



Draft Genome Sequence of *Dietzia* sp. Strain SYD-A1, Isolated from Coal Seam Formation Water

 Silas H. W. Vick,^{a,b,c} Paul Greenfield,^b  Sasha G. Tetu,^a David J. Midgley,^b  Ian T. Paulsen^a

^aDepartment of Molecular Sciences, Macquarie University, North Ryde, Australia

^bCommonwealth Scientific and Industrial Research Organisation (CSIRO), North Ryde, Australia

^cFaculty of Chemistry, Biotechnology, and Food Science, Norwegian University of Life Sciences, Ås, Norway

ABSTRACT Subsurface coal seams harbor an array of diverse microbial species subsisting as a community on the organic matter present in coal. Here, we present the annotated genome sequence of *Dietzia* sp. strain SYD-A1, a bacterium isolated from a terrestrial subsurface coal seam in New South Wales, Australia.

Microbial communities in coal seams are responsible for the transformation of fossilized coal organic matter to methane, which can then reenter the atmosphere or biosphere. This process is of interest to industry, as liberated methane can be used as a fuel source for energy generation, as well as to our understanding of global carbon cycling. As such, microbes in subsurface coal seams, as well as the genomes they harbor, are critical to our understanding of how this biogeochemical transformation is carried out.

Subsurface coal seams in eastern Australia are host to an array of microbial taxa. A sample of anoxic formation water from a Sydney Basin coal seam gas well (Sydney Basin well 2 [1]) was obtained from a depth of ~650 m and subjected to enrichment culturing and isolation. One of these isolation strategies sought to obtain facultatively aerobic taxa by plating 50 μ l of formation water onto tryptone soy agar (TSA) plates and incubating this at reservoir temperature (37°C) for 72 h. Colonies were picked and streaked onto TSA plates until an axenic culture of a Gram-positive, nonmotile, short (0.8 to 1.0 by 1.0 to 2.2 μ m) rod-shaped bacterium was obtained.

Genomic DNA was extracted from cells grown aerobically in tryptone soya broth (37°C) using the FastDNA spin kit (MP Biomedicals) according to the manufacturer's instructions and subjected to library preparation using the Nextera XT kit and subsequent paired-end (250 bp) Illumina sequencing (The Ramaciotti Centre for Genomics, University of New South Wales, Australia). The resultant reads (7,457,466 paired-end reads) were corrected using Blue v1 (2) prior to assembly with Velvet v1.2.10 (3). The draft genome sequence was ~3.6 Mbp long and comprised 595 contigs with a kmer coverage of ~64 \times and a GC content of 68.9%. The mean, median, and N_{50} lengths for the assembly were 5,990, 4,038, and 9,280 bp, respectively. The genome was annotated using the IMG-ER pipeline v4.0 and contains 3,676 putative protein-coding genes (4). BLASTn v2.7.1 comparison of the annotated 16S rRNA gene against the NCBI reference sequence database identified the sequence as most closely related to that of members of the genus *Dietzia*. Further phylogenetic analysis is outlined in Fig. 1, and the isolate was termed *Dietzia* sp. strain SYD-A1. Comparison of the annotated 16S rRNA gene sequence to a previously published database of coal seam-derived operational taxonomic units (OTUs) matched *Dietzia* sp. SYD-A1 to the OTU CSMB200 (1).

In order to better understand the potential metabolism of the strain, the genome was submitted to the TransportDB v2.0 (<http://membranetransport.org/transportDB2/>) and dbCAN v7 (<http://bcb.unl.edu/dbCAN2/>) Web servers to interrogate its transporters

Citation Vick SHW, Greenfield P, Tetu SG, Midgley DJ, Paulsen IT. 2021. Draft genome sequence of *Dietzia* sp. strain SYD-A1, isolated from coal seam formation water. Microbiol Resour Announc 10:e01341-20. <https://doi.org/10.1128/MRA.01341-20>.

Editor Frank J. Stewart, Georgia Institute of Technology

Copyright © 2021 Vick et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Silas H. W. Vick, silas.vick@nmbu.no.

