

SUPPLEMENTARY TABLES AND FIGURES

The Complete Genome Sequence of *Moorella thermoacetica* (f. *Clostridium thermoaceticum*)

Elizabeth Pierce¹, Gary Xie^{2,3}, Ravi D. Barabote^{2,3}, Elizabeth Saunders^{2,3}, Cliff S. Han^{2,3}, John C. Detter^{2,3}, Paul Richardson^{2,4}, Thomas S. Brettin^{2,3}, Amaresh Das⁵, Lars G. Ljungdahl⁵, and Stephen W. Ragsdale¹

¹*Department of Biological Chemistry, University of Michigan, Ann Arbor, Michigan;* ²*Los Alamos National Laboratory, Bioscience Division, Los Alamos, New Mexico;* ³*Department of Energy Joint Genome Institute, Walnut Creek, CA;* ⁴*Lawrence Berkeley National Laboratory, Berkeley, CA;* ⁵*Department of Biochemistry and Molecular Biology, University of Georgia*

Supplementary Table 1. Pseudogenes found in the *M. thermoacetica* genome.

Locus Tag	Product Name	AA Seq Length	Remnants	Redundance
Moth_0037	FAD/FMN-containing dehydrogenases	462	early stop codon, frameshift	2 paralogs
Moth_0138	purine-cytosine permease and related proteins	417	early stop codon	no paralog
Moth_0164	dTDP-4-dehydrorhamnose reductase	283	fragment, frameshift	2 paralogs
Moth_0166	glucose-1-phosphate thymidyltransferase	227	deletion in central region	no paralog
Moth_0180	helicase	77	C-terminal fragment	no paralog
Moth_0276	putative resolvase	484	N-terminal fragment	1 paralog
Moth_0410	response regulator receiver	45	N-terminal fragment	2 paralogs
Moth_0411	GGDEF	181	C-terminal fragment	7 paralogs
Moth_0413	phage integrase	97	fragment	1 paralog
Moth_0472	ATPase, E1-E2 type	351	C terminal left	1 paralog
Moth_0627	reverse transcriptase	97	C-terminal fragment	1 paralog
Moth_0634	Nucleotidyltransferase/DNA polymerase involved in DNA repair	41	C-terminal fragment	1 paralog
Moth_0659	dTDP-glucose pyrophosphorylase	67	N-terminal fragment	no paralog
Moth_0661	daunorubicin resistance ABC transporter ATP-binding subunit	256	central region fragment	2 paralogs
Moth_0667	Phosphoglucomutase/ phosphomannomutase C terminal	146	C-terminal fragment	no paralog
Moth_0677	UDP-glucose 4-epimerase	178	C-terminal fragment	1 paralog
Moth_0680	UDP-glucose 4-epimerase	125	C-terminal fragment	1 paralog
Moth_0682	conserved hypothetical protein	77	C-terminal fragment	no paralog
Moth_0683	UDP-glucose 4-epimerase	113	C-terminal fragment	1 paralog
Moth_0684	nucleoside-diphosphate sugar epimerases-like	44	central region fragment	no paralog
Moth_0691	transposase	73	central region fragment	no paralog
Moth_0702	5-methylthioribose kinase	421	early stop codon	no paralog
Moth_0813	Methyl-viologen-reducing hydrogenase, delta subunit	143	early stop codon	1 paralog
Moth_0939	conserved hypothetical protein, Patatin	95	central region fragment	1 paralog
Moth_0954	RNA polymerase sigma-70 factor, ECF family	179	N-terminal fragment	2 paralogs
Moth_1008	permease	56	C-terminal fragment	1 paralog
Moth_1301	amidohydrolase family protein	454	frameshift	3 paralogs
Moth_1429	hypothetical	104	central region fragment	no paralog
Moth_1433	Ferrous iron transport protein B	682	early stop codon	1 paralog
Moth_1445	Periplasmic solute binding protein	207	missing N-terminal, may be still functional	no paralog
Moth_1485	ABC transporter ATPase component	87	central region fragment	14 paralogs
Moth_1566	conserved hypothetical protein	96	deletion in central region	1 paralog
Moth_1567	HEPN protein	99	N-terminal fragment	no paralog
Moth_1578	transposase	93	C-terminal fragment	no paralog
Moth_1609	Transposase-like	229	N-terminal fragment	1 paralog
Moth_1611	transposase for insertion sequence element IS904	291	frameshift	no paralog
Moth_1613	Transposase	411	frameshift	3 paralogs
Moth_1618	Transposase	414	frameshift	3 paralogs
Moth_1725	ATPase involved in conjugal	129	fragments	1 paralog

Moth_1801	plasmid transfer ISChy2, transposase	425	frameshift	3 paralogs
Moth_1927	hypothetical	48	central region fragment	no paralog
Moth_1932	hypothetical	56	C-terminal fragment	no paralog
Moth_1987	B3/4	132	C-terminal fragment	no paralog
Moth_1988	Binding-protein-dependent transport systems inner membrane component	136	C-terminal fragment	1 paralog
Moth_2064	Transposase, IS4	369	Insertion by Moth_2062 Integrase	1 paralog
Moth_2076	cell division protein, rodA/ftsW/ spoVE family	250	C-terminal fragment	2 paralogs
Moth_2081	conserved hypothetical protein	161	C-terminal fragment	2 paralogs
Moth_2090	putative cobalt transport system permease protein	154	C-terminal fragment	1 paralog
Moth_2141	DNA integration/recombination/ inversion protein	372	early stop codon	1 paralog
Moth_2144	transposase	414	central region fragment	3 paralogs
Moth_2146	Phage integrase	309	early stop codon	1 paralog
Moth_2221	transcriptional regulator, ArsR family	70	early stop codon, C-terminal fragment	1 paralog
Moth_2222	putative efflux protein	237	C-terminal fragment	2 paralogs
Moth_2340	transposase, IS4	51	early stop codon, C-terminal fragment	1 paralog
Moth_2350	reverse transcriptase	76	early stop codon, C-terminal fragment	1 paralog
Moth_2356	response regulator	36	N-terminal fragment	4 paralogs
Moth_2357	ATP-binding region, ATPase-like	212	C-terminal fragment	4 paralogs
Moth_2361	putative UDP-glucose 4-epimerase	85	C-terminal fragment	1 paralog

Supplementary Table 2. Parameters for BLAST alignments of some *M. thermoacetica* sequences mentioned in the text with GenBank sequences

