Supplementary Information for:

Single-cell and metagenomic analyses indicate a fermentative, saccharolytic lifestyle for members of the OP9 lineage

Jeremy A. Dodsworth[1‡], Paul C. Blainey[2‡€], Senthil K. Murugapiran[1], Wesley D. Swingley[3*], Christian A. Ross[1], Susannah G. Tringe[4], Patrick S. G. Chain[4,5], Matthew B. Scholz[4,5], Chien-Chi Lo[4,5], Jason Raymond[6], Stephen R. Quake[2], and Brian P. Hedlund[1†]

[1]School of Life Sciences, University of Nevada, Las Vegas, Las Vegas, NV, 89154-4004, USA

[2]Department of Bioengineering, Stanford University, and Howard Hughes Medical Institute, Stanford, CA, 94305, USA

[3]School of Natural Sciences, University of California, Merced, Merced, CA, 95343, USA

[4]US Department of Energy Joint Genome Institute, Walnut Creek, CA, 94598, USA

[5]Genome Science Group, Los Alamos National Laboratory, Los Alamos, NM, 87545, USA

[6]School of Earth and Space Exploration, Arizona State University, Tempe, AZ, 85287, USA

[€]Current address: Broad Institute and Massachusetts Institute of Technology, Department of Biological Engineering, Cambridge, MA 02142, USA

[*]Current address: Northern Illinois University, Department of Biological Sciences, DeKalb, IL, 60115 USA

[†]Corresponding author: brian.hedlund@unlv.edu

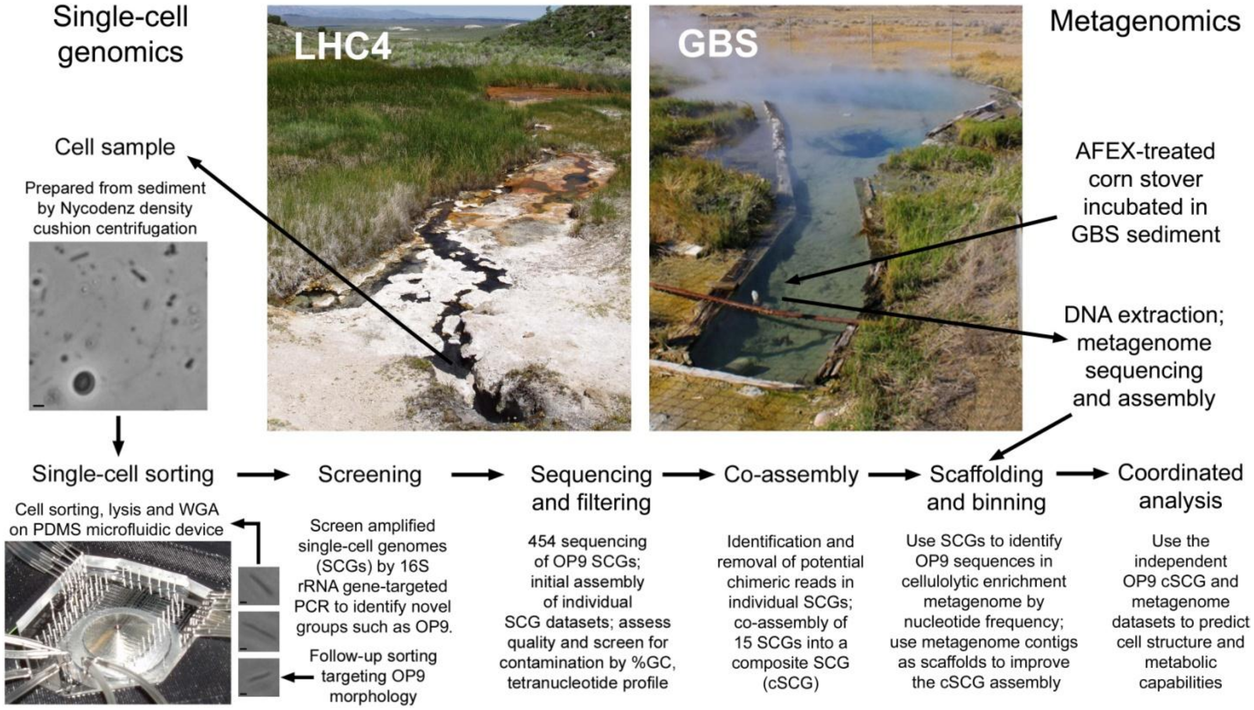
[‡]These authors contributed equally to this work

Supplementary Information includes:

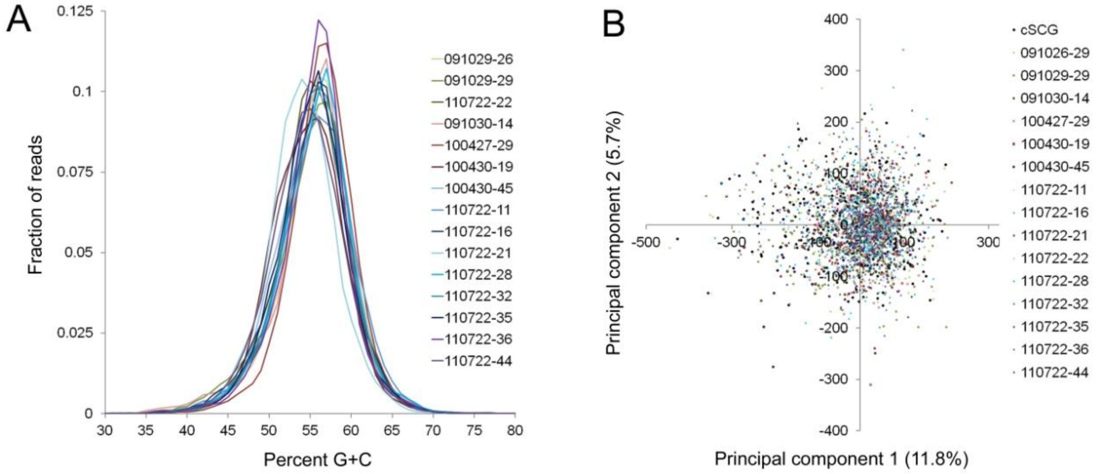
Supplementary Figures S1-S6

Supplementary Tables S1-S5

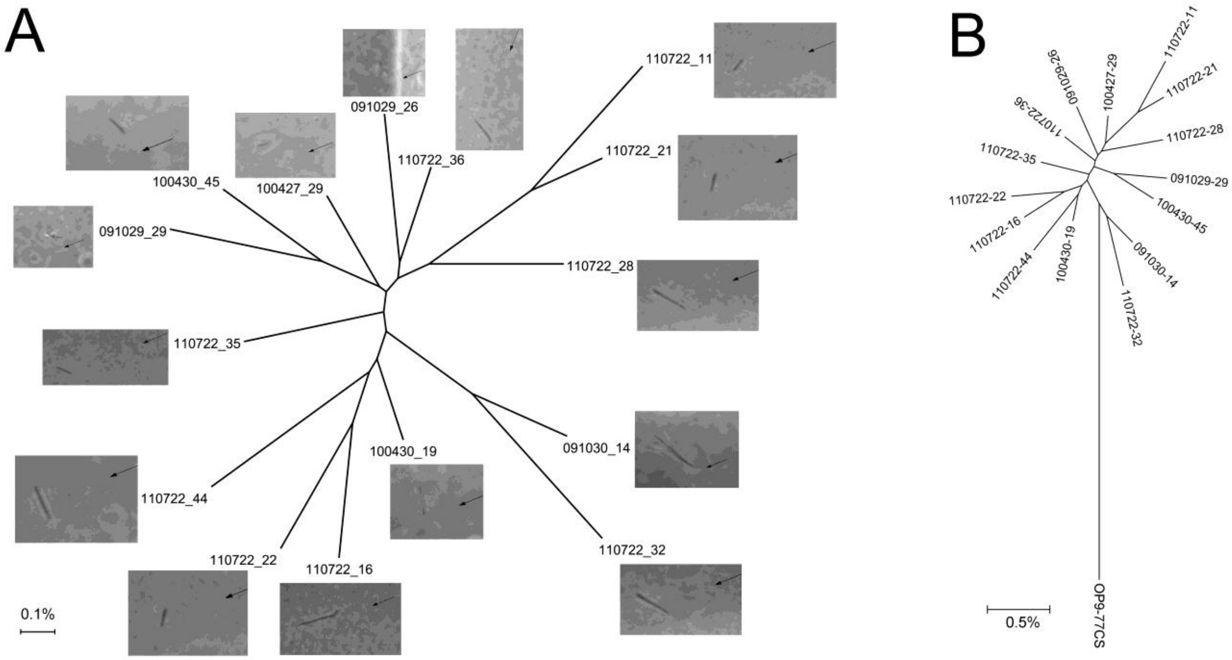
Supplementary References



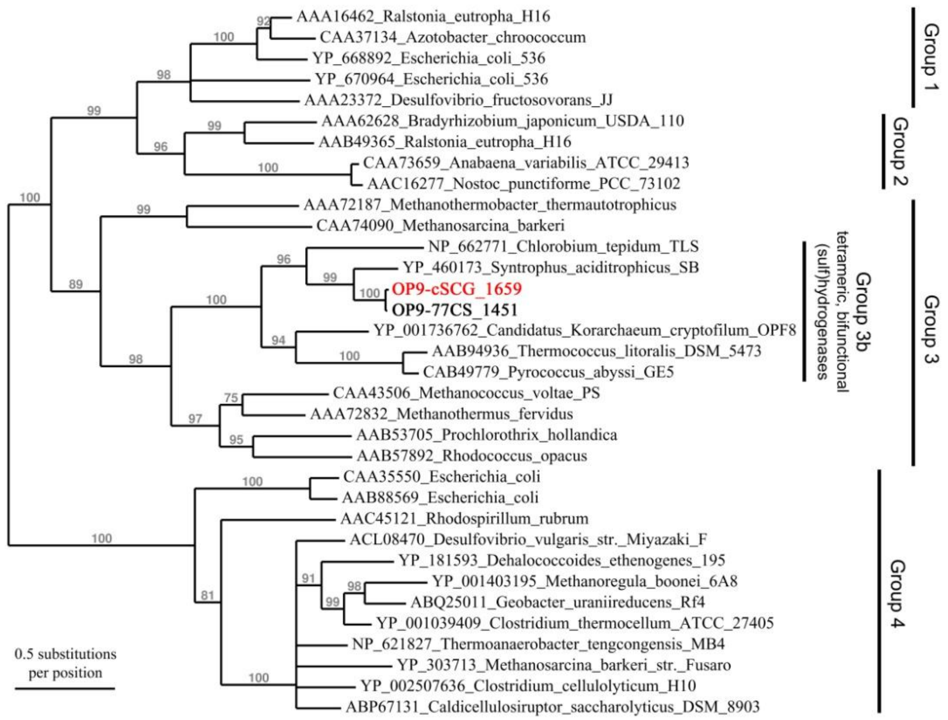
Supplementary Figure S1. Overview of the methods used to obtain distinct genomic datasets from two members of the OP9 lineage using coordinated single-cell genomics and metagenomics analyses. Black bars in phase-contrast photomicrograph insets of cell preparation and sorted OP9 cells represent 1 µm in scale.



