Supplementary Materials

Spiribacter halobius sp. nov., a novel halophilic Gammaproteobacterium with a relatively large genome

Ya Gong [1,2], Lu Ma [1], Zhao-Zhong Du [1], Wei-Shuang Zheng [3], De-Chen Lu [1] & Zong-Jun Du [1,2,*]

[a] Marine College, Shandong University, Weihai, Shandong, 264209, PR China.

[b] State key Laboratory of Microbial Technology, Shandong University, Qingdao, Shandong, 266237, PR China.

[c] PKU-HKUST Shenzhen-Hong Kong Institution, Shenzhen, Guangdong, 518057, China.

Contents

Figure S1, S2, S3, S4, S5, S6, and S7

Table S1, S2, S3, S4, S5, and S6

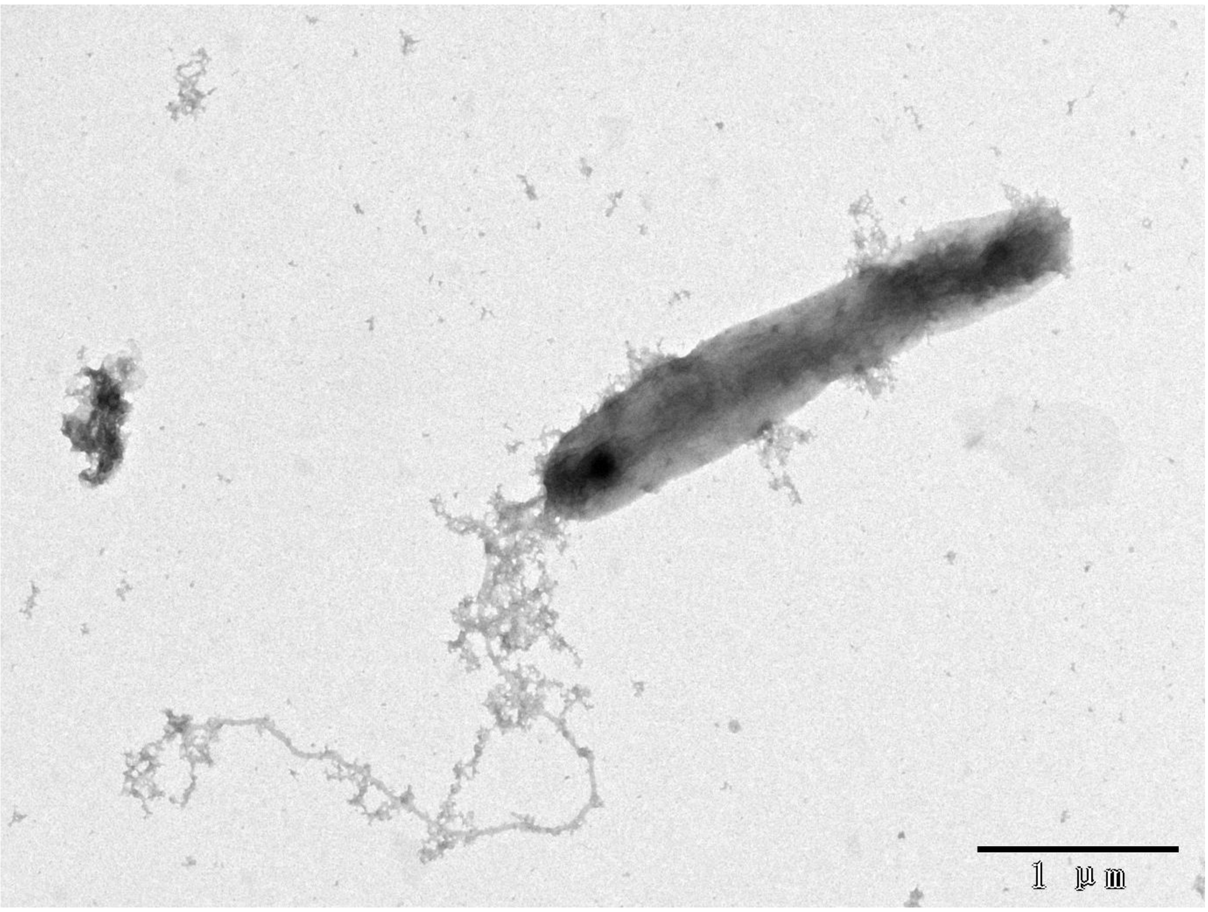

Figure S1 Transmission electron micrograph of E85^T. Cells grew aerobically on modified MA plate at 37°C for 3 days. Bar, 1 µm.

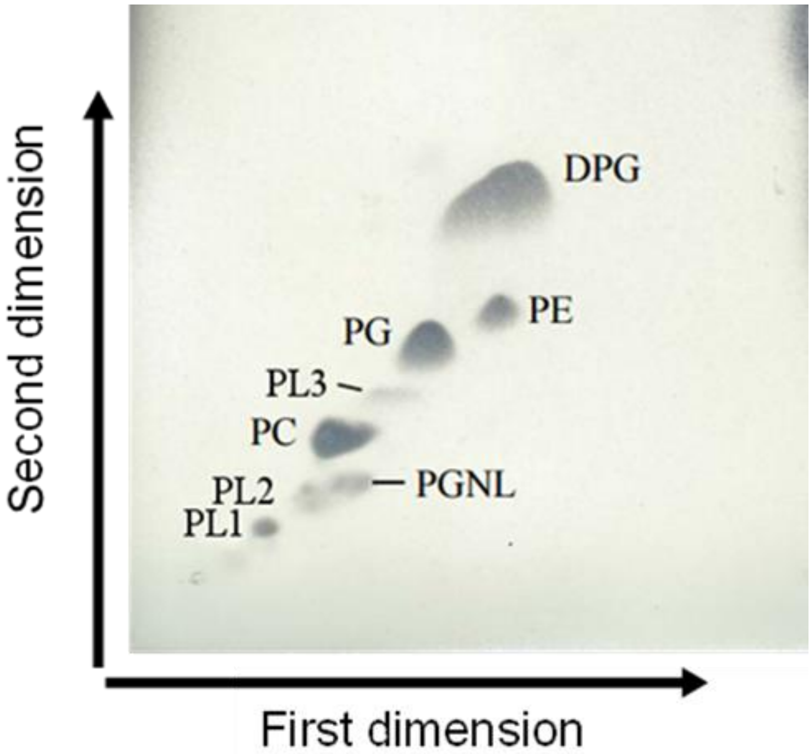

Figure S2 Two-dimensional thin-layer chromatogram of polar lipids of strain E85^T. PC, phosphatidylcholine; PG, phosphatidylglycerol; PE, phosphatidylethanolamine; DPG, diphosphatidylglycerol; PGNL, phosphoaminoglycolipid; PL, phospholipid.

