







One Complete and Seven Draft Genome Sequences of Subdivision 1 and 3 *Acidobacteria* Isolated from Soil

 Stephanie A. Eichorst,^a  Daniela Trojan,^a Marcel Huntemann,^b Alicia Clum,^b Manoj Pillay,^b Krishnaveni Palaniappan,^b Neha Varghese,^b Natalia Mikhailova,^b Dimitrios Stamatis,^b T. B. K. Reddy,^b Chris Daum,^b Lynne A. Goodwin,^c Nicole Shapiro,^b Natalia Ivanova,^b Nikos Kyrpides,^b  Tanja Woyke,^b  Dagmar Woebken^a

^aDivision of Microbial Ecology, Department of Microbiology and Ecosystem Science, Centre for Microbiology and Environmental Systems Science, University of Vienna, Vienna, Austria

^bDepartment of Energy, Joint Genome Institute, Berkeley, California, USA

^cMaterials Science and Technology, Engineered Materials, Los Alamos National Laboratory, Los Alamos, New Mexico, USA

ABSTRACT We report eight genomes from representatives of the phylum *Acidobacteria* subdivisions 1 and 3, isolated from soils. The genome sizes range from 4.9 to 6.7 Mb. Genomic analysis reveals putative genes for low- and high-affinity respiratory oxygen reductases, high-affinity hydrogenases, and the capacity to use a diverse collection of carbohydrates.

The *Acidobacteria* constitute a large, diverse, and phylogenetically distinct phylum with currently 26 defined subdivisions based on the 16S rRNA gene phylogeny (1). Typically, members of subdivisions 1, 2, 3, 4, 5, and 6 populate terrestrial environments at a high relative sequence abundance based on rRNA gene libraries (as high as 20 to 40% of the total bacterial community) (2, 3). Phylogenetically representative acidobacterium strains were chosen for genome sequencing, which were isolated from soils (meadow grassland, agricultural, or peat bog) (4–7). More specifically, six strains in subdivision 1 (*Terriglobus roseus* KBS 63, *Acidobacteriaceae* bacterium strain KBS 83, *Acidobacteriaceae* bacterium strain KBS 89, *Acidobacteriaceae* bacterium strain KBS 146, *Terriglobus* sp. strain TAA 43, and *Acidobacteriaceae* bacterium strain TAA 166) and two strains in subdivision 3 (*Acidobacteria* bacterium strain KBS 96 and *Bryobacter aggregatus* MPL3) were sequenced to better populate the phylogenetic branches in the phylum *Acidobacteria* with genomic representation.

Strains were grown on either a modified minimal medium or R2 medium as described previously (5–7). Genomic DNA (gDNA) was isolated using a modified cetyltrimethylammonium bromide (CTAB) DNA extraction protocol as recommended by the Department of Energy (DOE) Joint Genome Institute (JGI). Details on the isolation and growth conditions along with the sequencing can be found in Table 1. The genomes across all strains ranged in size from 4.9 to 6.7 Mb (Table 1), which is similar to the size range of the previously sequenced acidobacterial strains (8, 9). The G+C content ranged from 57 to 60 mol% (Table 1). Although many of these genomes are considered to be permanent drafts (Table 1), CheckM analysis indicated that they are ca. >95% complete (10).

The genomes of KBS 63 and KBS 96 harbored the catalytic subunit for the high-affinity *cbb*₃-type cytochrome *c* oxidase (heme-copper-oxygen reductase [HCO] type C, EC 1.9.3.1, KEGG orthology [KO] number K00404), while the genomes of KBS 146 and TAA 166 harbored the high-affinity cytochrome *bd*-type quinol oxidase (EC 7.1.1, KO number K00425) (10). All strains harbored at least one homologue of the low-affinity terminal oxidase HCO type A (EC 1.9.3.1, KO number K02274) (10). Numerous glycoside hydrolase (GH) families were detected across the genomes, with GH109, GH74, and GH13 being some of the most prevalent (10). The genomes of KBS 83 and KBS 96 harbored a high-affinity group 1h hydrogenase (EC 1.12.99.6, KO number K06281), presumably giving them the potential to scavenge atmospheric concentrations of hydrogen during periods of starvation

Citation Eichorst SA, Trojan D, Huntemann M, Clum A, Pillay M, Palaniappan K, Varghese N, Mikhailova N, Stamatis D, Reddy TBK, Daum C, Goodwin LA, Shapiro N, Ivanova N, Kyrpides N, Woyke T, Woebken D. 2020. One complete and seven draft genome sequences of subdivision 1 and 3 *Acidobacteria* isolated from soil. *Microbiol Resour Announc* 9:e01087-19. <https://doi.org/10.1128/MRA.01087-19>.