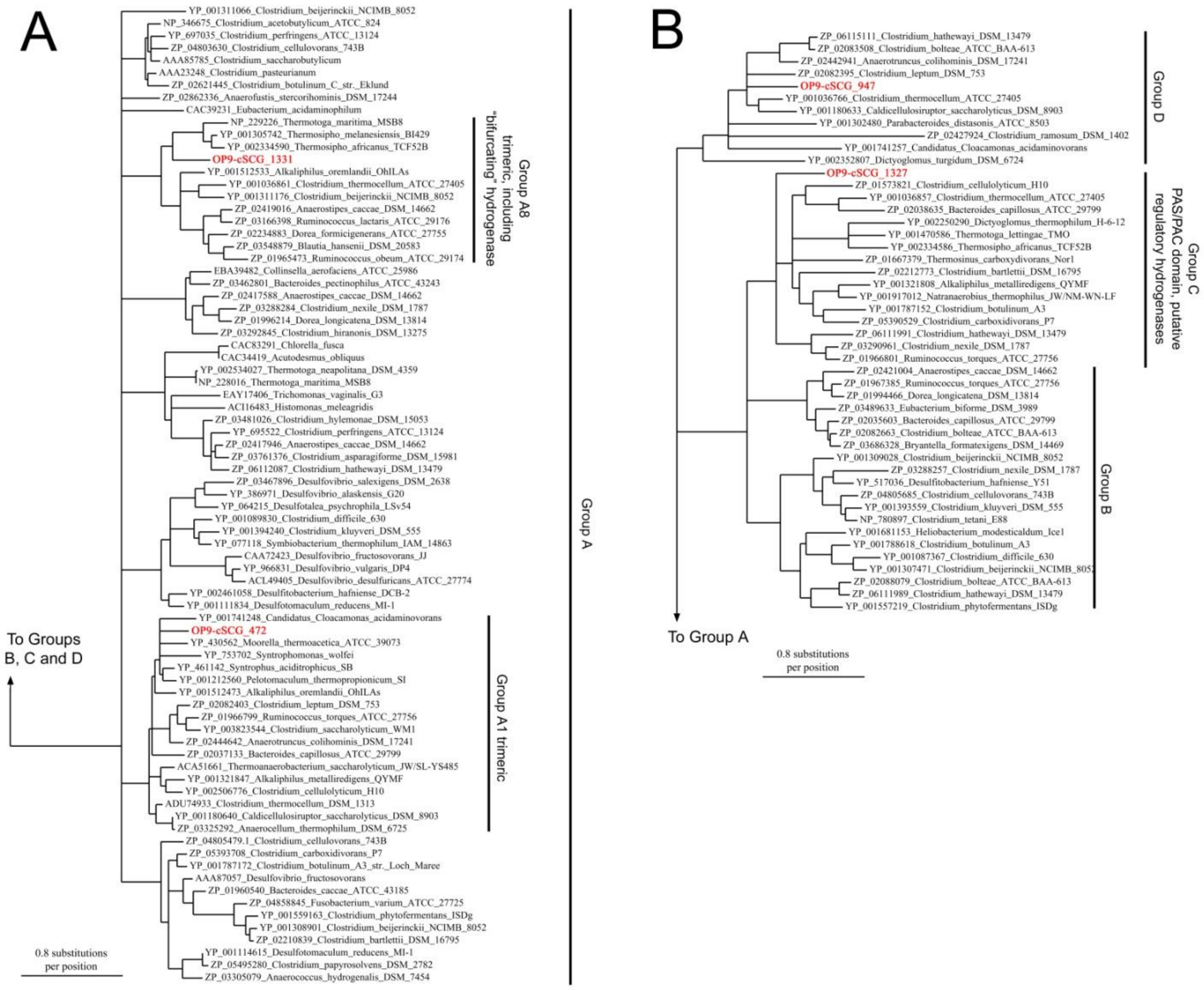
Supplementary Figure S2. Relative uniformity of individual OP9 SCG datasets based on (A) G+C% of individual reads or (B) TNF-PCA of assembled contigs.



Supplementary Figure S3. Neighbor-joining trees based on distance matricies of pairwise average nucleotide percentage identity. (A) Individual OP9 SCGs, identified by the sorting date in YYMMDD format followed by the chamber into which they were sorted. Scale bar indicates 0.1% sequence divergence. Photomicrographs of each sorted cell are shown, where the arrowhead represents ~1 µm. (B) OP9 SCGs rooted with the OP9-77CS assembly.

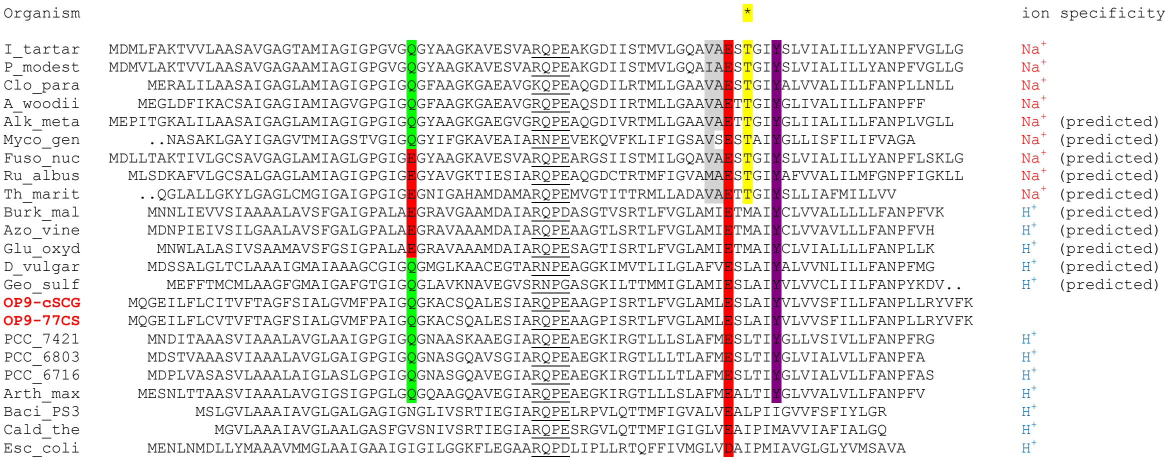


Supplementary Figure S4. Phylogeny of Group 1-4 NiFe hydrogenases, as defined in references 38 and 40 in main text, from OP9 and other organisms, based on 317 aligned positions. Sequences were submitted to phylogeny.fr[61], where multiple sequence alignment was performed using MUSCLE[62], phylogeny was inferred by maximum likelihood using PHYML[63], and the tree was visualized using TreeDyn[64]. Nodes with less than 70% support by approximate likelihood ratio test[65] are collapsed.



Supplementary Figure S5. Phylogeny of FeFe hydrogenases in (A) Group A and (B) Groups B-D, as defined in references 38 and 40 in main text, from OP9 and other organisms, based on 150 aligned positions. Sequences were submitted to phylogeny.fr[61], where multiple sequence alignment was performed using MUSCLE[62], phylogeny was inferred by maximum likelihood using PHYML[63], and the tree was visualized using TreeDyn[64]. Nodes with less than 70% support by approximate likelihood ratio test[65] are collapsed.