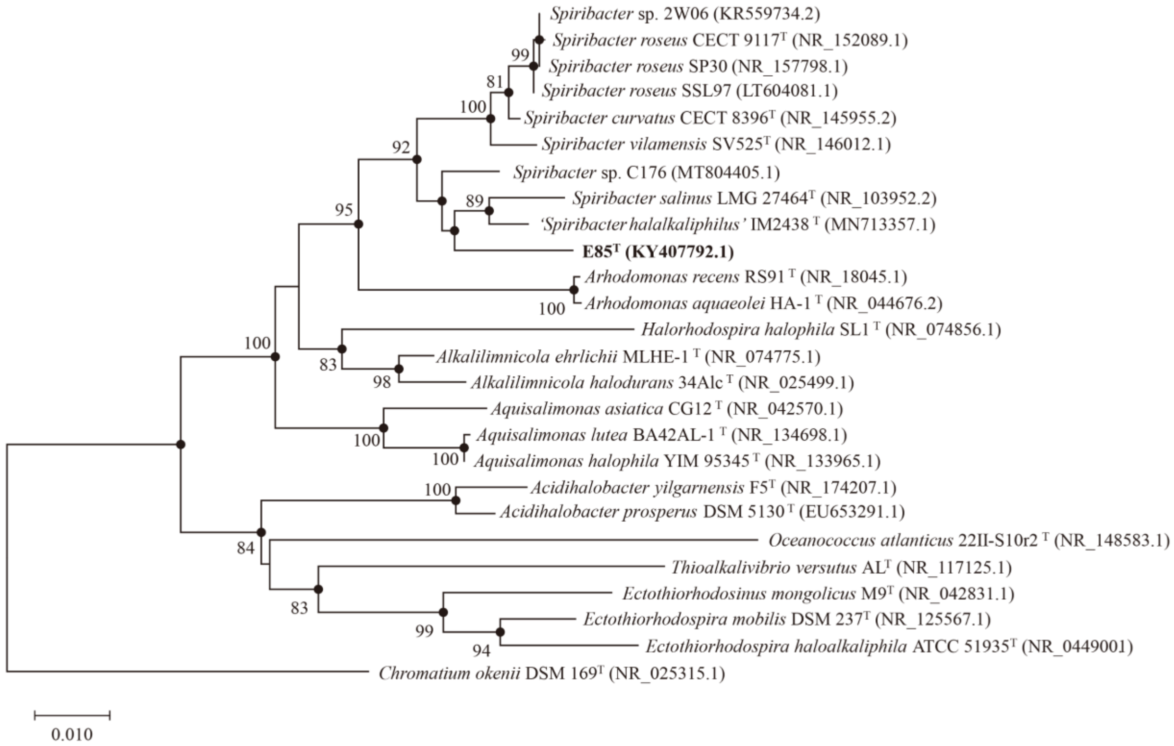

Figure S3 Neighbour-joining phylogenetic tree based on the 16S rRNA gene sequences showing the phylogenetic position of strain E85^{T} among members of the genus *Spiribacter*. Numbers on nodes represent bootstrap values (NJ) based on 1000 replications. Only bootstrap values higher than 70% are marked on the branches. The sequence of *Chromatium okenii* DSM 169^{T} (NR_025315.1) was used as an outgroup. Filled circles indicate nodes also obtained in both maximum-likelihood and maximum-parsimony trees. Bar, 0.01 substitutions per nucleotide position.

