Lawrence Berkeley National Laboratory
LBL Publications

Title

Complete Genome Sequences of Five Isolated Pseudomonas Strains that Catabolize
Pentose Sugars and Aromatic Compounds Obtained from Lignocellulosic Biomass

Permalink

https://escholarship.org/uc/item/0108v3d2

Journal

Microbiology Resource Announcements, 11(4)

ISSN

2169-8287

Authors

Park, Mee-Rye
Fong, Bonnie
Tofaha, Taqwa
et al.

Publication Date

2022-04-21

DOI

10.1128/mra.00987-21

Copyright Information

Peer reviewed

Complete Genome Sequences of Five Isolated *Pseudomonas* Strains that Catabolize Pentose Sugars and Aromatic Compounds Obtained from Lignocellulosic Biomass

Mee-Rye Park,[a,b] Bonnie Fong,[a,b] Taqwa Tofaha,[a,b,c] Blake A. Simmons,[a,b] Steven Singer[a,b]

[a]Joint BioEnergy Institute, Emeryville, California, USA
[b]Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA
[c]University of California, Berkeley, Berkeley, California, USA

ABSTRACT We report on complete genome sequences of five *Pseudomonas* soil isolates that are capable of metabolizing pentose sugars and aromatic monomers. These complete genome sequence data provide insight into possible alternative hosts for the production of biofuels and bio-based chemicals from lignocellulosic feedstock.

Lignocellulosic biomass from plants is the most abundant and renewable source available for bioconversion (1). *Pseudomonas putida* KT2440 is a promising host for the production of biofuels and bio-based chemicals, which are currently produced from lignocellulosic hydrolysates (2–5). There has been growing interest in maximizing the range of biomass components to include pentose sugars (e.g., xylose and arabinose), the most abundant components of hemicellulose from grasses (6, 7). However, *P. putida* KT2440 lacks the native ability to metabolize pentose sugars. While various approaches have been used to utilize pentose sugars through the heterologous expression of pentose sugar pathways in *P. putida* KT2440 (6–9), several limitations, such as low growth rate, long lag phase, and phenotypic instability, remain.

Here, we report five *Pseudomonas* isolates recovered from soils from different sites in Emeryville, California, that grow on pentose sugars. Soil samples were inoculated into M9 medium at approximately 2.5% (wt/vol). Serial dilutions were initially plated onto *Pseudomonas* isolation agar (PIA). Visible colonies were restreaked on M9 minimal media agar plates containing 0.5% (wt/vol) xylose and then on plates with 0.5% (wt/vol) *p*-coumarate as the sole carbon and energy source and were incubated at 30°C. Single bacterial colonies were picked and restreaked on the same medium several times for purification. The growth of each colony was monitored overnight at 30°C in liquid minimal medium (2) supplemented with 0.5% (wt/vol) glucose, xylose, or *p*-coumarate as the sole carbon source. Depending on the growth rate, end optical density at 600 nm (OD_{600}), and lag phase, five isolates were finally selected.

The soil isolates were grown overnight at 30°C in 5 mL LB broth with agitation for the isolation of high-molecular-weight genomic DNA as described previously (10). Pacific Biosciences (PacBio) SMRTbell library preparation (>10 kb, multiplexed) and long-read sequencing using the PacBio Sequel platform (11) were performed by the DOE Joint Genome Institute (JGI). The PacBio reads were filtered to remove reads missing dumbbells on the ends using BBTools (12). Reads of >5 kb were assembled with the Hierarchical Genome Assembly Process (HGAP) v4 (1.0) (smrtlink/8.0.0.80529) (13). Prodigal (14) was used to predict coding sequences (CDSs) on each contig, and the output protein sequences were aligned to the NCBI nonredundant database using DIAMOND (15). Contigs with a probability of being a plasmid were identified using TensorFlow (16). Gene annotations were completed within the JGI Integrated Microbial Genomes (IMG) platform (17) and KBase. Default parameters were used for all software. The

