

Supplementary Table 1 Some organisms in the family *Beijerinckiaceae*, with a brief description of growth substrates utilized. Data are from: Dedysh *et al.* 2000; Dunfield *et al.*, 2003; Dedysh *et al.*, 2005; Kennedy, 2005; Vorobev *et al.*, 2009; Dedysh and Dunfield, 2010; Dunfield *et al.*, 2010; Vorobev *et al.*, 2011; Berestovskaya *et al.*, 2012.

Organism	Growth on CH ₄	Growth on CH ₃ OH	Growth on substrates with C-C bonds	Description
<i>Methylocella</i> spp. (<i>silvestris</i> , <i>palustris</i> , <i>tundrae</i>)	+	+	acetate, ethanol, propane, pyruvate, succinate, malate	facultative methanotroph
<i>Methylocapsa acidiphila</i>	+	+	-	obligate methanotroph
<i>Beijerinckia indica</i>	-	-	many sugars, alcohols, organic acids	non-methylotrophic chemoorganotroph
<i>Beijerinckia mobilis</i>	-	+	many sugars, alcohols, organic acids	facultative methylotroph
<i>Methyloferula stellata</i>	+	+	-	obligate methanotroph
<i>Methylocapsa aurea</i>	+	+	acetate	facultative methanotroph
<i>Methylovirgula ligni</i>	-	+	many sugars, alcohols, organic acids	facultative methylotroph
<i>Methylorosula polaris</i>	-	+	many sugars, alcohols, organic acids	facultative methylotroph

Supplementary Table 2 16S rRNA gene sequence identities of some species in the *Beijerinckiaceae*.

Organism	<i>M. silvestris</i>	<i>M. acidiphila</i>	<i>B. indica</i>	<i>B. mobilis</i>	<i>M. aurea</i>	<i>M. stellata</i>
<i>Methylocella silvestris</i>						
<i>Methylocapsa acidiphila</i>	96.2					
<i>Beijerinckia indica</i>	97.2	96.3				
<i>Beijerinckia mobilis</i>	96.2	96.7	98.3			
<i>Methylocapsa aurea</i>	97.3	98.1	97.5	97.5		
<i>Methyloferula stellata</i>	97.0	96.5	96.7	96.8	97.8	

Supplementary Table 3 Methylotrophy genes assembled primarily from Chistoserdova *et al.* (2009), Vuilleumier *et al.* (2009), and Stein and Klotz (2011). The genes used to construct the phylogeny shown in Figure 5 are **in bold**. Presence of the genes in each of the three study genomes (Ms=*M. silvestris*; Mc=*M. acidiphila*; Bi=*B. indica*) is indicated. Positive detection of a homologue was based on annotations in JGI IMG, and on BLAST searches as described in the methods. A question mark indicates that there is a homologue (>30% identity) of uncertain function.

gene symbol	gene name	Ms	Mc	Bi
Methane oxidation				
<i>pmoC</i>	particulate methane monooxygenase (pMMO), subunit C	-	+	-
<i>pmoB</i>	particulate methane monooxygenase (pMMO), subunit B	-	+	-
<i>pmoA</i>	particulate methane monooxygenase (pMMO), subunit A	-	+	-
<i>mmoX</i>	soluble methane monooxygenase (sMMO), subunit alpha	+	-	-
<i>mmoY</i>	soluble methane monooxygenase (sMMO), subunit beta	+	-	-
<i>mmoB</i>	soluble methane monooxygenase (sMMO), regulatory component	+	-	-
<i>mmoZ</i>	soluble methane monooxygenase (sMMO), subunit gamma	+	-	-
Methanol oxidation				
<i>mxoF</i>	PQQ-dependent methanol dehydrogenase (MDH), large subunit	+	+	-
<i>mxoJ</i>	periplasmic protein, probable ABC transporter substrate-binding protein	+	+	-
<i>mxoG</i>	periplasmic c-type cytochrome, primary electron acceptor from PQQ-dependent MDH	+	+	-
<i>mxoI</i>	PQQ-dependent methanol dehydrogenase (MDH), small subunit	+	+	-
<i>mxoR</i>	ATPase, P-loop NTPase family	+	+	-
<i>mxoS</i>	a von Willebrand factor type A (vWA) domain involved in methanol oxidation	+	+	-
<i>mxoA</i>	essential for Ca ²⁺ insertion into MDH	+	+	-
<i>mxoC</i>	essential for Ca ²⁺ insertion into MDH	+	+	-
<i>mxoK</i>	essential for Ca ²⁺ insertion into MDH	+	+	-
<i>mxoL</i>	essential for Ca ²⁺ insertion into MDH	+	+	-
<i>mxoD</i>	essential for Ca ²⁺ insertion into MDH	+	+	-
<i>mxoE</i>	unknown	+	+	-
<i>mxoH</i>	unknown	+	-	-
<i>mxoB</i>	transcriptional regulator	-	-	-
<i>mxoD</i>	integral membrane sensor signal transduction histidine kinase	?	?	-
<i>mxoM</i>	response regulator	?	?	-
<i>mxoQ</i>	integral membrane sensor signal transduction histidine kinase	+	+	-
<i>mxoE</i>	response regulator	+	+	-
<i>pqqA</i>	coenzyme PQQ biosynthesis	+	+	-
<i>pqqB</i>	coenzyme PQQ biosynthesis	+	+	+
<i>pqqC</i>	coenzyme PQQ biosynthesis	+	+	+
<i>pqqD</i>	coenzyme PQQ biosynthesis	+	+	+
<i>pqqE</i>	coenzyme PQQ biosynthesis	+	+	+
<i>pqqF</i>	coenzyme PQQ biosynthesis	+	+	+
<i>pqqG</i>	coenzyme PQQ biosynthesis	+	+	+
<i>xoxF</i>	homolog of <i>mxoF</i>	+	+	+
Methylamine oxidation				
<i>mauF</i>	methylamine oxidation protein	-	-	-
<i>mauB</i>	methylamine dehydrogenase, large subunit	-	-	-
<i>mauE</i>	essential for small subunit maturation	-	-	-
<i>mauD</i>	essential for small subunit maturation	-	-	-
<i>mauA</i>	methylamine dehydrogenase, small subunit	-	-	-
<i>mauC</i>	amicyanin	-	-	-
<i>mauJ</i>	methylamine utilization protein	-	-	-
<i>mauG</i>	methylamine utilization protein	-	-	-
<i>mauM</i>	ferredoxin	-	-	-
<i>mauN</i>	ferredoxin	-	-	-
C1 transfer / formaldehyde oxidation				
<i>fae</i>	formaldehyde activating enzyme	+	+	-
<i>mtdB</i>	methylene-tetrahydromethanopterin (H₄MPT) dehydrogenase	+	+	-