<i>Moorella</i> gene	Annotation	Blasted against GenBank ID	Organism	e-value	percent identity	percent similarity
Moth_0037	GlcD	1141712	<i>Escherichia coli</i>	2.00E-96	41	62
Moth_0038	GlcD	1141712	<i>Escherichia coli</i>	5.00E-25	26	40
Moth_0038	GlcE	1141713	<i>Escherichia coli</i>	9.00E-17 ¹	29	50
Moth_0039	GlcF	1141714	<i>Escherichia coli</i>	1.00E-30	26	42
Moth_0439	HdrC	41017086	<i>M. marburgensis</i>	5.00E-19	34	49
Moth_0440	HdrB	41017087	<i>M. marburgensis</i>	6.00E-46	35	52
Moth_0441	HdrA	41017088	<i>M. marburgensis</i>	2.00E-56 ¹	32	48
Moth_0441	HdrA	72397227	<i>Methanosarcina barkeri</i>	3.00E-56 ¹	36	51
Moth_0441	HdrA	116698301	<i>S. fumaroxidans</i>	1.00E-05 ¹	24	37
Moth_0723	MoaD	2507065	<i>Escherichia coli</i>	2.7	30	44
Moth_0809	HdrC	41017086	<i>M. marburgensis</i>	5.00E-17 ¹	33	54
Moth_0810	HdrB	41017087	<i>M. marburgensis</i>	2.00E-46	31	53
Moth_0811 ²	HdrA	41017088	<i>M. marburgensis</i>	1.00E-65	58	74
Moth_0812 ²	HdrA	41017088	<i>M. marburgensis</i>	3.00E-126	49	65
Moth_0864	Phosphotransacetylase	5069455	<i>Salmonella enterica pduL</i>	4.00E-44	56	72
Moth_0977	NADH dehydrogenase I subunit A	77994891	<i>C. hydrogenoformans</i>	2.00E-28	62	75
Moth_0977	NADH dehydrogenase I subunit A	83575804	<i>Rhodospirillum rubrum</i>	8.00E-17	40	67
Moth_0977	NADH dehydrogenase I subunit A	72398012	<i>Methanosarcina barkeri</i>	5.00E-12	38	53
Moth_0978	NADH dehydrogenase I subunit B	77996073	<i>C. hydrogenoformans</i>	4.00E-59	71	86
Moth_0978	NADH dehydrogenase I subunit B	83575805	<i>Rhodospirillum rubrum</i>	1.00E-43	62	80
Moth_0978	NADH dehydrogenase I subunit B	1788624	<i>Escherichia coli</i>	5.00E-43 ¹	52	74
Moth_0978	NADH dehydrogenase I subunit B	72398011	<i>Methanosarcina barkeri</i>	7.00E-39	51	73
Moth_0979	NADH (or F420H ₂) dehydrogenase, subunit C	77995494	<i>C. hydrogenoformans</i>	2.00E-22	45	69
Moth_0979	NADH (or F420H ₂) dehydrogenase, subunit C	83575806	<i>Rhodospirillum rubrum</i>	2.00E-14 ¹	46	65
Moth_0979	NADH (or F420H ₂) dehydrogenase, subunit C	145693162	<i>Escherichia coli</i>	5.00E-14 ¹	36	52
Moth_0979	NADH (or F420H ₂) dehydrogenase, subunit C	72398010	<i>Methanosarcina barkeri</i>	2.00E-13 ¹	37	61
Moth_0980	NADH dehydrogenase I, D subunit	77994811	<i>C. hydrogenoformans</i>	5.00E-128	60	77
Moth_0980	NADH dehydrogenase I, D subunit	83575807	<i>Rhodospirillum rubrum</i>	1.00E-91	45	59
Moth_0980	NADH dehydrogenase I, D subunit	145693162	<i>Escherichia coli</i>	8.00E-79 ¹	39	57
Moth_0980	NADH dehydrogenase I, D subunit	72398009	<i>Methanosarcina barkeri</i>	1.00E-72	37	59
Moth_0981	NADH dehydrogenase, subunit H	77996987	<i>C. hydrogenoformans</i>	3.00E-99	56	74
Moth_0981	NADH dehydrogenase, subunit H	83575811	<i>Rhodospirillum rubrum</i>	4.00E-65	47	65
Moth_0981	NADH dehydrogenase, subunit H	1788618	<i>Escherichia coli</i>	3.00E-65	42	63
Moth_0981	NADH dehydrogenase, subunit H	72398008	<i>Methanosarcina barkeri</i>	4.00E-51	38	57
Moth_0982	NADH dehydrogenase I, subunit I	77995805	<i>C. hydrogenoformans</i>	2.00E-33	50	65

Moth_0982	NADH dehydrogenase I, subunit I	83575812	<i>Rhodospirillum rubrum</i>	9.00E-15	33	51
Moth_0982	NADH dehydrogenase I, subunit I	1788617	<i>Escherichia coli</i>	1.00E-15 ¹	36	54
Moth_0982	NADH dehydrogenase I, subunit I	72398007	<i>Methanosarcina barkeri</i>	4.00E-12	38	55
Moth_0983	NADH dehydrogenase I, subunit J	77995086	<i>C. hydrogenoformans</i>	2.00E-17	45	65
Moth_0983	NADH dehydrogenase I, subunit J	83575813	<i>Rhodospirillum rubrum</i>	4.00E-09	38	57
Moth_0983	NADH dehydrogenase I, subunit J	1788616	<i>Escherichia coli</i>	3.00E-06 ¹	29	49
Moth_0983	NADH dehydrogenase I, subunit J	72398006	<i>Methanosarcina barkeri</i>	3.00E-07 ¹	52	73
Moth_0984	NADH dehydrogenase I, subunit K	77996742	<i>C. hydrogenoformans</i>	7.00E-19	52	75
Moth_0984	NADH dehydrogenase I, subunit K	83575814	<i>Rhodospirillum rubrum</i>	1.00E-13	48	70
Moth_0984	NADH dehydrogenase I, subunit K	1788615	<i>Escherichia coli</i>	8.00E-09	37	62
Moth_0984	NADH dehydrogenase I, subunit K	72398004	<i>Methanosarcina barkeri</i>	2.00E-12	48	68
Moth_0985	NADH dehydrogenase I, subunit L	77996155	<i>C. hydrogenoformans</i>	2.00E-161	55	71
Moth_0985	NADH dehydrogenase I, subunit L	83575815	<i>Rhodospirillum rubrum</i>	1.00E-107	47	60
Moth_0985	NADH dehydrogenase I, subunit L	397909	<i>Escherichia coli</i>	7.00E-103	57	41
Moth_0985	NADH dehydrogenase I, subunit L	72398003	<i>Methanosarcina barkeri</i>	2.00E-101	42	57
Moth_0986	NADH dehydrogenase I, subunit M	77995931	<i>C. hydrogenoformans</i>	4.00E-136	53	74
Moth_0986	NADH dehydrogenase I, subunit M	83575816	<i>Rhodospirillum rubrum</i>	3.00E-80	36	58
Moth_0986	NADH dehydrogenase I, subunit M	1788613	<i>Escherichia coli</i>	4.00E-60 ¹	36	57
Moth_0986	NADH dehydrogenase I, subunit M	72398002	<i>Methanosarcina barkeri</i>	2.00E-66	35	55
Moth_0987	NADH dehydrogenase I, subunit M	77995299	<i>C. hydrogenoformans</i>	1e-122	48	67
Moth_0987	NADH dehydrogenase I, subunit M	83575817	<i>Rhodospirillum rubrum</i>	5.00E-63	41	63
Moth_0987	NADH dehydrogenase I, subunit M	145698289	<i>Escherichia coli</i>	7.00E-53	36	58
Moth_0987	NADH dehydrogenase I, subunit M	72398001	<i>Methanosarcina barkeri</i>	1.00E-66	37	57
Moth_1181	Phosphotransacetylase	5069455	<i>Salmonella enterica pduL</i>	1.00E-43	51	73
Moth_1194 ³	HdrA	41017088	<i>M. marburgensis</i>	2.00E-59	42	65
	HdrA	41017088	<i>M. marburgensis</i>	3.00E-59	47	64
	HdrA	41017088	<i>M. marburgensis</i>	5.00E-47	36	50
	HdrA	41017088	<i>M. marburgensis</i>	3.00E-42	51	72
Moth_1195	HdrB	41017087	<i>M. marburgensis</i>	2.00E-34	29	49
Moth_1196	HdrC	41017086	<i>M. marburgensis</i>	3.00E-13 ¹	26	44
Moth_1199	AcsF	6226573	<i>Rhodospirillum rubrum cooC</i>	2.00E-09	27	45
Moth_1204	CooC	6226573	<i>Rhodospirillum rubrum cooC</i>	3.00E-45	44	60
Moth_1278	TatA	57013123	<i>Bacillus subtilis</i>	7.00E-10 ¹	53	70
Moth_1279	TatC	2811036	<i>Bacillus subtilis</i>	5.00E-38	36	60
Moth_1379	TorD	471306	<i>Escherichia coli</i>	0.009	23	40
Moth_1390	TorD	471306	<i>Escherichia coli</i>	4.00E-04	23	38
Moth_1452	HdrC	41017086	<i>M. marburgensis</i>	4.00E-15 ¹	29	53
Moth_1453 ³	HdrA	41017088	<i>M. marburgensis</i>	2.00E-44	36	56
	HdrA	41017088	<i>M. marburgensis</i>	3.00E-44	32	51