TABLE 1 Genome sequence statistics and characteristics for the five isolates

Pseudomonas isolate	Raw sequencing results				Assembly results			Annotation results		
	No. of >5-kb reads	Mean read length for >5-kb reads (bp)	GC content (%)	Coverage (×)	Genome size (bp)	No. of contigs	No. of plasmids	No. of CDSs	No. of tRNAs	No. of rRNAs
M2	430,129	10,088	61.8	202.1	5,737,635	1	1	5,281	75	22
M5	659,815	10,907	61.9	219.0	5,442,015	1	0	4,903	76	22
BP6	837,793	12,136	61.6	198.2	5,928,556	1	1	5,312	77	22
BP7	763,684	11,060	61.6	202.6	5,979,470	1	1	5,397	77	22
BP8	655,709	10,770	61.8	197.3	6,004,477	1	0	5,341	70	22

sequence details are given in Table 1. Key structural features, including GC content, GC skew, and CDSs, are graphically depicted in Fig. 1.

The average nucleotide identities (ANIs) based on the whole-genome sequences were calculated using FastANI (18). One set of isolates (*Pseudomonas* sp. strains M2 and M5) and *Pseudomonas* sp. strain BP8 showed 85.6% and 84.6% ANI, respectively, to *P. putida* KT2440, whereas the second set of isolates (*Pseudomonas* sp. strains BP6 and BP7) showed 96.2% ANI to *P. putida* KT2440. The genome sequences of the isolates will contribute to the understanding and exploration of metabolic pathways of the main carbon sources derived from lignocellulosic biomass and will facilitate genetic engineering.

Data availability. The whole-genome sequences for each of the five *Pseudomonas* species have been deposited in GenBank under the following accession numbers:
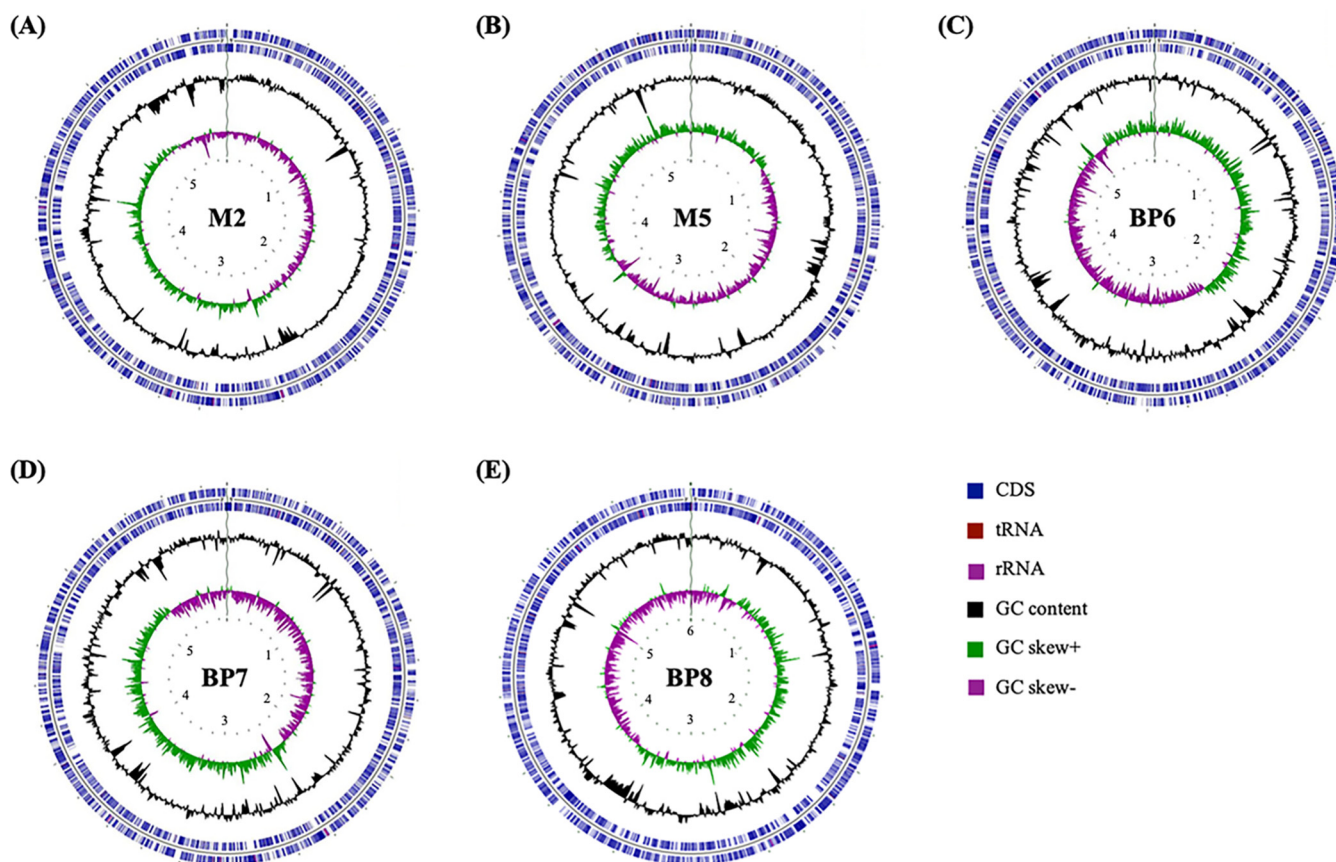