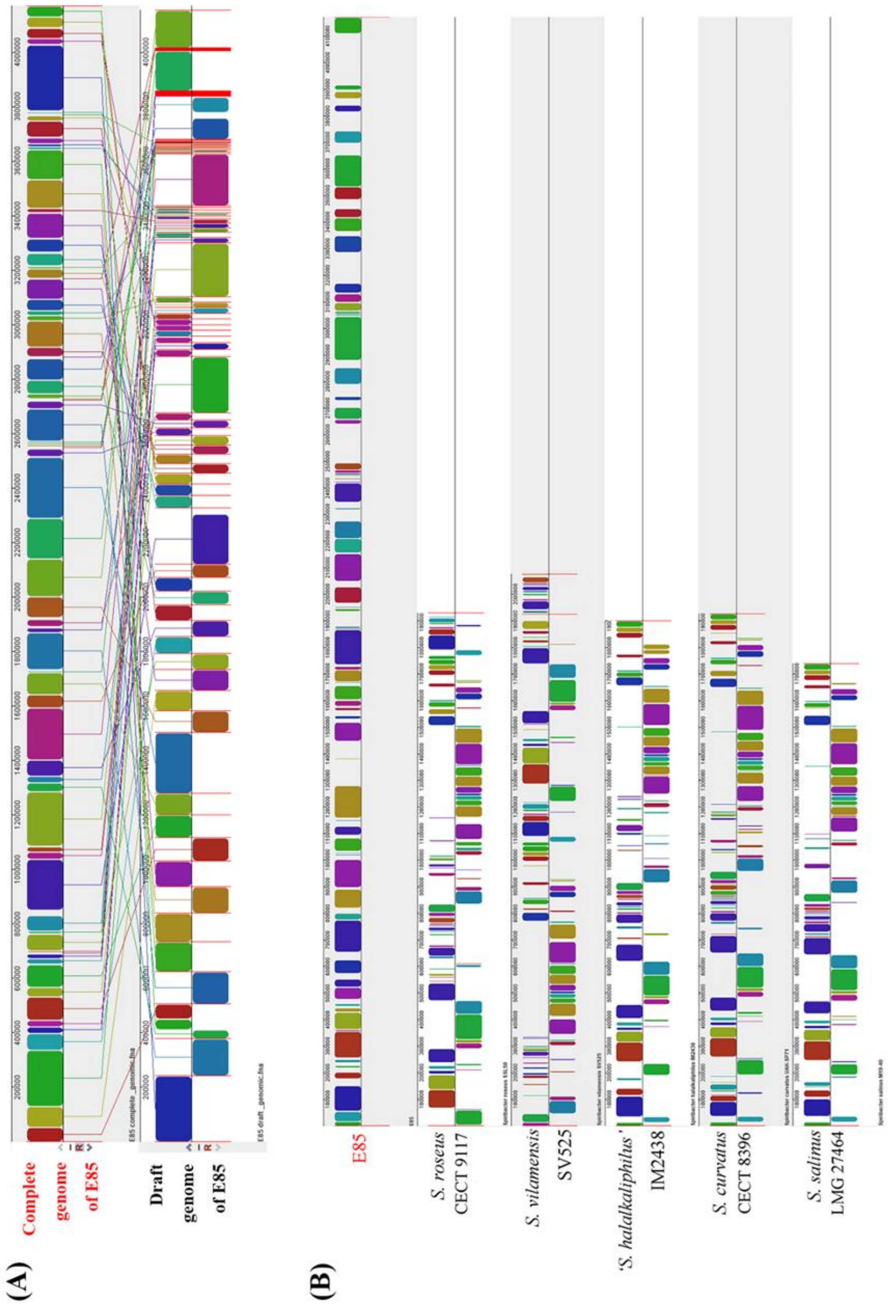

Figure S4 Multiple genome alignments of genomes of strain E85^T and closely related species. Each row is a genome and each colored block is genetically similar. When a block lies above the center line the aligned region is in the forward orientation relative to the first genome sequence. Blocks below the center line indicate regions that align in the reverse complement (inverse) orientation. **(A)** Whole-genome alignment of the draft and complete genome of strain E85^T. Lines connected similar colored blocks indicate which regions in each genome are homologous. **(B)** Multiple genome alignments of complete genomes of strain E85^T and other *Spiribacter* species. Regions outside blocks lack detectable homology among the input genomes.

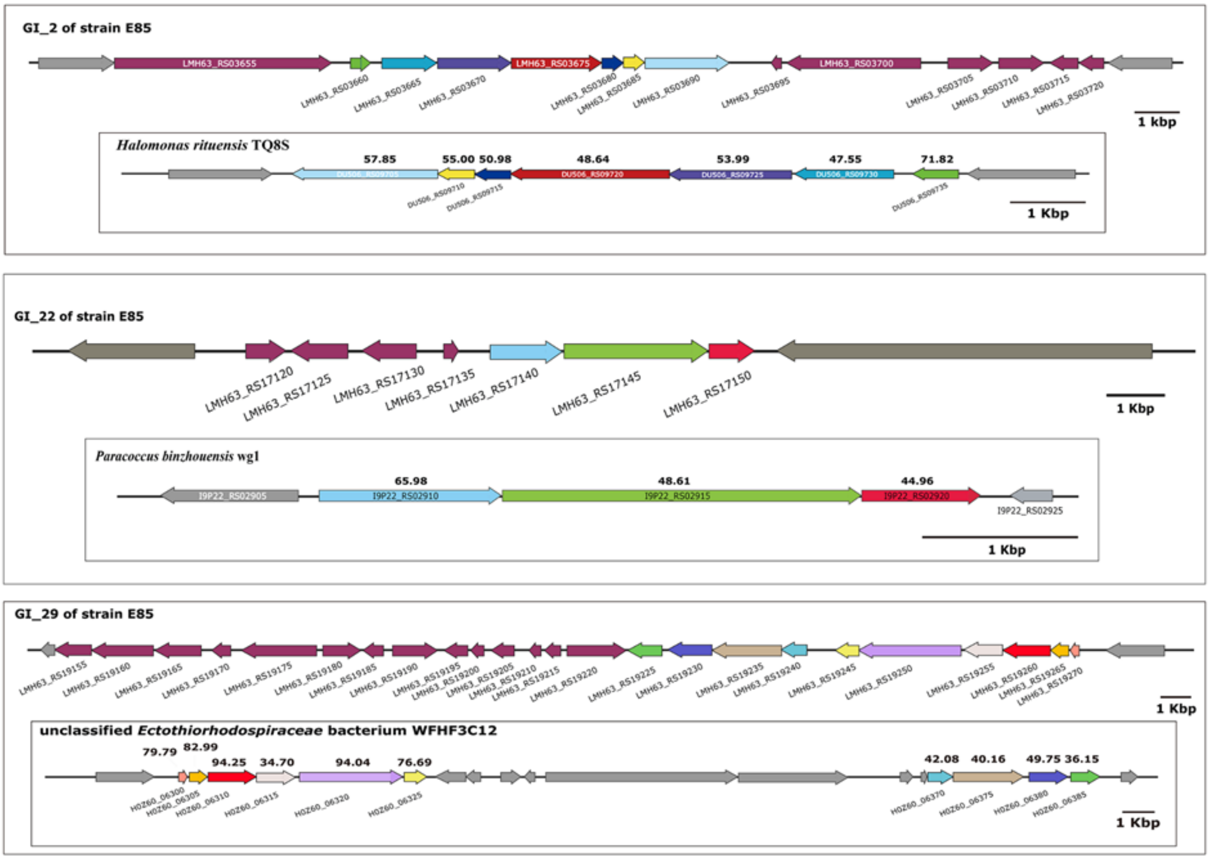

Figure S5 Comparative analysis of GIs, GI_2, GI_22, and GI_29, of strain E85^T and other bacterial species. Colors illustrate homologs as determined by BLASTp of the coding protein, with numbers designating % amino acid similarities.

Figure S6 Sample location of Yuncheng Salt Lakes. The six Salt Lake ponds are labeled as P170, P185, P205, P230, P300, and P340, respectively. The number in each pond means salinity. P170, 35°01'03.0" N 111°01'29.0" E; P185, 35°00'51.3" N 111°01'00.9" E; P205, 35°00'11.5" N 111°00'45.5" E; P230, 35°00'29.5" N 111°00'31.0" E; P300, 34°59'51.8" N 111°00'11.7" E; P340, 35°00'55.1" N 111°01'35.4" E.

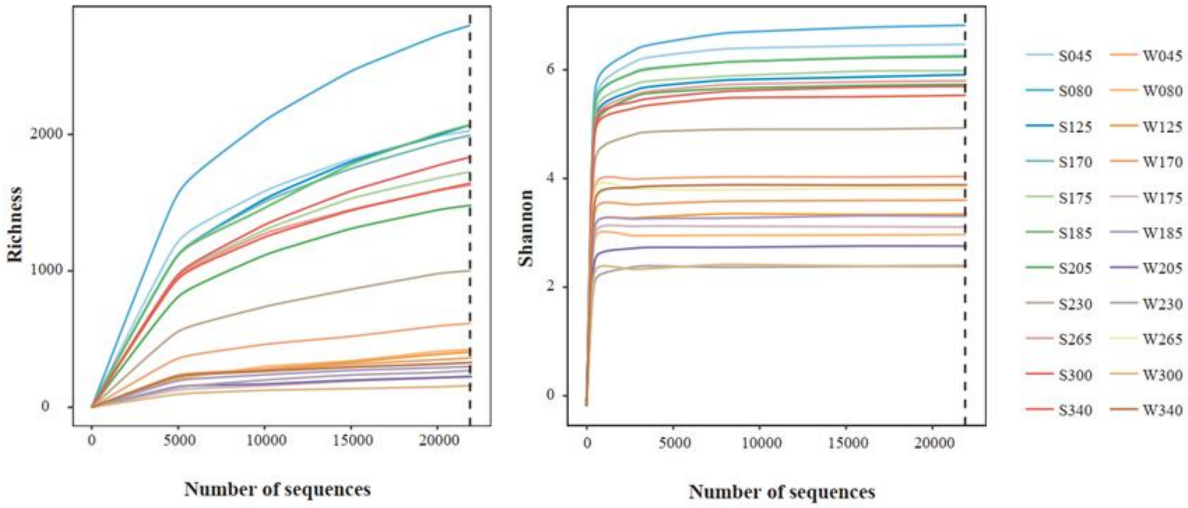

Figure S7 Rarefication curves based on richness and shannon index of each sample group.

