

in the *D. pachea nvd* coding region have caused the loss of NVD activity with cholesterol substrate. These mutations have turned *D. pachea* into an obligate specialist dependent on lathosterol, a compound that has been found in a single plant species in the Sonoran desert (5, 6).

Remarkably, *D. melanogaster nvd* RNAi flies expressing *D. pachea nvd* survive significantly better on lathosterol than on cholesterol (*t* test,  $t_{10,11} = 2.029$ ,  $P < 0.03$ ) (Fig. 2), but no effect on survival was detected with *nvd* RNAi flies expressing *D. pachea nvd* with the four ancestral amino acid changes (Fig. 2). This suggests that the mutations that abolished cholesterol conversion during *D. pachea* evolution provide a fitness advantage on lathosterol. The underlying mechanism remains unclear. Our *in vitro* assay does not uncover any benefit from the *D. pachea nvd* mutations: *D. pachea* NVD *in vitro* activity with lathosterol is not higher compared with other species (Fig. 3), and the NVD enzymes of related *Drosophila* species are already able to convert lathosterol into 7DHC. To assess population genetic forces at play on the *nvd* genomic region, we compared the 3-kb *nvd* locus and seven genes on the same 100-kb scaffold with nine control genes in 34 individuals from a single natural population. Our analysis reveals that *nvd* is in a genomic region of low nucleotide diversity, low recombination rate, and normal divergence rate (McDonald-Kreitman test,  $P > 0.85$ ; maximum likelihood extension of the Hudson-Kreitman-Aguadé test,  $P < 10^{-5}$ ) (Fig. 4 and tables S5 to S11). A signature of a selective sweep is detected [Kim and Nielsen omega (17)] over *nvd* and neighboring loci (Fig. 4), but nucleotide polymorphism is too low to infer whether this recent selection acted on the *nvd* mutations themselves. Tajima's *D* and Fu and Li tests are consistent with recovery from selective sweep in the *nvd* region (table S6).

A likely scenario is that *D. pachea* first evolved a resistance toward senita cactus toxic compounds (5) and slowly became restricted to this food source as it escaped competition with other fly species. Evolution of *D. pachea*'s resistance most likely did not involve NVD because *nvd* is not expressed in the midgut and fat body (fig. S3), the detoxification organs in insects (16). As lathosterol became *D. pachea*'s unique source of sterols for steroid hormone synthesis, mutations in *nvd* that abolished NVD activity on cholesterol appeared and were fixed rapidly due to their beneficial effect with lathosterol. As a result, *D. pachea* became an obligate specialist on the senita cactus. We point out that besides *nvd* mutations, mutation(s) in other genes might also have contributed to *D. pachea* dependence on lathosterol. Alternatively, the identified *nvd* mutations may have spread while *D. pachea* ancestors were still feeding on various plants and may thus have accelerated its ecological specialization. Our study, which uncovered several mutations underlying the obligate bond between a specialist species and its host, illustrates how a

few mutations in a single gene can restrict the ecological niche of a species.

#### References and Notes

1. A. Lwoff, *L'Évolution physiologique: Étude des pertes de fonctions chez les microorganismes* (Hermann, Paris, 1944).
2. P. Romero *et al.*, *Genome Biol.* **6**, R2 (2005).
3. D. Voet, *Biochemistry* (Wiley, Somerset, 1995).
4. X. Huang, J. T. Warren, L. I. Gilbert, *J. Genet. Genomics* **35**, 1 (2008).
5. W. B. Heed, H. W. Kircher, *Science* **149**, 758 (1965).
6. C. E. Campbell, H. W. Kircher, *Phytochemistry* **19**, 2777 (1980).
7. J. C. Fogleman, P. B. Danielson, *Am. Zool.* **41**, 877 (2001).
8. Materials and methods are available as supplementary materials on Science Online.
9. T. Yoshiyama, T. Namiki, K. Mita, H. Kataoka, R. Niwa, *Development* **133**, 2565 (2006).
10. V. Rottiers *et al.*, *Dev. Cell* **10**, 473 (2006).
11. S. Pitnick, W. B. Heed, *Ann. Entomol. Soc. Am.* **87**, 307 (1994).
12. R. Lafont, C. Dauphin-Villemant, J. T. Warren, H. H. Rees, in *Comprehensive Molecular Insect Science*, L. I. Gilbert, K. Iatrou, S. Gill, Eds. (Oxford, 2005), vol. 3, pp. 125–195.
13. C. Blais, T. Blasco, A. Maria, C. Dauphin-Villemant, R. Lafont, *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **878**, 925 (2010).
14. K. C. Goodnight, H. W. Kircher, *Lipids* **6**, 166 (1971).
15. T. Yoshiyama-Yanagawa *et al.*, *J. Biol. Chem.* **286**, 25756 (2011).
16. O. Mittapalli *et al.*, *PLoS ONE* **5**, e13708 (2010).
17. Y. Kim, R. Nielsen, *Genetics* **167**, 1513 (2004).