fhcA	formyltransferase/hydrolase complex, alpha subunit	+	+	-
fhcB	formyltransferase/hydrolase complex, beta subunit	+	+	-
fhcC	formyltransferase/hydrolase complex, gamma subunit	+	+	-
fhcD	formyltransferase/hydrolase complex, delta subunit	+	+	-
<i>orfY</i>	Unknown	+	+	-
<i>mtdA</i>	methylene-H ₄ MPT /methylene-tetrahydrofolate (H ₄ F) dehydrogenase	-	+	-
<i>fch</i>	methenyl-H ₄ F cyclohydrolase	-	+	-
<i>dmrA</i>	dihydropyrimethanopterin reductase	+	-	-
fflL	formyl-H₄F ligase	+	+	+
<i>folk</i>	2-amino-4-hydroxy-6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase	+	+	+
<i>folB</i>	dihydropyrimethanopterin aldolase	-	-	-
<i>folP</i>	dihydropyrimethanopterin synthase	+	+	+
<i>folE</i>	GTP cyclohydrolase	+	+	+
<i>folC</i>	dihydrofolate synthase	+	+	+
<i>orf5</i>	H ₄ MPT biosynthesis	+	+	-
<i>orf7</i>	Unknown	+	+	-
orf9	H₄MPT biosynthesis	+	+	-
<i>orf19</i>	H ₄ MPT biosynthesis	+	+	-
<i>orf20</i>	H ₄ MPT biosynthesis	+	+	-
<i>orf21</i>	H ₄ MPT biosynthesis	+	+	-
<i>orf22</i>	H ₄ MPT biosynthesis	+	+	-
mch	methenyl-H₄MPT cyclohydrolase	+	+	-
<i>dfrA</i>	dihydrofolate reductase	+	+	+
Formate oxidation				
<i>fdh1A</i>	tungsten-dependent formate dehydrogenase, alpha subunit	-	+	-
<i>fdh1B</i>	tungsten-dependent formate dehydrogenase, beta subunit	-	+	-
<i>fdh2A</i>	molybdenum-dependent formate dehydrogenase, alpha subunit	+	+	+
<i>fdh2B</i>	molybdenum-dependent formate dehydrogenase, beta subunit	+	+	+
<i>fdh2C</i>	molybdenum-dependent formate dehydrogenase, gamma subunit	+	+	+
<i>fdh2D</i>	molybdenum-dependent formate dehydrogenase, delta subunit	+	+	+
<i>fdh3A</i>	cytochrome-linked formate dehydrogenase, alpha subunit	+	?	?
<i>fdh3B</i>	cytochrome-linked formate dehydrogenase, beta subunit	+	-	-
<i>fdh3C</i>	cytochrome-linked formate dehydrogenase, gamma subunit	+	-	-
C1-assimilation: serine cycle				
sga	serine-glyoxylate aminotransferase	+	+	+
hpr	hydroxypyruvate reductase	+	+	+
gck	glycerate kinase	+	+	+
<i>eno</i>	enolase	+	+	+
ppcA	phosphoenolpyruvate carboxylase	+	+	+
<i>mdh</i>	malate dehydrogenase	+	+	+
mtkA	malate thiokinase, alpha subunit	+	+	?
mtkB	malate thiokinase, beta subunit	+	+	?
mcl	malyl-CoA lyase	+	+	?
<i>glyA</i>	serine hydroxymethyltransferase	+	+	+
<i>qscR</i>	serine cycle transcriptional regulator	+	-	-
C1-assimilation: EMC pathway, glyoxylate regeneration				
<i>phaA</i>	alpha-ketothiolase	+	+	+
<i>phaB</i>	acetoacetyl-CoA reductase	+	+	+
<i>phaC</i>	poly-beta-hydroxybutyrate polymerase	+	+	+
<i>phaR</i>	regulator of PHB synthesis and acetyl-CoA flux	+	+	+
<i>phaZ</i>	polyhydroxyalkanoate depolymerase	+	+	+
<i>icl</i>	isocitrate lyase	+	+	+
<i>croR</i>	crotonase	-	-	-
<i>ccr</i>	crotonyl-CoA reductase	-	-	-
<i>meaB</i>	essential for methylmalonyl-CoA mutase reaction	-	-	-
<i>meaC</i>	mesaconyl-CoA hydratase	-	-	-
<i>ibd</i>	methylsuccinyl-CoA dehydrogenase	-	-	-
<i>mclA2</i>	malyl-CoA lyase/beta-methylmalyl-CoA lyase	-	-	-
<i>ecm</i>	ethylmalonyl-CoA mutase	-	-	-
<i>pccA</i>	propionyl-CoA carboxylase, alpha subunit	+	+	+
<i>pccB</i>	propionyl-CoA carboxylase, beta subunit	+	+	+
<i>mcmA</i>	methylmalonyl-CoA mutase, alpha subunit	+	+	+

<i>mcmB</i>	methylmalonyl-CoA mutase, beta subunit	+	+	+
<i>epm</i>	methylmalonyl-CoA epimerase	+	+	+
<i>bdhA</i>	3-hydroxybutyrate dehydrogenase	+	+	+
<i>aad</i>	acetoacetate decarboxylase	+	+	+
<i>aas</i>	acetoacetyl-CoA synthetase	+	+	+
<i>sdhA</i>	succinate dehydrogenase flavoprotein subunit	+	+	+
<i>sdhB</i>	succinate dehydrogenase iron-sulfur binding subunit	+	+	+
<i>sdhC</i>	succinate dehydrogenase cytochrome B556 subunit	+	+	+
<i>sdhD</i>	succinate dehydrogenase membrane anchor subunit	+	+	+
<i>fumC</i>	fumarate hydratase	+	+	+
Nitrogen Metabolism				
<i>haoA</i>	hydroxylamine dehydrogenase	-	-	-
<i>nirK</i>	nitrite reductase, copper-containing	+	-	-
<i>nirS</i>	cytochrome cd1 nitrite reductase	-	-	-
<i>cytL</i>	cytochrome P460, nitric oxide : hydroxylamine dehydrogenase	+	-	-
<i>cytS</i>	putative nitric oxide reductase	-	-	-
<i>norB</i>	nitric oxide reductase, heme-copper oxidase subunit I	+	-	-

Supplementary Table 4 Amount of genes in some different COG categories in the three genomes, along with the proportions of putative laterally transferred genes in each calculated by a BLAST procedure at the taxonomic level of Order (from IMG) or Family. An asterisk indicates that the proportion of genes in a COG, or the proportion of predicted LGT in that COG, is significantly higher in *M. silvestris* or *B. indica* than in *M. acidiphila*, calculated with a Z-ratio test of proportions and a Bonferroni correction for three comparisons.

Organism	All protein-coding genes ^a	Carbohydrate transport and metabolism (G)	Energy production and conversion (C)	Signal transduction mechanisms (T)	Inorganic ion transport and metabolism (P)	Cell wall/ envelope/ membrane biogenesis (M)	Transcription (K)	Average other COGs ^b
<i>M. silvestris</i>								
Total genes	3917	129	251	154	213	183	188	1646
% LGT Order level	12.2*	7.8	8.8	7.1	11.7	4.3	13.8	6.6
% LGT Family level		27.2*	36.4*	34.2	34.4*	15.4	38.9	
<i>M. acidiphila</i>								
Total genes	3700	98	208	130	165	198	171	1688
% LGT Order level	7.75	7.1	3.4	4.6	6.7	10.1	10.5	6.2
% LGT Family level		13.6	20.3	23.8	22.4	21.3	28.3	
<i>B. indica</i>								
Total genes	3788	203*	259	127	199	199	226*	1593
% LGT Order level	17.8*	20.2*	18.5*	16.5*	16.1*	17.6	23.9*	9.2*
% LGT Family level		59.3*	42.2*	34.9	37.0*	36.5*	55.2*	

^a The numbers in individual COG categories do not sum to the total because not all the LGT genes are classified into COGs. Numbers do not include non-coding reference genes.

^b Comprises the 13 COG groups: Amino acid transport and metabolism (E); Coenzyme transport and metabolism (H); Cell cycle control, cell division, chromosome partitioning (D); Cell motility (N); Defense mechanisms (V); Function unknown (S); Intracellular trafficking, secretion, and vesicular transport; Lipid transport and metabolism (U); Nucleotide transport and metabolism (F); Posttranslational modification, protein turnover, chaperones (O); Replication, recombination and repair (L); Secondary metabolites biosynthesis, transport and catabolism (Q); and Translation, ribosomal structure and biogenesis (J).

Supplementary Table 5 Taxonomic distribution of the top BLAST hits of genes from Table 3 and Supplementary Table 4 that are predicted to be acquired via LGT.

Organism	Total predicted genes via LGT	other Rhizobiales	Rhodospirillales	Rhodobacteriales	Caulobacteriales	Sphingomonadales	Deltaproteobacteria	Betaproteobacteria (Burkholderiales)	Gammaproteobacteria	Firmicutes	Acidobacteria	Cyanobacteria	Thermus	Thermotoga
Transporters														
<i>M. silvestris</i>	15	10	2	0	0	0	0	2	0	1	0	0	0	0
<i>M. acidiphila</i>	4	1	1	0	0	0	0	0	1	0	0	0	1	0
<i>B. indica</i>	43	11	13	0	0	3	3	10	1	0	0	0	0	1
COG G: Carbohydrate transport and metabolism														
<i>M. silvestris</i>	33	20	1	1	1	0	0	3	1	1	2	0	0	0
<i>M. acidiphila</i>	13	5	3	0	0	0	0	1	1	1	1	0	1	0
<i>B. indica</i>	111	57	20	0	2	2	0	14	7	0	0	0	0	1
COG K: Transcription														
<i>M. silvestris</i>	72	38	14	0	2	1	0	5	6	0	1	2	0	0
<i>M. acidiphila</i>	47	16	7	2	2	2	6	3	4	0	1	1	0	0
<i>B. indica</i>	122	59	23	2	3	7	0	17	6	0	0	2	0	0

Supplementary Table 6 A section of the *B. indica* plasmid NC_010580 containing genes for Carbohydrate transport and metabolism (COG G), Amino acid transport and metabolism (COG E) and Energy production and conversion (COG C).