	HdrA	41017088	<i>M. marburgensis</i>	5.00E-33	32	52
	HdrA	41017088	<i>M. marburgensis</i>	9.00E-17	38	59
Moth_1599	DsrD	4028022	<i>D. thermocisternum</i>	3.00E-27	73	89
Moth_1600	DsrB	4028021	<i>D. thermocisternum</i>	0	79	91
Moth_1601	DsrA	4028020	<i>D. thermocisternum</i>	0	77	90
Moth_1602	NarI	146921	<i>Escherichia coli</i>	1.00E-03 ¹	33	45
Moth_1603	NfrD	404305	<i>Escherichia coli</i>	2.00E-01 ¹	19	39
Moth_1629	DsrB	4028021	<i>D. thermocisternum</i>	7.00E-37	30	48
Moth_1630	DsrA	4028020	<i>D. thermocisternum</i>	2.00E-40 ¹	31	48
Moth_1631	DsrC	2648296	<i>Archaeoglobus fulgidus</i>	5.00E-07 ¹	35	50
Moth_1632	DsrC	2648296	<i>Archaeoglobus fulgidus</i>	3.00E-25	48	70
Moth_1717	NADH dehydrogenase, subunit G	145693161	<i>Escherichia coli</i>	3.00E-21 ¹	28	42
Moth_1718	NADH dehydrogenase, subunit F	1788620	<i>Escherichia coli</i>	2.00E-100 ¹	42	65
Moth_1719	NADH dehydrogenase, subunit E	1788621	<i>Escherichia coli</i>	2.00E-22	35	52
Moth_1819	MobB	1170986	<i>Escherichia coli</i>	1.00E-23	38	56
Moth_1820	MoeA	127232	<i>Escherichia coli</i>	1.00E-28 ¹	30	46
Moth_1822	ModC	32172422	<i>Escherichia coli</i>	6.00E-31 ¹	40	56
Moth_1823	ModB	84028144	<i>Escherichia coli</i>	7.00E-26	35	55
Moth_1824	ModA	585493	<i>Escherichia coli</i>	7.00E-23	34	52
Moth_1826	L-Lactate dehydrogenase	15893559	<i>Clostridium acetobutylicum</i>	1.00E-89	49	72
Moth_1886	NADH dehydrogenase, subunit E	1788621	<i>Escherichia coli</i>	3.00E-28 ¹	39	54
Moth_1887	NADH dehydrogenase, subunit F	1788620	<i>Escherichia coli</i>	5.00E-30	28	43
Moth_1888	NADH dehydrogenase, subunit G	145693161	<i>Escherichia coli</i>	6.00E-01 ¹	45	63
Moth_1954	Glyoxylate reductase	47116739	<i>Thermococcus litoralis</i>	1.00E-81	51	69
Moth_1954	Glyoxylate reductase	1049265	<i>Escherichia coli</i> D-lactate DH	3.00E-37	32	48
Moth_1972	CODH	118663634	<i>Clostridium cellulolyticum</i>	7.00E-133	39	58
Moth_1972	CODH	78044052	<i>C. hydrogenoformans</i> CODH IV	9.00E-126	39	57
Moth_2132	MogA	84028146	<i>Escherichia coli</i>	4.00E-26	36	55
Moth_2134	Moac	67468684	<i>Escherichia coli</i>	2.00E-43	55	70
Moth_2135	Moac	266540	<i>Escherichia coli</i>	9.00E-47	32	50
Moth_2136	MoeA	127232	<i>Escherichia coli</i>	2.00E-34	30	48
Moth_2137	MoeA	127232	<i>Escherichia coli</i>	2.00E-52	38	51
Moth_2176	HypE	1789085	<i>Escherichia coli</i>	4.00E-77	46	63
Moth_2177	HypD	1789084	<i>Escherichia coli</i>	2.00E-83	43	59
Moth_2178	HypC	1789083	<i>Escherichia coli</i>	1.00E-06	32	53
Moth_2179	HypF	2367152	<i>Escherichia coli</i>	3.00E-131	40	54
Moth_2180	HypB	1789082	<i>Escherichia coli</i>	2.00E-51	48	63
Moth_2181	HypA	1789081	<i>Escherichia coli</i>	5.00E-09	30	44
Moth_2182	HycI	1789072	<i>Escherichia coli</i>	5.00E-05	25	52

Moth_2183	HycH	1789073	<i>Escherichia coli</i>	6.00E-29	43	68
Moth_2184	Hydrogenase 3 hycG	77996053	<i>C. hydrogenoformans</i>	2.00E-30 ¹	43	65
Moth_2184	Hydrogenase 3 hycG	1789074	<i>Escherichia coli</i> hydrogenase 3	3.00E-69	58	69
Moth_2184	Hydrogenase 3 hycG	1788834	<i>Escherichia coli</i> hydrogenase 4	6.00E-69	52	67
Moth_2185	Hydrogenase 4 hyfH	77995239	<i>C. hydrogenoformans</i>	5.00E-15	30	44
Moth_2185	Hydrogenase 4 hyfH	1789075	<i>Escherichia coli</i> hydrogenase 3	2.00E-37	44	59
Moth_2185	Hydrogenase 4 hyfH	1788833	<i>Escherichia coli</i> hydrogenase 4	3.00E-34	50	66
Moth_2186	Hydrogenase 3 hycE	77995986	<i>C. hydrogenoformans</i>	5.00E-82 ¹	42	59
Moth_2186	Hydrogenase 3 hycE	1789076	<i>Escherichia coli</i> hydrogenase 3	0	64	78
Moth_2186	Hydrogenase 3 hycE	1788832	<i>Escherichia coli</i> hydrogenase 4	0	62	76
Moth_2187	Hydrogenase 4 hyfD	77996057	<i>C. hydrogenoformans</i>	1.00E-34 ¹	35	53
Moth_2187	Hydrogenase 4 hyfD	2367154	<i>Escherichia coli</i> hydrogenase 3	5.00E-13 ¹	24	47
Moth_2187	Hydrogenase 4 hyfD	1788829	<i>Escherichia coli</i> hydrogenase 4	2.00E-150	55	74
Moth_2188 ⁴	Hydrogenase 4 hyfF	77996057	<i>C. hydrogenoformans</i>	5.00E-40	31	51
Moth_2188	Hydrogenase 4 hyfF	2367154	<i>Escherichia coli</i> hydrogenase 3	2.00E-21 ¹	28	47
Moth_2188	Hydrogenase 4 hyfF	1788831	<i>Escherichia coli</i> hydrogenase 4	1.00E-144	54	72
Moth_2189	Hydrogenase 4 hyfE	1788830	<i>Escherichia coli</i> hydrogenase 4	4.00E-49	46	67
Moth_2190	Hydrogenase 4 hyfC	77996815	<i>C. hydrogenoformans</i>	4.00E-24 ¹	36	53
Moth_2190	Hydrogenase 4 hyfC	1789077	<i>Escherichia coli</i> hydrogenase 3	3.00E-81	50	69
Moth_2190	Hydrogenase 4 hyfC	87082115	<i>Escherichia coli</i> hydrogenase 4	1.00E-76	51	69
Moth_2191 ⁴	Hydrogenase 4 hyfB	77996057	<i>C. hydrogenoformans</i>	4.00E-51	39	58
Moth_2191	Hydrogenase 4 hyfB	2367154	<i>Escherichia coli</i> hydrogenase 3	2.00E-93	40	57
Moth_2191	Hydrogenase 4 hyfB	178882	<i>Escherichia coli</i> hydrogenase 4	7.00E-164	47	65
Moth_2192	Hydrogenase 4 hyfA	77996906	<i>C. hydrogenoformans</i>	4.00E-30	39	50
Moth_2192	Hydrogenase 4 hyfA	1789079	<i>Escherichia coli</i> hydrogenase 3	3.00E-42	45	58
Moth_2192	Hydrogenase 4 hyfA	87082114	<i>Escherichia coli</i> hydrogenase 4	4.00E-44	49	61
Moth_2303	GlcA	2495662	<i>Escherichia coli</i> glcA	3.00E-163	54	72
Moth_2303	GlcA	462489	<i>Escherichia coli</i> L-lactate permease	6.00E-157	52	71
Moth_2304	GlcC	1141711	<i>Escherichia coli</i> glcC	5.00E-23	31	50
Moth_2304	GlcC	82582240	<i>Escherichia coli</i> pyruvate dehydrogenase complex repressor	6.00E-34	38	59
Moth_2305	GlcF	1141714	<i>Escherichia coli</i>	2.00E-20 ¹	22	40
Moth_2307	GlcF	1141714	<i>Escherichia coli</i>	3.00E-26	25	44
Moth_2308	GlcD	1141712	<i>Escherichia coli</i>	1.00E-93	40	59

¹ These genes align over less than 80 percent of one or both sequences.

² These sequences align with the N- and C-terminal parts of HdrA, as described in the text.

³ These sequences align in more than one part with two regions of HdrA (roughly amino acids 1-300 and 320-650 of *M. marburgensis* HdrA); the ordering of these chunks of sequence is scrambled in the *M. thermoacetica* sequences, relative to the *M. marburgensis* sequences.

⁴ This aligns with the same part of the *C. hydrogenoformans* sequence as Moth_2187 does.

Abbreviations: *M. marburgensis*: *Methanothermobacter marburgensis*; *S. fumaroxidans*: *Syntrophobacter fumaroxidans*; *C. hydrogenoformans*: *Carboxydothemus hydrogenoformans*; *D. thermocisternum*: *Deulfotomaculum thermocisternum*

Supplementary Table 3. Genes belonging to alcohol and aldehyde dehydrogenase COGs and pfams.