Received 24 November 2020

Accepted 19 February 2021

Published 11 March 2021

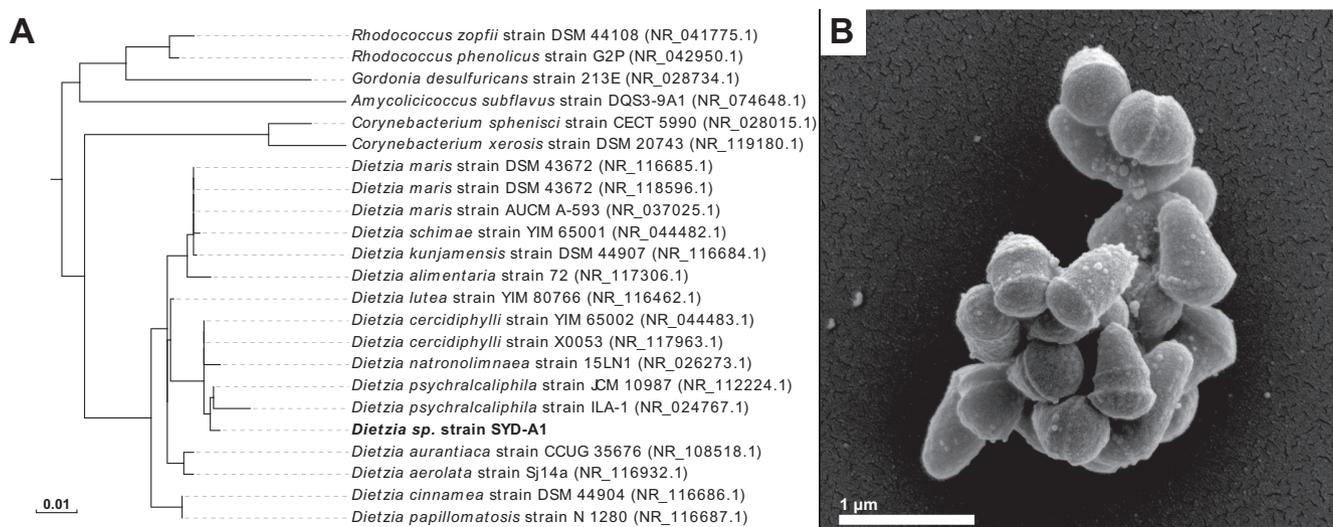


FIG 1 (A) Phylogenetic analysis of *Dietzia* species and outgroup organisms based on near-full-length 16S rRNA gene sequences obtained from the GenBank database. The 16S rRNA gene sequence for *Dietzia* sp. SYD-A1 was obtained from the whole-genome sequence of *Dietzia* sp. SYD-A1 presented here. Phylogenetic analysis was performed using the phylogeny.fr “one-click analysis” (<http://www.phylogeny.fr/>) (9). This includes alignment with MUSCLE v3.8.31 (10), refinement with Gblocks v0.91b (11), maximum likelihood phylogeny estimation with PhyML v3.1/3.0 aLRT (12, 13), and tree rendering with TreeDyn v198.3 (14). The scale bar for branch lengths indicates the number of substitutions per site. (B) Scanning electron microscopy image of *Dietzia* sp. SYD-A1; bar, 1 μm. Sample preparation and imaging were performed as previously described (15).

and carbohydrate active enzymes (5, 6). The genome encodes a modest number of carbohydrate active enzymes that appear to be transported to the cell surface via signal peptides. These include representatives with probable β -glucosidase activities (glycoside hydrolase 1 [GH1] and GH3), arabino/xylanosidase activity (GH43), and cellulolytic activity (GH5), along with a number of gene families with activity toward peptidoglycan (GH23 and GH25) (7). These latter glycoside hydrolases may be used for internal cell wall reorganization or have roles in scavenging carbon from peptidoglycan from moribund cellular material. Alkane degradation has been shown to be a key metabolic pathway for *Dietzia* species from petroleum-contaminated environments and is mediated by the alkane hydroxylase gene *alkB* and two cytochrome P450 CYP153 family genes (8). The presence of *alkB* and the two CYP153 genes in the genome of *Dietzia* sp. SYD-A1 was confirmed with BLASTn v2.7.1 matching of gene sequences from *Dietzia* sp. strain H0B to the *Dietzia* sp. strain SYD-A1 genome (*alkB* [GenBank accession number [FJ435355.1](https://www.ncbi.nlm.nih.gov/nuccore/FJ435355.1)], CYP153_1 [[FJ435360.1](https://www.ncbi.nlm.nih.gov/nuccore/FJ435360.1)], and CYP153_2 [[FJ435362.1](https://www.ncbi.nlm.nih.gov/nuccore/FJ435362.1)]). The presence of these genes suggests that SYD-A1 is capable of utilizing coal-derived alkanes during periods of oxic exposure in the terrestrial subsurface. Culture-based growth experiments on a range of substrates would be useful for clarifying the roles of these catabolic genes *in situ*. Default parameters were used for all software analyses unless otherwise stated.

