Science
AAAS

Environmental Genomics Reveals a Single-Species Ecosystem Deep Within Earth

Dylan Chivian, *et al.*

Science **322**, 275 (2008);
DOI: 10.1126/science.1155495

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005.

6. Y. Wu *et al.*, *Cell* **126**, 375 (2006).
7. A. Marson *et al.*, *Nature* **445**, 931 (2007).
8. Y. Zheng *et al.*, *Nature* **445**, 936 (2007).
9. M. Ono *et al.*, *Nature* **446**, 685 (2007).
10. P. Waterhouse *et al.*, *Science* **270**, 985 (1995).
11. E. A. Tivol *et al.*, *Immunity* **3**, 541 (1995).
12. The Wellcome Trust Case Control Consortium, *Nature* **447**, 661 (2007).
13. D. R. Leach, M. F. Krummel, J. P. Allison, *Science* **271**, 1734 (1996).
14. G. Q. Phan *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 8372 (2003).
15. S. Read *et al.*, *J. Immunol.* **177**, 4376 (2006).
16. D. M. Sansom, L. S. Walker, *Immunol. Rev.* **212**, 131 (2006).
17. Materials and methods are available as supporting material on *Science* Online.
18. S. Kawamoto *et al.*, *FEBS Lett.* **470**, 263 (2000).
19. M. Ono, J. Shimizu, Y. Miyachi, S. Sakaguchi, *J. Immunol.* **176**, 4748 (2006).
20. Y. Y. Wan, R. A. Flavell, *Nature* **445**, 766 (2007).
21. J. Shimizu, S. Yamazaki, S. Sakaguchi, *J. Immunol.* **163**, 5211 (1999).
22. L. Cederbom, H. Hall, F. Ivars, *Eur. J. Immunol.* **30**, 1538 (2000).
23. C. Oderup, L. Cederbom, A. Makowska, C. M. Cilio, F. Ivars, *Immunology* **118**, 240 (2006).
24. S. Yamazaki, K. Inaba, K. V. Tarbell, R. M. Steinman, *Immunol. Rev.* **212**, 314 (2006).
25. R. J. DiPaolo *et al.*, *J. Immunol.* **179**, 4685 (2007).
26. Y. Onishi *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 10113 (2008).

27. We thank M. Ono for discussion and R. Ishii and M. Matsushita for technical assistance. This work was supported by Grants-in-Aid from the Ministry of Education, Sports and Culture of Japan, Japan Science and Technology Agency. Z.F. was a Japan Society for the Promotion of Science fellow, and K.W. was granted a fellowship by Astra-Zeneca, Loughborough, UK.

Supporting Online Material

www.sciencemag.org/cgi/content/full/322/5899/271/DC1
Materials and Methods
SOM Text
Figs. S1 to S13
References

5 May 2008; accepted 15 August 2008
10.1126/science.1160062

Environmental Genomics Reveals a Single-Species Ecosystem Deep Within Earth

Dylan Chivian,[1,2]* Eoin L. Brodie,[2,3] Eric J. Alm,[2,4] David E. Culley,[5] Paramvir S. Dehal,[1,2] Todd Z. DeSantis,[2,3] Thomas M. Gihring,[6] Alla Lapidus,[7] Li-Hung Lin,[8] Stephen R. Lowry,[7] Duane P. Moser,[9] Paul M. Richardson,[7] Gordon Southam,[10] Greg Wanger,[10] Lisa M. Pratt,[11,12] Gary L. Andersen,[2,3] Terry C. Hazen,[2,3,12] Fred J. Brockman,[13] Adam P. Arkin,[1,2,14] Tullis C. Onstott[12,15]

DNA from low-biodiversity fracture water collected at 2.8-kilometer depth in a South African gold mine was sequenced and assembled into a single, complete genome. This bacterium, *Candidatus Desulforudis audaxviator*, composes >99.9% of the microorganisms inhabiting the fluid phase of this particular fracture. Its genome indicates a motile, sporulating, sulfate-reducing, chemoautotrophic thermophile that can fix its own nitrogen and carbon by using machinery shared with archaea. *Candidatus Desulforudis audaxviator* is capable of an independent life-style well suited to long-term isolation from the photosphere deep within Earth's crust and offers an example of a natural ecosystem that appears to have its biological component entirely encoded within a single genome.