A



```
Organism                                                                             *                            ion specificity

I_tartar   MDMLFAKTVVLAASAVGAGTAMIAGIGPGVGQGYAAGKAVESVARQPEAKGDIISTMVLGQAVAESTGIYSLVIALILLYANPFVGLLG      Na+
P_modest   MDMVLAKTVVLAASAVGAGAAMIAGIGPGVGQGYAAGKAVESVARQPEAKGDIISTMVLGQAIAESTGIYSLVIALILLYANPFVGLLG      Na+
Clo_para       MERALILAASAIGAGLAMIAGIGPGIGQGFAAGKGAEAVGKQPEAQGDILRTMLLGAAVAESTGIYALVVALILLFANPLLNLL       Na+
A_woodii      MEGLDFIKACSAIGAGIAMIAGVGPGIGQGFAAGKGAEAVGRQPEAQSDIIRTMLLGAAVAETTGIYGLIVALILLFANPFF          Na+
Alk_meta   MEPITGKALILAASAIGAGLAMIAGIGPGIGQGYAAGKGAEGVGRQPEAQGDIVRTMLLGAAVAETTGIYGLIIALILLFANPLVGLL       Na+ (predicted)
Myco_gen       ..NASAKLGAYIGAGVTMIAGSTVGIGQGYIFGKAVEAIARNPEVEKQVFKLIFIGSAVSESTAIYGLLISFILIFVAGA           Na+ (predicted)
Fuso_nuc   MDLLTAKTIVLGCSAVGAGLAMIAGLGPGIGEGYAAGKAVESVARQPEARGSIISTMILGQAVAESTGIYSLVIALILLYANPFLSKLG      Na+ (predicted)
Ru_albus     MLSDKAFVLGCSALGAGLAMIAGIGPGIGEGYAVGKTIESIARQPEAQGDCTRTMFIGVAMAESTGIYAFVVALILMFGNPFIGKLL      Na+ (predicted)
Th_marit      ..QGLALLGKYLGAGLCMGIGAIGPGIGEGNIGAHAMDAMARQPEMVGTITTRMLLADAVAETTGIYSLLIAFMILLVV             Na+ (predicted)
Burk_mal       MNNLIEVVSIAAAALAVSFGAIGPALAEGRAVGAAMDAIARQPDASGTVSRTLFVGLAMIETMAIYCLVVALLLLLFANPFVK        H+  (predicted)
Azo_vine       MDNPIEIVSILGAALAVSFGALGPALAEGRAVAAAMDAIARQPEAAGTLSRTLFVGLAMIETMAIYCLVVAVLLLFANPFVH         H+  (predicted)
Glu_oxyd        MNWLALASIVSAAMAVSFGSIGPALAEGRAVAAAMDAIARQPESAGTISRTLFVGLAMIETMAIYCLVIALLLLFANPLLK         H+  (predicted)
D_vulgar       MDSSALGLTCLAAAIGMAIAAAGCGIGQGMGLKAACEGTARNPEAGGKIMVTLILGLAFVESLAIYALVVNLILLFANPFMG         H+  (predicted)
Geo_sulf         MEFFTMCMLAAGFGMAIGAFGTGIGQGLAVKNAVEGVSRNPGASGKILTTMMIGLAMIESLAIYVLVVCLIILFANPYKDV..      H+  (predicted)
OP9-cSCG     MQGEILFLCITVFTAGFSIALGVMFPAIGQGKACSQALESIARQPEAAGPISRTLFVGLAMLESLAIYVLVVSFILLFANPLLRYVFK
OP9-77CS     MQGEILFLCVTVFTAGFSIALGVMFPAIGQGKACSQALESIARQPEAAGPISRTLFVGLAMLESLAIYVLVVSFILLFANPLLRYVFK
PCC_7421       MNDITAAASVIAAALAVGLAAIGPGIGQGNAASKAAEGIARQPEAEGKIRGTLLLSLAFMESLTIYGLLVSIVLLFANPFRG         H+
PCC_6803       MDSTVAAASVIAAALAVGLGAIGPGIGQGNASGQAVSGIARQPEAEGKIRGTLLLTLAFMESLTIYGLVIALVLLFANPFA          H+
PCC_6716       MDPLVASASVLAAALAIGLASLGPGIGQGNASGQAVEGIARQPEAEGKIRGTLLLTLAFMESLTIYGLVIALVLLFANPFAS         H+
Arth_max      MESNLTTAASVIAAALAVGIGSIGPGLGQGQAAGQAVEGIARQPEAEGKIRGTLLLSLAFMEALTIYGLVVALVLLFANPFV          H+
Baci_PS3            MSLGVLAAAIAVGLGALGAGIGNGLIVSRTIEGIARQPELRPVLQTTMFIGVALVEALPIIGVVFSFIYLGR              H+
Cald_the              MGVLAAAIAVGLAALGASFGVSNIVSRTIEGIARQPESRGVLQTTMFIGIGLVEAIPIMAVVIAFIALGQ              H+
Esc_coli       MENLNMDLLYMAAAVMMGLAAIGAAIGIGILGGKFLEGAARQPDLIPLLRTQFFIVMGLVDAIPMIAVGLGLYVMSAVA            H+
```

B

Accession number and organism	Amino acid position in *C.limicola* PPase					Potassium dependence	Coupling ion
	180	242	246	478	481		
NP627745 *Streptomyces coelicolor*	A	E	V	K	T	No	H^+
AAC38615 *Rhodospirillum rubrum*	S	E	V	K	T	No	H^+
NP632725 *Methanosarcina mazei*	S	E	V	K	T	No	H^+
AAF01029 *Pyrobaculum aerophilum*	S	E	V	K	T	No	H^+
AAC83018 *Arabidopsis thaliana* (AVP2)	S	E	A	K	T	No	H^+
YP359158 *Carboxydothermus hydrogenoformans*	S	E	A	A	A	Yes	H^+
ZP02417489 *Anaerostipes caccae*	S	E	G	A	G	Yes	Na^+
NP781083 *Clostridium tetani*	S	E	G	A	G	Yes	Na^+
ZP01313190 *Desulfuromonas acetoxidans*	S	E	G	A	G	Yes	Na^+
YP430205 *Moorella thermoacetica*	S	E	G	A	G	Yes	Na^+
OP9-cSCG and OP9-77CS	S	E	G	A	G		
NP632724 *Methanosarcina mazei*	S	E	G	A	G	Yes	Na^+
NP227989 *Thermotoga maritima*	S	E	G	A	G	Yes	Na^+
YP001193220 *Chlorobium limicola*	S	E	G	A	G	Yes	Na^+
YP001193830 *Flavobacterium johnsoniae*	E	G	A	A	G	Yes	H^+
YP001840784 *Leptospira biflexa*	S	G	E	A	G	Yes	H^+
AAK38076 *Toxoplasma gondii*	S	G	E	A	G	Yes	H^+
AAD17215 *Plasmodium falciparum*	S	G	E	A	G	Yes	H^+
XP814868 *Trypanosoma cruzi*	S	G	E	A	G	Yes	H^+
CAC44451 *Chlamydomonas reinhardtii*	S	G	E	A	G	Yes	H^+
BAA23649 *Vigna radiata*	S	G	E	A	G	Yes	H^+
AAA32754 *Arabidopsis thaliana* (AVP1)	S	G	E	A	G	Yes	H^+

Supplementary Figure S6. Potential ion specificity of OP9 F-type ATP synthases and pyrophosphatases. (A) Amino acid alignment of the C subunit from ATP synthases from OP9 and other organisms with known or predicted ion specificity. Selected sequences and highlignted amino acid residues, including a key threonine residue (marked with an asterisk, in yellow) involved in specificity for sodium ion transport, are after Figure 4 in Meier et al., 2009[43]. (B) Predicted membrane pyrophosphatases (PPases) from OP9-cSCG and OP9-77CS are 64% identical to the experimentally characterized Na^+-translocating PPase of *Moorella thermoacetica*, and share conserved residues involved in Na^+-translocation (positions 180, 242, and 246) and K^+-dependence (positions 478 and 481). Data from characterized membrane pyrophosphatases are from Figure 1 in Luoto et al., 2011[44]. Sequence abbreviations and accession numbers are as follows: I_tarta, *Ilyobacter tartaricus* (pdb|1YCE); P_modest, *Propionigenum modestum* (sp|P21905); Clo_para, *Clostridium paradoxum*, (ABB13421) A_woodii, *Acetobacterium woodii* (AFA47026); Alk_meta, *Alkaliphilus metalliredigens* (YP_001318237); Myco_gen, *Mycoplasma genitalium* (NP_073077); Fuso_nuc, *Fusobacterium nucleatum* (NP_603267); Ru_albus, *Ruminococcus albus* (BAA23683); Th_marit, *Thermotoga maritima* (NP_229415); Burk_mal, *Burkholderia mallei* (YP_990892); Azo_vine, *Azotobacter vinelandii* (YP_002799156); Glu_oxyd, *Gluconobacter oxydans* (YP_192564); D_vulgar, *Desulfovibrio vulgaris* (YP_967510); Geo_sulf, *Geobacter sulfurreducens* (NP_951393); PCC_7421, *Gloeobacter violaceus* PCC 7421 (NP_925855); PCC_6803, *Synechocystis* sp. PCC 6803 (|NP_440059); PCC_6716, *Synechococcus* sp. PCC 6716 (Q05366); Arth_max, *Arthrospira maxima* CS-328 ZP_03274297); Baci_PS3, *Bacillus* sp. PS3 (CAA30649); Cald_the, *Caldalkalibacillus thermarum* (ZP_08534118); Esc_coli, *Escherichia coli* (CAA23591).

Supplementary Table S1. Assembly statistics for individual OP9 SCGs.

Single cell datasets[a]	Total # reads	# reads assembled	total # contigs	Average contig size (bases)	N50 for contigs >500 bases	Assembly size (bases)
091029-26	94771	85466	532	1639	4021	871951
091029-29	25986	24589	354	1668	3165	590585
091030-14	18198	15125	74	1492	5254	110385
100427-29	113011	104107	282	2365	5700	667036
100430-19	103534	97348	161	2006	5059	323035
100430-45	93173	87300	189	2175	4402	411105
110722-11	34091	28769	155	1440	4306	223258
110722-16	45486	37861	268	1058	2966	283612
110722-21	38184	30513	121	933	3070	112855
110722-22	28715	24575	128	1508	2906	193032
110722-28	43240	36948	441	1338	2941	590254
110722-32	52342	45406	318	1392	3742	442776
110722-35	35788	30127	253	1401	3299	354450
110722-36	32590	27423	208	1239	3908	257618
110722-44	103030	87549	344	1438	4578	494684

[a] Single cell datasets are identified by the date in YYMMDD format followed by the sort number.

Supplementary Table S2. Conserved markers[a] in the OP9-cSCG and OP9-77CS for estimation of genome completeness.