**Acknowledgments:** We thank M. Joron for the linkage disequilibrium heat map; M.-A. Félix, N. Gompel, and C. Desplan for comments on the manuscript; C. Parada for help with field work; T. A. Markow for *D. pachea* samples; C. S. Thummel and the San Diego *Drosophila* Species Stock Center for flies; Y. Hiroimi for reagents; T. Blasco for liquid chromatography tandem mass spectrometry analyses; and M. Gho for hosting V.O. in 2009–2010. Supported by CNRS ATIP-AVENIR (to V.O.), French Foreign Ministry postdoctoral fellowship (to M.L.), NIH grant AI064950 (to A.G.C.), Japan-France Bilateral Cooperating Program from the Japan Society for the Promotion of Science (JSPS) (to H.K. and C.D.-V.), NSF award DEB-1020009 (to L.M.M.), JSPS postdoctoral fellowship (to T.Y.-Y.), Special Coordination Funds for Promoting Science and Technology from the Ministry of Education, Culture, Sports, Science and Technology in Japan (to R.N.), and Amylin Endowment (to T.A. Markow). The *nvd* sequences were deposited in GenBank under accession nos. JF764559 to JF764595 and JX066807 to JX067384. The authors declare no conflict of interest. V.O. conceived the study and wrote the paper. M.L., C.B., R.L., C.D.-V., L.M.M., H.K., and R.N. provided technical support and conceptual advice for designing the experiments. V.O., M.L., G.G., S.M., L.M.M., C.B., E.G., T.Y.-Y., C.D.-V., and R.N. performed the experiments. A.G.C. did the population genetics analysis.

#### Supplementary Materials

www.sciencemag.org/cgi/content/full/337/6102/1658/DC1  
Materials and Methods  
Supplementary Text  
Figs. S1 to S13  
Tables S1 to S11  
References (18–63)

17 May 2012; accepted 13 July 2012  
10.1126/science.1224829

## Fermentation, Hydrogen, and Sulfur Metabolism in Multiple Uncultivated Bacterial Phyla

Kelly C. Wrighton,<sup>1</sup> Brian C. Thomas,<sup>1</sup> Itai Sharon,<sup>1</sup> Christopher S. Miller,<sup>1</sup> Cindy J. Castelle,<sup>2</sup> Nathan C. VerBerkmoes,<sup>3</sup> Michael J. Wilkins,<sup>4</sup> Robert L. Hettich,<sup>3</sup> Mary S. Lipton,<sup>4</sup> Kenneth H. Williams,<sup>2</sup> Philip E. Long,<sup>2</sup> Jillian F. Banfield<sup>1,2\*</sup>

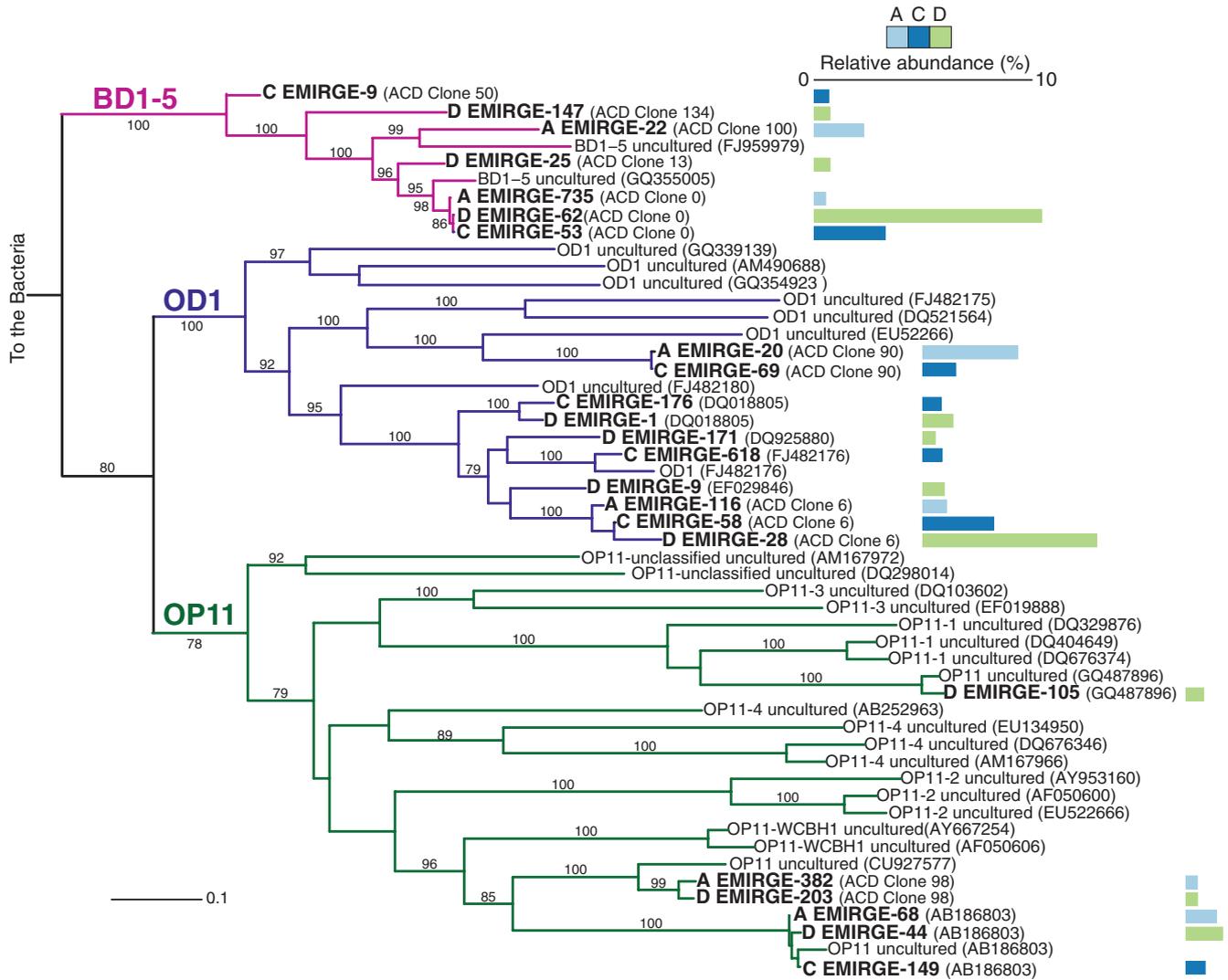
BD1-5, OP11, and OD1 bacteria have been widely detected in anaerobic environments, but their metabolisms remain unclear owing to lack of cultivated representatives and minimal genomic sampling. We uncovered metabolic characteristics for members of these phyla, and a new lineage, PER, via cultivation-independent recovery of 49 partial to near-complete genomes from an acetate-amended aquifer. All organisms were nonrespiring anaerobes predicted to ferment. Three augment fermentation with archaeal-like hybrid type II/III ribulose-1,5-bisphosphate carboxylase-oxygenase (RuBisCO) that couples adenosine monophosphate salvage with CO<sub>2</sub> fixation, a pathway not previously described in Bacteria. Members of OD1 reduce sulfur and may pump protons using archaeal-type hydrogenases. For six organisms, the UGA stop codon is translated as tryptophan. All bacteria studied here may play previously unrecognized roles in hydrogen production, sulfur cycling, and fermentation of refractory sedimentary carbon.