A more complete picture of life on, and even in, Earth has recently become possible by extracting and sequencing DNA from an environmental sample, a process called environmental genomics or metagenomics (*1*–*8*). This approach allows us to identify members of microbial communities and to characterize the abilities of the dominant members even when isolation of those organisms has proven intractable. However, with a few exceptions (*5*, *7*), assembling complete or even near-complete genomes for a substantial portion of the member species is usually hampered by the complexity of natural microbial communities.

In addition to elevated temperatures and a lack of O_2, conditions within Earth's crust at depths >1 km are fundamentally different from those of the surface and deep ocean environments. Severe nutrient limitation is believed to result in cell doubling times ranging from 100s to 1000s of years (*9*–*11*), and as a result subsurface microorganisms might be expected to reduce their reproductive burden and exhibit the streamlined genomes of specialists or spend most of their time in a state of semi-senescence, waiting for the return of favorable conditions. Such microorganisms are of particular interest because they permit insight into a mode of life independent of the photosphere.

One bacterium belonging to the *Firmicutes* phylum (Fig. 1A), which we herein name *Candidatus Desulforudis audaxviator*, is prominent in small subunit (SSU or 16*S*) ribosomal RNA (rRNA) gene clone libraries (*11*–*14*) from almost all fracture fluids sampled to date from depths greater than 1.5 km across the Witwatersrand basin (covering 150 km by 300 km near Johannesburg, South Africa). This bacterium was shown in a previous geochemical and 16*S* rRNA gene study (*11*) to dominate the indigenous microorganisms found in a fracture zone at 2.8 km below land surface at level 104 of the Mponeng mine (MP104). Although Lin *et al.* (*11*) discovered that this fracture zone contained the least-diverse natural free-living microbial community reported at that time, exceeding the ~80% dominance by the methanogenic archaeon IUA5/6 of a comparatively shallow subsurface community in Idaho (*15*), we were nonetheless surprised when the current environmental genomics study revealed only one species was actually present within the fracture fluid. Furthermore, we found that the genome of this organism appeared to possess all of the metabolic capabilities necessary for an independent life-style. This gene complement was consistent with the previous geochemical and thermodynamic analyses at the ambient ~60°C temperature and pH of 9.3, which indicated radiolytically generated chemical species as providing the energy and nutrients to the system (*11*), with formate and H_2 as possessing the greatest potential among candidate electron donors, and sulfate (SO_4^{2-}) reduction as the dominant electron-accepting process (*11*).

DNA was extracted from ~5600 liters of filtered fracture water by using a protocol that has been demonstrated to be effective on a broad range of bacterial and archaeal species, including recalcitrant organisms (*16*). A single, complete, 2.35–megabase pair (Mbp) genome was assembled with a combination of shotgun Sanger sequencing and 454 pyrosequencing (*16*). Similar to other studies that obtained near-complete consensus genomes from environmental samples (*5*, *17*), heterogeneity in the population of the dominant species as measured with single-nucleotide polymorphisms (SNP) was quite low, showing only 32 positions with a SNP observed

[1]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [2]Virtual Institute for Microbial Stress and Survival, Berkeley, CA 94720, USA. [3]Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [4]Departments of Biological and Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [5]Energy and Efficiency Technology Division, Pacific Northwest National Laboratory, Richland, WA 99352, USA. [6]Department of Oceanography, Florida State University, Tallahassee, FL 32306, USA. [7]Genomic Technology Program, U.S. Department of Energy (DOE) Joint Genomics Institute, Berkeley, CA 94598, USA. [8]Department of Geosciences, National Taiwan University, Taipei 106, Taiwan. [9]Division of Earth and Ecosystem Sciences, Desert Research Institute, Las Vegas, NV 89119, USA. [10]Department of Earth Sciences, University of Western Ontario, London, ON N6A 5B7, Canada [11]Department of Geological Sciences, Indiana University, Bloomington, IN 47405, USA. [12]Indiana Princeton Tennessee Astrobiology Initiative (IPTAI), NASA Astrobiology Institute, Bloomington, IN 47405, USA. [13]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99352, USA. [14]Department of Bioengineering, University of California, Berkeley, CA 94720, USA. [15]Department of Geosciences, Princeton University, Princeton, NJ 08544, USA.