Editor Irene L. G. Newton, Indiana University, Bloomington

Copyright © 2020 Eichorst 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 Stephanie A. Eichorst, stephanie.eichorst@univie.ac.at.

This article is Los Alamos unclassified report LA-UR-18-24028.

Received 1 October 2019

Accepted 17 December 2019

Published 30 January 2020

TABLE 1 Isolation conditions, library preparation information, genome sequencing statistics, and accession numbers for the described acidobacterial genomes

Strain	Isolation source; yr	Isolation and growth conditions (reference)	Library prepn	No. of reads (technology)	Avg read length (bp)	Assembler (reference)	No. of contigs (no. of scaffolds)	N/L_{50} of: Scaffolds	Contigs	Avg genome coverage (X)	Genome size (Mb)	G+C content (mol%)	Assembly level	GenBank accession no.	SRA accession no.
KBS 63	Kellogg Biological Station, Hickory Corners, MI, USA, grassland soil (treatment 8); 2003	VS8-7 mix organic C substrates in CO ₂ -enriched air (5)	454 paired-end library ^a ; Illumina std. shotgun library ^b	69,481,460 (Illumina Solexa); 514,324 (454 paired end)	76 (Illumina Solexa); 151 (454 paired end)	Newbler v2.3-Prerelease 6/30/2009; Velvet v1.0.13 (13)	33 (2)	1/5.2 Mb	1/5.2 Mb	1083.1 (Illumina Solexa); 8.5 (454 paired end)	5.23	60.3	Complete	CP003379	SRS1568612
KBS 83	Kellogg Biological Station, Hickory Corners, MI, USA, agricultural soil (treatment 1); 2006	V1.55 mix of plant polymeric C in air (7)	Illumina std. shotgun library ^b	12,985,884 (Illumina HiSeq 2000)	2 × 150	Velvet v1.1.04 (13), wgsim ^e , Allpaths-LG v r41043 (14)	25 (25)	5/504.7 kb	5/504.7 kb	390	6.25	59.2	Permanent draft	ARMDO00000000	SRS844143
KBS 89	Kellogg Biological Station, Hickory Corners, MI, USA, grassland soil (treatment 8); 2002	VS8-6.8 mix organic C substrates, acyl homoserine lactones in CO ₂ -enriched air (4, 5)	Illumina std. shotgun library ^b	12,475,762 (Illumina HiSeq 2000)	2 × 150	Velvet v1.1.04 (13), wgsim ^e , Allpaths-LG v r41043 (14)	14 (14)	3/943.6 kb	3/943.6 kb	122.6	6.01	57.6	Permanent draft	ARME0000000000	SRS844144
KBS 146	Kellogg Biological Station, Hickory Corners, MI, USA, grassland soil (treatment 8); 2006	V1.55 mix of organic C substrates, CO ₂ -enriched hypoxia (11)	PacBio SMRTbell library ^d	185,131 (PacBio RS platform)	3,493 ± 2,894	HGAP v2.0.0 (15)	2 (2)	1/5.0 Mb	1/5.0 Mb	64.1	5.00	56.7	Permanent draft	JHVA0000000000	SRS1534005
TAA 43	Hindgut of <i>Reikvultermes flavipes</i> (Kollar) (<i>Phanerochaete</i>), Dansville, MI, USA; 2002	VS8-7 yeast extract and peptone in CO ₂ -enriched air (4, 5)	Illumina std. shotgun library ^b	13,649,630 (Illumina HiSeq 2000)	2 × 151	Velvet v1.2.07 (13), wgsim ^e , Allpaths-LG v r46652 (14)	7 (7)	1/3.5 Mb	1/3.5 Mb	302.0	4.95	56.7	Permanent draft	JUGR0000000000	SRS1366045
TAA 166	Hindgut of <i>R. flavipes</i> (Kollar) (<i>Phanerochaete</i>), Dansville, MI, USA; 2002	VS8-7 yeast extract and peptone in CO ₂ -enriched air (4, 5)	Illumina std. shotgun library ^b and long-insert mate pair library ^c ; PacBio SMRTbell library ^d	40,920,398 (Illumina CLIP paired end, Illumina HiSeq 2000); 15,055,388 (Illumina std. paired end, Illumina HiSeq 2000); 233,258 (PacBio RS platform)	2 × 90; 2 × 150; 1 × 2,259	Allpaths-LG v r42328 (14)	3 (3)	1/4.7 Mb	1/4.7 Mb	973.9 (Illumina); 85.4 (PacBio)	6.14	58.8	Permanent draft	ATWD0000000000	SRS438888
KBS 96	Kellogg Biological Station, Hickory Corners, MI, USA, agricultural soil (treatment 1); 2006	V1.55 mix of plant polymeric C in air (7)	Illumina std. shotgun library ^b and long-insert mate pair library ^c	14,485,381 (Illumina std. paired end, HiSeq 2000); 58,296,769 (Illumina CLIP paired end, Illumina HiSeq 2000)	2 × 150; 2 × 89	Allpaths-LG v r41043 (14)	13 (2)	1/6.7 Mb	2/1.6 Mb	1,098.6	6.69	57.2	Permanent draft	ARMF0000000000	SRS438892