Table S1. Fatty acid composition of strain E85^T and the related type strains.

Fatty acids	**E85^T**	***Spiribacter salinus*** **LMG 27464^T**	***Spiribacter curvatus*** **CECT 8396^T**	***Spiribacter roseus*** **CECT 9117^T**
$C_{10:0}$	ND	tr	tr	1.7
$C_{10:0}$ 3-OH	3.4	4.0	3.8	5.8
$C_{12:0}$	4.6	4.1	5.5	10.5
$C_{12:0}$ 3-OH	ND	tr	1.1	3.3
$C_{13:1}$	1.2	ND	ND	ND
anteiso-$C_{15:0}$	1.9	ND	ND	ND
iso-$C_{15:0}$	2.1	ND	ND	ND
$C_{16:0}$	**19.2**	**14.8**	**17.6**	**25.95**
$C_{18:0}$	4.2	2.8	2.4	2.27
$C_{18:1}\omega7c$ †	**36.8**	**68.6**	**47.6**	**18.3**
$C_{18:1}\omega7c$ 11-methyl	2.2	ND	tr	2.2
$C_{19:0}$ cyclo $\omega8c$	4.7	ND	5.0	**17.1**
$C_{20:1}\omega7c$	ND	ND	1.1	ND
Summed features*	ND	ND	ND	ND
2	6.2	tr	ND	ND
3	4.1	3.3	4.5	tr
7	3.7	ND	7.3	**11.1**
unknown 11.799	3.7	tr	1.5	tr

Values are percentages of the total fatty acids. Data were obtained from this study. Major fatty acids (> 10 %) are highlighted in bold. Only fatty acids exceeding 1.0 % of the total cellular fatty acids of at least one of the strains are shown. tr, trace (< 1.0 %); ND, not detected. *: Summed features represent groups of two or three fatty acids that could not be separated by GLC with the MIDI system. Summed feature 2, $C_{14:0}$ 3-OH and/or iso-$C_{16:1}$; summed feature 3, $C_{16:1}$ ω7c and/or iso-$C_{15:0}$ 2-OH; summed feature 7, $C_{19:1}$ $\omega7c$ and/or un 18.846. †: This corresponds to summed feature 8 ($C_{18:1}$ $\omega6c$ and/or $C_{18:1}$ $\omega7c$) in other studies.

Table S2. Differential characteristics of strain E85^T and closely related species.

Characteristic	1	2	3	4	5	6
Cell morphology	Straight to curved rod	Thin curved rod, spiral cell	curved rods or short spirals	Thin curved rod, spiral cell	Curved rod to nearly closed ring, short spiral cell	Rod
Cell size (μm)	0.2-0.3 × 1.2-4.5	0.3-0.8 × 1.8	0.2-0.3 × 0.5-1.5	0.5 × 1.5	0.2-0.3 × 1.6-1.8	0.5-0.8 × 1.0-2.0
Mobility	Single, polar flagellum	not-motile	not-motile	not-motile	not-motile	not-motile
Colony colour	Cream-white	Pink	Light purple	Dark pink	Pink	Pink
Requirement for O_2	Facultative aerobic	Aerobic	Aerobic	Aerobic	Aerobic	Aerobic
NaCl for growth (%, w/v)						
range	0.5-16	10-25	3-27	5-20	7.5-20	7.5-25
optimum	3-6	15	10-11	10	15	10-15
Temperature for growth (°C)						
range	20-50	15-40	20-40	5-40	15-40	10-35
optimum	37-40	37	37	37	37	25-30
Nitrate reduction	+	−	−	−	−	−
Citrate utilization	+	−	−	−*	−	−
Urease	+	+	+	−	+	−
Phosphatase	+	−	+	+	w	−
DNA G+C content (mol%)†	70.2 (69.7)	60.0 (62.7)	(64.2)	60.4 (63.9)	64.2 (66.0)	60.4 (64.2)

Strains: 1, E85^T; 2, *Spiribacter salinus* LMG 27464^T (León, et al. 2014); 3, '*Spiribacter halalkaliphilus*' IM2438^T (Xue, et al. 2021); 4, *Spiribacter curvatus* CECT 8396^T (León, et al. 2015); 5, *Spiribacter roseus* CECT 9117^T (León, et al. 2016); 6, *Spiribacter vilamensis* SV525^T (León, et al. 2020; Menes, et al. 2016). Data for strains 2-6 are taken from corresponding studies. +, Positive; –, negative; w, weakly positive. All the type strains are identical for motility (−), Voges-Proskauer reaction (−), indole production (−; not determined for *S. vilamensis* SV525^T), oxidase (+), catalase (+), and hydrolysis of starch (−) and Tween 80 (−). †: data in the bracket were calculated from genome sequence.

Table S3. Pairwise 16S rRNA gene sequence similarities, The average nucleotide identity (ANI) values, the digital DNA-DNA hybridization (dDDH) values, and the average amino-acid identity (AAI) values between E85^T and type species of genus *Spiribacter* (all values as percentages).

Strain	**1**	**2**	**3**	**4**	**5**	**6**
			16S rRNA gene sequence **similarity**			
1	100					
2	97.3	100				
3	96.2	98.4	100			
4	97.1	96.6	97.1	100		
5	96.2	96.5	97.0	99.4	100	
6	96.0	95.6	96.7	98.8	98.5	100
			ANI value			
1	100					
2	73.2	100				
3	73.5	74.4	100			
4	73.4	73.1	73.4	100		
5	74.5	74.3	74.0	80.9	100	
6	73.9	73.3	73.5	83.8	80.0	100
			dDDH value			
1	100					
2	19.2	100				
3	19.1	18.4	100			
4	19.3	17.9	18.1	100		
5	19.4	18.4	18.6	22.5	100	
6	20.1	17.6	18.0	26.7	22.0	100
			AAI value			
1	100					
2	70.8	100				
3	69.2	75.2	100			
4	69.2	74.0	72.4	100		
5	69.5	74.1	72.4	83.7	100	
6	68.8	73.4	72.2	87.6	82.7	100

Strains: 1, E85^T; 2, *Spiribacter salinus* LMG 27464^T; 3, '*Spiribacter halalkaliphilus*' IM2438^T ; 4, *Spiribacter curvatus* CECT 8396^T ; 5, *Spiribacter roseus* CECT 9117^T; 6, *Spiribacter vilamensis* SV525^T.