Locus Tag	gene description	COG category	detected as foreign via IMG BLAST
Bind_3793	extracellular solute-binding protein family 1	G	
Bind_3794	binding-protein-dependent transport systems inner membrane component	G	
Bind_3795	binding-protein-dependent transport systems inner membrane component	G	
Bind_3796	ABC transporter related	G	
Bind_3797	Mannitol dehydrogenase domain (EC:1.1.1.67)	G	
Bind_3800	sorbitol dehydrogenase small subunit (EC:1.1.99.21)	-	x
Bind_3801	glucose-methanol-choline oxidoreductase	E	x
Bind_3802	Gluconate 2-dehydrogenase (acceptor)	C	x
Bind_3803	putative insertion sequence transposase-like protein	L	
Bind_3804	hypothetical protein	-	
Bind_3805	transketolase (EC:2.2.1.1)	G	
Bind_3806	transaldolase (EC:2.2.1.2)	G	
Bind_3807	xylulokinase (EC:2.7.1.17)	G	
Bind_3808	transcriptional regulator, AraC family	K	
Bind_3809	hypothetical protein	-	x
Bind_3810	Alcohol dehydrogenase GroES domain protein (EC:1.1.1.9)	E	
Bind_3811	periplasmic binding protein/LacI transcriptional regulator	G	

Supplementary Table 7 The total amount of LGT in the three genomes predicted by different methods. The letters indicate significant differences among the three genomes in the proportion of LGT at $P=0.05$, calculated with a Z-ratio test of proportions and a Bonferroni correction for three comparisons.

Organism	ORFs	LGT genes predicted via BLAST at the order level	LGT genes predicted by Alien Hunter	LGT genes predicted by IslandViewer
<i>M. silvestris</i>	3917	481 ^A	545 ^A	209 ^A
<i>M. acidiphila</i>	3700	287 ^B	660 ^B	179 ^A
<i>B. indica</i>	3788	678 ^C	641 ^B	165 ^A

Supplementary Table 8 COG distributions of the genes predicted in islands identified by IslandViewer. Most genes were hypothetical and not assigned to any COG category. The letters indicate significant differences among the three genomes for genes in each COG category at $P=0.05$, calculated with a Z-ratio test of proportions and a Bonferroni correction for three comparisons.

Organism	All genes in Islands	Carbohydrate transport and metabolism (G)	Energy production and conversion (C)	Signal transduction mechanisms (T)	Inorganic ion transport and metabolism (P)	Cell wall/ envelope/ membrane biogenesis (M)	Transcription (K)
<i>M. silvestris</i>	209	4 ^{AB}	10 ^A	0 ^A	3 ^A	0 ^A	6 ^A
<i>M. acidiphila</i>	179	0 ^A	0 ^B	2 ^A	1 ^A	10 ^B	15 ^{AB}
<i>B. indica</i>	165	9 ^B	8 ^A	3 ^A	3 ^A	7 ^B	20 ^B

Supplementary Table 9 Some potential islands of LGT in the *B. indica* genome, identified as a set of contiguous genes all predicted as foreign by at least one method. Islands predicted primarily by AlienHunter alone are not shown because of the imprecision of this method in targeting individual genes. An x indicates a putative foreign gene. COG categories G, C, T, P, M and K, where most LGT in *B. indica* are evident, are in bold (See Figure 2). For genes without an assigned COG, or assigned to COG R (general function prediction only) KOGs are also given in parentheses where identified.

Locus Tag	gene	COG category	IMG BLAST	IslandViewer	AlienHunter
Bind_0087	glycosyltransferase sugar-binding region containing DXD motif	M	x		x
Bind_0088	glycosyl transferase family 2	-			x
Bind_0089	NADH:flavin oxidoreductase/NADH oxidase	C	x		
Bind_0090	NAD-dependent epimerase/dehydratase	G	x		
Bind_0503	4Fe-4S ferredoxin, iron-sulfur binding	-	x		
Bind_0504	cyclic nucleotide-binding protein	T	x		
Bind_0505	oxidoreductase FAD/NAD(P)-binding domain protein	C	x		
Bind_0506	NADH ubiquinone oxidoreductase 20 kDa subunit	C	x		
Bind_0507	nickel-dependent hydrogenase large subunit	C	x		
Bind_0508	hydrogenase maturation protease	C	x		
Bind_0619	transcriptional regulator, LysR family	K	x		x
Bind_0620	short-chain dehydrogenase/reductase SDR(EC:1.1.1.69)	I			x
Bind_0621	hypothetical protein	-	x		x
Bind_0622	D-isomer specific 2-hydroxyacid dehydrogenase NAD-binding	C	x		x
Bind_0623	putative 2-keto-4-pentenoate hydratase-like protein	Q	x		x
Bind_0624	transcriptional regulator, LacI family	K			x
Bind_0625	carbohydrate kinase, thermoresistant glucokinase family	G	x		x
Bind_0626	oxidoreductase domain protein	R (G, Q)	x		x
Bind_0627	major facilitator superfamily MFS_1	G	x		x
Bind_0791	transcriptional regulator, LysR family	K	x		
Bind_0792	NAD-dependent epimerase/dehydratase	G	x		
Bind_0793	hypothetical protein	-	x		
Bind_0838	glycosyl transferase family 2	R	x		
Bind_0839	glycosyl transferase group 1	M	x		
Bind_0840	glycosyl transferase group 1	M	x		
Bind_0841	glycosyl transferase group 1	M	x		
Bind_0842	hypothetical protein	-	x		x
Bind_0843	hypothetical protein	-			x
Bind_0844	glycosyl transferase family 2	R	x		x
Bind_1824	Glutathione S-transferase domain	O	x		x
Bind_1825	short-chain dehydrogenase/reductase	R (Q)			x
Bind_1826	alpha/beta hydrolase fold	R (I)	x		x
Bind_1827	transcriptional regulator, MarR family	K	x	x	x
Bind_1828	NmrA family protein	G		x	x
Bind_1829	transcriptional regulator, AraC family	K	x	x	x
Bind_1830	short-chain dehydrogenase/reductase SDR	R (Q)	x	x	x
Bind_1832	protein of unknown function	S		x	x
Bind_1833	transcriptional regulator, LysR family	K	x	x	x