(Continued on next page)

TABLE 1 (Continued)

Strain	Isolation source; yr	Isolation and growth conditions (reference)	Library prepn	No. of reads (technology)	Avg read length (bp)	Assembler (reference)	No. of contigs (no. of scaffolds)	N_{50} of scaffolds	Contigs	Avg genome coverage (x)	Genome size (Mb)	G+C content (mol%)	Assembly level	GenBank accession no.	SRA accession no.
MPL3	Acidic <i>Sphagnum</i> peat bog, Bakchar, Tomsk region, West Siberia; 2004	Biofilm-mediated enrichment approach (6, 12)	PacBio SMRTbell library ^a	208,346 (PacBio RS platform)	3,087 ± 2,325	HGAP v2.0.0 (15)	4 (4)	1/4.3 Mb	1/4.3 Mb	176.8	5.75	57.0	Permanent draft	JNFF00000000	SRS1520682

^a454 Titanium, paired ends, 8 kb; 15 μ g genomic DNA are sheared by the Hydroshear to ~8-kb size fragments. The sheared samples are then gel selected for the 8-kb bands, purified, and ligated to the 42-bp loxP linkers on either end. These loxP linkers are labeled by biotin. The loxP linker-ligated fragments are then circularized by the Cre recombinase. As a result, the ends of 20-kb fragments are brought together and bridged by a single loxP linker. These circular DNAs are further sheared to 500-bp fragments, and the fragments carrying the loxP linkers are recovered by the Streptavidin-coated magnetic beads. Consequently, the loxP linker-containing fragments are ligated to the 454 Titanium adapters A and B in the same way that the shotgun libraries are created. The 454 library fragments are then clonally amplified in bulk by capturing them through hybridization on microparticle beads and subjecting them to emulsion-based PCR. This results in beads that are covered with millions of copies of a single DNA fragment (range, 400–800 bp), where each bead contains a different clonally amplified library fragment. After amplification, the beads are recovered from the emulsions and are loaded into the wells of a PicoTiterPlate (PTP) device such that wells contain single DNA beads. The PTP device is then inserted into the 454 genome sequencer FLX-Titanium instrument for sequencing where sequencing reagents are sequentially flowed over the PTP wells. Each incorporation of a nucleotide complementary to the template strand results in a chemiluminescent light signal that is recorded by a camera, and the sequence of the DNA fragments is determined. This sequencing-by-synthesis method is known as pyrosequencing.