Table S4. Genomic information of strain E85^T and other *Ectothiorhodospiraceae* strains included in this study.

Strain	Size (bp)	GC%	Protein	rRNA	tRNA	Other RNA	Gene	Pseudogene	Assembly	Assembly level	NCBI Reference Sequence
Spiribacter sp. E85	4,172,350	69.7	3771	3	46	4	3874	50	ASM2088345v1	Complete Genome	NZ_CP086615.1
Spiribacter roseus CECT 9117	1,930,413	66	1860	3	45	4	1918	6	ASM281363v1	Complete Genome	NZ_CP016382.1
Spiribacter sp. SSL99	2,007,791	65.9	1911	3	45	4	1987	24	ASM980592v1	Contig	NZ_MAOM00000000.1
Spiribacter roseus SSL 97	2,062,548	65.7	1959	3	45	4	2039	28	ASM980588v1	Contig	NZ_MAOL00000000.1
Spiribacter roseus SSL25	1,876,226	65.7	1857	3	45	4	1966	57	ASM980585v1	Contig	NZ_MAOK00000000.1
Spiribacter roseus SSL4	1,992,107	66	1946	3	45	4	2018	20	ASM980584v1	Contig	NZ_MAOJ00000000.1
Spiribacter roseus SP30	1,990,024	66	1941	3	45	4	2017	24	ASM762521v1	Scaffold	NZ_VMKP00000000.1
Spiribacter sp. 2W06	2,132,736	64.7	1958	3	45	4	2043	33	ASM171528v2	Scaffold	NZ_MDVM00000000.2
Spiribacter curvatus CECT 8396	1,926,631	63.9	1845	3	45	5	1921	23	ASM48590v1	Complete Genome	NC_022664.1
Spiribacter vilamensis SV525	2,076,385	64.2	1985	3	45	4	2057	20	ASM762516v1	Scaffold	NZ_VMKO00000000.1
Spiribacter halalkaliphilus' IM2438	1,900,931	64.2	1828	3	45	4	1900	20	ASM967670v1	Complete Genome	NZ_CP046046.1
Spiribacter sp. T1Sed10_76	1,366,796	63.8	-	-	-	-	-	-	ASM355354v1	Contig	PWEP00000000.1
uncultured *Spiribacter* sp. SRR4030052_bin.5	1,867,007	65.5	-	-	-	-	-	-	SRR4030052_bin.5_MetaBAT_v2.12.1_MAG	Contig	CAJXMI000000000.1
Spiribacter salinus SL48-SHIP-2	4,021,157	61.3	3928	-	41	6	4101	126	ASM656919v1	Scaffold	VIFK00000000.1
Spiribacter salinus LMG 27464	1,739,487	62.7	1683	3	45	4	1744	9	ASM31957v2	Complete Genome	NC_021291.1
Spiribacter salinus M7-304	1,890,714	62.3	1771	3	45	4	1831	8	ASM1980062v1	Contig	NZ_MBFX00000000.1
Spiribacter sp. C176	2,235,242	54.1	2060	3	45	4	2152	40	ASM964922v1	Contig	NZ_WJPP00000000.1
Arhodomonas aquaeolei DSM 8974	3,957,769	69.4	3519	8	46	4	3618	41	ASM37464v1	Scaffold	NZ_ARGF00000000.1
Alkalilimnicola ehrlichii MLHE-1	3,275,944	67.5	2873	6	48	4	2964	33	ASM1478v1	Complete Genome	NC_008340.1
Halorhodospira halophila SL1	2,678,452	68	2429	6	46	4	2518	33	ASM1558v1	Complete Genome	NC_008789.1
Aquisalimonas asiatica CGMCC 1.6291	4,037,907	66	3662	3	46	5	3756	40	IMG-taxon 2617270766 annotated assembly	Scaffold	NZ_FOEG00000000.1
Oceanococcus atlanticus 22II-S10r2	3,652,004	60.6	3280	3	44	4	3346	15	ASM208823v1	Contig	NZ_AQQV00000000.1
Acidihalobacter prosperus DSM 5130	3,359,675	64.4	3206	3	45	4	3303	45	ASM75409v2	Contig	NZ_JQSG00000000.2

Strain	Size (bp)	GC%	Protein	rRNA	tRNA	Other RNA	Gene	Pseudogene	Assembly	Assembly level	NCBI Reference Sequence
Acidihalobacter yilgarnensis F5	3,566,941	59.9	3291	3	47	4	3478	133	ASM175324v1	Complete Genome	NZ_CP017415.1
Ectothiorhodosinus mongolicus M9	1,990,961	55.6	1820	3	44	4	1895	24	IMG-taxon 2681812864 annotated assembly	Contig	NZ_FTPK00000000.1
Ectothiorhodospira haloalkaliphila A	3,460,134	63	3105	4	46	4	3308	149	ASM58381v1	Complete Genome	NZ_CP007268.1
Ectothiorhodospira mobilis DSM 4180	2,624,954	68.2	2375	5	44	4	2489	61	IMG-taxon 2671180234 annotated assembly	Contig	NZ_FOUO00000000.1
Thioalkalivibrio versutus D301	2,969,361	66	2699	3	46	7	2805	50	ASM102095v1	Complete Genome	NZ_CP011367.1

Table S5. Annotation results of strain E85^{T} and other five *Spiribacter* species.