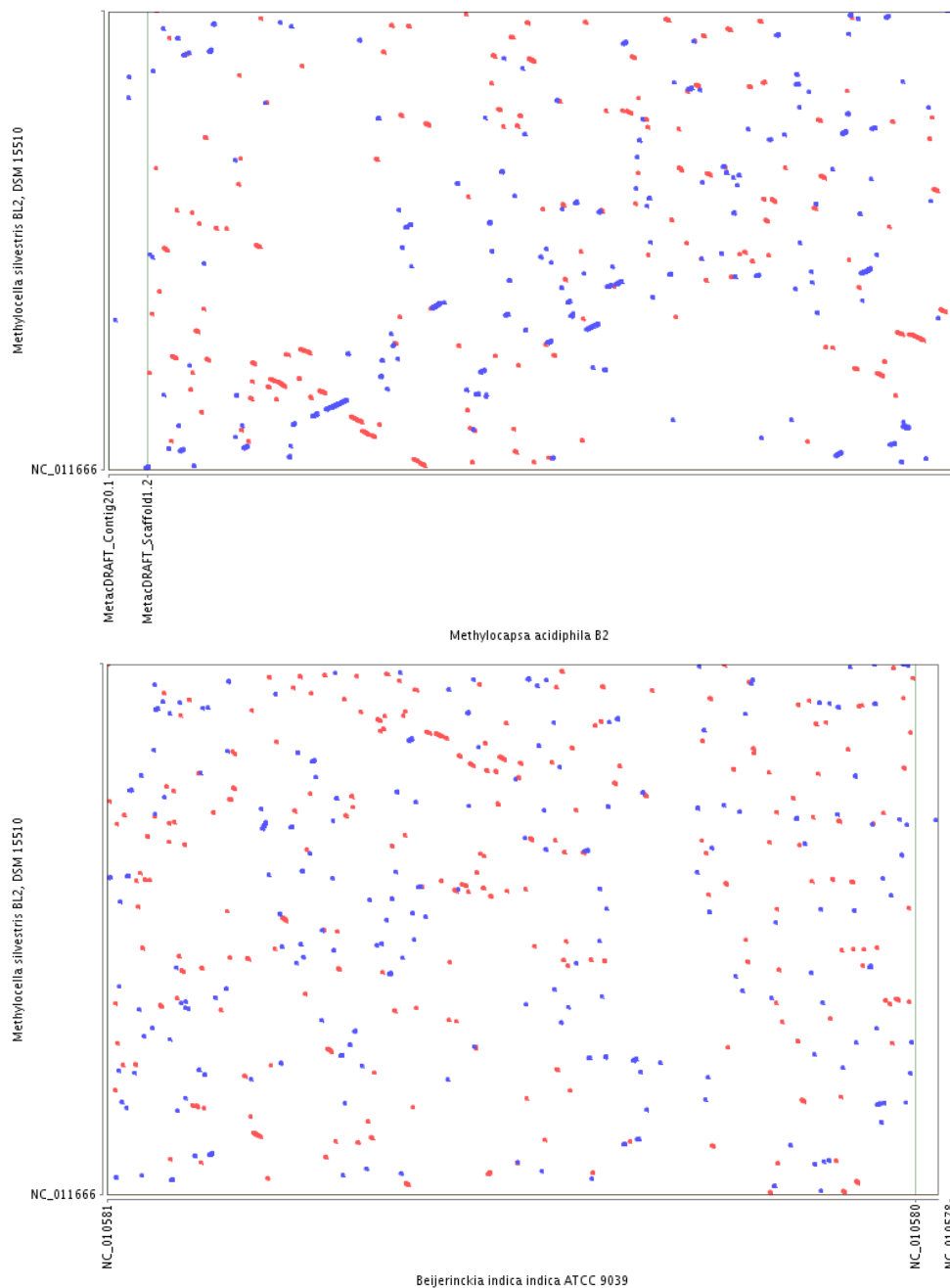
Bind_1834	NADH:flavin oxidoreductase/NADH oxidase	C	x	x
Bind_1835	Alcohol dehydrogenase zinc-binding domain protein	C	x	x
Bind_1881	Isoprenylcysteine carboxyl methyltransferase	O	x	
Bind_1882	multi-sensor signal transduction histidine kinase	T	x	
Bind_1883	response regulator receiver protein	T	x	
Bind_1884	response regulator receiver sensor signal transduction histidine kinase	T	x	
Bind_1932	flavoprotein WrbA	R (R)	x	
Bind_1933	Ycel family protein	S	x	
Bind_1934	cytochrome B561	C	x	
Bind_1935	Ycel family protein	S	x	
Bind_1937	Pirin domain protein	R	x	
Bind_1966	molybdopterin-guanine dinucleotide biosynthesis protein B	H	x	
Bind_1967	ABC-type nitrate/sulfonate/bicarbonate transport systems periplasmic component-like protein	P	x	
Bind_1968	response regulator receiver and ANTAR domain protein	T	x	
Bind_1969	molybdopterin oxidoreductase (EC:1.8.1.2)	R, P (C)	x	
Bind_1970	nitrite reductase (NAD(P)H), large subunit (EC:1.7.1.4)	C	x	
Bind_1971	major facilitator superfamily MFS_1	P	x	
Bind_2004	putative ABC-type Fe³⁺ transport system periplasmic component	P	x	x
Bind_2005	transport system permease protein	P	x	
Bind_2006	ABC transporter related	H	x	
Bind_2015	polysaccharide biosynthesis protein	R (-)	x	x
Bind_2016	FAD linked oxidase domain protein	C		x
Bind_2017	Cyclopropane-fatty-acyl-phospholipid synthase	M		x
Bind_2018	4-oxalocrotonate tautomerase	-		x
Bind_2019	Mn²⁺/Fe²⁺ transporter, NRAMP family	P		x
Bind_2020	Levanase	G	x	x
Bind_2021	Levansucrase	-	x	
Bind_2150	hypothetical protein	-	x	x
Bind_2151	outer membrane efflux protein	M	x	x
Bind_2152	efflux transporter, RND family, MFP subunit	M	x	
Bind_2153	heavy metal efflux pump, CzcA family	P	x	
Bind_2498	flavin reductase domain protein, FMN-binding	R (-)	x	
Bind_2499	transcriptional regulator, LysR family	K	x	
Bind_2500	major facilitator superfamily MFS_1	-	x	
Bind_2501	secretion protein HlyD family protein	V	x	
Bind_2502	RND efflux system, outer membrane lipoprotein, NodT family	M, U	x	
Bind_2503	transcriptional regulator, PadR-like family	K	x	
Bind_2505	short-chain dehydrogenase/reductase SDR	R (-)	x	
Bind_2506	ThiJ/Pfpl domain protein	R (-)	x	
Bind_2507	transcriptional regulator, LuxR family	K	x	x
Bind_2508	short-chain dehydrogenase of various substrate specificities	R (Q)	x	x
Bind_2510	Na/Picotransporter	P		x
Bind_2512	hypothetical protein	S	x	x
Bind_2513	hypothetical protein	-	x	x
Bind_2514	putative transcriptional regulator	-	x	x
Bind_2516	NAD-dependent epimerase/dehydratase	G	x	x
Bind_2519	ABC transporter related	G	x	x
Bind_2520	binding-protein-dependent transport systems inner membrane component	P	x	x
Bind_2521	extracellular solute-binding protein family	P	x	x

Bind_2522	hypothetical protein	-	x	x
Bind_2523	chromate transporter, chromate ion transporter (CHR) family	P		x
Bind_2530	amino acid permease-associated region	E	x	x
Bind_2531	Siderophore synthetase component-like protein	Q		x
Bind_2532	putative nickel ABC transporter, permease protein	R (-)		x
Bind_2533	MaoC domain protein dehydratase	I	x	x
Bind_2534	Citryl-CoA lyase	G	x	x
Bind_2535	major facilitator superfamily MFS_1	G	x	x
Bind_2536	L-carnitine dehydratase/bile acid-inducible protein F	F		x
Bind_2537	transcriptional regulator, LysR family	K		x
Bind_2538	prevent-host-death family protein	-		x
Bind_2539	PilT protein domain protein	R (-)		x
Bind_2540	transcriptional regulator	K		x
Bind_2541	Malate/L-lactate dehydrogenase	C	x	x
Bind_2542	major facilitator superfamily MFS_1	G		x
Bind_2543	3-carboxymuconate cyclase-like protein	G	x	x
Bind_2544	porin	-		x
Bind_2545	isocitrate dehydrogenase, NADP-dependent	R (C)		x
Bind_2546	extracellular solute-binding protein family	E		x
Bind_2573	cytochrome-c peroxidase	P	x	x
Bind_2576	major facilitator superfamily MFS_1	P	x	x
Bind_2577	hypothetical protein	-	x	x
Bind_2578	hypothetical protein	-	x	x
Bind_2579	metal-dependent hydrolase	-	x	x
Bind_2580	regulatory protein MerR	C	x	x
Bind_2582	protein of unknown function	-	x	x
Bind_2583	putative insertion sequence	L		x
Bind_2584	hypothetical protein	-	x	x
Bind_2585	short-chain dehydrogenase/reductase	R (Q)		x
Bind_2586	heme oxygenase family	H		x
Bind_2587	AMP-dependent synthetase and ligase	I		x
Bind_2588	hypothetical protein	-		x
Bind_2589	two component transcriptional regulator	K	x	x
Bind_2590	integral membrane sensor signal	T	x	x
Bind_2591	hypothetical protein	-		x
Bind_2594	VacJ family lipoprotein	M	x	x
Bind_2595	hypothetical protein	S		x
Bind_2596	tetratricopeptide TPR_2 repeat protein	N		x
Bind_2597	Methyltransferase type 11	H		x
Bind_2598	hypothetical protein	-		x
Bind_2599	hypothetical protein	S	x	x
Bind_2600	tetratricopeptide TPR_2 repeat protein	N		x
Bind_2601	aminotransferase class-III	E	x	x
Bind_2602	3-Oxoacyl-(acyl-carrier-protein (ACP))	Q	x	x
Bind_2603	monooxygenase FAD-binding	C	x	x
Bind_2604	Transposase and inactivated derivatives-like protein	L		x
Bind_2606	hypothetical protein	-	x	x
Bind_2607	hypothetical protein	-		x
Bind_2609	transcriptional regulator, AraC family	K	x	x
Bind_2611	transcriptional regulator, AraC family	K	x	x
Bind_2612	aldo/keto reductase	C	x	x
Bind_2613	NAD-dependent epimerase/dehydratase	G	x	x
Bind_2614	NAD-dependent epimerase/dehydratase	G	x	x
Bind_2615	X-Pro dipeptidyl-peptidase domain protein	R (-)	x	x
Bind_2619	Transposase and inactivated derivatives-like protein	L		x
Bind_2620	transcriptional regulator, AraC family	K	x	x
Bind_2622	FMN-dependent alpha-hydroxy acid dehydrogenase	C		x
Bind_2623	ThiJ/Pfpl domain protein	R (V, R)	x	x
Bind_2624	Alkanesulfonate monooxygenase	C		x