*To whom correspondence should be addressed. E-mail: DCChivian@lbl.gov

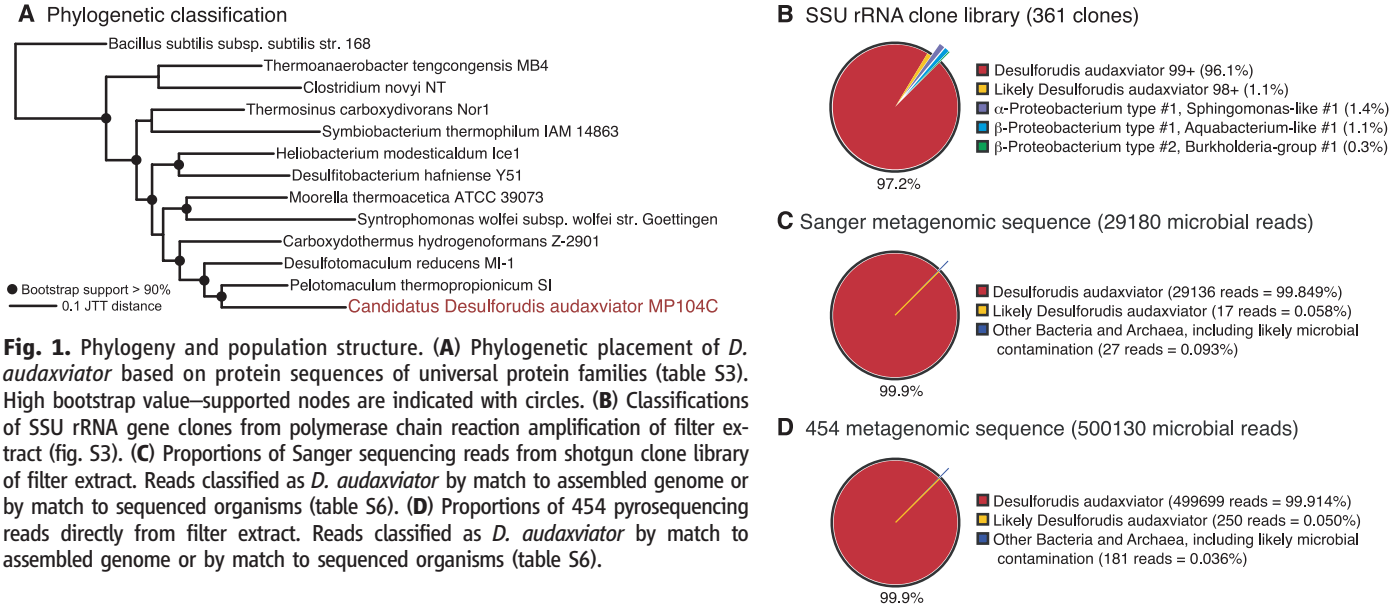

Fig. 1. Phylogeny and population structure. (**A**) Phylogenetic placement of *D. audaxviator* based on protein sequences of universal protein families (table S3). High bootstrap value–supported nodes are indicated with circles. (**B**) Classifications of SSU rRNA gene clones from polymerase chain reaction amplification of filter extract (fig. S3). (**C**) Proportions of Sanger sequencing reads from shotgun clone library of filter extract. Reads classified as *D. audaxviator* by match to assembled genome or by match to sequenced organisms (table S6). (**D**) Proportions of 454 pyrosequencing reads directly from filter extract. Reads classified as *D. audaxviator* by match to assembled genome or by match to sequenced organisms (table S6).

more than once (table S7), suggesting a recent selective sweep or other population bottleneck.