Strain	**E85^{T}**	***Spiribacter curvatus* CECT 8396^{T}**	**‘*Spiribacter halalkaliphilus*’ IM2438^{T}**	***Spiribacter roseus* CECT 9117^{T}**	***Spiribacter salinus* LMG 27464^{T}**	***Spiribacter vilamensis* SV525^{T}**
COG Category						
Information storage and processing						
COG-A (RNA processing and modification)	1	0	0	0	0	0
COG-B (Chromatin structure and dynamics)	5	3	3	3	3	3
COG-J (Translation, ribosomal structure and biogenesis)	191	158	161	158	162	160
COG-K (Transcription)	244	93	82	88	77	91
COG-L (Replication, recombination and repair)	172	110	122	112	101	119
Cellular processes and signaling						
COG-D (Cell cycle control, cell division, chromosome partitioning)	54	39	41	42	41	45
COG-M (Cell wall/membrane/envelope biogenesis)	186	139	118	140	122	156
COG-N (Cell motility)	85	23	17	19	14	23
COG-O (Posttranslational modification, protein turnover, chaperones)	139	83	86	87	84	82
COG-T (Signal transduction mechanisms)	211	67	50	63	48	77
COG-U (Intracellular trafficking, secretion, and vesicular transport)	75	73	60	69	62	72
COG-V (Defense mechanisms)	43	25	22	26	22	25
Metabolism						
COG-C (Energy production and conversion)	241	143	147	140	132	147
COG-E (Amino acid transport and metabolism)	302	177	170	176	176	171
COG-F (Nucleotide transport and metabolism)	88	63	64	63	63	65
COG-G (Carbohydrate transport and metabolism)	170	98	82	87	73	87
COG-H (Coenzyme transport and metabolism)	150	107	112	110	102	109
COG-I (Lipid transport and metabolism)	130	62	66	63	59	68
COG-P (Inorganic ion transport and metabolism)	211	114	105	113	95	120
COG-Q (Secondary metabolites biosynthesis, transport and catabolism)	107	43	37	38	36	38
Poorly characterized						
COG-S (Function unknown)	635	272	281	263	243	309
KEGG Module Completeness						
M00001 Glycolysis (Embden-Meyerhof pathway), glucose => pyruvate	complete	complete	complete	complete	complete	complete
M00002 Glycolysis, core module involving three-carbon compounds	complete	complete	complete	complete	complete	complete
M00003 Gluconeogenesis, oxaloacetate => fructose-6P	deficient	deficient	deficient	deficient	deficient	deficient
M00004 Pentose phosphate pathway (Pentose phosphate cycle)	deficient	deficient	deficient	deficient	deficient	deficient
M00005 PRPP biosynthesis, ribose 5P => PRPP	complete	complete	complete	complete	complete	complete
M00006 Pentose phosphate pathway, oxidative phase, glucose 6P => ribulose 5P	deficient	deficient	deficient	deficient	deficient	deficient
M00007 Pentose phosphate pathway, non-oxidative phase, fructose 6P => ribose 5P	complete	complete	complete	complete	complete	complete
M00008 Entner-Doudoroff pathway, glucose-6P => glyceraldehyde-3P + pyruvate	complete	complete	complete	complete	complete	complete
M00009 Citrate cycle (TCA cycle, Krebs cycle)	complete	complete	complete	complete	complete	complete
M00010 Citrate cycle, first carbon oxidation, oxaloacetate => 2-oxoglutarate	complete	complete	complete	complete	complete	complete
M00011 Citrate cycle, second carbon oxidation, 2-oxoglutarate => oxaloacetate	complete	complete	complete	complete	complete	complete

M00012 Glyoxylate cycle	deficient	complete	complete	complete	complete	complete
M00013 Malonate semialdehyde pathway, propanoyl-CoA => acetyl-CoA	deficient	deficient	deficient	none	deficient	deficient
M00014 Glucuronate pathway (uronate pathway)	deficient	deficient	deficient	deficient	deficient	deficient
M00015 Proline biosynthesis, glutamate => proline	complete	complete	complete	complete	complete	complete
M00016 Lysine biosynthesis, succinyl-DAP pathway, aspartate => lysine	complete	complete	complete	complete	complete	complete
M00017 Methionine biosynthesis, apartate => homoserine => methionine	deficient	deficient	deficient	deficient	deficient	deficient
M00018 Threonine biosynthesis, aspartate => homoserine => threonine	complete	complete	complete	complete	complete	complete
M00019 Valine/isoleucine biosynthesis, pyruvate => valine / 2-oxobutanoate => isoleucine	complete	complete	complete	complete	complete	complete
M00020 Serine biosynthesis, glycerate-3P => serine	complete	complete	complete	complete	complete	complete
M00021 Cysteine biosynthesis, serine => cysteine	deficient	deficient	deficient	deficient	deficient	deficient
M00022 Shikimate pathway, phosphoenolpyruvate + erythrose-4P => chorismate	complete	complete	complete	complete	complete	complete
M00023 Tryptophan biosynthesis, chorismate => tryptophan	complete	complete	deficient	complete	complete	deficient
M00024 Phenylalanine biosynthesis, chorismate => phenylpyruvate => phenylalanine	deficient	deficient	deficient	deficient	deficient	deficient
M00025 Tyrosine biosynthesis, chorismate => HPP => tyrosine	deficient	deficient	deficient	deficient	deficient	deficient
M00026 Histidine biosynthesis, PRPP => histidine	deficient	deficient	deficient	deficient	deficient	deficient
M00027 GABA (gamma-Aminobutyrate) shunt	deficient	deficient	deficient	deficient	deficient	deficient
M00028 Ornithine biosynthesis, glutamate => ornithine	complete	complete	complete	complete	complete	complete
M00029 Urea cycle	deficient	deficient	deficient	deficient	deficient	deficient
M00032 Lysine degradation, lysine => saccharopine => acetoacetyl-CoA	deficient	none	none	none	none	none
M00033 Ectoine biosynthesis, aspartate => ectoine	complete	deficient	deficient	deficient	deficient	deficient
M00034 Methionine salvage pathway	deficient	deficient	deficient	deficient	deficient	deficient
M00035 Methionine degradation	deficient	deficient	deficient	deficient	deficient	deficient
M00036 Leucine degradation, leucine => acetoacetate + acetyl-CoA	deficient	deficient	deficient	deficient	deficient	deficient
M00038 Tryptophan metabolism, tryptophan => kynurenine => 2-aminomuconate	deficient	none	none	none	none	none
M00040 Tyrosine biosynthesis, chorismate => arogenate => tyrosine	deficient	deficient	deficient	deficient	deficient	deficient
M00046 Pyrimidine degradation, uracil => beta-alanine, thymine => 3-aminoisobutanoate	deficient	none	none	none	none	none
M00048 Inosine monophosphate biosynthesis, PRPP + glutamine => IMP	complete	complete	complete	complete	complete	complete
M00049 Adenine ribonucleotide biosynthesis, IMP => ADP,ATP	complete	complete	complete	complete	complete	complete
M00050 Guanine ribonucleotide biosynthesis, IMP => GDP,GTP	complete	complete	complete	complete	complete	complete
M00051 Uridine monophosphate biosynthesis, glutamine (+ PRPP) => UMP	deficient	deficient	deficient	deficient	deficient	deficient
M00052 Pyrimidine ribonucleotide biosynthesis, UMP => UDP/UTP,CDP/CTP	complete	complete	complete	complete	complete	complete
M00053 Pyrimidine deoxyribonuleotide biosynthesis, CDP => dCTP	deficient	deficient	deficient	deficient	deficient	complete
M00060 KDO2-lipid A biosynthesis, Raetz pathway, LpxL-LpxM type	deficient	deficient	deficient	deficient	deficient	deficient
M00061 D-Glucuronate degradation, D-glucuronate => pyruvate + D-glyceraldehyde 3P	deficient	deficient	deficient	deficient	deficient	deficient
M00063 CMP-KDO biosynthesis	complete	deficient	deficient	deficient	deficient	deficient
M00064 ADP-L-glycero-D-manno-heptose biosynthesis	deficient	complete	deficient	complete	none	complete
M00082 Fatty acid biosynthesis, initiation	complete	complete	complete	complete	complete	complete
M00083 Fatty acid biosynthesis, elongation	complete	complete	complete	complete	complete	complete
M00086 beta-Oxidation, acyl-CoA synthesis	complete	none	none	none	none	none
M00087 beta-Oxidation	complete	none	none	none	none	none
M00088 Ketone body biosynthesis, acetyl-CoA => acetoacetate/3-hydroxybutyrate/acetone	deficient	deficient	deficient	deficient	deficient	deficient
M00089 Triacylglycerol biosynthesis	deficient	deficient	deficient	deficient	deficient	deficient
M00091 Phosphatidylcholine (PC) biosynthesis, PE => PC	complete	complete	complete	complete	complete	complete