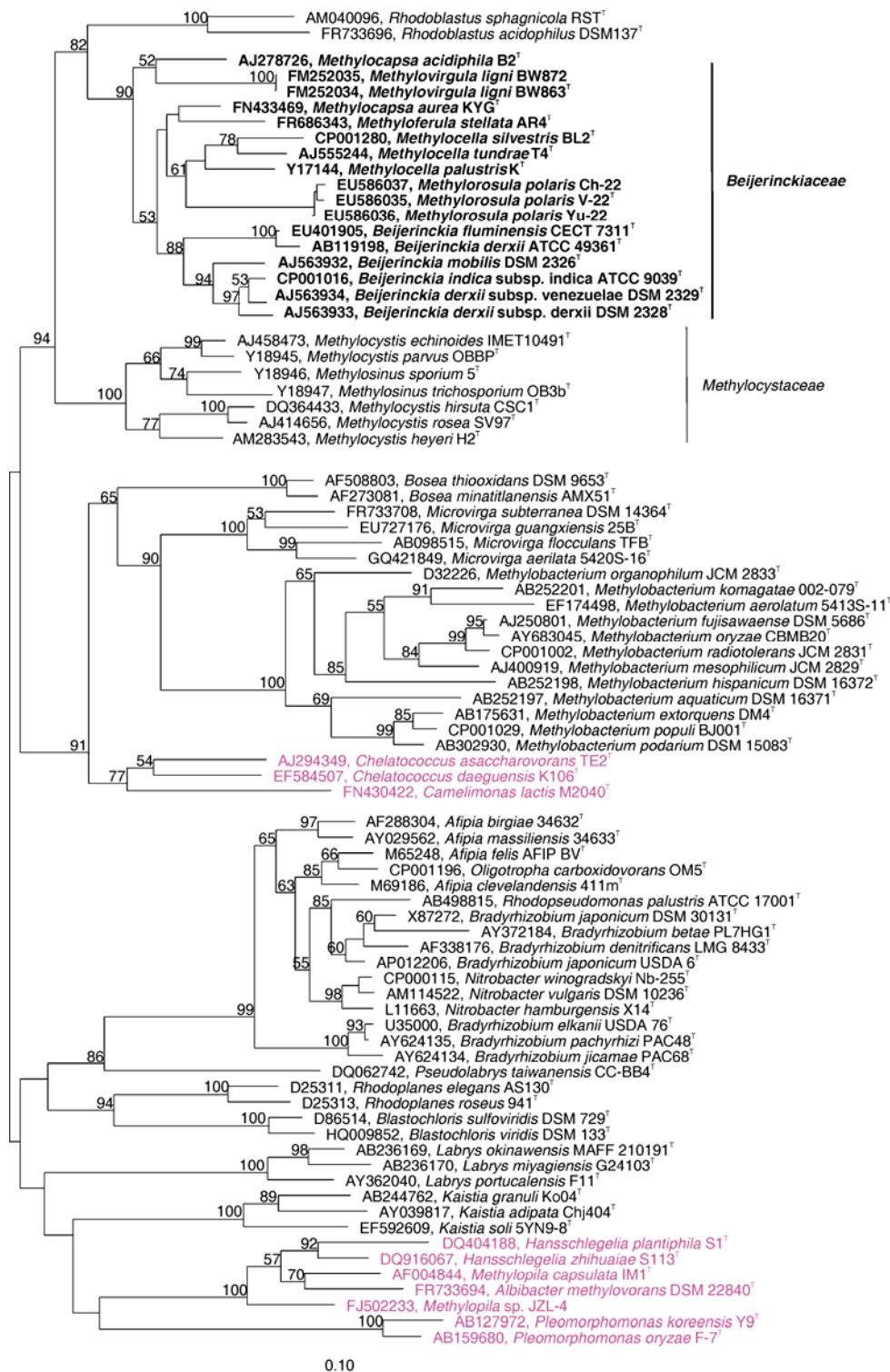
Bind_2625	flavin reductase domain protein FMN-binding	R, C	x	x
Bind_2626	5-carboxymethyl-2-hydroxymuconate Delta-isomerase	Q	x	x
Bind_2627	3,4-dihydroxy-2-butanone 4-phosphate synthase	H	x	x
Bind_2628	extracellular ligand-binding receptor	E	x	x
Bind_2629	hypothetical protein	-	x	x
Bind_2633	Alcohol dehydrogenase zinc-binding domain protein	C	x	x
Bind_2634	major facilitator superfamily MFS_1	-(G)	x	x
Bind_2635	transcriptional regulator, LysR family	K	x	x
Bind_2637	peptidase M10A and M12B matrixin and adamalysin	-(O, W)	x	x
Bind_2639	nucleotide sugar dehydrogenase	M	x	x
Bind_2640	transcriptional regulator, ModE family	R (-)	x	x
Bind_2641	hypothetical protein	-	x	x
Bind_2642	mannose-1-phosphate	M	x	x
Bind_2643	Endo-1,4-beta-xylanase	G	x	x
Bind_2644	capsular exopolysaccharide family	M	x	x
Bind_2645	glycosyl transferase group 1	M	x	x
Bind_2647	hypothetical protein	-	x	x
Bind_2648	glycosyl transferase, WecB/TagA/CpsF family	M	x	x
Bind_2650	glycosyl transferase group 1	M	x	x
Bind_2651	acyltransferase 3	I	x	x
Bind_2653	polysaccharide biosynthesis protein	R (-)	x	x
Bind_2655	acyltransferase 3	G	x	x
Bind_2656	NAD-dependent epimerase/dehydratase	G	x	x
Bind_2657	hypothetical protein	-	x	x
Bind_2658	hypothetical protein	-	x	x
Bind_2661	hypothetical protein	-	x	x
Bind_2665	integrase family protein	L	x	x
Bind_2666	YcfA family protein	N	x	x
Bind_2667	protein of unknown function	S	x	x
Bind_2756	Galactose oxidase	-	x	x
Bind_2757	glycosyl transferase group 1	M		x
Bind_2758	hypothetical protein	-		x
Bind_2759	glycoside hydrolase family 5	G	x	x
Bind_2760	hypothetical protein	-	x	x
Bind_2761	hypothetical protein	-		x
Bind_2762	lipopolysaccharide biosynthesis protein	M		x
Bind_2763	hypothetical protein	-	x	x
Bind_2764	hypothetical protein	R		x
Bind_2775	putative H(+)-transporting ATP synthase, gene 1 protein	-	x	
Bind_2776	ATP synthase F0, A subunit(EC:3.6.3.14)	C	x	
Bind_2777	ATP synthase F0, C subunit(EC:3.6.3.14)	C	x	
Bind_2827	oxidoreductase domain protein	R (G, Q)	x	
Bind_2828	Inositol 2-dehydrogenase(EC:1.1.1.18)	R (G, Q)	x	
Bind_2829	D-isomer specific 2-hydroxyacid dehydrogenase NAD-binding (EC:1.1.1.95)	E, H	x	
Bind_2830	HpcH/Hpal aldolase(EC:4.1.2.-)	G	x	
Bind_2831	Aldehyde Dehydrogenase(EC:1.2.1.16)	C	x	
Bind_2832				
Bind_2833	transcriptional regulator, RpiR family	K	x	
Bind_2834	thiamine pyrophosphate protein central region (EC:3.7.1.-)	E	x	
Bind_2835	Xylose isomerase domain protein TIM barrel (EC:4.2.1.44)	G	x	
Bind_2961	Alcohol dehydrogenase GroES domain protein	C	x	
Bind_2962	transcriptional regulator, LysR family	K	x	
Bind_2963	aldo/keto reductase	C	x	x
Bind_2964	beta-lactamase domain protein	R	x	x
Bind_2965	hypothetical protein	-		x
Bind_2966	hypothetical protein	-		x