The DNA recovered from the filter, assuming the capture of cells and extraction of DNA from those cells was indeed comprehensive, revealed that this genome represented the only species present in the fluid phase of the fracture. Of the ~0.1% of microbial reads not belonging to *D. audaxviator* (Fig. 1, C and D, and tables S5 and S6), about one-half represented clear laboratory contamination (table S6), the removal of which resulted in only 22 of 29,179 Sanger reads (0.075%) and 59 of 500,008 pyrosequencing reads (0.012%) that could be from other microorganisms. Despite precautions taken in collecting the sample, some of the trace reads could come from microbial contaminants in the mine. An upper-bound estimate of the contribution of any microorganism other than *D. audaxviator* to the community (table S6) offered at most only five Sanger reads (0.017%) corresponding to γ-Proteobacteria and at most nine pyrosequencing reads (0.0018%) corresponding to α-Proteobacteria, both of which are common in the mining water (*11*, *14*). Even if these Proteobacteria were not contaminants, it is unlikely that *D. audaxviator*, and indeed the functioning of the ecosystem, is metabolically dependent on organisms that would be outnumbered by about 5000 to 1 (or about 50,000 to 1 from the pyrosequencing data). However, we could not rule out the presence of organisms that might adhere to the surfaces of the fracture or that were smaller than the 0.2-μm filter pore that might play a role in the MP104 ecosystem, perhaps as reservoirs of genetic variation (*18*).

We analyzed the genome of *D. audaxviator* with use of MicrobesOnline (www.microbesonline.org) (*19*). If *D. audaxviator* is indeed the solitary resident of this habitat, then its genome should contain the complete genetic complement for maintaining the biological component of the ecosystem, which would prohibit the extreme reduction of its genome. The genome (Table 1) at 2.35 Mbp was smaller than the 3 Mbp of its nearest sequenced relative, *Pelotomaculum thermopropionicum*. It contained 2157 predicted protein coding genes, more than found in streamlined free-living microorganisms, which typically have fewer than 2000 genes (*20*). We found all of the processes necessary for life encoded within the genome, including energy metabolism, carbon fixation, and nitrogen fixation.

Consistent with the thermodynamic evaluation (*11*) that SO_4^{2-} offers the most energetically favorable electron acceptor, the genome possesses the capacity for dissimilatory sulfate reduction (DSR) (Figs. 2 and 3 and table S13) with a gene repertoire like that of other SO_4^{2-}-reducing microorganisms (*21*). These genes are present in a set of operons (labeled SR1 to SR11 in Fig. 2) and include an extra copy of an archael-type sulfate adenylyltransferase (Sat) (fig. S5) and a H^+-translocating pyrophosphatase, both of which appear to be a consequence of horizontal gene transfer (HGT). High potential electrons probably enter primarily via the activity of a variety of hydrogenases acting upon H_2 (table S24).

Carbon assimilation may be from a variety of sources depending on local conditions. The genome contains sugar and amino acid transporters (Fig. 3 and table S20), suggesting that, at locations where biodensity is high, heterotrophic sources could be used, including recycling of dead cells. At MP104, where biodensity is low, carbon is fixed from inorganic sources. *D. audaxviator* appeared not to use the reverse TCA cycle (table S23) but did have all the machinery of the acetyl–coenzyme A (CoA) synthesis (Wood-Ljungdahl) pathway (*22*, *23*), which uses carbon monoxide dehydrogenase (CODH) for the assimilation of inorganic carbon (Figs. 2 and 3, fig. S7, and table S14). Entry of CO_2 substrate into the cell may be accomplished by its anionic species through a putative carbonate adenosine triphosphate (ATP)–binding cassette transporter or a putative bicarbonate/Na^+ symporter (Fig. 3 and table S20). Formate and CO may serve as alternate, more direct, carbon sources in other fractures when sufficiently abundant (table S2).

The ambient concentration of ammonia in the fracture water ($[NH_3] + [NH_4^+]$ = ~100 μM) (*11*) appears sufficient for *D. audaxviator* (which has an ammonium transporter as well as glutamine

Table 1. General features of the *D. audaxviator* genome.