**S**equencing of total DNA recovered directly from natural systems (metagenomics) often reveals previously unknown genes (1, 2)

<sup>1</sup>Department of Earth and Planetary Science, University of California, Berkeley, Berkeley, CA 94720, USA. <sup>2</sup>Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. <sup>3</sup>Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA. <sup>4</sup>Pacific Northwest National Laboratory, Richland, WA 99352, USA.

\*To whom correspondence should be addressed. E-mail: jbanfield@berkeley.edu

and has the potential to yield near-complete genomes suitable for metabolic and phylogenetic analyses (3–5). Numerous bacteria are known exclusively through cultivation-independent recovery of their ribosomal RNA (rRNA) genes and thus are important targets for this approach (6). Here, we sequenced DNA from three microbial communities from an acetate-amended aquifer to reconstruct genomes of organisms that may contribute to biogeochemical cycling in anoxic



**Fig. 1.** Maximum likelihood 16S rRNA gene phylogenetic tree showing the placement of the uncultivated phyla recovered by EMIRGE (bold text). The closest representative to each EMIRGE sequence is denoted in parentheses (clone library or Silva accession number). The bar chart

indicates the relative abundance of each sequence in the A, C, and D samples (maximum is 11%), with bootstrap support >80 noted. The 16S rRNA tree from all organisms and details are provided in fig. S2 and (7).

subsurface environments. We recovered 49 genomes from members of candidate phyla widely encountered in rRNA microbial surveys and found evidence for metabolic strategies not previously described in Bacteria. Overall, this study contributes new insights into the physiology and diversity across several major branches of the tree of life.

Groundwater samples (A, C, and D) were collected 5, 7, and 10 days after the start of addition of acetate to an anoxic aquifer in Colorado, USA (fig. S1) (7). From each sample, we recovered microbial cells that passed through a 1.2- $\mu$ m pre-filter to be retained on a 0.2- $\mu$ m filter. The samples were immediately frozen on site for DNA extraction and for mass spectrometry-based proteomics to verify the activity of organisms in situ (7).

Illumina sequences from DNA extracted from each sample were coassembled by using strategies optimized for community genomics (7).

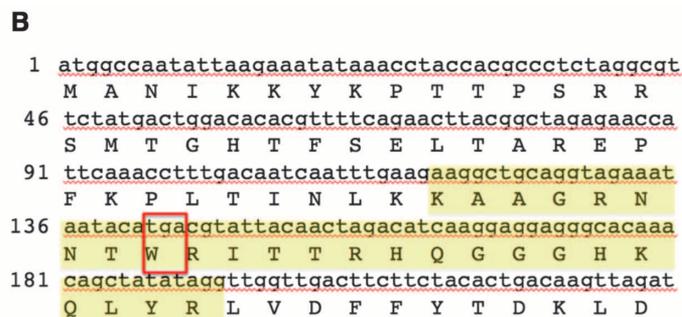
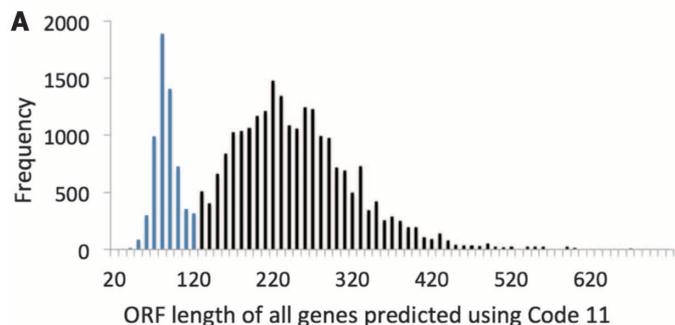
**Table 1.** Overview of genome recovery. Individual genome completion information is in fig. S8.