M00093 Phosphatidylethanolamine (PE) biosynthesis, PA => PS => PE	complete	complete	complete	complete	complete	complete
M00095 C5 isoprenoid biosynthesis, mevalonate pathway	deficient	deficient	deficient	deficient	deficient	deficient
M00096 C5 isoprenoid biosynthesis, non-mevalonate pathway	deficient	deficient	deficient	deficient	deficient	deficient
M00097 beta-Carotene biosynthesis, GGAP => beta-carotene	deficient	deficient	deficient	deficient	deficient	deficient
M00098 Acylglycerol degradation	complete	none	deficient	none	none	deficient
M00115 NAD biosynthesis, aspartate => quinolinate => NAD	complete	complete	deficient	complete	deficient	complete
M00116 Menaquinone biosynthesis, chorismate (+ polyprenyl-PP) => menaquinol	deficient	deficient	deficient	deficient	deficient	deficient
M00117 Ubiquinone biosynthesis, prokaryotes, chorismate (+ polyprenyl-PP) => ubiquinol	deficient	deficient	deficient	deficient	deficient	deficient
M00118 Glutathione biosynthesis, glutamate => glutathione	complete	complete	complete	complete	complete	complete
M00119 Pantothenate biosynthesis, valine/L-aspartate => pantothenate	deficient	deficient	deficient	deficient	deficient	deficient
M00120 Coenzyme A biosynthesis, pantothenate => CoA	complete	complete	complete	complete	complete	complete
M00121 Heme biosynthesis, plants and bacteria, glutamate => heme	complete	complete	complete	complete	complete	complete
M00122 Cobalamin biosynthesis, cobyrinate a,c-diamide => cobalamin	complete	deficient	deficient	deficient	deficient	deficient
M00124 Pyridoxal-P biosynthesis, erythrose-4P => pyridoxal-P	deficient	deficient	deficient	deficient	deficient	deficient
M00125 Riboflavin biosynthesis, plants and bacteria, GTP => riboflavin/FMN/FAD	complete	complete	complete	complete	complete	complete
M00126 Tetrahydrofolate biosynthesis, GTP => THF	complete	complete	deficient	complete	deficient	deficient
M00127 Thiamine biosynthesis, prokaryotes, AIR (+ DXP/tyrosine) => TMP/TPP	deficient	deficient	deficient	deficient	deficient	deficient
M00129 Ascorbate biosynthesis, animals, glucose-1P => ascorbate	deficient	deficient	deficient	deficient	deficient	deficient
M00131 Inositol phosphate metabolism, Ins(1,3,4,5)P4 => Ins(1,3,4)P3 => myo-inositol	deficient	deficient	deficient	deficient	deficient	deficient
M00133 Polyamine biosynthesis, arginine => agmatine => putrescine => spermidine	deficient	none	none	none	none	deficient
M00135 GABA biosynthesis, eukaryotes, putrescine => GABA	deficient	none	none	none	none	none
M00136 GABA biosynthesis, prokaryotes, putrescine => GABA	deficient	deficient	none	deficient	none	deficient
M00140 C1-unit interconversion, prokaryotes	deficient	deficient	deficient	deficient	deficient	deficient
M00141 C1-unit interconversion, eukaryotes	deficient	deficient	deficient	deficient	deficient	deficient
M00144 NADH:quinone oxidoreductase, prokaryotes	complete	complete	complete	complete	complete	complete
M00149 Succinate dehydrogenase, prokaryotes	complete	complete	complete	complete	complete	complete
M00151 Cytochrome bc1 complex respiratory unit	complete	complete	complete	complete	complete	complete
M00155 Cytochrome c oxidase, prokaryotes	complete	complete	complete	complete	complete	complete
M00156 Cytochrome c oxidase, cbb3-type	complete	none	none	none	none	none
M00157 F-type ATPase, prokaryotes and chloroplasts	complete	complete	complete	complete	complete	complete
M00165 Reductive pentose phosphate cycle (Calvin cycle)	deficient	deficient	deficient	deficient	deficient	deficient
M00166 Reductive pentose phosphate cycle, ribulose-5P => glyceraldehyde-3P	deficient	deficient	deficient	deficient	deficient	deficient
M00167 Reductive pentose phosphate cycle, glyceraldehyde-3P => ribulose-5P	deficient	deficient	deficient	deficient	deficient	deficient
M00168 CAM (Crassulacean acid metabolism), dark	complete	complete	complete	complete	complete	complete
M00169 CAM (Crassulacean acid metabolism), light	deficient	deficient	deficient	none	deficient	none
M00170 C4-dicarboxylic acid cycle, phosphoenolpyruvate carboxykinase type	deficient	deficient	deficient	deficient	deficient	deficient
M00171 C4-dicarboxylic acid cycle, NAD - malic enzyme type	deficient	deficient	deficient	deficient	deficient	deficient
M00172 C4-dicarboxylic acid cycle, NADP - malic enzyme type	deficient	deficient	deficient	deficient	deficient	deficient
M00173 Reductive citrate cycle (Arnon-Buchanan cycle)	deficient	deficient	deficient	deficient	deficient	deficient
M00176 Assimilatory sulfate reduction, sulfate => H2S	deficient	none	deficient	none	none	none
M00307 Pyruvate oxidation, pyruvate => acetyl-CoA	complete	none	complete	complete	complete	complete
M00308 Semi-phosphorylative Entner-Doudoroff pathway, gluconate => glycerate-3P	deficient	deficient	deficient	deficient	deficient	deficient
M00346 Formaldehyde assimilation, serine pathway	deficient	deficient	deficient	deficient	deficient	deficient