Feature	Value
Genome size (bp)	2,349,476
G+C content (%)	60.9
Predicted protein coding genes (CDS/ORF)	2157
Genes without homology to other organisms (ORFans)	210
Pseudogenes derived from a protein coding gene	83
Average CDS/ORF length (bp)	910
Longest CDS/ORF length (bp)	5601
Percent of genome protein coding (%)	86.8
rRNA operons (16*S*-23*s*-5*S*)	2
Transfer RNAs (all amino acids represented, including selenocysteine)	45
Other nonprotein coding RNAs	7
CRISPR regions	2
Mobile elements (transposons/integrases)	
Gene groups	30
Gene count	83
Other phage-associated genes	18

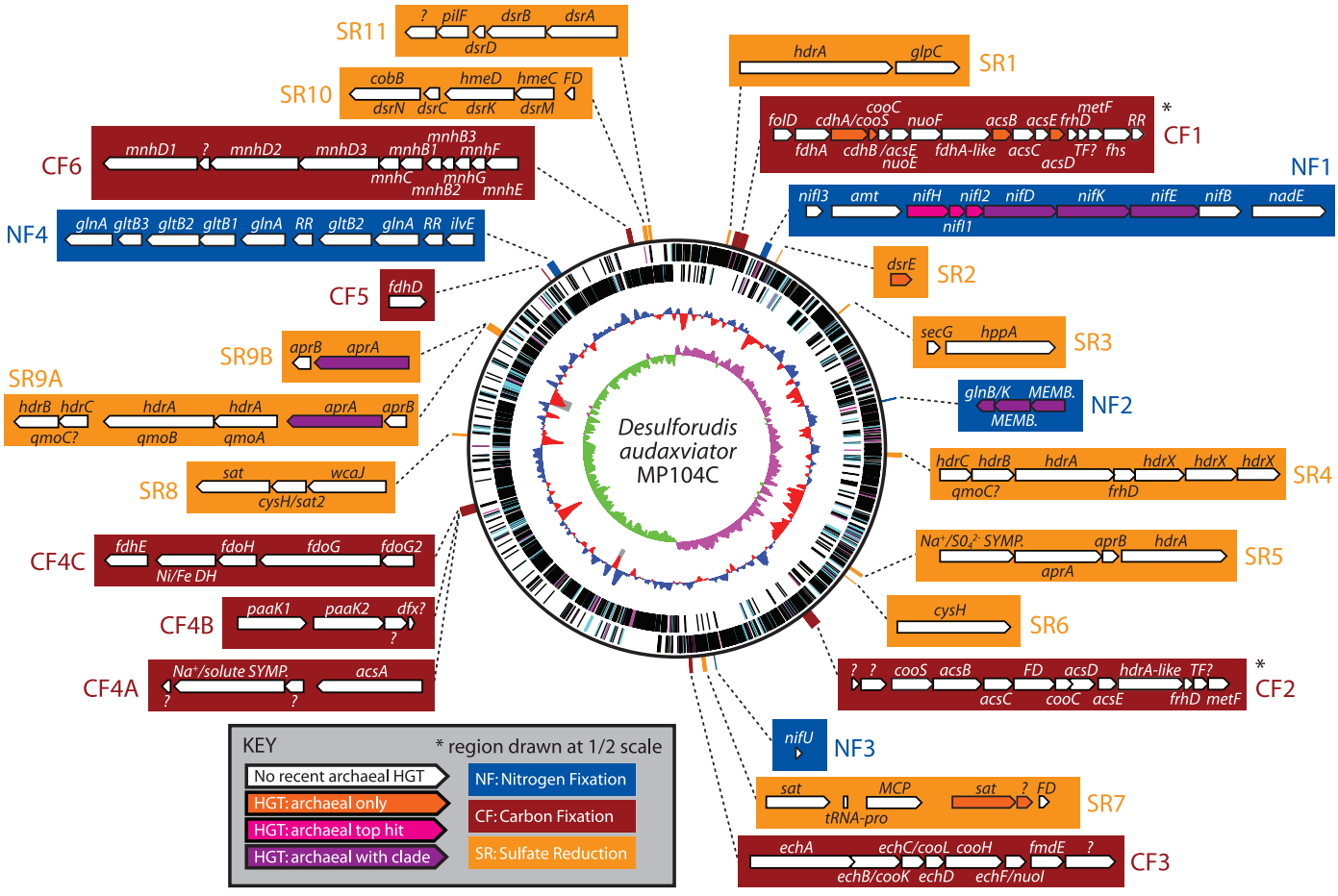

Fig. 2. Genome of *D. audaxviator*, with key genes highlighted. (Innermost ring) GC skew [average of (G-C)/(G+C) over 10,000 bases, plotted every 1000 bases]. Transition at the top (near dnaA) is origin of replication. (Second ring) G+C content [average of (G+C) over 10,000 bases, plotted every 1000 bases], with greater-than-average values (61%) in blue and below average in red. Below-average G+C regions that result from CRISPR sequences are indicated in gray. (Third and fourth rings) Predicted protein coding genes on each strand. Genes with homologs only found within closest clade species [including open reading frame (ORF)an genes] are in cyan, genes that are found only within closest clade species and within archaea (resulting from horizontal transfer) in magenta, and all other genes in black. (Outer boxes) Genes of interest are shown around the ring as operons for sulfate reduction (SR), carbon fixation via acetyl-CoA synthesis pathway (CF), and nitrogen fixation (NF). Horizontally acquired genes shared with archaea specific to *D. audaxviator* and its nearest relatives are colored according to the key.

synthetase) to obtain its nitrogen from ammonia without resorting to an energetically costly nitrogenase conversion of N_2 to ammonia. Nonetheless, a nitrogenase is present in the genome (Fig. 2 and table S15) with a nifH subunit that is more similar to archaeal types, including high-temperature variants (*24*), than to the nitrogenase of *Desulfotomaculum reducens* (figs. S4 and S8). It may be that *D. audaxviator* is not always presented with sufficient amounts of ammonia, so the versatility provided by the horizontally acquired nitrogenase may have contributed substantially to the success of *D. audaxviator* in colonizing such habitats.

