate degradation	Fumarylacetoacetate hydrolase family protein
Protocatechuate branch of beta-ketoadipate pathway	dicarboxylic acid transporter PcaT
N-heterocyclic aromatic compound degradation	Isoquinoline 1-oxidoreductase alpha subunit (EC 1.3.99.16)
N-heterocyclic aromatic compound degradation	Isoquinoline 1-oxidoreductase beta subunit (EC 1.3.99.16)
n-Phenylalkanoic acid degradation	3-Hydroxybutyryl-CoA epimerase (EC 5.1.2.3)
n-Phenylalkanoic acid degradation	Long-chain-fatty-acid–CoA ligase (EC 6.2.1.3)
n-Phenylalkanoic acid degradation	Enoyl-CoA hydratase (EC 4.2.1.17)
n-Phenylalkanoic acid degradation	3-Ketoacyl-CoA thiolase (EC 2.3.1.16)
n-Phenylalkanoic acid degradation	Delta(3)-cis-delta(2)-trans-enoyl-CoA isomerase (EC 5.3.3.8)
n-Phenylalkanoic acid degradation	3-Hydroxyacyl-CoA dehydrogenase (EC 1.1.1.35)
n-Phenylalkanoic acid degradation	enoyl-CoA hydratase, R-specific
p-Hydroxybenzoate degradation	4-Hydroxybenzoate transporter
p-Hydroxybenzoate degradation	P-hydroxybenzoate hydroxylase (EC 1.14.13.2)
Protocatechuate branch of beta-ketoadipate pathway	4-Carboxymuconolactone decarboxylase (EC 4.1.1.44)
Protocatechuate branch of beta-ketoadipate pathway	Succinyl-CoA:3-ketoacid-coenzyme A transferase (EC 2.8.3.5)

(continued on next page)

Table 2 *(continued)*

Category	Representative genes
Protocatechuate branch of beta-ketoadipate pathway	Beta-ketoadipyl CoA thiolase (EC 2.3.1.-)
Protocatechuate branch of beta-ketoadipate pathway	PhaK-like protein
Protocatechuate branch of beta-ketoadipate pathway	Pca regulon regulatory protein PcaR
Protocatechuate branch of beta-ketoadipate pathway	Protocatechuate 3,4-dioxygenase alpha chain (EC 1.13.11.3)
Protocatechuate branch of beta-ketoadipate pathway	Protocatechuate 3,4-dioxygenase beta chain (EC 1.13.11.3)
Protocatechuate branch of beta-ketoadipate pathway	Beta-ketoadipate enol-lactone hydrolase (EC 3.1.1.24)
Protocatechuate branch of beta-ketoadipate pathway	3-Carboxy-cis,cis-muconate cycloisomerase (EC 5.5.1.2)
Quinate degradation	Quinate/shikimate dehydrogenase [Pyrroloquinoline-quinone] (EC 1.1.99.25)
Quinate degradation	3-Dehydroquinate dehydratase II (EC 4.2.1.10)
Salicylate and gentisate catabolism	Fumarylacetoacetase (EC 3.7.1.2)
Salicylate and gentisate catabolism	4-Hydroxybenzoate transporter
Salicylate and gentisate catabolism	Maleylacetoacetate isomerase (EC 5.2.1.2)
Salicylate and gentisate catabolism	Salicylate hydroxylase (EC 1.14.13.1)
Salicylate and gentisate catabolism	Fumarylacetoacetate hydrolase family protein
Salicylate ester degradation	Salicylate hydroxylase (EC 1.14.13.1)

Author contributions

RI helped prepare samples and draft the manuscript. AD carried out the *de novo* assembly and gene annotation. BI constructed the phylogenetic tree and helped with drafting the manuscript. All authors read and approved the final text.

Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgments

Funding for whole genome sequencing of *P. nitroreducens* DF05 was provided by the National Institute of Standards and Technology (NIST) (G110008 58106).

References

[1] R. Iyer, J. Aggarwal, B. Iken, A review of the Texas, USA San Jacinto Superfund site and the deposition of polychlorinated dibenzo-p-dioxins and dibenzofurans in the San Jacinto River and Houston Ship Channel, Environ. Sci. Pollut. Res. 23 (2016) 23321–23338.

[2] B. Bioinformatics, FastQC a Quality Control Tool for High Throughput Sequence Data, Babraham Institute, Cambridge, United Kingdom, 2011http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

[3] B. Bushnell, BBMap Short Read Aligner. http://sourceforge.net/projects/bbmap, (2016).

[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] B.D. Ondov, T.J. Treangen, P. Melsted, A.B. Mallonee, N.H. Bergman, S. Koren, A.M. Phillippy, Mash: fast genome and metagenome distance estimation using MinHash, Genome Biol. 17 (1) (2016) 132.

[6] A. Gurevich, V. Saveliev, N. Vyahhi, G. Tesler, QUAST: quality assessment tool for genome assemblies, Bioinformatics 29 (8) (2013) 1072–1075.

[7] H. Wickham, J. Hester, R. Francois, Readr: Read Tabular Data. https://CRAN.R-project.org/package=readr, (2016).

[8] A. de Vries, B.D. Ripley, ggdendro: Create Dendrograms and Tree Diagrams Using 'ggplot2'. R package version 0.1–20. https://CRAN.R-project.org/package=ggdendro, (2016).

[9] 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.

[10] 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.

[11] 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.

[12] E. Lang, B. Griese, C. Spröer, P. Schumann, M. Steffen, S. Verbarg, Characterization of '*Pseudomonas azelaica*' DSM 9128, leading to emended descriptions of *Pseudomonas citronellolis* Seubert 1960 (approved lists 1980) and *Pseudomonas nitroreducens* Iizuka and Komagata 1964 (approved lists 1980), including *Pseudomonas multiresinivorans* as its later heterotypic synonym, Int. J. Syst. Evol. Microbiol. 57 (4) (2007) 878–882.

[13] C.O. Onwosi, F.J.C. Odibo, Effects of carbon and nitrogen sources on rhamnolipid biosurfactant production by *Pseudomonas nitroreducens* isolated from soil, World J. Microbiol. Biotechnol. 28 (3) (2012) 937–942.

[14] S.L. Huang, H. Chen, A. Hu, N.N. Tuan, C.P. Yu, Draft genome sequence of *Pseudomonas nitroreducens* strain TX1, which degrades nonionic surfactants and estrogen-like alkylphenols, Genome Announc. 2 (1) (2014) e01262–13.