FIG 1 Circular maps representing the genomes of M2 (A), M5 (B), BP6 (C), BP7 (D), and BP8 (E). Forward-strand and reverse-strand CDSs (blue) are depicted on the outermost two circles of the map, and RNA genes (tRNA, red; rRNA, violet) are shown on the same circles. GC content (black) and GC skew (positive GC skew, green; negative GC skew, violet) are drawn on the third and fourth circles, respectively. The scale (in mega-based pairs, mbp) is indicated on the innermost circle. CGView software (19) was used to construct the genome map.

Pseudomonas sp. strain BP6, JAGINI000000000; *Pseudomonas* sp. strain BP7, JAGINJ000000000; *Pseudomonas* sp. strain BP8, JAGINK00000000; *Pseudomonas* sp. strain M2, JADOUD010000001; *Pseudomonas* sp. strain M5, JAFBBH000000000. The SRA accession numbers for the raw reads are as follows: *Pseudomonas* sp. strain BP6, SRX13609329; *Pseudomonas* sp. strain BP7, SRX13609331; *Pseudomonas* sp. strain BP8, SRX13609332; *Pseudomonas* sp. strain M2, SRX9632768; *Pseudomonas* sp. strain M5, SRX10105427.

REFERENCES

1. Isikgor FH, Becer CR. 2015. Lignocellulosic biomass: a sustainable platform for the production of bio-based chemicals and polymers. Polym Chem 6: 4497–4559. https://doi.org/10.1039/C5PY00263J.

2. Park M-R, Chen Y, Thompson M, Benites VT, Fong B, Petzold CJ, Baidoo EEK, Gladden JM, Adams PD, Keasling JD, Simmons BA, Singer SW. 2020. Response of *Pseudomonas putida* to complex aromatic-rich fractions from biomass. ChemSusChem 13:4455–4467. https://doi.org/10.1002/cssc.202000268.

3. Nogales J, Mueller J, Gudmundsson S, Canalejo FJ, Duque E, Monk J, Feist AM, Ramos JL, Niu W, Palsson BO. 2020. High-quality genome-scale metabolic modelling of *Pseudomonas putida* highlights its broad metabolic capabilities. Environ Microbiol 22:255–269. https://doi.org/10.1111/1462-2920.14843.