Data availability. This whole-genome sequence has been deposited in GenBank and is available under the accession number [JADDKI000000000](https://www.ncbi.nlm.nih.gov/nuccore/JADDKI000000000). The raw sequencing reads are available from the Sequence Read Archive under accession number [SRR12786059](https://www.ncbi.nlm.nih.gov/sra/SRR12786059).

ACKNOWLEDGMENTS

Microscopy was undertaken at the Macquarie University Microscopy Department with training and assistance provided by Sue Lindsay and Nadia Suarez-Bosche.

Silas H. W. Vick was supported by a Macquarie University postgraduate scholarship, and the work carried out in the current study was supported by the CSIRO Energy strategic research initiative.

REFERENCES

1. Vick SH, Greenfield P, Tran-Dinh N, Tetu SG, Midgley DJ, Paulsen IT. 2018. The COAL SEAM Microbiome (CSMB) reference set, a *lingua franca* for the microbial coal-to-methane community. *Int J Coal Geol* 186:41–50. <https://doi.org/10.1016/j.coal.2017.12.003>.
2. Greenfield P, Duesing K, Papanicolaou A, Bauer DC. 2014. Blue: correcting sequencing errors using consensus and context. *Bioinformatics* 30:2723–2732. <https://doi.org/10.1093/bioinformatics/btu368>.
3. Zerbino DR, Birney E. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res* 18:821–829. <https://doi.org/10.1101/gr.074492.107>.
4. Markowitz VM, Mavromatis K, Ivanova NN, Chen I-MA, Chu K, Kyrpides NC. 2009. IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* 25:2271–2278. <https://doi.org/10.1093/bioinformatics/btp393>.
5. Elbourne LD, Tetu SG, Hassan KA, Paulsen IT. 2017. TransportDB 2.0: a database for exploring membrane transporters in sequenced genomes from all domains of life. *Nucleic Acids Res* 45:D320–D324. <https://doi.org/10.1093/nar/gkw1068>.
6. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, Busk PK, Xu Y, Yin Y. 2018. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* 46:W95–W101. <https://doi.org/10.1093/nar/gky418>.
7. Henrissat B, Davies G. 1997. Structural and sequence-based classification of glycoside hydrolases. *Curr Opin Struct Biol* 7:637–644. [https://doi.org/10.1016/s0959-440x\(97\)80072-3](https://doi.org/10.1016/s0959-440x(97)80072-3).
8. Alonso-Gutiérrez J, Teramoto M, Yamazoe A, Harayama S, Figueras A, Novoa B. 2011. Alkane-degrading properties of *Dietzia* sp. H0B, a key player in the Prestige oil spill biodegradation (NW Spain). *J Appl Microbiol* 111:800–810. <https://doi.org/10.1111/j.1365-2672.2011.05104.x>.
9. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard J-F, Guindon S, Lefort V, Lescot M, Claverie J-M, Gascuel O. 2008. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res* 36:W465–W469. <https://doi.org/10.1093/nar/gkn180>.
10. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797. <https://doi.org/10.1093/nar/gkh340>.
11. Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 17:540–552. <https://doi.org/10.1093/oxfordjournals.molbev.a026334>.
12. Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52:696–704. <https://doi.org/10.1080/10635150390235520>.
13. Anisimova M, Gascuel O. 2006. Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Syst Biol* 55:539–552. <https://doi.org/10.1080/10635150600755453>.
14. Chevenet F, Brun C, Bañuls A-L, Jacq B, Christen R. 2006. TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinformatics* 7:439. <https://doi.org/10.1186/1471-2105-7-439>.
15. Vick SH, Greenfield P, Tetu SG, Midgley DJ, Paulsen IT. 2019. Genomic and phenotypic insights point to diverse ecological strategies by facultative anaerobes obtained from subsurface coal seams. *Sci Rep* 9:1–13. <https://doi.org/10.1038/s41598-019-52846-7>.