Candidate divisions	Number of genomes	Percent complete	Bins with >1 genome	Genomes >90%	Genomes >50%	Estimated genes/genome
OD1	21	75 $\pm$ 24	0	10	18	1353 $\pm$ 350
OP11	19	67 $\pm$ 26	5	3	15	1364 $\pm$ 252
BD1-5	5	93 $\pm$ 3	0	5	5	1540 $\pm$ 222
PER	3	80 $\pm$ 17	0	2	3	1666 $\pm$ 221
ACD80	1	95	0	1	1	1301

Using EMIRGE (8), 16S rRNA genes were reconstructed and confirmed by clone library analysis (7). By linking 16S rRNA (Fig. 1 and fig. S2) to phylogenetically informative genes in the assembly (fig. S3 and table S1), we demonstrated genomic sampling of organisms affiliated with the phylum-level groups OD1, OP11, and BD1-5 (7). Another bacterial group, although it formed a monophyletic clade, did display wandering be-

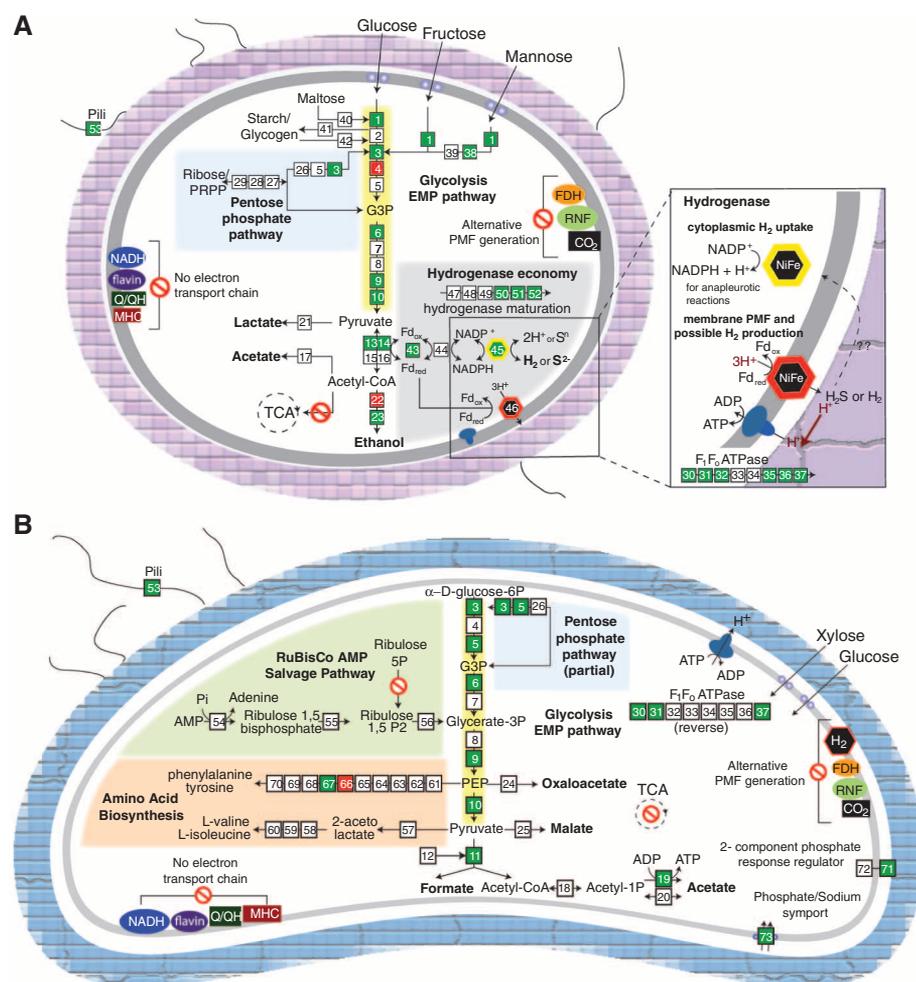
havior in terms of relative phylogenetic position in protein-coding phylogenetic analyses; it is referred to as the Peregrines (PERs). On the basis of protein-coding trees and partial 16S rRNA gene information, we suggest that PERs may represent a previously unknown phylum-level branch within the bacterial domain (7).

Genome fragments from the coassembly (termed ACD) were clustered into 87 organism



**Fig. 2.** Detection of alternative coding. **(A)** Histogram of average open reading frame (ORF) length achieved with ORF predictions using the standard bacterial genetic code. The peak with unusually small gene lengths is as-

sociated with ORFs that should have been predicted using code 4. **(B)** Peptides identified by proteomics were mapped onto proteins with code 4 predictions to verify that UGA codes for tryptophan (W).



**Fig. 3.** Metabolic models with detected genes (white box), genes with proteins confirmed by proteomics (green box), and genes missing from pathways (red box). For full gene information for box numbers see table S4. **(A)** OD1-i may produce acetate, ethanol, lactate, and hydrogen as fermentation end products. Hydrogen could be generated via NiFe type 4 membrane-bound hydrogenases (red outline) and cytoplasmic type 3b hydrogenase (yellow outline). The expanded view shows type 4 hydrogenase-generated PMF and ATP synthesis (red line) and possible hydrogen cycling to type 3b hydrogenases (black dashed line). FDH, formate dehydrogenase; Fd<sub>ox</sub> and Fd<sub>red</sub>, oxidized and reduced ferredoxin, respectively; G3P, glyceraldehyde-3-phosphate; MHC, multiheme c-type cytochrome; NADP<sup>+</sup>, nicotinamide adenine dinucleotide phosphate; Q/QH, ubiquinone–reduced ubiquinone; rhodobacter nitrogen-fixing (RNF) complex. **(B)** PERs produce acetate and formate from pyruvate (no formate dehydrogenases or hydrogenases were identified). PER genomes lack detectable mechanisms for recovering additional energy via membrane-potential coupled to ATP synthase but may use RuBisCO, analogous to Archaea, for salvaging ATP by shunting nucleotides through central-carbon metabolism. PEP, phosphoenolpyruvate.