D. audaxviator shares other genes with archaea that may confer benefits in extreme environments. In addition to the unusual nitrogenase and sulfate adenylyltransferase, acquisitions by ancestors of *D. audaxviator* (table S10) include a second CODH system (CODH1 in Fig. 2 and fig. S7), cobalamin biosynthesis protein CobN, and genes for the formation of gas vesicles. It also has two clustered regularly interspaced short palindromic repeat (CRISPR) regions (table S12) that are used for viral defense (*25*) and that occur in the genome with adjacent CRISPR-associated genes (CAS), some of which are horizontally shared between *D. audaxviator* and archaea.

D. audaxviator's ability to colonize independently is also assisted by its possession of all of the amino acid synthesis pathways (table S21). Other factors that may confer fitness in this environment are the ability to form endospores (table S16) and the potential for it to grow in deeper, hotter conditions (table S9) than provided by MP104. *D. audaxviator* appears capable of sensing nutrients (table S19) in its environment and possesses flagellar genes (table S18) to permit motility along chemical gradients, such as those that occur at the mineral surfaces of the fracture (*26*). One ability that *D. audaxviator* is lacking is a complete system for oxygen resistance (table S25), suggesting the long-term isolation from O_2.

The MP104 fracture contains the simplest natural environmental microbial community yet described and has yielded a single, complete genome of an uncultured microorganism with the use of environmental genomics. *D. audaxviator*'s ability to reduce SO_4^{2-} grants access to the most energetically favorable electron acceptor in the fracture zones of the Witwatersrand basin (*27*). Additionally, inherited characteristics of *D. audaxviator*, such as motility, sporulation, and carbon fixation, have been complemented by horizontally acquired systems frequently found in archaea. These abilities have enabled *D. audaxviator* to colonize the deep subsurface, a process that, unlike surface habitats which permit more immediate access, has required fitness throughout the history of the colonization. This bold traveler (*audax viator*) has revealed a mode of life isolated from the photosphere, capturing all of the roles necessary for an independent life-style and showing that it is possible to encode the entire biological component of a simple ecosystem within a single genome.