Functional class	COG	Function (annotation)	OP9-cSCG CDS[b,c]	OP9-77CS CDS [c]
Translation	COG0080	Ribosomal protein L11	1345	1818
Translation	COG0081	Ribosomal protein L1	1346	1819
Translation	COG0087	Ribosomal protein L3	0525	0870
Translation	COG0088	Ribosomal protein L4	0526	0871
Translation	COG0089	Ribosomal protein L23	0527	0872
Translation	COG0090	Ribosomal protein L2	0528	0873
Translation	COG0091	Ribosomal protein L22	0530	0875
Translation	COG0093	Ribosomal protein L14	0535	0880
Translation	COG0094	Ribosomal protein L5	0537	0882
Translation	COG0097	Ribosomal protein L6P/L9E	0540	0885
Translation	COG0102	Ribosomal protein L13	0140	0164
Translation	COG0197	Ribosomal protein L16/L10E	0532	0877
Translation	COG0198	Ribosomal protein L24	0536	0881
Translation	COG0200	Ribosomal protein L15	0544	0889
Translation	COG0203	Ribosomal protein L17	0554	0899
Translation	COG0211	Ribosomal protein L27	0485	0429
Translation	COG0222	Ribosomal protein L7/L12	1348	1821
Translation	COG0227	Ribosomal protein L28	0983	2018
Translation	COG0230	Ribosomal protein L34	Partial, uncalled ORF (positions 12569-12435 of contig155_S19, 95% nt identity to ORF in 77CS)	1277
Translation	COG0244	Ribosomal protein L10	1347	1820
Translation	COG0254	Ribosomal protein L31	1559	0032
Translation	COG0255	Ribosomal protein L29	0533	0878
Translation	COG0256	Ribosomal protein L18	0541	0886
Translation	COG0257	Ribosomal protein L36	0549	0894
Translation	COG0261	Ribosomal protein L21	0484	0430
Translation	COG0267	Ribosomal protein L33	0316	2363
Translation	COG0291	Ribosomal protein L35	0102	0120
Translation	COG0292	Ribosomal protein L20	0101	0119
Translation	COG0333	Ribosomal protein L32	Partial, uncalled ORF (positions 67-1 contig010_S1, 100% nt identity to ORF in 77CS)	0101
Translation	COG0335	Ribosomal protein L19	0071	0087
Translation	COG0359	Ribosomal protein L9	0500	0415
Translation	COG1825	Ribosomal protein L25 (general stress protein Ctc)	Not present	0592
Translation	COG1841	Ribosomal protein L30/L7E	0543	0888
Translation	COG0048	Ribosomal protein S12	1167	1087
Translation	COG0049	Ribosomal protein S7	1166	1086
Translation	COG0051	Ribosomal protein S10	0524	0869
Translation	COG0052	Ribosomal protein S2	0674	1196

Translation	COG0092	Ribosomal protein S3	0531	0876
Translation	COG0096	Ribosomal protein S8	0539	0884
Translation	COG0098	Ribosomal protein S5	0542	0887
Translation	COG0099	Ribosomal protein S13	0550	0895
Translation	COG0100	Ribosomal protein S11	0551	0896
Translation	COG0103	Ribosomal protein S9	0139	0163
Translation	COG0184	Ribosomal protein S15P/S13E	0700	0493
Translation	COG0185	Ribosomal protein S19	0529	0874
Translation	COG0186	Ribosomal protein S17	0534	0879
Translation	COG0199	Ribosomal protein S14	0538	0883
Translation	COG0228	Ribosomal protein S16	0075	0091
Translation	COG0238	Ribosomal protein S18	0498	0417
Translation	COG0268	Ribosomal protein S20	0162	1376
Translation	COG0360	Ribosomal protein S6	0495	0419
Translation	COG0522	Ribosomal protein S4 and related proteins	0552	0897
Translation	COG0539	Ribosomal protein S1	1688	0400
Translation	COG0828	Ribosomal protein S21	0812	0519
Translation	COG0008	Glutaminyl-tRNA synthetase	233, 771	566, 810
Translation	COG0008	Glutamyl-tRNA synthetase	233, 771	566, 810
Translation	COG0013	Alanyl-tRNA synthetase	1690	0402
Translation	COG0016	Phenylalanyl-tRNA synthetase alpha subunit	0100	0118
Translation	COG0017	Aspartyl/asparaginyl-tRNA synthetases	0029	0957
Translation	COG0018	Arginyl-tRNA synthetase	1176	0980
Translation	COG0060	Isoleucyl-tRNA synthetase	0322	0285
Translation	COG0072	Phenylalanyl-tRNA synthetase beta subunit	0099	0117
Translation	COG0124	Histidyl-tRNA synthetase	1686	0398
Translation	COG0162	Tyrosyl-tRNA synthetase	0093	0112
Translation	COG0172	Seryl-tRNA synthetase	1152	0936
Translation	COG0173	Aspartyl-tRNA synthetase	1687	0399
Translation	COG0180	Tryptophanyl-tRNA synthetase	0660	0856
Translation	COG0215	Cysteinyl-tRNA synthetase, class Ia	Partial, uncalled ORF (positions 1424-1 contig030_S02, 92% nt identity to ORF in 77CS)	0809
Translation	COG0215	Cysteinyl-tRNA synthetase	Partial, uncalled ORF (positions 1424-1 contig030_S02, 92% nt identity to ORF in 77CS)	0809
Translation	COG0441	Threonyl-tRNA synthetase, class IIa	1217	1067
Translation	COG0441	Threonyl-tRNA synthetase	1217	1067
Translation	COG0442	Prolyl-tRNA synthetase	0683	1187
Translation	COG0495	Leucyl-tRNA synthetase	0880	0580
Translation	COG0525	Valyl-tRNA synthetase	0942	1506
Translation	COG0751	Glycyl-tRNA synthetase, beta subunit[d]	1725	1016
Translation	COG0752	Glycyl-tRNA synthetase, alpha subunit[d]	1724	1015
Translation	COG1190	Lysyl-tRNA synthetase (class II)	0245	0820
Translation	COG0290	Translation initiation factor 3 (IF-3)	0103	0121

Translation	COG0361	Translation initiation factor 1 (IF-1)	0548	0893
Translation	COG0532	Translation initiation factor 2 (IF-2; GTPase)	1537	0011
Histidine biosynthesis	COG0040	ATP phosphoribosyltransferase	0300	0260
Histidine biosynthesis	COG0079	Histidinol-phosphate/aromatic aminotransferase and cobyric acid decarboxylase	0263	0714
Histidine biosynthesis	COG0106	Phosphoribosylformimino-5-aminoimidazole carboxamide ribonucleotide (ProFAR) isomerase	1138	0921
Histidine biosynthesis	COG0107	Imidazoleglycerol-phosphate synthase	1137	0921
Histidine biosynthesis	COG0118	Glutamine amidotransferase	1139	0923
Histidine biosynthesis	COG0131	Imidazoleglycerol-phosphate dehydratase	1140	0924
Histidine biosynthesis	COG0139	Phosphoribosyl-AMP cyclohydrolase	1136	0920
Histidine biosynthesis	COG0140	Phosphoribosyl-ATP pyrophosphohydrolase	1136	0920
Histidine biosynthesis	COG0141	Histidinol dehydrogenase	0301	0261
Histidine biosynthesis	COG0241	Histidinol phosphatase and related phosphatases	0013	1104
Histidine biosynthesis	COG0462	Phosphoribosylpyrophosphate synthetase	Not present	0593
Chorismate biosynthesis	COG0082	Chorismate synthase	1577	0047
Chorismate biosynthesis	COG0128	5-enolpyruvylshikimate-3-phosphate synthase	0666	0850
Chorismate biosynthesis	COG0169	Shikimate 5-dehydrogenase	1695	0406
Chorismate biosynthesis	COG0337	3-dehydroquinate synthetase	0639	1036
Chorismate biosynthesis	COG0703	Shikimate kinase	0638	1037
Chorismate biosynthesis	COG0710	3-dehydroquinate dehydratase	0640	1035
Chorismate biosynthesis	COG0722	3-deoxy-D-arabino-heptulosonate 7-phosphate (DAHP) synthase	1580	50-51
Chorismate biosynthesis	COG1605	Chorismate mutase	0665	0851
Threonine biosynthesis	COG0083	Homoserine kinase	0257	0709
Threonine biosynthesis	COG0136	Aspartate-semialdehyde dehydrogenase	0255	0707
Threonine biosynthesis	COG0460	Homoserine dehydrogenase	0017	1101
Threonine biosynthesis	COG0498	Threonine synthase	0258	0710
Threonine biosynthesis	COG0527	Aspartokinases	0256	0708
Tryptophan biosynthesis	COG0133	Tryptophan synthase beta chain	0306	0266
Tryptophan biosynthesis	COG0134	Indole-3-glycerol phosphate synthase	0308	0268
Tryptophan biosynthesis	COG0135	Phosphoribosylanthranilate isomerase	0307	0267
Tryptophan biosynthesis	COG0147	Anthranilate/para-aminobenzoate synthases component I	0312	0272
Tryptophan biosynthesis	COG0159	Tryptophan synthase alpha chain	0305	0265
Tryptophan biosynthesis	COG0512	Anthranilate/para-aminobenzoate synthases component II	0310	0270
Tryptophan biosynthesis	COG0547	Anthranilate phosphoribosyltransferase	0309	0269
Coenzyme A biosynthesis	COG0237	Dephospho-CoA kinase	0114	0138
Coenzyme A biosynthesis	COG0413	Ketopantoate hydroxymethyltransferase	Not present	Not present
Coenzyme A biosynthesis	COG0414	Panthothenate synthetase	Not present	Not present
Coenzyme A biosynthesis	COG0452	Phosphopantothenoylcysteine synthetase/decarboxylase	0975	0635
Coenzyme A biosynthesis	COG0669	Phosphopantetheine adenylyltransferase	0086	0105
Coenzyme A biosynthesis	COG0853	Aspartate 1-decarboxylase	Not present	Not present
Coenzyme A biosynthesis	COG1072	Panthothenate kinase	0052	1790
Coenzyme A biosynthesis	COG1893	Ketopantoate reductase	Not present	Not present
FAD biosynthesis	COG0054	Riboflavin synthase beta-chain	Not present	Not present