“bins” from emergent self-organizing map analysis of their tetranucleotide sequence composition (9). Organism abundance ratios between samples were used to further refine binning (7). Here, we focus on 49 genomes from BD1-5, OP11, OD1, and PER, relevant to carbon, sulfur, and hydrogen cycling (additional analyses of these and other genomes will be reported separately). From the inventory of conserved, single-copy genes (7, 10), we estimated that 21 of the 49 genomes were >90% complete (Table 1). Note that the majority of these organisms each represented ~1% of the assembled community (table S2).

Previously, only 33 protein-coding genes have been reported for the OD1 phylum (11). We recovered more than 24,000 OD1 gene sequences (with an average of 1119 genes per genome) for 21 species, on genome fragments up to 358 kilobase pairs. Phylogenetically, we resolved multiple OD1 lineages and recognized one sublineage, OD1-i (fig. S3). For OP11, there has only been one fragmented partial sampling (~270 kilobases) from a single cell (16S rRNA OP11 class “unclassified”) (12). Here, we recovered more than 25,000 genes (with an average of 1337 genes per genome) for 19 organisms from OP11 classes I and WCBH1-64 (Fig. 1). There has been no previous genomic sampling of PER or BD1-5 (7).

BD1-5 genes predicted using the standard bacterial genetic code were anomalously short (7), with a low overall coding density. We deduced, and confirmed using proteomic analyses (Fig. 2), that they use genetic code 4, where the normal stop codon (UGA) is translated as tryptophan (W). Recoding of UGA to W in Bacteria is rare but has been noted within the Firmicutes and Proteobacteria phyla (some Mollicutes and Alphaproteobacteria). It is often associated with small genomes and low GC content and may be a consequence of genome reduction (13). Our code 4 genomes are estimated to be <2 Mb and have low but variable GC contents (27 to 43%). One genotype, ACD80 is phylogenetically very distantly related to the other BD1-5 organisms (fig. S3), and it may ultimately be recognized as a separate lineage (7). Future sampling may resolve whether code 4 usage is an ancestral trait or arose independently.

Genomic analyses indicate that all 49 organisms were nonrespiring. Given the lack of a tricarboxylic acid (TCA) cycle, subunits of the reduced form of nicotinamide adenine dinucleotide (NADH) dehydrogenase, and most other electron-transport chain complexes including terminal oxidases, we infer a strictly anaerobic fermentation-based lifestyle. All have a glycolysis (Embden-Meyerhof-Parnas) pathway and convert pyruvate to acetyl-coenzyme A (acetyl-CoA) without pyruvate dehydrogenase, instead they use pyruvate-formate lyase (PFL in PER) or pyruvate ferredoxin oxidoreductase (PFOR in OD1, OP11, and BD1-5) (Fig. 3). Most generate adenosine triphosphate (ATP) by converting acetyl-CoA to acetate via two enzymes (acetate kinase and phosphate acetyltransferase) (Fig. 3B). However, the OD1-i are predicted to use a single enzyme, acetyl-CoA synthetase (EC: 6.2.1.13) (Fig. 3A) for ATP generation, which is common in some fermentative sulfur-reducing Archaea (e.g., *Pyrococcus* spp.) but rare in Bacteria (14). OD1-i may reoxidize NADH produced during glycolysis by converting pyruvate to D-lactate and acetyl-CoA to ethanol (Fig. 3A). Excreted fermentation end products could sustain Geobacteraceae and sulfate-reducing bacteria that bloom when acetate is added to the aquifer to promote uranium bioremediation (fig. S2) (15).

The BD1-5, OP11, and OD1 phyla contain members that augment substrate-level phosphorylation from fermentation by coupling a mem-