Locus tag	pfam	COG	Annotation
Moth_0464	pfam00465	COG0371	Iron-containing alcohol dehydrogenase
Moth_0426	pfam00106	COG1028	Short-chain dehydrogenase/reductase SDR
Moth_0948	pfam00106	COG1028	3-oxoacyl-(acyl-carrier-protein) reductase
Moth_1258	pfam00106	COG1028	Short-chain SDR
Moth_0475	pfam00107	COG1063	Alcohol dehydrogenase zinc-containing
Moth_2268	pfam00107	COG1063	Alcohol dehydrogenase zinc-containing
Moth_1024	pfam00465	COG1454	Iron-containing alcohol dehydrogenase
Moth_1911	pfam00465	COG1454	Iron-containing alcohol dehydrogenase
Moth_1776	pfam02396	COG4569	Semialdehyde dehydrogenase, -binding
NAD Moth_0154			
Moth_0722			
Moth_2300	pfam02730 and pfam01314	COG2414	Aldehyde:ferredoxin oxidoreductase

Supplementary Table 4. Five *M. thermoacetica* gene clusters made up of COGs found in pyruvate:ferredoxin oxidoreductase.

Locus Tag	Annotation	COG
Moth_0033	Pyruvate ferredoxin oxidoreductase alpha subunit-like	COG0674
Moth_0034	2-oxoglutarate synthase	COG1013
Moth_0035	pyruvate:ferredoxin oxidoreductase, gamma subunit-like	COG1014
Moth_0064	Pyruvate:ferredoxin oxidoreductase, three COG fusion	COG0674 COG1014 COG1013
Moth_0376	pyruvate ferredoxin oxidoreductase, gamma subunit	COG1014
Moth_0377	pyruvate:ferredoxin oxidoreductase, delta subunit	COG1144
Moth_0378	pyruvate ferredoxin oxidoreductase, alpha subunit	COG0674
Moth_0379	ketoisovalerate ferredoxin oxidoreductase, beta subunit	COG1013
Moth_0934	pyruvate ferredoxin oxidoreductase, gamma subunit	COG1014
Moth_1591	pyruvate ferredoxin oxidoreductase, beta subunit	COG1013
Moth_1592	pyruvate ferredoxin oxidoreductase, alpha subunit	COG0674
Moth_1593	pyruvate:ferredoxin oxidoreductase, gamma subunit, delta subunit fusion	COG1014 COG1144
Moth_1921	pyruvate ferredoxin oxidoreductase, beta subunit	COG1013
Moth_1922	pyruvate ferredoxin oxidoreductase, alpha subunit	COG0674
Moth_1923	pyruvate:ferredoxin oxidoreductase, delta subunit	COG1144
Moth_1924	pyruvate ferredoxin oxidoreductase, gamma subunit	COG1014
Moth_1984	2-oxoglutarate ferredoxin oxidoreductase, alpha subunit	COG0674
Moth_1985	2-oxoglutarate ferredoxin oxidoreductase, beta subunit	COG1013
Moth_1986	2-oxoglutarate ferredoxin oxidoreductase, gamma subunit	COG1014
Moth_2276	pyruvate ferredoxin oxidoreductase, gamma subunit	COG1014

Supplementary Table 5. Dissimilatory sulfite reductase gene cluster.

Locus tag	Annotation	COG
Moth_1599	DsrD	no COG
Moth_1600	Dissimilatory sulfite reductase beta subunit	COG2221
Moth_1601	Dissimilatory sulfite reductase alpha subunit	COG2221
Moth_1602	Similar to <i>E. coli</i> nitrate reductase A γ subunit	COG2181
Moth_1603	Similar to NrfD	COG5557
Moth_1604	Iron sulfur cluster containing protein	COG0437
Moth_1605	Cytochrome c protein	no COG
Moth_1606	Iron sulfur oxidoreductase of unknown function	COG0247
Moth_1629	Dissimilatory sulfite reductase subunit	COG2221
Moth_1630	Dissimilatory sulfite reductase subunit	COG2221
Moth_1631	DsrC-like	COG2920
Moth_1632	DsrC-like	COG2920

Supplementary Table 6. COGs with greater representation in the *M. thermoacetica* genome than in genomes of two closely related bacteria.

COG, annotation	<i>M. thermoacetica</i>	<i>T. ethanolicus</i>	<i>T. tengcongensis</i>
DMSO reductase-type molybdopterin oxidoreductases			
COG0243, Typically cysteine-containing	6	0	0
COG3383, Uncharacterized	4	0	0
Xanthine dehydrogenase-type molybdopterin oxidoreductases			
COG1529, Molybdopterin binding subunit	4	0	0
COG1319, FAD binding subunit	2	0	0
COG2080, Fe-S subunit	2	0	0
COG0402, Amidohydrolase	5	1	1
Heterodisulfide reductases			
COG1148, Subunit A	4	1	0
COG2048, Subunit B	3	0	0
COG1150, Subunit C	3	0	0
COG1908 Similar to archaeal F420 reducing hydrogenases, δ subunit	3	0	0
Dissimilatory sulfite reductases			
COG2221, Alpha and beta subunits	6	1	0
COG2920, DsrC-like	2	0	0
COG2181, Similar to <i>E. coli</i> nitrite reductase, γ subunit	2	0	0
Miscellaneous electron transfer related COGs			
COG0247, Fe-S oxidoreductase	5	0	0
COG0277, FAD-linked oxidase	3	0	1
COG1251, NAD(P)H dependent nitrite reductase similar to <i>E. coli</i> nirB	4	1	0
COG0437, Ferredoxin	8	0	0
COG2414, Aldehyde ferredoxin oxidoreductase	3	1	0

Supplementary Table 7. Transporters encoded in the genome of *M. thermoacetica*.

Transporter Classification (TC) Family	Number of transport systems	Component proteins ¹
I. Channels		
1.A.11. The Ammonia Transporter Channel (Amt) Family	1	[Moth_0192]
1.A.23. The Small Conductance Mechanosensitive Ion Channel (MscS) Family	1	[Moth_0082]
1.A.30.1. The H ⁺ - or Na ⁺ -translocating Bacterial Flagellar Motor (Mot) Family	3	[Moth_0458, Moth_0459] [Moth_0781, Moth_0782] [Moth_1740, Moth_1741]
1.A.35. The CorA Metal Ion Transporter (MIT) Family	1	[Moth_0267]
Ila. MFS-type permeases		
2.A.1.1. The Sugar Porter (SP) Family	1	[Moth_0632]
2.A.1.2. The Drug:H ⁺ Antiporter-1 (12 Spanner) (DHA1) Family	1	[Moth_2226]
2.A.1.3. The Drug:H ⁺ Antiporter-2 (14 Spanner) (DHA2) Family	3	[Moth_0999] [Moth_1211] [Moth_2324]
2.A.1.4. The Organophosphate:Pi Antiporter (OPA) Family	1	[Moth_2345]
2.A.1.8. The Nitrate/Nitrite Porter (NNP) Family	1	[Moth_1315]
2.A.1.11. The Oxalate:Formate Antiporter (OFA) Family	1	[Moth_1590]
2.A.1.14. The Anion:Cation Symporter (ACS) Family	1	[Moth_0477]
2.A.1.15. The Aromatic Acid:H ⁺ Symporter (AAHS) Family	1	[Moth_2413]
2.A.1.32. The Putative Aromatic Compound/Drug Exporter (ACDE) Family	1	[Moth_0116]
2.A.1.35. The Fosmidomycin Resistance (Fsr) Family	1	[Moth_1519]
2.A.1.36. The Acriflavin-sensitivity (YnfM) Family	1	[Moth_1860]
2.A.1.40. The Purine Transporter (AzcA) Family	1	[Moth_1232]
2.A.1.45. The 2,4-diacetylphloroglucinol (PHL) Exporter (PHL-E) Family	1	[Moth_1257]

Unclassified MFS transporters	1	[Moth_2103]
IIb. Other permeases		
2.A.3. The Amino Acid-Polyamine-Organocation (APC) Family	6	[Moth_0156] [Moth_0737] [Moth_1345] [Moth_1528] [Moth_1920] [Moth_1930]
2.A.3.9. The Spore Germination Protein (SGP) Family	3	[Moth_1012] [Moth_1460] [Moth_2041]
2.A.4. The Cation Diffusion Facilitator (CDF) Family	1	[Moth_1683]
2.A.6.2. The (Largely Gram-negative Bacterial) Hydrophobe/Amphiphile Efflux-1 (HAE1) Family	1	[Moth_2369, Moth_2370]
2.A.6.4. The SecDF (SecDF) Family	1	[Moth_1690, Moth_1691]
2.A.7.22. The 4 TMS Small Multidrug Resistance-2 (SMR2) Family	1	[Moth_2476]
2.A.7.23. The Putative Tryptophan Efflux (Trp-E) Family	1	[Moth_1107]
2.A.7.3. The 10 TMS Drug/Metabolite Exporter (DME) Family	4	[Moth_1125] [Moth_1215] [Moth_1849] [Moth_2311]
2.A.8. The Gluconate:H ⁺ Symporter (GntP) Family	2	[Moth_0728] [Moth_0895]
2.A.9. The Cytochrome Oxidase Biogenesis (Oxa1) Family	1	[Moth_2521]
2.A.14. The Lactate Permease (LctP) Family	1	[Moth_2303]
2.A.19. The Ca ²⁺ :Cation Antiporter (CaCA) Family	1	[Moth_0818]
2.A.20. The Inorganic Phosphate Transporter (PiT) Family	1	[Moth_1237]
2.A.21. The Solute:Sodium Symporter (SSS) Family	1	[Moth_2279]
2.A.28. The Bile Acid:Na ⁺ Symporter (BASS) Family	1	[Moth_0574]
2.A.37. The Monovalent Cation:Proton Antiporter-2 (CPA2) Family	1	[Moth_2065]
2.A.38. The K ⁺ Transporter (Trk) Family	1	[Moth_1754, Moth_1755]