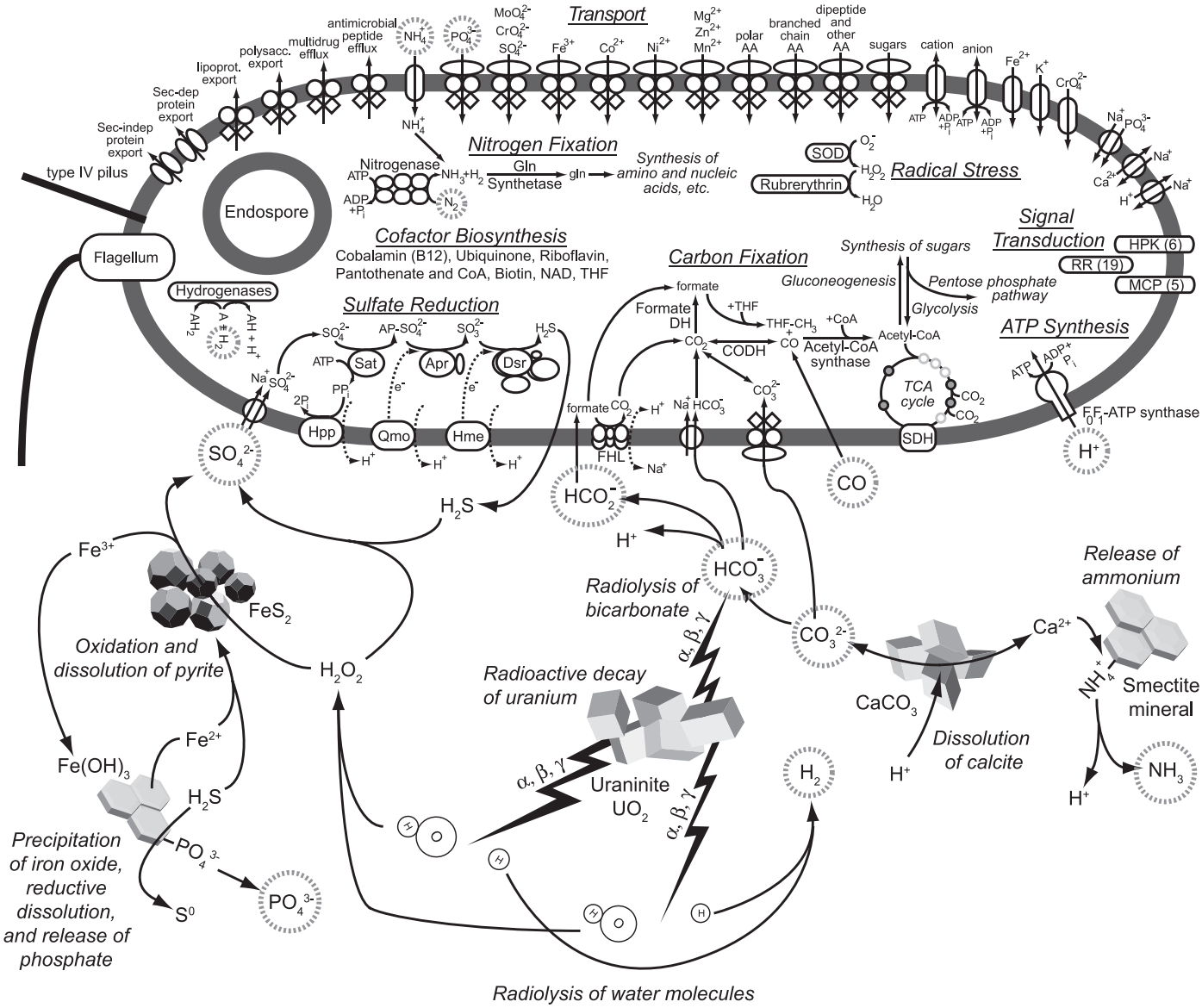

Fig. 3. Model of the single-species ecosystem at MP104. *D. audaxviator*'s machinery is shown in a cartoon representation, including pathways for sulfate reduction, nitrogen fixation, and carbon fixation. Signal transduction proteins are reported including the number found in parentheses, with MCP indicating methyl-accepting chemotaxis proteins; HPK, histidine protein kinases; and RR, response regulators. Transporters include approximate substrates. Also shown are the radiolytically generated sources of energy and nutrients for the ecosystem, as detailed in Lin *et al.* (*11*), shown experimentally by Lefticariu *et al.* (*28*), and described in (*16*).