FAD biosynthesis	COG0108	3,4-dihydroxy-2-butanone 4-phosphate synthase	Not present	Not present
FAD biosynthesis	COG0117	Pyrimidine deaminase	Not present	Not present
FAD biosynthesis	COG0196	FAD synthase	1531	0006
FAD biosynthesis	COG0307	Riboflavin synthase alpha chain	Not present	Not present
FAD biosynthesis	COG0807	GTP cyclohydrolase II	Not present	Not present
FAD biosynthesis	COG1985	Pyrimidine reductase, riboflavin biosynthesis	Not present	Not present
Isoprenoid biosynthesis	COG0245	2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	0236	0812
Isoprenoid biosynthesis	COG0743	1-deoxy-D-xylulose 5-phosphate reductoisomerase	0680	1190
Isoprenoid biosynthesis	COG1154	Deoxyxylulose-5-phosphate synthase	0347	1486
Isoprenoid biosynthesis	COG1211	4-diphosphocytidyl-2-methyl-D-erithritol synthase	0237	0813
Isoprenoid biosynthesis	COG1947	4-diphosphocytidyl-2C-methyl-D-erythritol 2-phosphate synthase	1173	1093
Isoprenoid biosynthesis	COG0761	Penicillin tolerance protein	0668	0848
Isoprenoid biosynthesis	COG0821	Enzyme involved in the deoxyxylulose pathway of isoprenoid biosynthesis	0682	1188
Isoprenoid biosynthesis	COG0020	Undecaprenyl pyrophosphate synthase	0678	1192
Purine biosynthesis	COG0015	Adenylosuccinate lyase	1244	2014
Purine biosynthesis	COG0026	Phosphoribosylaminoimidazole carboxylase (NCAIR synthetase)	1282	2305
Purine biosynthesis	COG0027	Formate-dependent phosphoribosylglycinamide formyltransferase (GAR transformylase)	1868	1212
Purine biosynthesis	COG0034	Glutamine phosphoribosylpyrophosphate amidotransferase	0290	0248
Purine biosynthesis	COG0041	Phosphoribosylcarboxyaminoimidazole (NCAIR) mutase	1283	2306
Purine biosynthesis	COG0046	Phosphoribosylformylglycinamidine (FGAM) synthase, synthetase domain	0292	0249
Purine biosynthesis	COG0047	Phosphoribosylformylglycinamidine (FGAM) synthase, glutamine amidotransferase domain	Partial, uncalled ORF (positions 31031-30585 contig031_S02, 91% nt identity to ORF in 77CS)	0250
Purine biosynthesis	COG0104	Adenylosuccinate synthase	0123	0147
Purine biosynthesis	COG0138	AICAR transformylase/IMP cyclohydrolase PurH (only IMP cyclohydrolase domain in Aful)	1683	0395
Purine biosynthesis	COG0150	Phosphoribosylaminoimidazole (AIR) synthetase	1869	1213
Purine biosynthesis	COG0151	Phosphoribosylamine-glycine ligase	1867	1210
Purine biosynthesis	COG0152	Phosphoribosylaminoimidazolesuccinocarboxamide (SAICAR) synthase	0297	0258
Purine biosynthesis	COG0299	Folate-dependent phosphoribosylglycinamide formyltransferase PurN	1868	1212
Purine biosynthesis	COG0516	IMP dehydrogenase/GMP reductase	0125	0149
Purine biosynthesis	COG0518	GMP synthase - Glutamine amidotransferase domain	0124	0148
Purine biosynthesis	COG0519	GMP synthase, PP-ATPase domain/subunit	0124	0148
Purine biosynthesis	COG0563	Adenylate kinase and related kinases	0546	0891
Pyrimidine biosynthesis	COG0105	Nucleoside diphosphate kinase	Not present	2215
Pyrimidine biosynthesis	COG0167	Dihydroorotate dehydrogenase	0359	1549
Pyrimidine biosynthesis	COG0283	Cytidylate kinase	0667	0849
Pyrimidine biosynthesis	COG0284	Orotidine-5-phosphate decarboxylase	Partial, uncalled ORF (positions 782-4 contig041_S03, 88% nt identity to ORF in 77CS)	1547
Pyrimidine biosynthesis	COG0418	Dihydroorotase	0361	1546
Pyrimidine biosynthesis	COG0458	Carbamoylphosphate synthase large subunit (split gene in MJ)	0709	0501
Pyrimidine biosynthesis	COG0461	Orotate phosphoribosyltransferase	0358	1550
Pyrimidine biosynthesis	COG0504	CTP synthase (UTP-ammonia lyase)	1175	0979
Pyrimidine biosynthesis	COG0505	Carbamoylphosphate synthase small subunit	0711	0502

Pyrimidine biosynthesis	COG0528	Uridylate kinase	0676	1194
Pyrimidine biosynthesis	COG0540	Aspartate carbamoyltransferase, catalytic chain	0363	1545
Pyrimidine biosynthesis	COG0125	Thymidylate kinase	1862	1205
Pyrimidine biosynthesis	COG0207	Thymidylate synthase	1558	0031
Pyrimidine biosynthesis	COG0717	Deoxycytidine deaminase	Not present	Not present
Pyrimidine biosynthesis	COG0756	dUTPase	0986	1179
Sec translocase	COG0201	Preprotein translocase subunit SecY	0545	0890
Sec translocase	COG0341	Preprotein translocase subunit SecF	0375	0724
Sec translocase	COG0342	Preprotein translocase subunit SecD	0374	0723
Sec translocase	COG0653	Preprotein translocase subunit SecA (ATPase, RNA helicase)	0755	0550
Sec translocase	COG0690	Preprotein translocase subunit SecE	1343	2356
Sec translocase	COG0706	Preprotein translocase subunit YidC	1367	1921
Sec translocase	COG1314	Preprotein translocase subunit SecG	Not present	Not present
Sec translocase	COG1862	Preprotein translocase subunit YajC	0373	0722
Sec translocase	COG1952	Preprotein translocase subunit SecB	Not present	Not present
Transcription	COG0085	DNA-directed RNA polymerase, beta subunit/140 kD subunit	1170	1090
Transcription	COG0086	DNA-directed RNA polymerase, beta' subunit/160 kD subunit	1169	1089
Transcription	COG0202	DNA-directed RNA polymerase, alpha subunit/40 kD subunit	0553	0898
Transcription	COG0568	DNA-directed RNA polymerase, sigma subunit RpoD	1731	1022
Transcription	COG0568	DNA-directed RNA polymerase, sigma subunit RpoS	1731	1022
Transcription	COG1758	DNA-directed RNA polymerase, subunit K/omega	0974	0636

All markers	OP9-cSCG	OP9-77CS
Markers present	164	167
Total number of markers	181	181
% of total markers present	91.2	92.8

All markers except those involved in FAD and Co-A synthesis	OP9-cSCG	OP9-77CS
Markers present	160	163
Total number of markers	166	166
% of total markers present	96.4	98.2

[a] Markers used for estimation of genome coverage by Martin et al. (2006)[29], were identified in the OP9 genomes by COGs and RAST annotation.

[b] Several markers identified as uncalled ORFs (partial genes) in the OP9-cSCG by tBLASTn of the corresponding marker in the OP9-77CS were counted as present.

[c] Numbers correspond to the gene number in the assemblies on RAST and submitted to Genbank.

[d] Class II glycyl-tRNA synthetase (COG0423) was not included in the analysis because of the presence of the Class I glycyl-tRNA synthetases in OP9-cSCG and OP9-77CS.

Supplementary Table S3. Organisms and 16S rRNA gene sequence accession numbers used for phylogenetic inference.

Phylum[a]	Genbank Accession number	Organism/clone name
Firmicutes	AADT03000052	*Moorella thermoacetica*
Firmicutes	AF022248	*Syntrophomonas wolfei*
Firmicutes	D26185	*Bacillus subtilis*
Firmicutes	CP001785	*Ammoniafex degensii*
Firmicutes	FR749972	*Caldanaerobacter subterraneus*
Firmicutes	NC_009012	*Clostridium thermocellum*
Actinobacteria	X60514	*Streptomyces coelicolor*
Actinobacteria	U75647	*acidimicrobium ferrooxidans*
Actinobacteria	AJ243871	*Rubrobacter xylanophilus*
Actinobacteria	AE014295	*Bifidobacterium longum*
Acidobacteria	X77215	*Holophaga foetida*
Acidobacteria	D26171	*Acidobacterium capsulatum*
Armatimonadetes	AB529679	*Armatimonas rosea*
Armatimonadetes	AM749780	*Chthonomonas calidirosea*
Armatimonadetes	GQ339893	*Fimbriimonas ginsengisoli*
Verrucomicrobia	X90515	*Verrucomicrobium spinosum*
Verrucomicrobia	NR_028890	*Opitutus terrae*
Aquificae	M83548	*Aquifex pyrophilus*
Aquificae	D16296	*Hydrogenobaculum acidophilus*
Aquificae	AJ278895	*Thermocrinis albus*
Thermotogae	M59176	*Fervidobacterium icelandicum*
Thermotogae	L10659	*Geotoga subterranea*
Thermotogae	CP000702	*Thermotoga petrophila*
Chloroflexi	CP001275	*Thermomicrobium roseum*
Chloroflexi	CP000909	*Chloroflexus aurantiacus*
Chloroflexi	RN_040878	*Caldilinea aerophila*
Thermi	L09663	*Thermus aquaticus*
Thermi	NR_027614	*Meiothermus ruber*
Thermi	M21413	*Deinococcus radiodurans*

Bacteroidetes	EU823302	*Leeuwenhoekiella marinoflava*
Bacteroidetes	NR_042496	*Flavobacterium johnsoniae*
Bacteroidetes	L11703	*Thermonema lapsum*
Proteobacteria	D14516	*Sinorhizobium fredii*
Proteobacteria	M11223	*Agrobacterium tumefaciens*
Proteobacteria	AY268175	*Pseudomonas aeruginosa*
Proteobacteria	AF353160	*Nitrosomonas europaea*
Cyanobacteria	D83715	*Synechococcus elongatus*
Cyanobacteria	BA000045	*Gloeobacter violaceus*
Cyanobacteria	NR_028762	*Prochlorococcus marinus*
Cyanobacteria	AB428654	*Nostoc commune*
Synergistetes	AF073521	*Aminobacterium mobile*
Synergistetes	NR_036784	*Anaerobaculum thermoterrnum*
Synergistetes	CP001818	*Thermanaerovibrio acidaminovorans*
Synergistetes	DQ071273	*Thermovirga lienlii*
Dictyoglomi	NR_029235	*Dictyoglomus thermophilum*
Dictyoglomi	NR_043385	*Dictyoglomus turgidum*
Cand. phylum JS1	AF029050	Clone SB-45
Cand. phylum JS1	AB015271	Clone JTB243
Cand. phylum JS1	AY053496	Clone AT425Eub_A5
Cand. phylum JS1	AY093469	Nankai Forearc Basin clone MB-B2-103
Cand. phylum JS1	AF154106	Hydrocarbon seep clone GCA025
Cand. phylum JS1	AJ535219	Hydrate Ridge clone Hyd01-10
Cand. phylum JS1	AF029043	Benzene-mineralizing consortium clone SB-15
Cand. phylum OP9	AB710366.1	Uncultured bacterium gene for 16S ribosomal RNA, partial sequence, clone: N2>CO2B5
Cand. phylum OP9	EF205555.1	Uncultured candidate division OP9 bacterium clone TP29 16S ribosomal RNA gene, partial sequence
Cand. phylum OP9	FN356241.1	Uncultured bacterium 16S rRNA gene, clone Dan_Bac32
Cand. phylum OP9	FR744612.1	Uncultured bacterium partial 16S rRNA gene, clone PWB042
Cand. phylum OP9	AF027082.1	Candidate division OP9 clone OPB47 16S ribosomal RNA gene, complete sequence
Cand. phylum OP9	AF027081.1	Candidate division OP9 clone OPB46 16S ribosomal RNA gene, complete sequence
Cand. phylum OP9	EU635950.1	Uncultured bacterium clone SSW_L1_H02 16S ribosomal RNA gene, partial sequence
Cand. phylum OP9	EU924241.1	Uncultured bacterium clone LHC4_L1_A09 16S