2.A.40. The Nucleobase:Cation Symporter-2 (NCS2) Family	2	[Moth_2118] [Moth_2478]
2.A.44. The Formate-Nitrite Transporter (FNT) Family	1	[Moth_2194]
2.A.46. The Benzoate:H ⁺ Symporter (BenE) Family	1	[Moth_1772]
2.A.47. The Divalent Anion:Na ⁺ Symporter (DASS) Family	2	[Moth_0383] [Moth_2267]
2.A.49. The Chloride Carrier/Channel (CIC) Family	1	[Moth_1588]
2.A.56. The Tripartite ATP-independent Periplasmic Transporter (TRAP-T) Family	1	[Moth_0422, Moth_0423, Moth_0424]
2.A.58. The Phosphate:Na ⁺ Symporter (PNaS) Family	1	[Moth_1817]
2.A.59. The Arsenical Resistance-3 (ACR3) Family	1	[Moth_2208]
2.A.66.1. The Multi Antimicrobial Extrusion (MATE) Family	1	[Moth_2080]
2.A.66.4. The Mouse Virulence Factor (MVF) Family	2	[Moth_0216] [Moth_0997]
2.A.69. The Auxin Efflux Carrier (AEC) Family	1	[Moth_0872]
2.A.76. The Resistance to Homoserine/Threonine (RhtB) Family	1	[Moth_0820]
2.A.78. The Branched Chain Amino Acid Exporter (LIV-E) Family	1	[Moth_1241, Moth_1242]
2.A.86. The Autoinducer-2 Exporter (AI-2E) Family (Formerly the PerM Family, TC #9.B.22)	1	[Moth_1647]
IIIa. ABC-type transporters		
3.A.1.1. The Carbohydrate Uptake Transporter-1 (CUT1) Family	1	[Moth_1481 (C), Moth_1482 (M), Moth_1483 (M), Moth_1484 (R)]
3.A.1.2. The Carbohydrate Uptake (CUT2) Family Moth_0614 (M) Moth_0700 (C+C), [Moth_1950 (M), Moth_1951 (M), Moth_1952 (C+C)]	4	[Moth_0612 (R), Transporter-2 Moth_0613 (C+C), [Moth_0699 (R), Moth_0701 (M)]

		[Moth_2020 (M), Moth_2021 (C+C), Moth_2022 (R)]
3.A.1.3. The Polar Amino Acid Uptake Transporter (PAAT) Family	1	[Moth_1126 (R), Moth_1127 (M), Moth_1128 (C)]
3.A.1.4. The Hydrophobic Amino Acid Uptake Transporter (HAAT) Family	1	[Moth_2498 (C), Moth_2499 (C), Moth_2500 (M), Moth_2501 (M), Moth_2502 (R)]
3.A.1.6. The Sulfate/Tungstate Uptake Transporter (SulT) Family	1	[Moth_0067 (R), Moth_0068 (M), Moth_0069 (C)]
3.A.1.7. The Phosphate Uptake Transporter (PhoT) Family	1	[Moth_0112 (M), Moth_0113 (M), Moth_0114 (C), Moth_0473 (R)]
3.A.1.8. The Molybdate Uptake Transporter (MolT) Family	1	[Moth_1822 (C), Moth_1823 (M), Moth_1824 (R)]
3.A.1.12. The Quaternary Amine Uptake Transporter (QAT) Family (Similar to 3.A.1.16 and 3.A.1.17)	1	[Moth_1685 (R), Moth_1686 (M), Moth_1687 (M), Moth_1688 (C)]
3.A.1.14. The Iron Chelate Uptake Transporter (FeCT) Family (Similar to 3.A.1.13 and 3.A.1.15)	2	[Moth_1422 (R), Moth_1423 (C), Moth_1424 (M)] [Moth_1464 (C), Moth_1465 (M), Moth_1466 (R)]
3.A.1.15. The Manganese/Zinc/Iron Chelate Uptake Transporter (MZT) Family (Similar to 3.A.1.12, 3.A.1.14 and 3.A.1.16)	1	[Moth_0928 (M), Moth_0929 (C), Moth_0930 (R)]
3.A.1.16. The Nitrate/Nitrite/Cyanate Uptake Transporter (NitT) Family (Similar to 3.A.1.12 and 3.A.1.17)	2	[Moth_0405 (M), Moth_0406 (C)] [Moth_1973 (R), Moth_1974 (M), Moth_1975 (C)]
3.A.1.17. The Taurine Uptake Transporter (TauT) Family (Similar to 3.A.1.12 and 3.A.1.16)	3	[Moth_0098 (C), Moth_0099 (M)] [Moth_0466 (M), Moth_0467 (C)] [Moth_1979 (C), Moth_1981 (R), Moth_1982 (M)]
3.A.1.18. The Cobalt Uptake Transporter (CoT) Family	1	[Moth_1219 (C), Moth_1220 (M), Moth_1221 (R), Moth_1222 (M)]
3.A.1.23. The Nickel/Cobalt Uptake Transporter (NiCoT) Family (M)	5	[Moth_0707 (C+C), [Moth_0708 (M)] [Moth_1456 (C+C), Moth_1457 [Moth_1861 (C), Moth_1862 (M), Moth_1863 (M)] [Moth_2058 (M), Moth_2059 (C+C)] [Moth_2427 Moth_2428 (C), Moth_2429 (C)]
3.A.1.102. The Lipooligosaccharide Exporter (LOSE) Family	1	[Moth_0874 (C), Moth_0875 (M)]
3.A.1.105. The Drug Exporter-1	1	[Moth_2297 (C), Moth_2298 (M)]

(DrugE1) Family

3.A.1.122. The Macrolide Exporter (MacB) Family	1	[Moth_2353 (M), Moth_2354 (C), Moth_2355 (MFP)]
3.A.1.129. The CydDC Cysteine Exporter (CydDC-E) Family	2	[Moth_2097 (M+C)] [Moth_2098 (M+C)]
Unclassified ABC-type transporters	4	[Moth_0960 (C), Moth_0961 (M)] [Moth_1171 (M), Moth_1172 (C), Moth_1174 (C)] [Moth_1899 (M), Moth_1900 (C), Moth_1901 (C), Moth_1902 (MFP)] [Moth_2328 (MFP), Moth_2329 (C), Moth_2330 (M), Moth_2331 (M)]
Orphan ABC transporter components	5	[Moth_1164 (C)] [Moth_1265 (C)] [Moth_1419 (R)] [Moth_1444 (M)] [Moth_0660 (M)]
IIIb. Other Active transporters		
3.A.3. The P-type ATPase (P-ATPase) Superfamily	3	[Moth_0887] [Moth_1994] [Moth_2203]
3.A.10. The H ⁺ -translocating Pyrophosphatase (H ⁺ -PPase) Family	2	[Moth_0418] [Moth_1349]
IV. Phosphotransferase system		
4.A.2. The PTS Fructose-Mannitol (Fru) Moth_0015, Family V. Accessory transport proteins	1	[Moth_0013, Moth_0014, Moth_0016]
8.A.1. The Membrane Fusion Protein (MFP) Family VI. Putative or Unassigned	2	[Moth_2322] [Moth_2323]
9.A.8. The Ferrous Iron Uptake (FeoB) Family	1	[Moth_1182] [Moth_1405]
9.B.3. The Putative Bacterial Murein Precursor Exporter (MPE) Family	3	[Moth_0544] [Moth_0842] [Moth_0910]
9.B.10. The 6 TMS Putative MarC Transporter (MarC) Family	1	[Moth_1933]
9.B.14. The Putative Heme Handling Protein (HHP) Family	1	[Moth_2200]
9.B.22. The Putative Thiamin Transporter (PTT) Family	1	[Moth_1040]
9.B.26. The PF27 (PF27) Family	1	[Moth_0395]
9.B.27. The YdjX-Z (YdjX-Z) Family	1	[Moth_0719]
9.B.31. The YqiH (YqiH) Family	2	[Moth_1217] [Moth_1321]
9.B.32. The Putative Vectorial Glycosyl	2	[Moth_1831] [Moth_2154]

Polymerization (VGP) Family

9.B.53. The Unknown IT-6 (UIT6) Family	1	[Moth_0569]
9.B.56. The Unknown IT-9 (UIT9) Family	1	[Moth_2124]
9.B.59. The Putative Peptide Transporter Carbon Starvation CstA (CstA) Family	1	[Moth_1963]
9.B.63. The 10 TMS Putative Sulfate Exporter (PSE) Family	1	[Moth_1878]
Unclassified putative transporters	2	[Moth_1233] [Moth_2507]

¹All proteins that constitute a single transport system are enclosed in square ([])brackets. For the ABC-type transporters, C denotes the cytoplasmic ATPase, M is the membrane-spanning integral membrane transporter, R denotes the substrate-binding receptor, and MFP is the membrane fusion protein.