^bIllumina regular fragment, 300 bp; 100 ng of DNA was sheared to 300 bp using the Covaris LE220 instrument and size selected using solid-phase reversible immobilization (SPRI) beads (Beckman Coulter). The fragments were treated with end repair, A-tailing, and ligation of Illumina-compatible adapters (IDT, Inc.) using the KAPA-Illumina library creation kit (Kapa Biosystems). The prepared libraries were quantified using Kapa Biosystem's next-generation sequencing library quantitative PCR (qPCR) kit and run on a Roche LightCycler 480 real-time PCR instrument. The quantified libraries were then multiplexed with other libraries, and the pool of libraries was then prepared for sequencing on the Illumina HiSeq or the Illumina GALIX (KBS 63) sequencing platform utilizing a TruSeq paired-end cluster kit v3 or a paired-end cluster generation kit, v4 (KBS 63) and Illumina's cBot instrument to generate a clustered flow cell for sequencing. Sequencing of the flow cell was performed on the Illumina HiSeq 2000 sequencer using a TruSeq SBS sequencing kit v3 following a 2×150 indexed run recipe and on the Illumina GALIX sequencer using SBS sequencing kits, v4, following a 2×76 run recipe (KBS 63).

^cIllumina regular long mate pair, 8 kb, Cre-Lox random shear mate pair sequences; 5 μ g of DNA was sheared using = Covaris g-TUBES and was gel size selected for 8 kb. The sheared DNA was treated with end repair and ligated with biotinylated adapters containing loxP. The adapter-ligated DNA fragments were circularized via recombination with a Cre excision reaction (NEB). The circularized DNA templates were then randomly sheared using the Covaris LE220 instrument. The sheared fragments were treated with end repair and A-tailing using the KAPA-Illumina library creation kit (Kapa Biosystems) followed by immobilization of mate pair fragments on streptavidin beads (Invitrogen). Illumina-compatible adapters (IDT, Inc.) were ligated to the mate pair fragments, and 14 cycles of PCR were used to enrich for the final library (Kapa Biosystems). The prepared libraries were quantified using Kapa Biosystem's next-generation sequencing library qPCR kit and run on a Roche LightCycler 480 real-time PCR instrument. The quantified libraries were then multiplexed with other libraries, and the pool of libraries was then prepared for sequencing on the Illumina HiSeq sequencing platform utilizing a TruSeq paired-end cluster kit v3 and Illumina's cBot instrument to generate a clustered flow cell for sequencing. Sequencing of the flow cell was performed on the Illumina HiSeq 2000 sequencer using a TruSeq SBS sequencing kit v3 following a 2×100 indexed run recipe.

^dPacBio >10-kb libraries with AMPure bead size selection; unamplified libraries were generated using the Pacific Biosciences standard template preparation protocol for creating >10-kb libraries. A total of 5 μ g of genomic DNA was used to generate each library, and the DNA was sheared using Covaris g-TUBES to generate sheared fragments >10 kb in length. The sheared DNA fragments were then prepared using Pacific Biosciences SMRTbell template preparation kit v1.0, where the fragments were treated with DNA damage repair, had their ends repaired so that they were blunt ended, and were 5' phosphorylated. Pacific Biosciences hairpin adapters were then ligated to the fragments to create the SMRTbell template for sequencing. The SMRTbell templates were then purified using exonuclease treatments and size selected using AMPure PB beads. Sequencing primer was then annealed to the SMRTbell templates, and version XL sequencing polymerase was bound to them. The prepared SMRTbell template libraries were then sequenced on a Pacific Biosciences RS II sequencer using version C2 chemistry and 2-h sequencing movie run times.

^e<https://github.com/lh3/avgsim>.

(10). The addition of these genomes in the public databases will help to further our understanding of this prevalent and diverse phylum.

Data availability. The genome sequences are available in GenBank; accession numbers can be found in Table 1.

ACKNOWLEDGMENTS

We thank Svetlana Dedysh for donating her strain *Bryobacter aggregatus* MPL3^T and the Division of Computational Systems Biology for providing and maintaining the Life Science Compute Cluster (LiSC) at the University of Vienna.

This work was supported by an Austrian Science Fund FWF project grant (P26392-B20 to D.W. and S.A.E.) and a Joint Genome Institute Community Sequencing Program, “Populating the Branches of the Phylum *Acidobacteria*” (number 386 to S.A.E., Cheryl R. Kuske, and Jonathan Eisen). The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy under contract number DE-AC02-05CH11231.