Bind_2967	transcriptional regulator, LysR family	K		x
Bind_2973	transcriptional regulator, LuxR family	K	x	x
Bind_2974	autoinducer synthesis protein	Q	x	x
Bind_2975	Phytanoyl-CoA dioxygenase	Q	x	x
Bind_2976	transcriptional regulator, LuxR family	K	x	x
Bind_2977	2OG-Fe(II) oxygenase	R (Q, R)	x	x
Bind_2979	hypothetical protein	E	x	x
Bind_2980	transcriptional regulator, LysR family	K	x	x
Bind_2981	transcriptional regulator, TetR family	K	x	x
Bind_2982	hypothetical protein	-		x
Bind_2983	hypothetical protein	-		x
Bind_2984	protein of unknown function	S		x
Bind_2985	hypothetical protein	-		x
Bind_2986	isochorismatase hydrolase	Q		x
Bind_2987	inorganic pyrophosphatase	C		x
Bind_2989	alpha/beta hydrolase fold	R (R)	x	x
Bind_2990	short-chain dehydrogenase/reductase SDR	R (Q)		x
Bind_2991	hypothetical protein	-		x
Bind_2992	short-chain dehydrogenase/reductase SDR	I		x
Bind_2993	hypothetical protein	-	x	x
Bind_2994	hypothetical protein	-	x	x
Bind_2995	protein of unknown function	R (-)		x
Bind_2996	hypothetical protein	-		x
Bind_2997	integral membrane sensor signal transduction histidine kinase	T		x
Bind_2998	two component, sigma54 specific,transcriptional regulator, Fis family	T		x
Bind_3000	hypothetical protein	-		x
Bind_3001	hypothetical protein	-		x
Bind_3002	Arsenical pump membrane protein	P		x
Bind_3003	hypothetical protein	S	x	x
Bind_3004	major facilitator superfamily MFS_1	G		x
Bind_3078	Aldehyde Dehydrogenase	C		x
Bind_3079	major facilitator superfamily MFS_1	G	x	x
Bind_3080	dihydrodipicolinate synthetase	M, E		x
Bind_3081	Dihydroxy-acid dehydratase	G, E		x
Bind_3388	hypothetical protein	-	x	
Bind_3389	thiamine pyrophosphate protein TPP binding domain protein	G	x	
Bind_3390	major facilitator superfamily MFS_1	G	x	
Bind_3431	Glucan 1,4-alpha-glucosidase	G	x	
Bind_3432	FMN-binding domain protein	-	x	x
Bind_3433	ApbE family lipoprotein	H	x	x
Bind_3434	putative transmembrane protein	-	x	x
Bind_3435	conserved hypothetical secreted protein	-	x	x

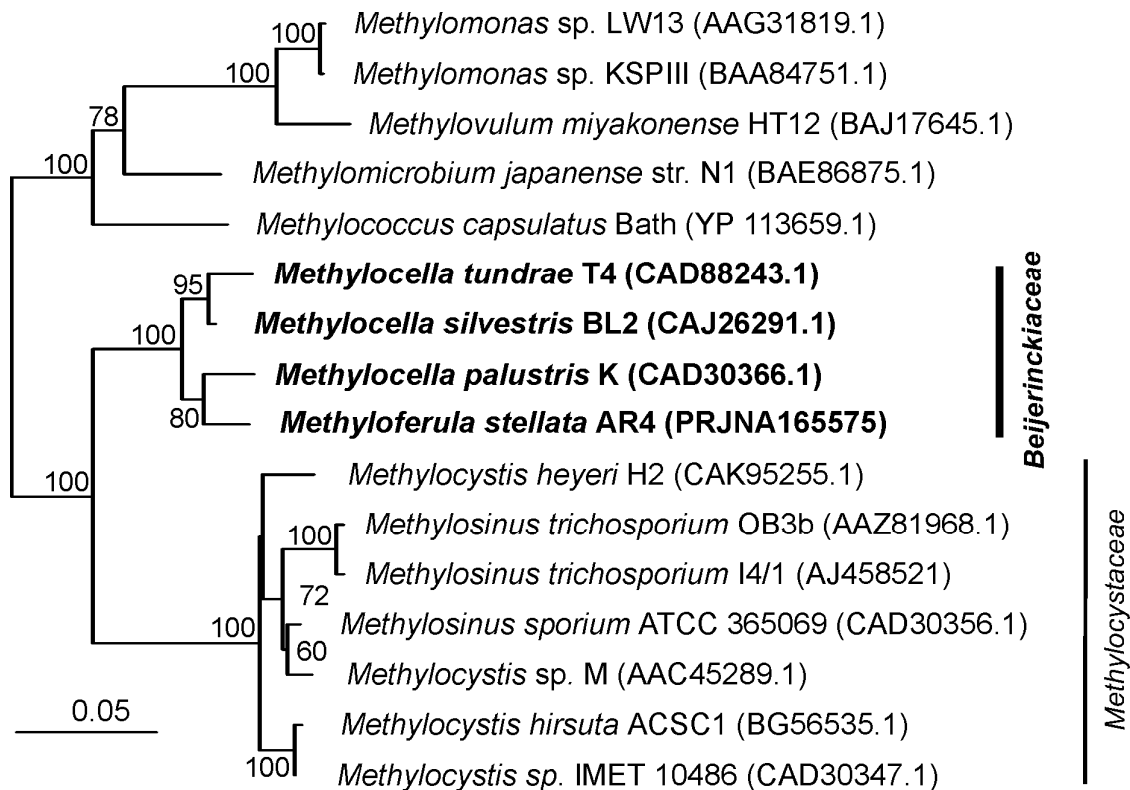
Supplementary Figure 1 Synteny dot plots of *M. silvestris* versus *M. acidiphila* and *M. silvestris* versus *B. indica* generated using the Promer program in IMG.



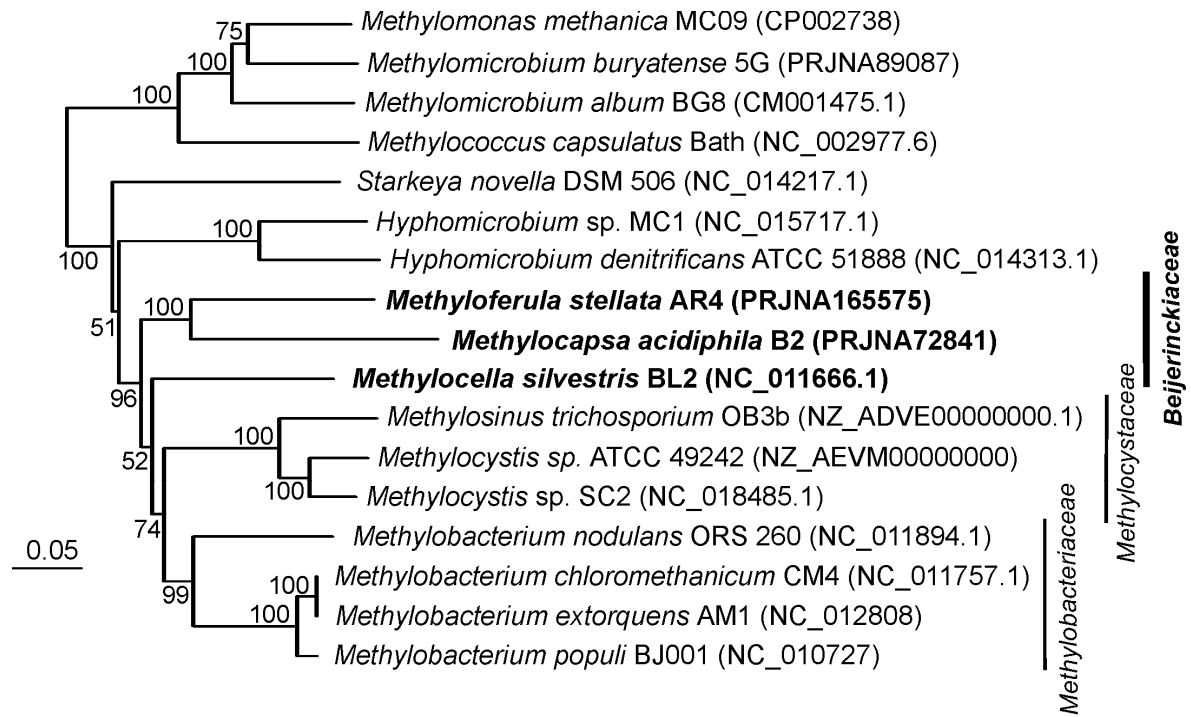
Supplementary Figure 2 Unrooted 16S rRNA gene-based tree showing the relationships of *Beijerinckiaceae* and *Methylocystaceae* to other *Rhizobiales*. The tree was constructed in the ARB platform from nearly full length sequences (>1400 nt) via Neighbour-joining with a Jukes-Cantor correction (Ludwig *et al.*, 2004). A consensus tree of 1000 bootstrap replicates is shown, only bootstrap values greater than 50% are indicated. The scale bar represents 0.1 change per nucleotide site. Species that have been incorrectly classified as members of the *Beijerinckiaceae* or *Methylocystaceae* are shown in pink. Sequence alignments were taken from the ARB-Silva reference database Version 111 (Quast *et al* 2013).