Supplementary Figure legends

Supplementary Figure 1. Circular representations of the genome of *M.*

thermoacetica ATCC 39073. The outermost two circles indicate start sites of genes and assigned function (colored by COG categories). Circle 1 consists of forward-strand gene products. Circle 2 consists of reverse-strand gene products. Colors represent the following functional categories: amino acid biosynthesis, cyan; biosynthesis of cofactors, brown; cell envelope, light gray; cellular processes, light blue; central intermediary metabolism, dark salmon; energy metabolism, green; fatty acid and phospholipid metabolism, orange; other categories, salmon; protein fate, dark gray; purines, pyrimidines, nucleosides, and nucleotides, light green; regulatory functions, light salmon; replication, blue; transcription and translation, magenta; transport and binding proteins, yellow; unassigned, black; unknown function, red; Circle 3, RNA genes (tRNAs green, sRNAs red); circle 4, pseudogenes, circle 5, IS elements; circle 6, G+C content; circle 7, GC skew ((G-C/G+C), khaki indicates values >1, purple <1).

Supplementary Figure 2. Pathway proposed for metabolism of xylose. Locus tags for genes encoding individual steps are shown.

Supplementary Figure 3. Pathway proposed for metabolism of glucose and fructose.

Locus tags for genes encoding individual steps are shown.

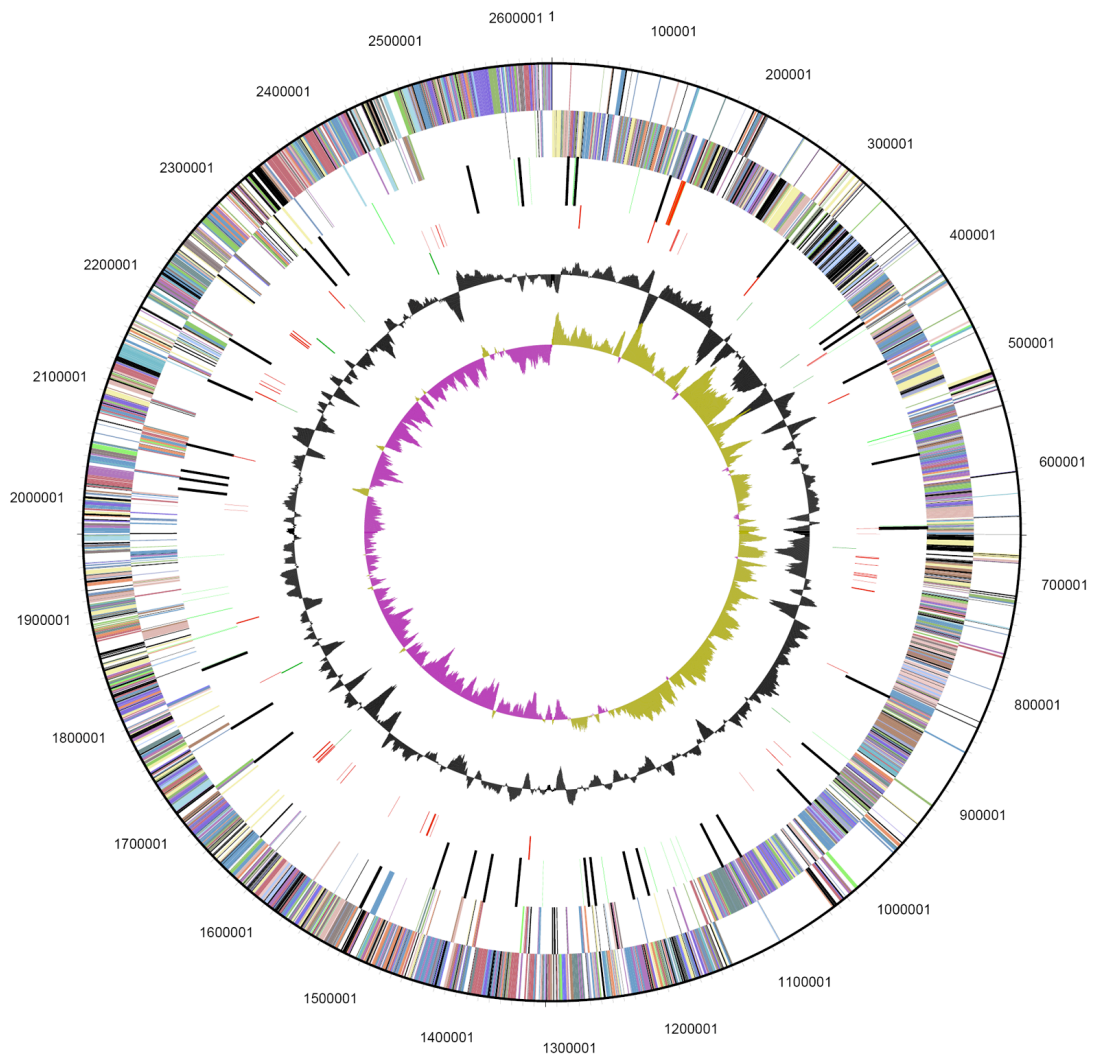
Supplementary Figure 4. Pyruvate from sugar metabolism provides electrons and carbon for acetate synthesis by the Wood-Ljungdahl pathway.

Supplementary Figure 5. Tetrahydrofolate synthesis. *M. thermoacetica* genes which could be involved in synthesis of tetrahydrofolate from guanosine triphosphate and chorismate. The sequences of genes encoding enzymes which catalyze the transformations of 7,8-dihydroneopterin 3'-triphosphate to dihydroneopterin phosphate and dihydroneopterin phosphate to dihydroneopterin are not known.

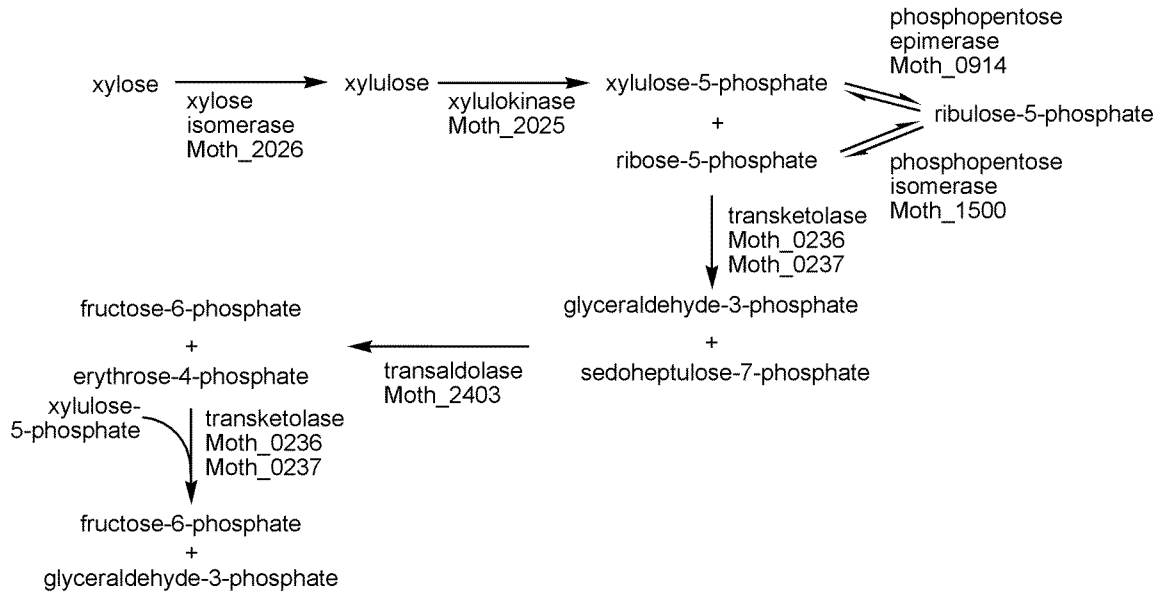
Supplementary Figure 6. Synthesis of molybdenum cofactor and bis-molybdopterin guanine dinucleotide cofactor from guanine triphosphate. No *M. thermoacetica* genes were found with homology to the large subunit of molybdopterin synthase, MoaE.