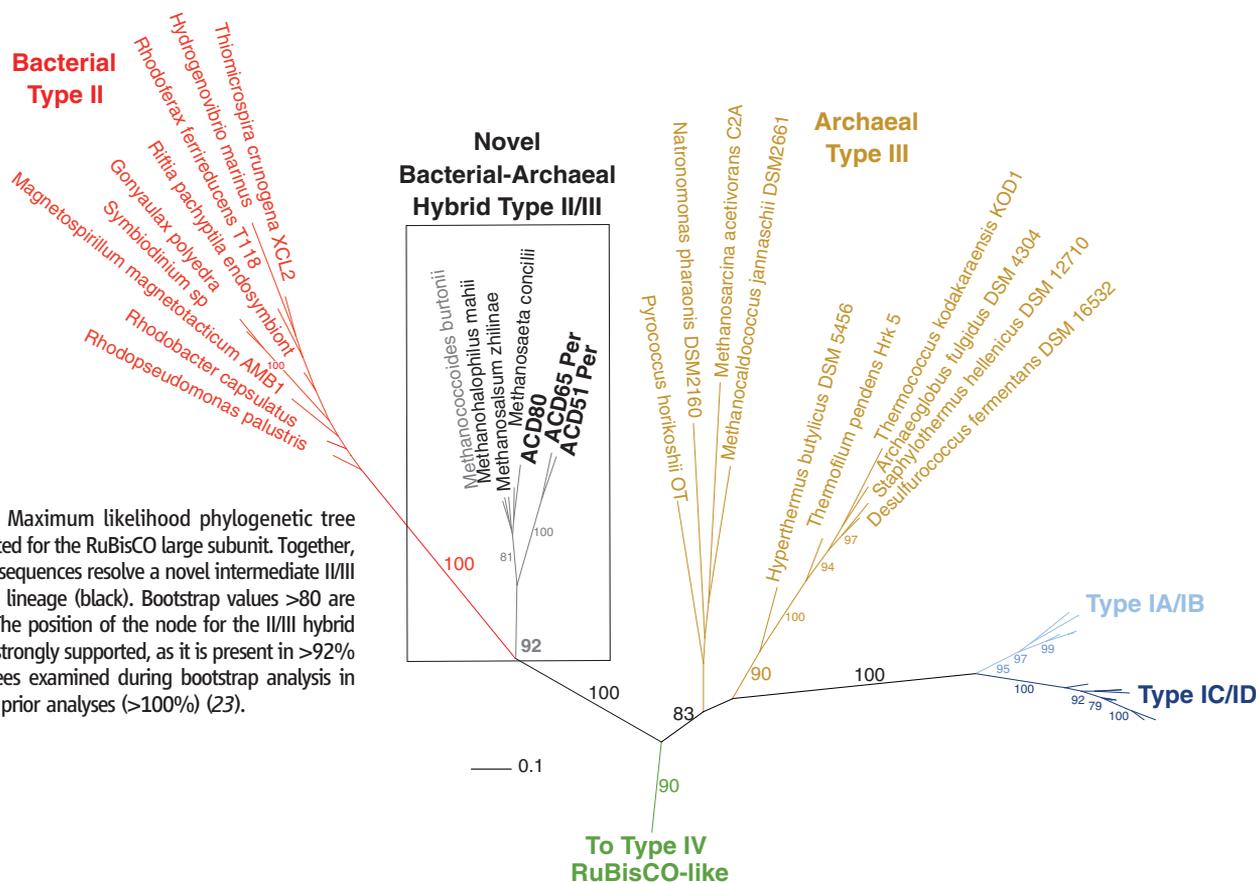
brane proton-motive force (PMF) (independent of a canonical electron-transport chain) to ATP generation via a  $F_1F_0$  ATP synthase. In some BD1-5 and OP11, this appears to rely on proton-pumping by membrane-bound pyrophosphatases ( $H^+$ -PPases). Similar  $H^+$ -PPases studied in syntrophic bacteria (16) translocate protons across the membrane for energy conservation via ATP synthase. We find proteomic support for  $H^+$ -PPase (BD1-5 and OP11), alcohol dehydrogenase (OD1), glycolysis, PFOR or PFL, and ATP synthase from all lineages; this confirmed fermentation in situ (Fig. 3 and table S3).

Some fermentative anaerobes produce  $H_2$  to dispose of excess reductant (17). In OD1 and OP11, we identified three Fe-only hydrogenases and 23 NiFe hydrogenases (table S5). Phylogenetic analyses of the NiFe hydrogenase catalytic subunits revealed that 17 are type 3b cytoplasmic hydrogenases most closely related to those of fermentative, sulfur-reducing Thermococcales Archaea (figs. S4 and S5) (18). The type 3b hydrogenases may produce  $H_2$  during fermentation or  $H_2S$  when polysulfide is available. Alternatively, they may consume  $H_2$  to produce the reduced form of nicotinamide adenine dinucleotide phosphate (NADPH) for anabolic metabolism (18, 19). Proteomic detection of hydrogenase-related proteins in conjunction with proteins for fermentation, as well as the abundance of organisms with this capacity when sulfide was detected in groundwater (Figs. 1 and 3A), supports a role in  $H_2$  or  $H_2S$

generation rather than  $H_2$  consumption. We also identified type 4 membrane-bound hydrogenase genes in some OD1 genomes (fig. S5). The dual hydrogenase system may function in intracellular hydrogen cycling, as in Thermococcales (7, 20–22), where PMF and  $H_2$  are produced by the membrane-bound hydrogenase, and  $H_2$  is shuttled to the cytoplasmic hydrogenase where it is oxidized to produce NADPH (Fig. 3A).

In ACD80 and two PER genomes, we identified putative ribulose-1,5-bisphosphate carboxylase-oxygenase (RuBisCO) genes potentially involved in  $CO_2$  assimilation. All residues for catalytic activity are present, except one for substrate binding that is absent in the ACD80 version (fig. S6). This finding and structural modeling (7) suggest carboxylase-oxygenase activity. Our RuBisCO sequences clade separately from the bacterial type II sequences and are most closely related to a sequence from *Methanococcoides burtonii* (MBR) and a global ocean sampling (GOS)-derived sequence of unknown affiliation that may represent a II/III hybrid RuBisCO (Fig. 4) (23). We expanded the membership of this clade to eight sequences by identifying three genes in publicly available methanogenic archaeal genomes (*Methanohalophilus zhilinae*, *Methanosaeta concilii*, and *Methanohalophilus mahii*) (table S1C). Our data show that the RuBisCO hybrid occurs in Bacteria.

The MBR type II/III RuBisCO function is comparable to the type III archaeal RuBisCO (23). It does not function in the Calvin-Benson-



**Fig. 4.** Maximum likelihood phylogenetic tree constructed for the RuBisCO large subunit. Together, the new sequences resolve a novel intermediate II/III RuBisCO lineage (black). Bootstrap values >80 are shown. The position of the node for the II/III hybrid clade is strongly supported, as it is present in >92% of all trees examined during bootstrap analysis in this and prior analyses (>100%) (23).