REFERENCES

- Barns SM, Cain EC, Sommerville L, Kuske CR. 2007. *Acidobacteria* phylum sequences in uranium-contaminated subsurface sediments greatly expand the known diversity within the phylum. *Appl Environ Microbiol* 73:3113–3116. <https://doi.org/10.1128/AEM.02012-06>.
- Lipson DA, Schmidt SK. 2004. Seasonal changes in an alpine soil bacterial community in the Colorado Rocky Mountains. *Appl Environ Microbiol* 70:2867–2879. <https://doi.org/10.1128/aem.70.5.2867-2879.2004>.
- Janssen PH. 2006. Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. *Appl Environ Microbiol* 72:1719–1728. <https://doi.org/10.1128/AEM.72.3.1719-1728.2006>.
- Stevenson BS, Eichorst SA, Wertz JT, Schmidt TM, Breznak JA. 2004. New strategies for cultivation and detection of previously uncultured microbes. *Appl Environ Microbiol* 70:4748–4755. <https://doi.org/10.1128/AEM.70.8.4748-4755.2004>.
- Eichorst SA, Breznak JA, Schmidt TM. 2007. Isolation and characterization of soil bacteria that define *Terriglobus* gen. nov., in the phylum *Acidobacteria*. *Appl Environ Microbiol* 73:2708–2717. <https://doi.org/10.1128/AEM.02140-06>.
- Kulichevskaya IS, Suzina NE, Liesack W, Dedysh SN. 2010. *Bryobacter aggregatus* gen. nov., sp. nov., a peat-inhabiting, aerobic chemo-organotroph from subdivision 3 of the *Acidobacteria*. *Int J Syst Evol Microbiol* 60:301–306. <https://doi.org/10.1099/ijs.0.013250-0>.
- Eichorst SA, Kuske CR, Schmidt TM. 2011. Influence of plant polymers on the distribution and cultivation of bacteria in the phylum *Acidobacteria*. *Appl Environ Microbiol* 77:586–596. <https://doi.org/10.1128/AEM.01080-10>.
- Ward NL, Challacombe JF, Janssen PH, Henrissat B, Coutinho PM, Wu M, Xie G, Haft DH, Sait M, Badger J, Barabote RD, Bradley B, Brettin TS, Brinkac LM, Bruce D, Creasy T, Daugherty SC, Davidsen TM, DeBoy RT, Detter JC, Dodson RJ, Durkin AS, Ganapathy A, Gwinn-Giglio M, Han CS, Khouri H, Kiss H, Kothari SP, Madupu R, Nelson KE, Nelson WC, Paulsen I, Penn K, Ren Q, Rosovitz MJ, Selengut JD, Shrivastava S, Sullivan SA, Tapia R, Thompson LS, Watkins KL, Yang Q, Yu C, Zafar N, Zhou L, Kuske CR. 2009. Three genomes in the phylum *Acidobacteria* provide insight into their lifestyles in soils. *Appl Environ Microbiol* 74:2046–2056. <https://doi.org/10.1128/AEM.02294-08>.
- Rawat SR, Männistö MK, Bromberg Y, Häggblom MM. 2012. Comparative genomic and physiological analysis provides insights into the role of *Acidobacteria* in organic carbon utilization in Arctic tundra soils. *FEMS Microbiol Ecol* 82:341–355. <https://doi.org/10.1111/j.1574-6941.2012.01381.x>.
- Eichorst SA, Trojan D, Roux S, Herbold C, Rattei T, Woebken D. 2018. Genomic insights into the *Acidobacteria* reveal strategies for their success in terrestrial environments. *Environ Microbiol* 20:1041–1063. <https://doi.org/10.1111/1462-2920.14043>.
- Eichorst SA. 2007. Isolation and characterization of members of the phylum *Acidobacteria* from soils. PhD thesis. Michigan State University, East Lansing, MI. <https://doi.org/10.25335/M5804XZ94>.
- Dedysh SN, Pankratov TA, Belova SE, Kulichevskaya IS, Liesack W. 2006. Phylogenetic analysis and *in situ* identification of *Bacteria* community composition in an acidic *Sphagnum* peat bog. *Appl Environ Microbiol* 72:2110–2117. <https://doi.org/10.1128/AEM.72.3.2110-2117.2006>.
- 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>.
- Gnerre S, MacCallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S, Berlin AM, Aird D, Costello M, Daza R, Williams L, Nicol R, Gnirke A, Nusbaum C, Lander ES, Jaffe DB. 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc Natl Acad Sci U S A* 108:1513–1518. <https://doi.org/10.1073/pnas.1017351108>.
- Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Non-hybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 10:563–569. <https://doi.org/10.1038/nmeth.2474>.