References and Notes

1. A. M. Deutschbauer, D. Chivian, A. P. Arkin, *Curr. Opin. Biotechnol.* **17**, 229 (2006).
2. O. Beja *et al.*, *Environ. Microbiol.* **2**, 516 (2000).
3. M. R. Rondon *et al.*, *Appl. Environ. Microbiol.* **66**, 2541 (2000).
4. J. C. Venter, *Science* **304**, 66 (2004); published online 4 March 2004 (10.1126/science.1093857).
5. G. W. Tyson *et al.*, *Nature* **428**, 37 (2004).
6. S. G. Tringe, *Science* **308**, 554 (2005).
7. M. Strous *et al.*, *Nature* **440**, 790 (2006).
8. D. B. Rusch *et al.*, *PLoS Biol.* **5**, e77 (2007).
9. T. J. Phelps, E. M. Murphy, S. M. Pfiffer, D. C. White, *Microb. Ecol.* **28**, 335 (1994).
10. B. B. Jørgensen, S. D'Hondt, *Science* **314**, 932 (2006).
11. L. H. Lin *et al.*, *Science* **314**, 479 (2006).
12. D. P. Moser *et al.*, *Appl. Environ. Microbiol.* **71**, 8773 (2005).
13. D. P. Moser *et al.*, *Geomicrobiol. J.* **20**, 517 (2003).
14. T. M. Gihring *et al.*, *Geomicrobiol. J.* **23**, 415 (2006).
15. F. H. Chapelle *et al.*, *Nature* **415**, 312 (2002).
16. Materials and methods are available as supporting material on *Science* Online.
17. V. Zverlov *et al.*, *J. Bacteriol.* **187**, 2203 (2005).
18. M. L. Sogin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 12115 (2006).
19. E. J. Alm *et al.*, *Genome Res.* **15**, 1015 (2005).
20. S. J. Giovannoni *et al.*, *Science* **309**, 1242 (2005).
21. M. Mussmann *et al.*, *J. Bacteriol.* **187**, 7126 (2005).
22. H. L. Drake, S. L. Daniel, *Res. Microbiol.* **155**, 869 (2005).
23. M. Wu *et al.*, *PLoS Genet.* **1**, e65 (2005).
24. M. P. Mehta, J. A. Baross, *Science* **314**, 1783 (2006).
25. R. Barrangou *et al.*, *Science* **315**, 1709 (2007).
26. G. Wanger, T. C. Onstott, G. Southam, *Geomicrobiol. J.* **23**, 443 (2006).
27. T. C. Onstott *et al.*, *Geomicrobiol. J.* **23**, 369 (2006).
28. L. Lefticariu, L. M. Pratt, E. M. Ripley, *Geochim. Cosmochim. Acta* **70**, 4889 (2006).

29. We thank J. Banfield and G. Tyson for helpful discussion; J. Bruckner and B. Baker for assistance with microscopy; F. Warnecke for advice on 16S fluorescent in situ hybridization; T. Kieft, G. Zane, and the MicrobesOnline team (M. Price, K. Keller, and K. Huang) for advice; and D. Kershaw and colleagues at the Mponeng mine and AngloGold Ashanti Limited, RSA. This work was part of the Virtual Institute for Microbial Stress and Survival (http://vimss.lbl.gov), supported by DOE, Office of Science, Office of Biological and Environmental Research, Genomics Program:GTL through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and DOE. This work was also supported by the NASA Astrobiology Institute through award NNA04CC03A to the IPTAI Team co-directed by L.M.P. and T.C.O. A.P.A. received support from the Howard Hughes Medical Institute. The genome sequence and 16S library sequences reported in this study have been deposited in GenBank under the accession numbers CP000860 and EU730965 to EU731008, respectively.

Supporting Online Material

www.sciencemag.org/cgi/content/full/322/5899/275/DC1
Materials and Methods
Figs. S1 to S8
Tables S1 to S26
References

22 January 2008; accepted 11 September 2008
10.1126/science.1155495