Bassham (CBB) pathway but fixes CO<sub>2</sub> and contributes to adenosine monophosphate (AMP) recycling (7, 24). We predict that the bacterial type II/III RuBisCO genes share this function. Homologs of DeoA and E2b2, key enzymes of the CO<sub>2</sub>-fixing AMP-recycling pathway (24), were identified (Fig. 3B), and no essential CBB cycle enzymes (e.g., phosphoribulokinase) were detected. Salvaging purine-pyrimidine products to produce RuBisCO generates 3-phosphoglycerates and perhaps pyruvate (ACD80), which could ultimately be fermented for ATP production (7).

From this study and rRNA gene survey information indicating prevalence in anoxic, organic carbon-rich environments (25, 26), we predict widespread fermentation-based metabolism in the 49 OD1, OP11, BD1-5, and PER genomes sampled here. We find it intriguing that several pathways for anoxic carbon, hydrogen, and sulfur cycling in these organisms share features previously documented only in Archaea. Some OD1 may contribute to sulfur cycling, on the basis of their previous association with sulfur-rich environments (11, 12, 26–28). Given the absence of genes for sulfur respiration in our near-complete OD1-i genomes, the link may involve hydrogenase-mediated sulfur-reductase activity. Notably, these insights were obtained through cultivation-independent analyses and have contributed more near-complete (>90%) genomic sampling for OD1 than is available for almost half of all of the genomically characterized bacterial phyla (29).

## References and Notes

- M. Hess *et al.*, *Science* **331**, 463 (2011).
- D. Wu *et al.*, *PLoS ONE* **6**, e18011 (2011).
- G. W. Tyson *et al.*, *Nature* **428**, 37 (2004).
- V. Iversen *et al.*, *Science* **335**, 587 (2012).
- P. G. Falkowski, T. Fenchel, E. F. Delong, *Science* **320**, 1034 (2008).
- M. S. Rappé, S. J. Giovannoni, *Annu. Rev. Microbiol.* **57**, 369 (2003).
- Materials and methods are available as supplementary materials on Science Online.
- C. S. Miller, B. J. Baker, B. C. Thomas, S. W. Singer, J. F. Banfield, *Genome Biol.* **12**, R44 (2011).
- G. J. Dick *et al.*, *Genome Biol.* **10**, R85 (2009).
- J. Raes, J. O. Korbel, M. J. Lercher, C. von Mering, P. Bork, *Genome Biol.* **8**, R10 (2007).
- M. S. Elshahed *et al.*, *Appl. Environ. Microbiol.* **71**, 7598 (2005).
- N. H. Youssef, P. C. Blainey, S. R. Quake, M. S. Elshahed, *Appl. Environ. Microbiol.* **77**, 7804 (2011).
- J. P. McCutcheon, B. R. McDonald, N. A. Moran, *PLoS Genet.* **5**, e1000565 (2009).
- X. Mai, M. W. Adams, *J. Bacteriol.* **178**, 5897 (1996).
- R. T. Anderson *et al.*, *Appl. Environ. Microbiol.* **69**, 5884 (2003).
- L. Schöcke, B. Schink, *Eur. J. Biochem.* **256**, 589 (1998).
- P. M. Vignais, B. Billoud, *Chem. Rev.* **107**, 4206 (2007).
- K. Ma, R. N. Schicho, R. M. Kelly, M. W. Adams, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 5341 (1993).
- D. J. van Haaster, P. J. Silva, P. L. Hagedoorn, J. A. Jongejans, W. R. Hagen, *J. Bacteriol.* **190**, 1584 (2008).
- R. Sapra, K. Bagramyan, M. W. W. Adams, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 7545 (2003).
- P. J. Silva *et al.*, *Eur. J. Biochem.* **267**, 6541 (2000).
- T. Kanai *et al.*, *J. Bacteriol.* **193**, 3109 (2011).
- F. R. Tabita, T. E. Hanson, S. Satagopan, B. H. Witte, N. E. Kreef, *Philos. Trans. R. Soc. B* **363**, 2629 (2008).
- T. Sato, H. Atomi, T. Imanaka, *Science* **315**, 1003 (2007).

- C. Briée, D. Moreira, P. López-García, *Res. Microbiol.* **158**, 213 (2007).
- S. Peura *et al.*, *ISME J.* **6**, 1640 (2012).
- P. Hugenholtz, C. Pittule, K. L. Hershberger, N. R. Pace, *J. Bacteriol.* **180**, 366 (1998).
- J. K. Harris, S. T. Kelley, N. R. Pace, *Appl. Environ. Microbiol.* **70**, 845 (2004).
- I. Pagani *et al.*, *Nucleic Acids Res.* **40** (Database issue), D571 (2012).

**Acknowledgments:** We thank F. Larimer for RuBisCO analyses input; M. Shah for proteomic support; C. Thrash, P. Hugenholtz, and J. Eisen for manuscript suggestions; and the U.S. Department of Energy (DOE) Subsurface Biogeochemistry Program for funding, the DOE Knowledgebase Program for funding Ggkbase, and EMBO for a fellowship to I.S. The Rifle, Colorado, Integrated Field Research Center Project is managed by Lawrence Berkeley National Laboratory for the U.S. DOE (contract no. DE-AC02-05CH11231). Portions of this work were performed in the Environmental Molecular Science Laboratory, a DOE national scientific user facility at Pacific Northwest National Laboratory. The sequences were deposited in the National Center for Biotechnology Information Sequence Read Archive (accession no. SRA050978.1). This Whole-Genome Shotgun project has been deposited at GenBank under the accession no. AMFJ00000000. The version described in this paper is the first version, AMFJ01000000. Genomic and proteomic data can be accessed via [http://geomicrobiology.berkeley.edu/rifle/acd\\_ggkbase.html](http://geomicrobiology.berkeley.edu/rifle/acd_ggkbase.html).