Supplementary Figure 7. Components of the bacterial flagellar proteins in *M. thermoacetica*. The figure was derived from the KEGG web server (Kanehisa et al, 2006; <http://www.genome.jp/kegg/>). The gene IDs that correspond to the labeled flagellar proteins are as follows: FlgB (Moth_0768), FlgC (Moth_0769), FlgD (Moth_0777), FlgE (Moth_0779), FlgG (Moth_0764), FlgK (Moth_0746), FlgL (Moth_0747), FlgL (Moth_0760), FlgM (Moth_0744), FlgN (Moth_0745), FlhA (Moth_0790), FlhB (Moth_0789), FliD (Moth_0765), FliE (Moth_0770), FliF (Moth_0771), FliG (Moth_0772), FliH (Moth_0773), FliI (Moth_0774), FliJ (Moth_0775), FliK (Moth_0776), FliM (Moth_0804), FliN (Moth_0784), FliO (Moth_0785), FliP (Moth_0786), FliQ (Moth_0787), FliR (Moth_0788), FliS (Moth_0766). Genes encoding FlgA, FlhC, FlhD, FliT proteins could not be identified in the *M. thermoacetica* genome. However, additional putative flagellar genes with unknown functions are encoded in the genome, such as Moth_0780 (FlbD), Moth_0783 (FliL), Moth_0791 (FlhF), Moth_0797 (FlgE-like), Moth_0798 (FlgE-like).

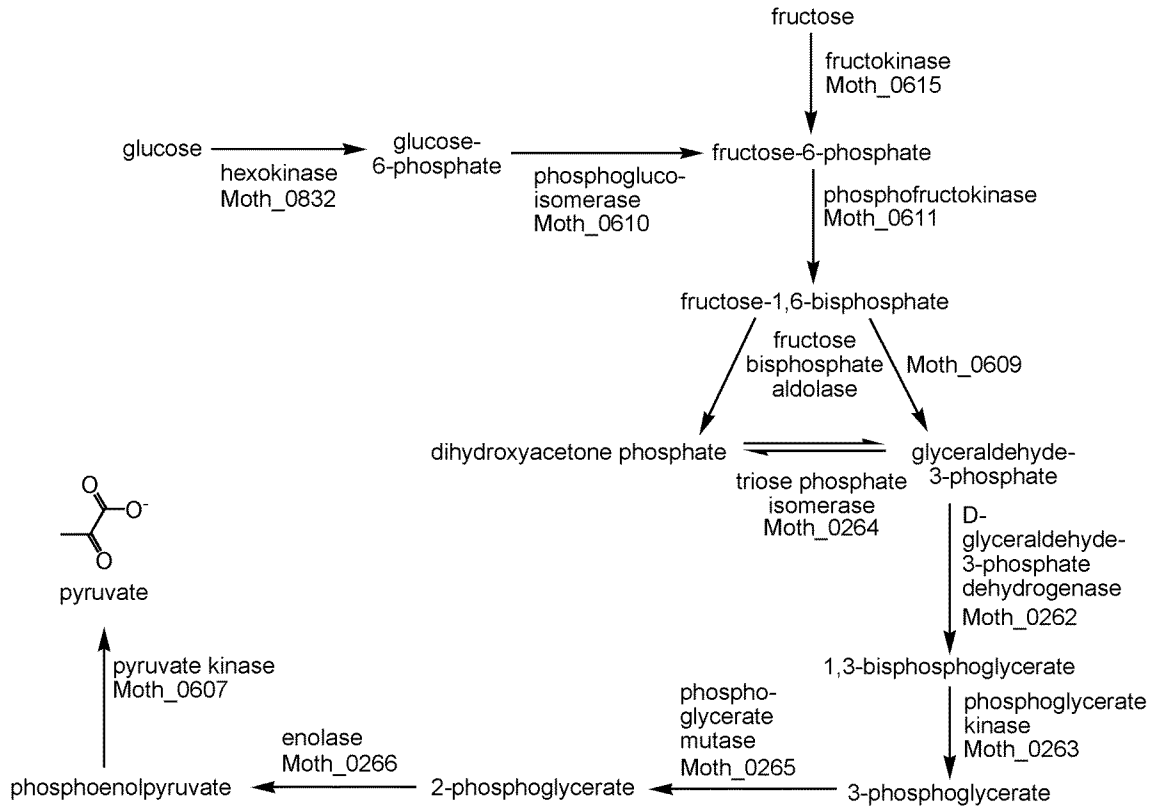
Supplementary Figure 1. Circular representations of the genome of *M. thermoacetica* ATCC 39073.



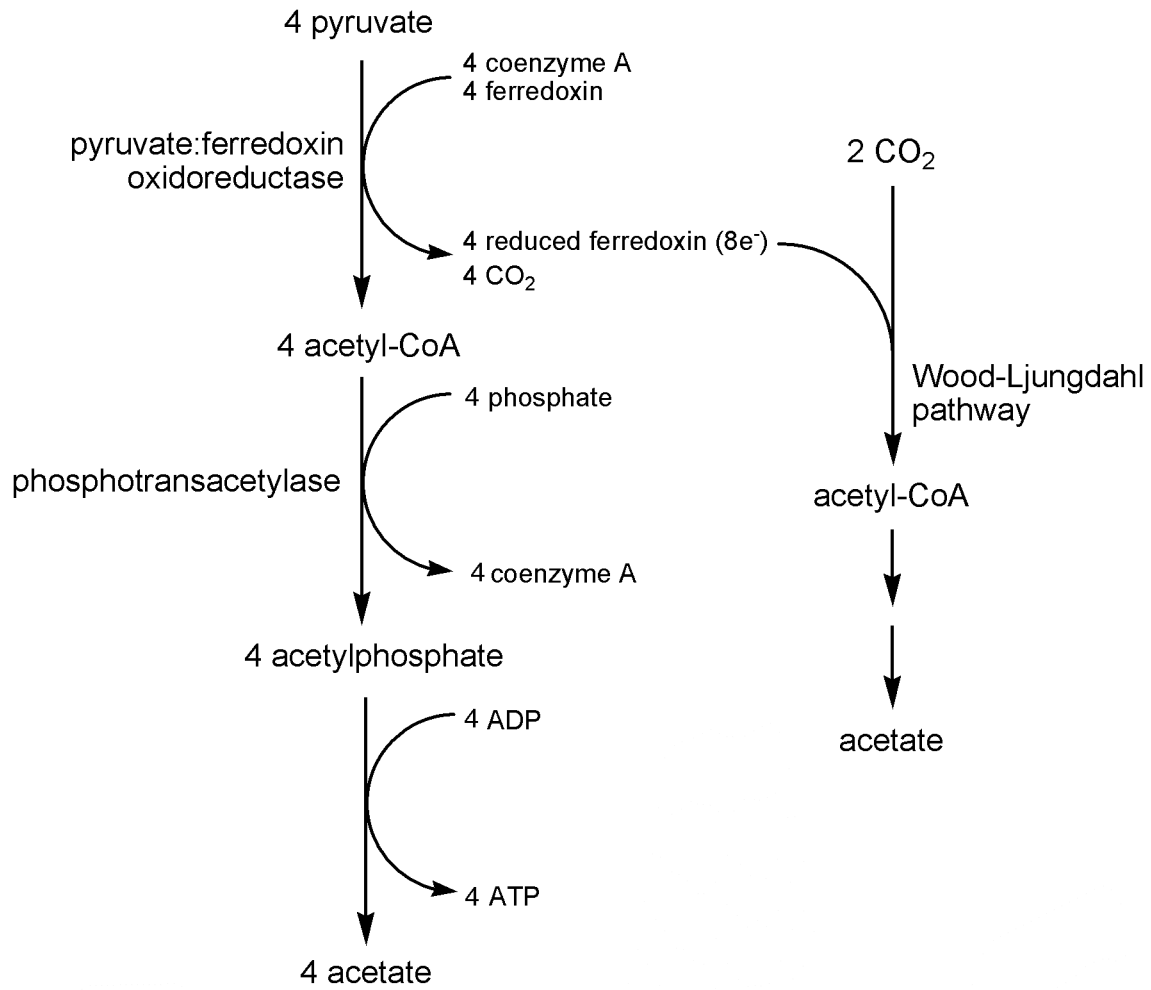
Supplementary Figure 2. Pathway proposed for metabolism of xylose.