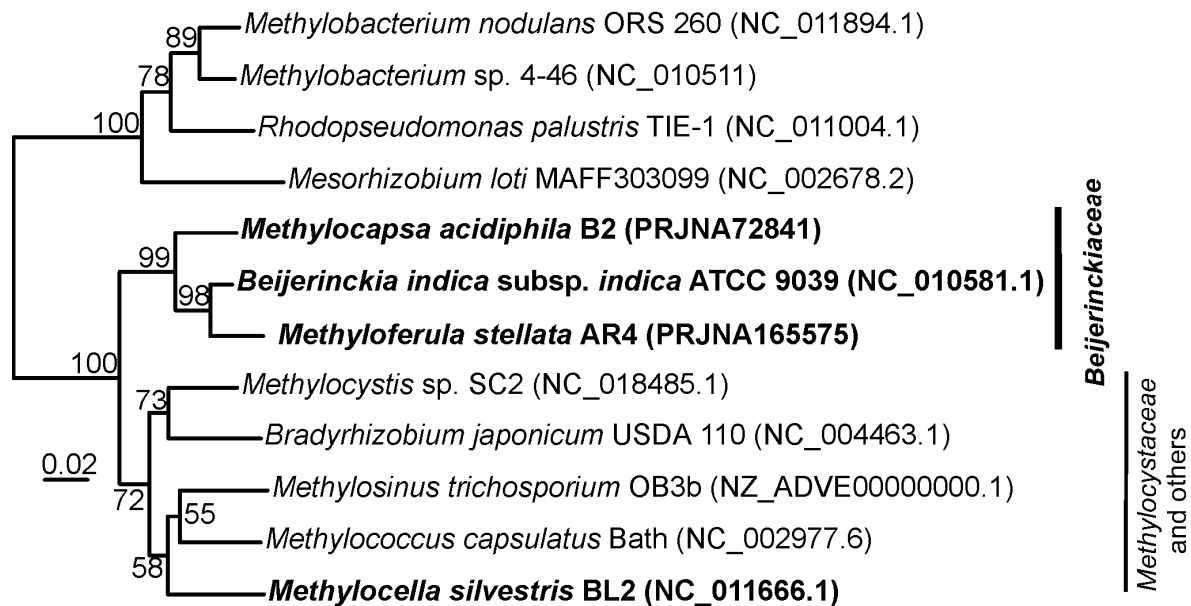
Supplementary Figure 3 Phylogenetic tree showing the relationship of derived MmoX proteins, the hydroxylase component of sMMO. The tree was constructed using Neighbour-Joining and a Kimura distance correction, based on 390 amino acids total length. The scale bar represents 0.05 changes per amino acid position. Bootstrap values of 1000 constructions are shown at the nodes when greater than 50%. Gene accession numbers are given in parentheses.



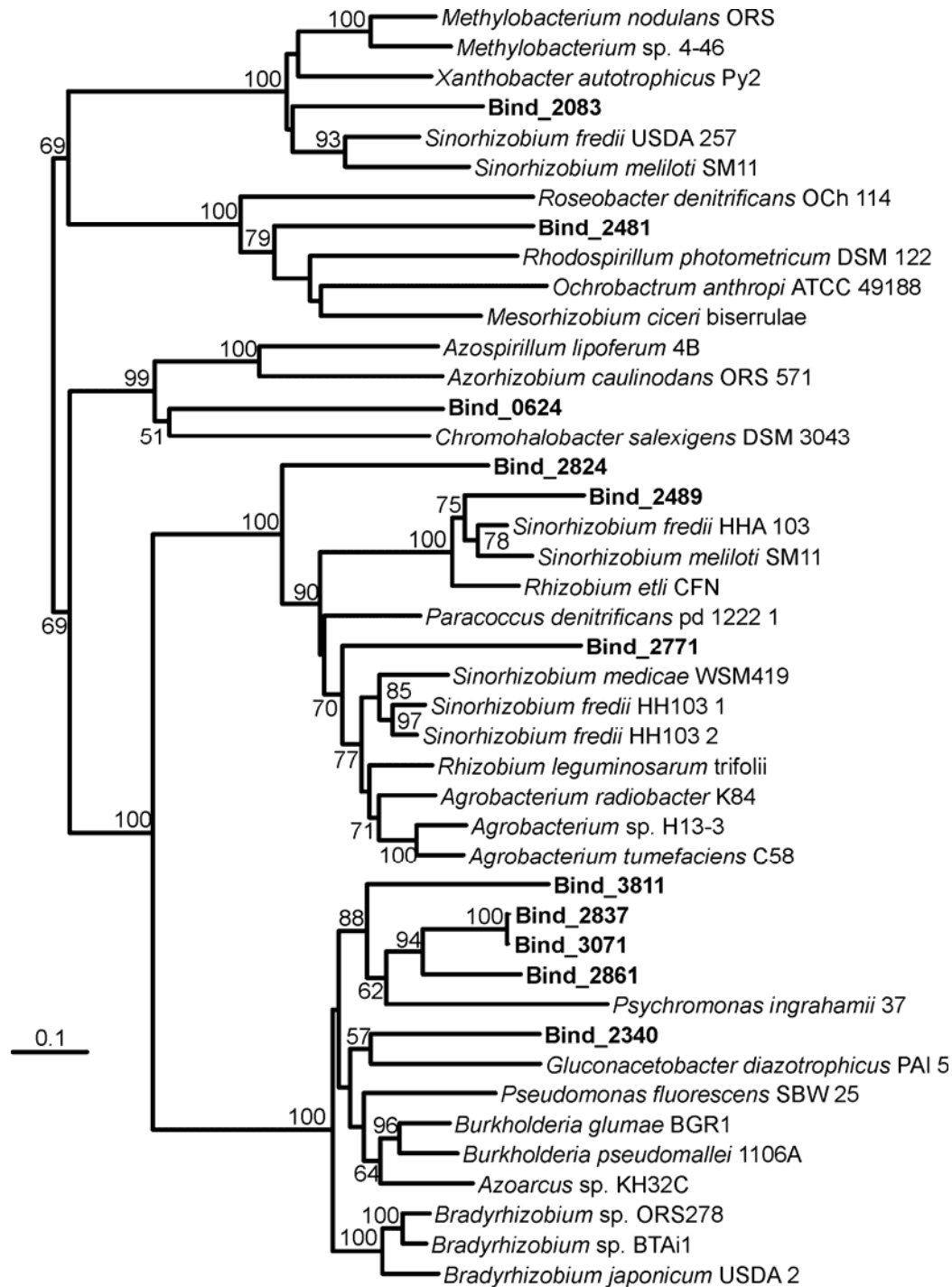
Supplementary Figure 4 Phylogenetic tree showing the relationship of concatenated derived MxaFJI proteins. The tree was constructed using Neighbour-Joining and a Kimura distance correction, based on 851 amino acids total length. The scale bar represents 0.05 changes per amino acid position. Bootstraps values of 1000 constructions are shown at the nodes. Genome accession numbers are given in parentheses.