4. Dong J, Chen Y, Benites VT, Baidoo EEK, Petzold CJ, Beller HR, Eudes A, Scheller HV, Adams PD, Mukhopadhyay A, Simmons BA, Singer SW. 2019. Methyl ketone production by *Pseudomonas putida* is enhanced by plant-derived amino acids. Biotechnol Bioeng 116:1909–1922. https://doi.org/10.1002/bit.26995.

5. Horlamus F, Wang Y, Steinbach D, Vahidinasab M, Wittgens A, Rosenau F, Henkel M, Hausmann R. 2019. Potential of biotechnological conversion of lignocellulose hydrolyzates by *Pseudomonas putida* KT2440 as a model organism for a bio-based economy. Glob Change Biol Bioenergy 11: 1421–1434. https://doi.org/10.1111/gcbb.12647.

6. Dvořák P, de Lorenzo V. 2018. Refactoring the upper sugar metabolism of *Pseudomonas putida* for co-utilization of cellobiose, xylose, and glucose. Metab Eng 48:94–108. https://doi.org/10.1016/j.ymben.2018.05.019.

7. Wang Y, Horlamus F, Henkel M, Kovacic F, Schläfle S, Hausmann R, Wittgens A, Rosenau F. 2019. Growth of engineered *Pseudomonas putida* KT2440 on glucose, xylose, and arabinose: hemicellulose hydrolysates and their major sugars as sustainable carbon sources. Glob Change Biol Bioenergy 11:249–259. https://doi.org/10.1111/gcbb.12590.

8. Bator I, Wittgens A, Rosenau F, Tiso T, Blank LM. 2019. Comparison of three xylose pathways in *Pseudomonas putida* KT2440 for the synthesis of valuable products. Front Bioeng Biotechnol 7:480. https://doi.org/10.3389/fbioe.2019.00480.

9. Meijnen J-P, de Winde JH, Ruijssenaars HJ. 2008. Engineering *Pseudomonas putida* S12 for efficient utilization of D-xylose and L-arabinose. Appl Environ Microbiol 74:5031–5037. https://doi.org/10.1128/AEM.00924-08.

10. Mayjonade B, Gouzy J, Donnadieu C, Pouilly N, Marande W, Callot C, Langlade N, Muños S. 2016. Extraction of high-molecular-weight genomic DNA for long-read sequencing of single molecules. Biotechniques 61: 203–205. https://doi.org/10.2144/000114460.

11. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, Dewinter A, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong X, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma C, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomaney A, Travers K, Trulson M, Vieceli J, Wegener J, Wu D, Yang A, Zaccarin D, Zhao P, Zhong F, Korlach J, Turner S. 2009. Real-time DNA sequencing from single polymerase molecules. Science 323:133–138. https://doi.org/10.1126/science.1162986.

12. Bushnell B. 2014. BBTools software package. https://sourceforge.net/projects/bbmap.

13. Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods 10:563–569. https://doi.org/10.1038/nmeth.2474.

14. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119. https://doi.org/10.1186/1471-2105-11-119.

15. Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. Nat Methods 12:59–60. https://doi.org/10.1038/nmeth.3176.

16. Rampasek L, Goldenberg A. 2016. TensorFlow: biology's gateway to deep learning? Cell Syst 2:12–14. https://doi.org/10.1016/j.cels.2016.01.009.

17. Chen I-MA, Chu K, Palaniappan K, Pillay M, Ratner A, Huang J, Huntemann M, Varghese N, White JR, Seshadri R, Smirnova T, Kirton E, Jungbluth SP, Woyke T, Eloe-Fadrosh EA, Ivanova NN, Kyrpides NC. 2019. IMG/M v. 5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. Nucleic Acids Res 47:D666–D677. https://doi.org/10.1093/nar/gky901.

18. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun 9:8. https://doi.org/10.1038/s41467-018-07641-9.

19. Grant JR, Stothard P. 2008. The CGView Server: a comparative genomics tool for circular genomes. Nucleic Acids Res 36:W181–W184. https://doi.org/10.1093/nar/gkn179.