## Supplementary Materials

[www.sciencemag.org/cgi/content/full/337/6102/1661/DC1](http://www.sciencemag.org/cgi/content/full/337/6102/1661/DC1)  
Materials and Methods  
Supplementary Text  
Figs. S1 to S9  
Tables S1 to S6  
References (30–58)  
Database 1

3 April 2012; accepted 18 July 2012  
10.1126/science.1224041

# Disulfide Rearrangement Triggered by Translocon Assembly Controls Lipopolysaccharide Export

Shu-Sin Chng,<sup>1,2\*†</sup> Mingyu Xue,<sup>1†</sup> Ronald A. Garner,<sup>1</sup> Hiroshi Kadokura,<sup>2,3</sup> Dana Boyd,<sup>2</sup> Jonathan Beckwith,<sup>2</sup> Daniel Kahne<sup>1,4‡</sup>

The presence of lipopolysaccharide (LPS) on the cell surface of Gram-negative bacteria is critical for viability. A conserved  $\beta$ -barrel membrane protein LptD (lipopolysaccharide transport protein D) translocates LPS from the periplasm across the outer membrane (OM). In *Escherichia coli*, this protein contains two disulfide bonds and forms the OM LPS translocon with the lipoprotein LptE. Here, we identified seven *in vivo* states on the oxidative-folding pathway of LptD. Proper assembly involved a nonfunctional intermediate containing non-native disulfides. Intermediate formation required the oxidase DsbA, and subsequent maturation to the active form with native disulfides was triggered by LptE. Thus, disulfide bond-dependent protein folding of LptD requires the proper assembly of a two-protein complex to promote disulfide bond rearrangement.

**A** defining feature of Gram-negative organisms is the presence of lipopolysaccharide (LPS) on the cell surface (1). LPS must be properly assembled in the outer leaflet of the outer membrane (OM) to establish a permeability barrier against toxic compounds, including antibiotics (2, 3). In *Escherichia coli*, a translocon responsible for LPS movement across the OM is composed of two essential OM proteins: an in-

tegral  $\beta$ -barrel protein, LptD (lipopolysaccharide transport protein D), and a lipoprotein, LptE (4–6). The LptD/E complex forms part of the trans-envelope LPS exporter, which contains five other essential Lpt proteins that collectively move LPS from the inner membrane (IM) to the cell surface (7). Assembly of the OM LPS translocon presents a challenging protein-folding problem, because LptE resides inside LptD (5, 6) and

formation of the correct disulfide bonds in LptD is required for this translocon to function (8). How the cell coordinates assembly of the OM complex with the formation of the rest of the trans-envelope exporter is unknown.

To understand the assembly of the functional OM LPS translocon, we examined the biogenesis of LptD. *E. coli* LptD contains an N-terminal periplasmic domain (amino acids 25 to 202) and a C-terminal integral  $\beta$ -barrel domain (amino acids 203 to 784) (5), which is folded and inserted into the OM by the Bam complex ( $\beta$ -barrel assembly machine) (9–11). LptD has four cysteine residues, two in the N-terminal domain (Cys<sup>31</sup> and Cys<sup>173</sup>) and two very near the C terminus (Cys<sup>724</sup> and Cys<sup>725</sup>). In its mature form, LptD contains two long-range nonconsecutive disulfide bonds connecting the N- and C-terminal domains, one

<sup>1</sup>Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA. <sup>2</sup>Department of Microbiology and Immunobiology, Harvard Medical School, Boston, MA 02115, USA. <sup>3</sup>Graduate School of Biological Sciences, Nara Institute of Science and Technology, 8916-5, Takayama, Ikoma, Nara 630-0192, Japan. <sup>4</sup>Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA.

\*Present address: Department of Chemistry, National University of Singapore, Singapore 117543.

†These authors contributed equally to this work.

‡To whom correspondence should be addressed. E-mail: kahne@chemistry.harvard.edu

---

*This copy is for your personal, non-commercial use only.*

---

**If you wish to distribute this article to others**, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

**Permission to republish or repurpose articles or portions of articles** can be obtained by following the guidelines [here](#).

**The following resources related to this article are available online at [www.sciencemag.org](http://www.sciencemag.org) (this information is current as of August 24, 2015 ):**

A correction has been published for this article at:  
<http://www.sciencemag.org/content/338/6108/742.3.full.html>

**Updated information and services**, including high-resolution figures, can be found in the online version of this article at:  
<http://www.sciencemag.org/content/337/6102/1661.full.html>

**Supporting Online Material** can be found at:  
<http://www.sciencemag.org/content/suppl/2012/09/26/337.6102.1661.DC1.html>

A list of selected additional articles on the Science Web sites **related to this article** can be found at:  
<http://www.sciencemag.org/content/337/6102/1661.full.html#related>

This article **cites 57 articles**, 34 of which can be accessed free:  
<http://www.sciencemag.org/content/337/6102/1661.full.html#ref-list-1>

This article has been **cited by 27** articles hosted by HighWire Press; see:  
<http://www.sciencemag.org/content/337/6102/1661.full.html#related-urls>

This article appears in the following **subject collections**:  
Microbiology  
<http://www.sciencemag.org/cgi/collection/microbio>