		ribosomal RNA gene, partial sequence
Cand. phylum OP9	HM041949.1	Uncultured candidate division OP9 bacterium clone NRB32 16S ribosomal RNA gene, partial sequence
Cand. phylum OP9	FN356272.1	Uncultured bacterium 16S rRNA gene, clone Dan_Bac65
Cand. phylum OP9	EU334519.1	Uncultured bacterium clone CB05 16S ribosomal RNA gene, partial sequence
Cand. phylum OP9	AJ420322.1	uncultured bacterium partial 16S rRNA gene, clone SHA-110
Cand. phylum OP9	EF559042.1	Uncultured bacterium clone A55_D21_L_B_C04 16S ribosomal RNA gene, partial sequence
Cand. phylum OP9	FN436137.1	Uncultured bacterium partial 16S rRNA gene, clone HAW-R60-B-924d-G
Cand. phylum OP9	JF417929.1	Uncultured bacterium clone 1-1B-38 16S ribosomal RNA gene, partial sequence
Cand. phylum OP9	FN436061.1	Uncultured bacterium partial 16S rRNA gene, clone HAW-R60-B-727d-S
Cand. phylum OP9	EF558973.1	Uncultured bacterium clone B55_F_B_E11 16S ribosomal RNA gene, partial sequence
Cand. phylum OP9	FJ535531.1	Uncultured bacterium clone SWADLP3-17 16S ribosomal RNA gene, partial sequence
Cand. phylum OP9	FN436208.1	Uncultured bacterium partial 16S rRNA gene, clone HAW-R60-B-1249d-AH
Cand. phylum OP9	AB546037.1	Uncultured bacterium gene for 16S rRNA, partial sequence, clone: ARWH-BH02
Cand. phylum OP9	EF688189.1	Uncultured bacterium clone 4B10_cons 16S ribosomal RNA gene, partial sequence
Cand. phylum OP9	EU592420.1	Uncultured bacterium clone SSS11N 16S ribosomal RNA gene, partial sequence
Cand. phylum OP9	AB332115.1	Uncultured bacterium gene for 16S rRNA, partial sequence, clone: EPE303
Cand. phylum OP9	EU639006.1	Uncultured bacterium clone SHBZ497 16S ribosomal RNA gene, partial sequence
Cand. phylum OP9	EU638743.1	Uncultured bacterium clone SHBZ1212 16S ribosomal RNA gene, partial sequence
Cand. phylum OP9	CU924348.1	Uncultured OP9 bacterium 16S rRNA gene from clone QEDP3DA09

[a] 16S rRNA gene sequences representing members of the candidate phylum OP9, both those included in this analysis and those excluded due to their shorter length or similarity to included sequences. In addition to references 1 and 19-23 in the main text, OP9 sequences have been recovered from environments including terrestrial geothermal systems[66-69], eutrophic saline lakes[70], petroleum reservoirs[71-74], thermal bioreactors and digesters[75-80], and wastewater treatment facilities[81-83].

Supplementary Table S4. Markers diagnostic for a diderm cell envelope structure[a]

Protein name	Function/annotation	Pfam number	cSCG CDS[b]	cSCG e-value	77CS CDS[b]	77CS e-value
BamA (YaeT)	OM[c] assembly	PF01103	1613	8.2E-49	0186	4.7E-50
BamA (YaeT)	OM[c] assembly	PF07244	1613	3.6E-07	0186	7.4E-12
OmpH	Periplasm chaperone/OM[c] porin	PF03938	1612	2.2E-05	0185	1.4E-12
TolC	Type 1 secretion	PF02321	1526	9.4E-14	0001	5.6E-14
Secretin	Type 2, 3 secretion	PF00263	0636	8.6E-39	1039	5.4E-22
Secretin	Type 2, 3 secretion	PF00263	1380	1.0E-40	1286	6.9E-42
TonB	Type 2, 3 secretion	PF07660	0636	1.1E-07	1040	1.0E-07
FlgH	Flagellar L-ring	PF02107	0734	2.7E-31	0605	2.8E-29
FlgI	Flagellar P-ring	PF02119	Present[d]	n.d.[e]	0605	1.9E-139

[a] After Sutcliffe (2011)[31].

[b] Numbers correspond to the gene number in the assemblies on RAST and submitted to Genbank.

[c] OM, outer membrane.

[d] Uncalled CDS; has first 104 bp of 1113 bp gene, positions 12267-12370 of cSCG contig077_S08.

[e] n.d., not determined

Supplementary Table S5. List of CDSs[a] in OP9-cSCG and OP9-77CS discussed in the text and Figure 4

Pathway and functional annotation in RAST	77CS CDS	cSCG CDS	Predicted amino acid % identity[d]
Oligosaccharide utilization			
Beta-glucosidase (EC 3.2.1.21)	1643	1812	95.45
Beta-glucosidase (EC 3.2.1.21)	2453	2276	93.88
Alpha-galactosidase (EC 3.2.1.22)	1881	1082	94.35
Alpha-xylosidase (EC 3.2.1.-)	1007	1197	94.41
Endo-1,4-beta-glucanase	1856	0402	90.34
Alpha-N-arabinofuranosidase (EC 3.2.1.55)	1750	1423	95.87
Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8)	1945	1420	95.8
Beta-mannosidase (EC 3.2.1.25)	0690	0859	96.79
4-alpha-glucanotransferase (amylomaltase) (EC 2.4.1.25)	1053	0520	87.76
Amylopullulanase	0297	0334	93.89
Alpha-L-rhamnosidase (EC 3.2.1.40)	Absent	1984	0
Beta-galactosidase (3.2.1.23)	1215	Absent	0
Alpha-L-fucosidase (3.2.1.51)	1219	Absent	0
Glycolysis			
Glucokinase, ROK family protein (EC 2.7.1.2)	0659	1019	85.63
Glucose-6-phosphate isomerase (EC 5.3.1.9)	1096	0021	91.07
Glucose-6-phosphate isomerase (EC 5.3.1.9), archaeal	0985	1179	97.88
6-phosphofructokinase (EC 2.7.1.11)	0342	0165	98.61
Fructose-bisphosphate aldolase class II (EC 4.1.2.13)	0513	0805	99.38
NAD-dependent glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.12)	0131	Not called[b]	92% nt id to 743-1 cSCG contig017[c]
Phosphoglycerate kinase (EC 2.7.2.3)	0167	0143	97.73
2,3-bisphosphoglycerate-independent phosphoglycerate mutase (EC 5.4.2.1)	1482	0349	97.02
Enolase (EC 4.2.1.11)	0582	0882	96.96
Pyruvate kinase (EC 2.7.1.40)	1594	0456	95.34
Pentose phosphate pathway and xylose utilization (key genes)			
Transaldolase (EC 2.2.1.2)	0514	0806	98.16
Transketolase (EC 2.2.1.1)	0341	0166	97.07
6-phosphogluconate dehydrogenase, decarboxylating (EC 1.1.1.44)	1598	Not called[b]	93% nt id to cSCG contig050 712-1, contig051 1-297[c]
Glucose-6-phosphate 1-dehydrogenase (EC 1.1.1.49)	0535	0825	95.77
Glucose-6-phosphate isomerase (EC 5.3.1.9)	1096	0021	91.07
Glucose-6-phosphate isomerase, archaeal (EC 5.3.1.9)	0985	1179	97.88

Ribulose-phosphate 3-epimerase (EC 5.1.3.1)	0328	0179	97.75
Ribulose-phosphate 3-epimerase (EC 5.1.3.1)	2019	0982	97.27
Ribose 5-phosphate isomerase B (EC 5.3.1.6)	0027	1554	96.64
Ribose 5-phosphate isomerase B (EC 5.3.1.6)	0341	0304	97.33
Ribose-phosphate pyrophosphokinase (EC 2.7.6.1)	0593	Absent	0
Putative xylose isomerase xylA	0968	1712	96.9
Xylulose kinase (EC 2.7.1.17)	00969	1713	92.41
Pyruvate and acetate catabolism			
Pyruvate:ferredoxin oxidoreductase, gamma subunit (EC 1.2.7.1)	0740	0394	97.37
Pyruvate:ferredoxin oxidoreductase, delta subunit (EC 1.2.7.1)	0741	0395	95.79
Pyruvate:ferredoxin oxidoreductase, alpha subunit (EC 1.2.7.1)	0742	0396	96.28
Pyruvate:ferredoxin oxidoreductase, beta subunit (EC 1.2.7.1)	0743	0398	98.7
Acetate kinase (EC 2.7.2.1)	0103	0085	99.12
Acetate kinase (EC 2.7.2.1)	1121	1740	96.97
Phosphate acetyltransferase (EC 2.3.1.8)	0104	Not called[b]	88% nt id to 752-1 of cSCG contig011[c]
Aldehyde dehydrogenase (EC 1.2.1.3)	1397	1916	98.55
Alcohol dehydrogenase (EC 1.1.1.1)	0206	1634	95.81
Alcohol dehydrogenase (EC 1.1.1.1)	1647	1824	93.96
Bidirectional NiFe (sulf)hydrogenase			
heterodisulfide reductase, iron-sulfur binding subunit, putative	1449	1662	96.84
Heterodisulfide reductase, cytochrome reductase subunit	1448	1661	97.84
NAD-reducing hydrogenase subunit HoxY (EC 1.12.1.2)	1447	1660	98.06
Uptake hydrogenase large subunit (EC 1.12.99.6)	1446	1659	97.09
Bifurcating FeFe hydrogenase			
NAD-reducing hydrogenase subunit HoxE (EC 1.12.1.2)	1257	1329	98.21
NAD-reducing hydrogenase subunit HoxF (EC 1.12.1.2)	1258	1330	97.17
[Fe] hydrogenase large subunit (EC 1.12.7.2)	1259	1331	96.75
Additional Group A, B and D FeFe hydrogenases			
[Fe] hydrogenase large subunit (EC 1.12.7.2)	0442	0472	97.84
[Fe] hydrogenase large subunit, Group C	1255	1327	85.4
[Fe] hydrogenase large subunit, Group D	1502	0946	96.67
Energy conservation			
Pyrophosphate-energized ion pump	1020	1729	98.93
F-type ATP synthase A chain (EC 3.6.3.14)	0016	1542	96.38
F-type ATP synthase C chain (EC 3.6.3.14)	0017	1543	98.86
F-type ATP synthase B chain (EC 3.6.3.14)	0018	1544	92.98