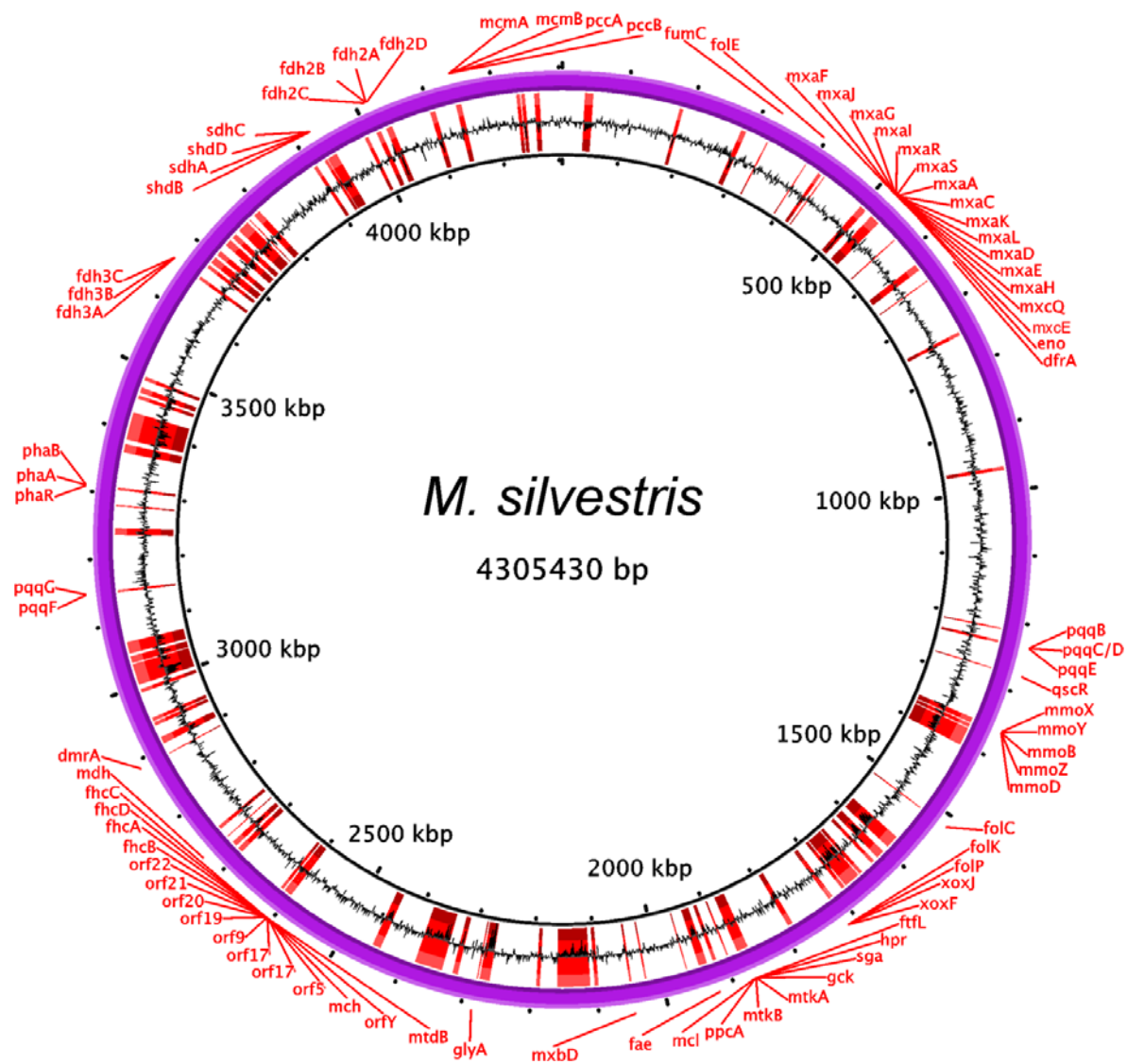
Supplementary Figure 5 Phylogenetic tree showing the relationship of concatenated derived products of *nifH* and *nifD*. The tree was constructed using Neighbour-Joining and a Kimura distance correction, based on a concatenation of 290 amino acids total length. The scale bar represents 0.02 changes per amino acid position. Bootstraps values of 1000 constructions are shown at the nodes. Genome accession numbers are given in parentheses.

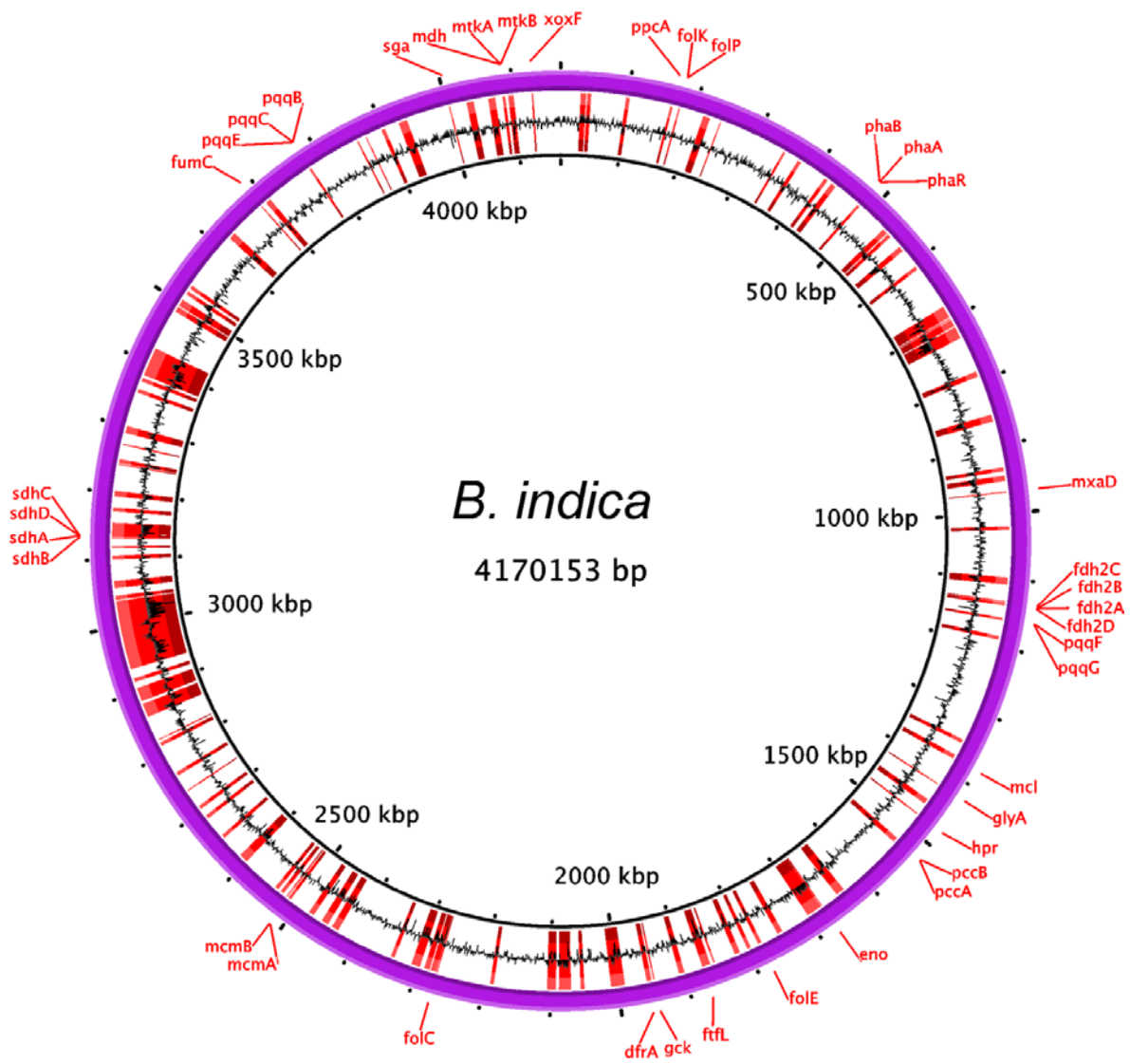


Supplementary Figure 6 Phylogenetic tree of derived proteins from genes identified as periplasmic binding protein/LacI transcriptional regulators. Proteins derived from the *B. indica* genome (in bold) are compared to the closest public-domain sequences identified via BLAST. The tree was constructed using Neighbour-Joining and a Kimura distance correction, based on 538 amino acids. The scale bar represents 0.1 changes per amino acid position. Bootstraps values of 1000 constructions are shown at the nodes when >50%.



Supplementary Figure 7 Methylo-trophy islands and potential LGT islands in the three genomes. Methylo-trophy genes as described in Supplementary Table 3 are indicated on the outer rings. On the inner rings GC content is shown and shaded areas represent islands identified by IslandViewer or AlienHunter. The figure was drawn using Ring Image Generator (Alikhan *et al.*, 2011). Methylo-trophy genes were identified by BLAST, therefore they do not necessarily represent genes for methylo-trophy functions, only homologues of these.





Supplementary Figure 8 Clustering of some genes involved in the serine cycle of C1 assimilation in some *Alphaproteobacteria*. Clustered genes directly involved in the serine cycle (orange color) are *sga*, serine cycle glyoxylate aminotransferase; *hpr*, hydroxypyruvate reductase; *mtkA*, malate thiokinase, β large subunit; *mtkB*, malate thiokinase, α small subunit; *ppc*, phosphoenolpyruvate carboxylase; *mcl*, malyl-CoA lyase; *gck*, glycerate kinase. Clustered genes involved indirectly in the serine cycle by activating formate (blue color) are *mtdA*, methylene- H_4 MPT/methylene- H_4 F dehydrogenase; *fch*, methenyl- H_4 F cyclohydrolase; *ftL*, formyl- H_4 F ligase. Genes indicated by gray are not known to be related to the serine cycle.