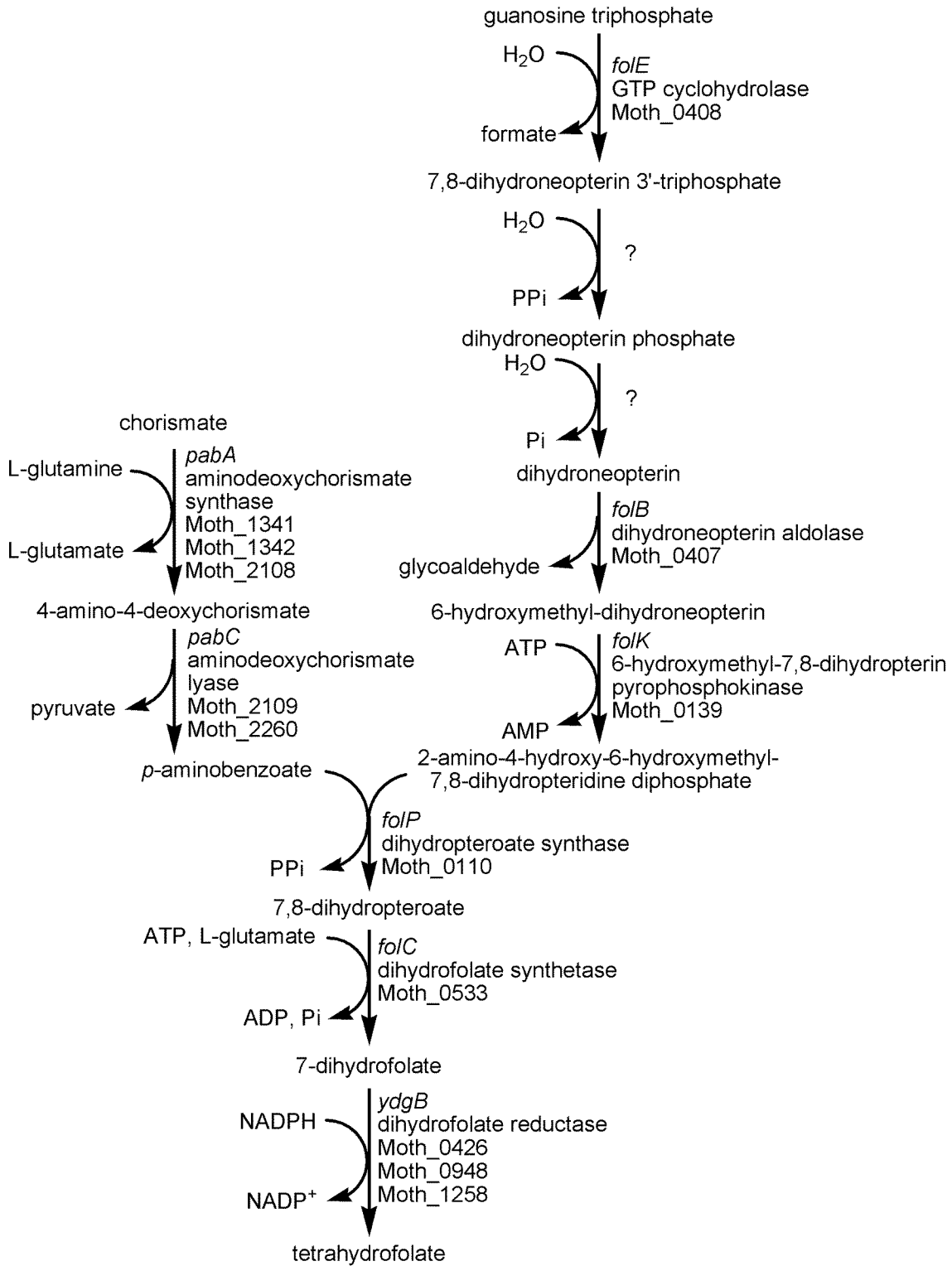
Supplementary Figure 3. Pathway proposed for metabolism of glucose and fructose.



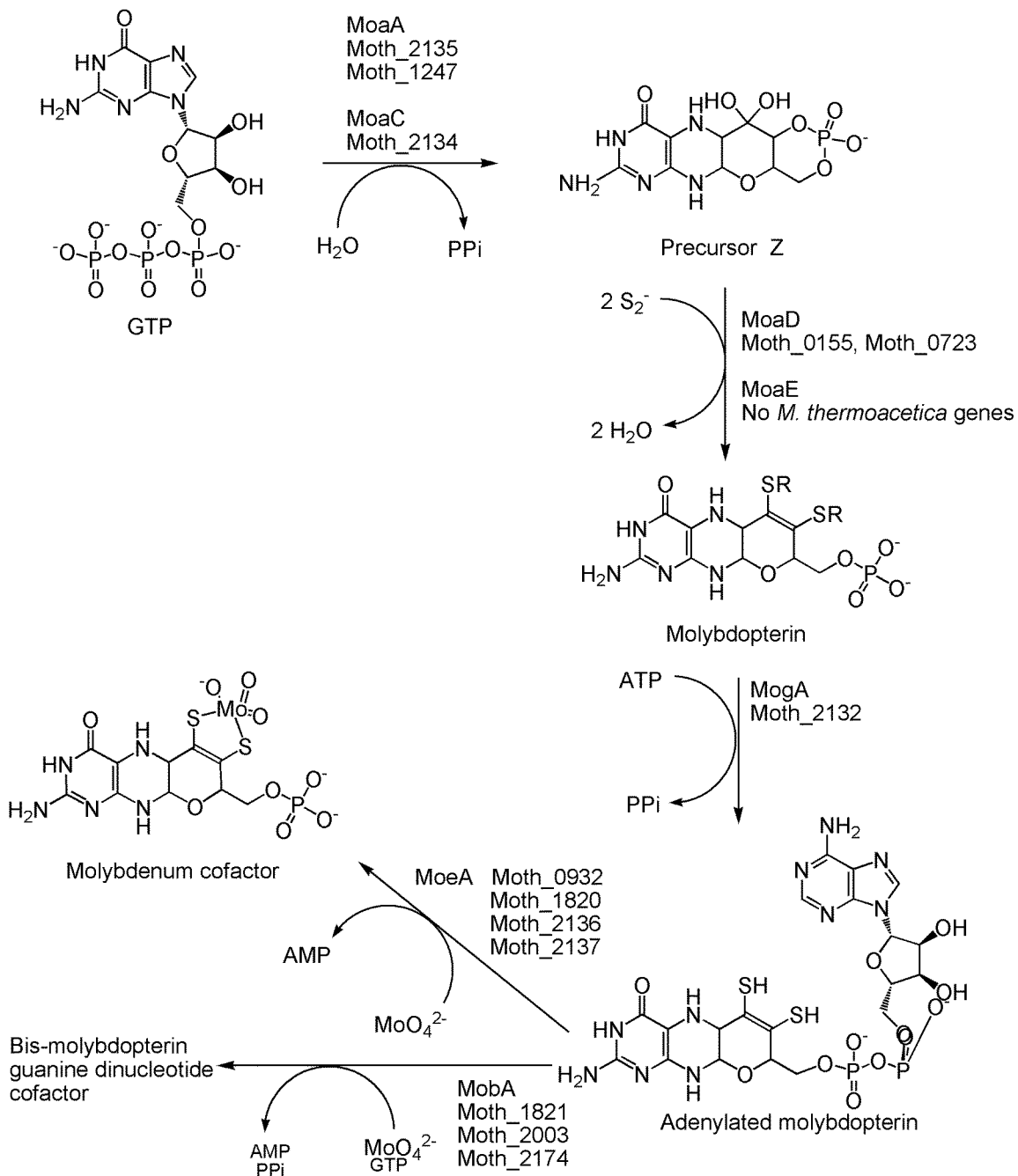
Supplementary Figure 4. Pyruvate from sugar metabolism provides electrons and carbon for acetate synthesis by the Wood-Ljungdahl pathway.



Supplementary Figure 5. Tetrahydrofolate synthesis.



Supplementary Figure 6. Synthesis of molybdenum cofactor and bis-molybdopterin guanine dinucleotide cofactor from guanine triphosphate.



Supplementary Figure 7. Components of the bacterial flagellar proteins in *M. thermoacetica*.