F-type ATP synthase delta chain (EC 3.6.3.14)	0019	1545	95.24
F-type ATP synthase alpha chain (EC 3.6.3.14)	0020	1546	94.86
F-type ATP synthase gamma chain (EC 3.6.3.14)	0021	1547	92.96
F-type ATP synthase beta chain (EC 3.6.3.14)	0022	1548	97.65
F-type ATP synthase epsilon chain (EC 3.6.3.14)	0023	1549	91.24
V-type ATP synthase subunit D (EC 3.6.3.14)	0235	0279	99.02
V-type ATP synthase subunit B (EC 3.6.3.14)	0236	0280	99.14
V-type ATP synthase subunit A (EC 3.6.3.14)	0237	0281	98.15
V-type ATP synthase subunit F (EC 3.6.3.14)	0238	0282	95.54
V-type H+-transporting two-sector ATPase, C (AC39) subunit	0239	0283	94.28
V-type H+-transporting ATP synthase, subunit E	0240	0284	97.46
V-type ATP synthase subunit K (EC 3.6.3.14)	0241	0285	99.37
V-type ATP synthase subunit I (EC 3.6.3.14)	0242	0286	97.37
Electron transport complex protein RnfC	0474	0599	96.73
Electron transport complex protein RnfD	0475	0598	89.63
Electron transport complex protein RnfG	0476	0597	94.09
Electron transport complex protein RnfE	0477	0596	97.99
Electron transport complex protein RnfA	0478	0595	97.96
Electron transport complex protein RnfB	0479	0594	95.99
ECF transporters			
ATPase component of general energizing module of ECF transporters	0900	0555	96.74
ATPase component of general energizing module of ECF transporters	0901	0556	97.95
Transmembrane component of general energizing module of ECF transporters	0902	0558	95.15
Substrate-specific component PanT of predicted pantothenate ECF transporter	1789	0053	94.77
Substrate-specific component QueT of predicted queuosine-regulated ECF transporter	0113	0095	96.05
Substrate-specific component FolT of folate ECF transporter	0329	0178	97.09
Substrate-specific component BioY of biotin ECF transporter	2053	1132	94.54
Transmembrane component CbiQ of energizing module of cobalt ECF transporter	1912	1470	92.42
Additional substrate-specific component CbiN of cobalt ECF transporter	1913	1471	95.45
Substrate-specific component CbiM of cobalt ECF transporter	1914	1472	96.12
Substrate-specific component RibU of riboflavin ECF transporter	0067	1594	95.96
Substrate-specific component CblT of predicted ECF transporter for dimethylbenzimidazole	1645	1822	96.08
Nitrogen assimilation			
Ammonium transporter	1938	1700	96.86
Nitrogen regulatory protein P-II	1939	1701	100
Glutamine synthetase type I (EC 6.3.1.2)	1752	0219	97.52
Glutamine amidotransferase, class-II	1753	0218	98.09
Glutamate synthase [NADPH] large chain (EC 1.4.1.13)	1754	0217	99.21
Nitrite reductase probable electron transfer 4Fe-S subunit (EC 1.7.1.4)	1755	0216	95.8
Nitrite reductase probable [NAD(P)H] subunit (EC 1.7.1.4)	1756	0215	95.51
Glutamate synthase, alpha subunit domain protein	1757	0214	97.55

Motility and chemotaxis			
flagellar basal body rod protein	0602	0731	93.23
Flagellar basal-body rod protein FlgG	0603	0732	99.62
Flagellar basal-body P-ring formation protein FlgA	0604	0733	89.52
Flagellar L-ring protein FlgH	0605	0734	91.37
Flagellar P-ring protein FlgI	0606	Not called[b]	95% nt id to 12267-12370 of contig077[c]
Flagellar hook-associated protein FlgK	0611	Absent	0
Flagellar hook-associated protein FlgL	0612	0735	92.01
Flagellar assembly factor FliW	0613	Absent	0
Methyl-accepting chemotaxis protein	0512	0804	88.36
Methyl-accepting chemotaxis protein	0703	0251	95.39
Methyl-accepting chemotaxis protein	2121	1467	97.54
Methyl-accepting chemotaxis protein	2133	1509	95.72
Positive regulator of CheA protein activity (CheW)	0468	0604	97.45
Signal transduction histidine kinase CheA (EC 2.7.3.-)	1460	1293	85.61
Chemotaxis regulator CheY	1459	1292	98.35
Chemotaxis protein CheD	1291	1251	96.2
Chemotaxis protein CheC -- inhibitor of MCP methylation	1292	1252	99.03
Chemotaxis protein methyltransferase CheR (EC 2.1.1.80)	1293	1253	96.99
Chemotaxis regulator protein-glutamate methylesterase CheB (EC 3.1.1.61)	1294	1254	92.61
conserved hypothetical protein related to flagellar biosynthesis	1295	1255	97.67
Flagellar biosynthesis protein FlhA	1296	1256	98.24
Flagellar biosynthesis protein FlhB	1297	1257	95.22
Flagellar biosynthesis protein FliR	1298	1258	94.12
Flagellar biosynthesis protein FliQ	1299	1259	95.51
Flagellar biosynthesis protein FliP	1300	1260	96.65
hypothetical protein	1301	1261	89.05
Flagellar motor switch protein FliN	1302	1262	92.73
Flagellar motor switch protein FliM	1303	1263	97.87
Flagellar motor rotation protein MotB	1746	1265	97.91
Flagellar motor rotation protein MotA	1745	1266	98.55
Flagellar protein FlbD	1744	1267	91.89
Flagellar hook protein FlgE	1742	1268	90.28
Putative flagellar hook associated protein	1741	1269	89.23
Flagellar basal-body rod modification protein FlgD	1740	1270	93.8
Flagellar protein FliJ	1736	1274	97.35
Flagellum-specific ATP synthase FliI	1930	1275	98.55
Flagellar assembly protein FliH	1931	1276	93.89
Flagellar motor switch protein FliG	1932	1277	98.94
Flagellar M-ring protein FliF	1933	1278	96.23
Flagellar hook-basal body complex protein FliE	1934	1279	75.53

Flagellar basal-body rod protein FlgC	1935	1280	93.48
Flagellar basal-body rod protein FlgB	2304	1281	93.33
Flagellin protein FlaA	2070	1768	95.15
flagellar protein FlaG protein	2071	1769	94.87
Flagellar hook-associated protein FliD	2072	1770	92.67
Flagellar biosynthesis protein FliS	2073	1771	98.47
ABC transporters, amino acids and peptides			
Methionine ABC transporter ATP-binding protein	0062	1589	97.74
ABC transporter permease protein	0063	1590	98.07
ABC transporter substrate-binding protein	0064	1591	96.72
Glutamine ABC transporter, periplasmic glutamine-binding protein (TC 3.A.1.3.2)	1098	0020	98.02
glutamine ABC transporter, permease protein (glnP)	1099	0019	97.25
glutamine ABC transporter ATP-binding protein	1100	0018	99.19
Branched-chain amino acid transport ATP-binding protein LivF (TC 3.A.1.4.1)	0108	0089	95.58
Branched-chain amino acid transport ATP-binding protein LivG (TC 3.A.1.4.1)	0109	0090	91.57
High-affinity branched-chain amino acid transport system permease protein LivH (TC 3.A.1.4.1)	0110	0091	95.74
Branched-chain amino acid ABC transporter, amino acid-binding protein (TC 3.A.1.4.1)	0111	0092	97.11
Branched-chain amino acid transport ATP-binding protein LivF (TC 3.A.1.4.1)	0229	0273	96.6
Branched-chain amino acid transport ATP-binding protein LivG (TC 3.A.1.4.1)	0230	0274	97.68
Branched-chain amino acid transport system permease protein LivM (TC 3.A.1.4.1)	0231	0275	95.21
High-affinity branched-chain amino acid transport system permease protein LivH (TC 3.A.1.4.1)	0232	0276	97.31
Branched-chain amino acid ABC transporter, amino acid-binding protein (TC 3.A.1.4.1)	0233	0277	97.06
Oligopeptide transport ATP-binding protein OppF (TC 3.A.1.5.1)	0122	0104	92.24
Oligopeptide transport ATP-binding protein OppD (TC 3.A.1.5.1)	0123	0105	94.64
Dipeptide transport system permease protein DppC (TC 3.A.1.5.2)	0124	0106	97.83
Dipeptide transport system permease protein DppB (TC 3.A.1.5.2)	0125	0412	40.82
Dipeptide-binding ABC transporter, periplasmic substrate-binding component (TC 3.A.1.5.2)	0127	0416	32.76
Dipeptide transport system permease protein DppB (TC 3.A.1.5.2)	0353	0412	98.2
Dipeptide transport system permease protein DppC (TC 3.A.1.5.2)	0354	0413	95.38
Oligopeptide transport ATP-binding protein OppD (TC 3.A.1.5.1)	0355	0414	84.86
Oligopeptide transport ATP-binding protein OppF (TC 3.A.1.5.1)	0356	0415	77.81
Dipeptide-binding ABC transporter, periplasmic substrate-binding component (TC 3.A.1.5.2)	0357	0416	96.32
ABC transporters, zinc and phosphate			
Zinc ABC transporter, inner membrane permease protein ZnuB	0732	0383	91.42
Zinc ABC transporter, ATP-binding protein ZnuC	0733	0384	96.18
Zinc ABC transporter, periplasmic-binding protein ZnuA	0734	0385	92.5
Phosphate transport system permease protein PstA (TC 3.A.1.7.1)	0345	0404	93.47
Phosphate transport system permease protein PstC (TC 3.A.1.7.1)	0346	0405	96.35
Phosphate ABC transporter, periplasmic phosphate-binding protein PstS (TC 3.A.1.7.1)	0347	0406	97.27
Phosphate transport ATP-binding protein PstB (TC 3.A.1.7.1)	1713	1851	97.24

Phosphate transport system permease protein PstA (TC 3.A.1.7.1)	1714	1852	95.42
Phosphate transport system permease protein PstC (TC 3.A.1.7.1)	1715	1853	96.06
ABC transporters, saccharides			
Maltose/maltodextrin ABC transporter, permease protein MalF	0369	0426	97.73
sugar ABC transporter permease protein	0370	0427	97.45
SugC	0371	0428	95.09
ABC transporter, ATP-binding protein SugC	0372	0429	92.93
Maltose/maltodextrin ABC transporter, permease protein MalF	0688	0857	96.85
Maltose/maltodextrin ABC transporter, permease protein MalG	0689	0858	93.97
Ribose ABC transport system, periplasmic ribose-binding protein RbsB (TC 3.A.1.2.1)	0838	1049	95.94
Ribose ABC transport system, permease protein RbsC (TC 3.A.1.2.1)	0839	1050	99.08
Ribose ABC transport system, ATP-binding protein RbsA (TC 3.A.1.2.1)	0840	1051	96.23
Ribose ABC transport system, permease protein RbsC (TC 3.A.1.2.1)	0826	1034	95.88
ABC transporter related	0827	1035	91.6
ABC-type sugar transport system periplasmic component-like protein	0828	1036	98.84
Multiple sugar-binding transport atp-binding protein msmk	1311	0463	95.35
Dihydroxyacetone ABC transport system, permease protein	1312	0462	93.41
probable sugar ABC transporter, permease protein	1313	0461	98.66
probable sugar ABC transporter, substrate-binding protein	1308	0468	86.01
Maltose/maltodextrin ABC transporter, permease protein MalG	1345	1390	94.68
Multiple sugar ABC transporter, membrane-spanning permease protein MsmF	1346	1389	98.3
Multiple sugar ABC transporter, substrate-binding protein	1347	1388	98.11
sugar ABC transporter, ATP-binding protein	1400	1914	98.8
sugar ABC transporter, ATP-binding protein	1401	1913	97.89
Maltose/maltodextrin ABC transporter, permease protein MalG	1403	1911	98.17
ABC sugar transporter, permease component	1404	1910	97.74
ABC transporter, periplasmic substrate-binding protein	1405	1908	98.7
ABC transporter periplasmic binding protein	1522	1055	94.99
inner-membrane translocator	1523	1056	95.77
sugar ABC transporter, permease protein	1524	1057	98.02
ribose ABC transporter ATP-binding	1525	1058	96.91
Xylose oligosaccharides ABC transporter, sugar-binding protein	1638	1807	98.75
Oligopeptide transport system permease protein OppB (TC 3.A.1.5.1)	1639	1808	99.4
Xylose oligosaccharides ABC transporter, permease protein 2	1640	1809	98.6
oligopeptide ABC transporter, ATP-binding protein	1641	1810	96.99
Xylose oligosaccharides ABC transporter, ATP-binding protein 2	1642	1811	97.48
Ribose ABC transporter, periplasmic ribose-binding protein RbsB (TC 3.A.1.2.1)	1810	0026	98.75
Ribose ABC transport system, permease protein RbsC (TC 3.A.1.2.1)	1811	0027	97.92
Putative ribose/galactose/methyl galactoside import ATP-binding protein 1 (EC 3.6.3.17)	1812	0028	97.41
Ribose ABC transport system, ATP-binding protein RbsA (TC 3.A.1.2.1)	1827	1453	97.41
Ribose ABC transport system, permease protein RbsC (TC 3.A.1.2.1)	1828	1454	97.2
Ribose ABC transport system, periplasmic ribose-binding protein RbsB (TC 3.A.1.2.1)	1829	1455	97.89

Ribose ABC transport system, permease protein RbsC (TC 3.A.1.2.1)	1830	1456	90.06
COG1172: Ribose/xylose/arabinose/galactoside ABC-type transport systems, permease	2039	1901	96.76
Ribose/xylose/arabinose/galactoside ABC-type transport systems, permease component 1	2040	1900	98.04
ABC transporter, ATP-binding protein	2041	1897	96.12
ribose ABC transporter, periplasmic binding protein	2066	1988	98.78
ribose ABC transporter permease	2067	1987	99.43
Monosaccharide-transporting ATPase(EC:3.6.3.17)	2068	2288	100

[a] Numbers correspond to the gene number in the assemblies on RAST and submitted to Genbank.

[b] ORF was not called in cSCG, but was detected by BLASTn of corresponding gene in OP9-77CS.

[c] Nucleotide identity and region of cSCG detected by BLASTn of corresponding gene in OP9-77CS.

[d] Percent identity of predicted amino acid sequences for a given gene in the OP9-cSCG and OP9-77CS.

Supplementary References

61. Dereeper, A. *et al.* Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* **36**, W465–469 (2008).
62. Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113 (2004).
63. Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696–704 (2003).
64. Chevenet, F., Brun, C., Bañuls, A.-L., Jacq, B. & Christen, R. TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinformatics* **7**, 439 (2006).
65. Anisimova, M. & Gascuel, O. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst. Biol.* **55**, 539–552 (2006).
66. Lau, M., Aitchison, J. & Pointing, S. Bacterial community composition in thermophilic microbial mats from five hot springs in central Tibet. *Extremophiles* **13**, 139–149 (2009).
67. Sayeh, R. *et al.* Microbial diversity in Tunisian geothermal springs as detected by molecular and culture-based approaches. *Extremophiles* **14**, 501–514 (2010).
68. Burgess, E. A., Unrine, J. M., Mills, G. L., Romanek, C. S. & Wiegel, J. Comparative geochemical and microbiological characterization of two thermal pools in the Uzon Caldera, Kamchatka, Russia. *Microb. Ecol.* **63**, 471–489 (2012).

69. Sievert, S. M., Kuever, J. & Muyzer, G. Identification of 16S ribosomal DNA-defined bacterial populations at a shallow submarine hydrothermal vent near Milos Island (Greece). *Appl. Environ. Microbiol.* **66**, 3102–3109 (2000).
70. Dillon, J. G., McMath, L. M. & Trout, A. L. Seasonal changes in bacterial diversity in the Salton Sea. *Hydrobiologia* **632**, 49–64 (2009).
71. Gittel, A., Kofoed, M. V. W., Sørensen, K. B., Ingvorsen, K. & Schramm, A. Succession of Deferribacteres and Epsilonproteobacteria through a nitrate-treated high-temperature oil production facility. *Syst. Appl. Microbiol.* **35**, 165–174 (2012).
72. Kobayashi, H. *et al.* Phylogenetic diversity of microbial communities associated with the crude-oil, large-insoluble-particle and formation-water components of the reservoir fluid from a non-flooded high-temperature petroleum reservoir. *J. Biosci. Bioeng.* **113**, 204–210 (2012).
73. Yamane, K., Hattori, Y., Ohtagaki, H. & Fujiwara, K. Microbial diversity with dominance of 16S rRNA gene sequences with high GC contents at 74 and 98 °C subsurface crude oil deposits in Japan. *FEMS Microbiol. Ecol.* **76**, 220–235 (2011).
74. Pham, V. D. *et al.* Characterizing microbial diversity in production water from an Alaskan mesothermic petroleum reservoir with two independent molecular methods. *Environ. Microbiol.* **11**, 176–187 (2009).
75. Sugihara, T. *et al.* Unique diversity content of a thermophilic anaerobic microbial consortium that utilizes propionate in a synthetic medium. *J. Gen. Appl. Microbiol.* **53**, 363–369 (2007).
76. Wrighton, K. C. *et al.* A novel ecological role of the Firmicutes identified in thermophilic microbial fuel cells. *ISME J* **2**, 1146–1156 (2008).

77. Krakat, N., Westphal, A., Schmidt, S. & Scherer, P. Anaerobic digestion of renewable biomass: thermophilic temperature governs methanogen population dynamics. *Appl. Environ. Microbiol.* **76**, 1842–1850 (2010).
78. Worm, P., Fermoso, F. G., Lens, P. N. L. & Plugge, C. M. Decreased activity of a propionate degrading community in a UASB reactor fed with synthetic medium without molybdenum, tungsten and selenium. *Enzyme and Microbial Technology* **45**, 139–145 (2009).
79. Chouari, R. *et al.* Novel predominant archaeal and bacterial groups revealed by molecular analysis of an anaerobic sludge digester. *Environ. Microbiol.* **7**, 1104–1115 (2005).
80. Ito, T., Yoshiguchi, K., Ariesyady, H. D. & Okabe, S. Identification of a novel acetate-utilizing bacterium belonging to Synergistes group 4 in anaerobic digester sludge. *ISME J* **5**, 1844–1856 (2011).
81. Levén, L., Eriksson, A. R. B. & Schnürer, A. Effect of process temperature on bacterial and archaeal communities in two methanogenic bioreactors treating organic household waste. *FEMS Microbiol. Ecol.* **59**, 683–693 (2007).
82. Li, P., Wang, Y., Wang, Y., Liu, K. & Tong, L. Bacterial community structure and diversity during establishment of an anaerobic bioreactor to treat swine wastewater. *Water Science & Technology* **61**, 243 (2010).
83. Li, T., Mazéas, L., Sghir, A., Leblon, G. & Bouchez, T. Insights into networks of functional microbes catalysing methanization of cellulose under mesophilic conditions. *Environ. Microbiol.* **11**, 889–904 (2009).