M00357 Methanogenesis, acetate => methane	deficient	deficient	deficient	deficient	deficient	deficient
M00358 Coenzyme M biosynthesis	deficient	none	none	none	none	none
M00364 C10-C20 isoprenoid biosynthesis, bacteria	deficient	deficient	deficient	deficient	deficient	deficient
M00366 C10-C20 isoprenoid biosynthesis, plants	deficient	deficient	deficient	deficient	deficient	deficient
M00368 Ethylene biosynthesis, methionine => ethylene	deficient	deficient	deficient	deficient	deficient	deficient
M00373 Ethylmalonyl pathway	deficient	deficient	deficient	deficient	deficient	deficient
M00374 Dicarboxylate-hydroxybutyrate cycle	deficient	deficient	deficient	deficient	deficient	deficient
M00375 Hydroxypropionate-hydroxybutylate cycle	deficient	deficient	deficient	deficient	deficient	deficient
M00376 3-Hydroxypropionate bi-cycle	deficient	deficient	deficient	deficient	deficient	deficient
M00377 Reductive acetyl-CoA pathway (Wood-Ljungdahl pathway)	deficient	deficient	deficient	deficient	deficient	deficient
M00432 Leucine biosynthesis, 2-oxoisovalerate => 2-oxoisocaproate	complete	complete	complete	complete	complete	complete
M00525 Lysine biosynthesis, acetyl-DAP pathway, aspartate => lysine	deficient	deficient	deficient	deficient	deficient	deficient
M00526 Lysine biosynthesis, DAP dehydrogenase pathway, aspartate => lysine	deficient	deficient	deficient	deficient	deficient	deficient
M00527 Lysine biosynthesis, DAP aminotransferase pathway, aspartate => lysine	deficient	deficient	deficient	deficient	deficient	deficient
M00529 Denitrification, nitrate => nitrogen	deficient	none	none	none	none	none
M00530 Dissimilatory nitrate reduction, nitrate => ammonia	deficient	none	none	none	none	none
M00531 Assimilatory nitrate reduction, nitrate => ammonia	deficient	none	none	none	none	none
M00532 Photorespiration	deficient	deficient	deficient	deficient	deficient	deficient
M00535 Isoleucine biosynthesis, pyruvate => 2-oxobutanoate	deficient	deficient	deficient	deficient	deficient	deficient
M00546 Purine degradation, xanthine => urea	deficient	none	none	deficient	none	none
M00549 Nucleotide sugar biosynthesis, glucose => UDP-glucose	complete	complete	complete	complete	complete	complete
M00552 D-galactonate degradation, De Ley-Doudoroff pathway, D-galactonate => glycerate-3P	deficient	deficient	deficient	deficient	deficient	deficient
M00555 Betaine biosynthesis, choline => betaine	complete	none	none	complete	none	complete
M00565 Trehalose biosynthesis, D-glucose 1P => trehalose	deficient	none	none	none	none	none
M00569 Catechol meta-cleavage, catechol => acetyl-CoA / 4-methylcatechol => propanoyl-CoA	deficient	none	none	none	none	none
M00570 Isoleucine biosynthesis, threonine => 2-oxobutanoate => isoleucine	complete	complete	complete	complete	complete	complete
M00572 Pimeloyl-ACP biosynthesis, BioC-BioH pathway, malonyl-ACP => pimeloyl-ACP	deficient	deficient	deficient	deficient	deficient	deficient
M00575 Pertussis pathogenicity signature, T1SS	deficient	deficient	deficient	deficient	deficient	deficient
M00579 Phosphate acetyltransferase-acetate kinase pathway, acetyl-CoA => acetate	complete	none	none	none	none	none
M00580 Pentose phosphate pathway, archaea, fructose 6P => ribose 5P	deficient	deficient	deficient	deficient	deficient	deficient
M00596 Dissimilatory sulfate reduction, sulfate => H2S	deficient	deficient	deficient	deficient	deficient	deficient
M00609 Cysteine biosynthesis, methionine => cysteine	deficient	deficient	deficient	deficient	deficient	deficient
M00615 Nitrate assimilation	deficient	none	deficient	none	none	none
M00618 Acetogen	deficient	none	none	none	none	none
M00620 Incomplete reductive citrate cycle, acetyl-CoA => oxoglutarate	deficient	deficient	deficient	deficient	deficient	deficient
M00622 Nicotinate degradation, nicotinate => fumarate	deficient	none	none	none	none	none
M00631 D-Galacturonate degradation (bacteria), D-galacturonate => pyruvate + D-glyceraldehyde 3P	deficient	deficient	deficient	deficient	deficient	deficient
M00632 Galactose degradation, Leloir pathway, galactose => alpha-D-glucose-1P	deficient	none	none	deficient	deficient	deficient
M00698 Multidrug resistance, efflux pump BpeEF-OprC	deficient	deficient	deficient	deficient	deficient	deficient
M00740 Methylaspartate cycle	deficient	deficient	deficient	deficient	deficient	deficient
M00744 Cationic antimicrobial peptide (CAMP) resistance, protease PgtE	deficient	none	none	deficient	none	deficient
M00793 dTDP-L-rhamnose biosynthesis	complete	deficient	none	deficient	none	complete
M00804 Complete nitrification, comammox, ammonia => nitrite => nitrate	deficient	none	none	none	none	none

M00835 Pyocyanine biosynthesis, chorismate => pyocyanine	deficient	none	none	none	none	none
M00840 Tetrahydrofolate biosynthesis, mediated by ribA and trpF, GTP => THF	deficient	deficient	deficient	deficient	deficient	deficient
M00841 Tetrahydrofolate biosynthesis, mediated by PTPS, GTP => THF	deficient	deficient	deficient	deficient	deficient	deficient
M00842 Tetrahydrobiopterin biosynthesis, GTP => BH4	deficient	deficient	deficient	deficient	deficient	deficient
M00843 L-threo-Tetrahydrobiopterin biosynthesis, GTP => L-threo-BH4	deficient	deficient	deficient	deficient	deficient	deficient
M00844 Arginine biosynthesis, ornithine => arginine	complete	complete	complete	complete	complete	complete
M00845 Arginine biosynthesis, glutamate => acetylcitrulline => arginine	deficient	deficient	deficient	deficient	deficient	deficient
M00846 Siroheme biosynthesis, glutamyl-tRNA => siroheme	deficient	complete	complete	complete	deficient	deficient
M00849 C5 isoprenoid biosynthesis, mevalonate pathway, archaea	deficient	deficient	deficient	deficient	deficient	deficient
M00855 Glycogen degradation, glycogen => glucose-6P	deficient	deficient	deficient	deficient	deficient	deficient
M00866 KDO2-lipid A biosynthesis, Raetz pathway, non-LpxL-LpxM type	deficient	deficient	deficient	deficient	deficient	deficient
M00868 Heme biosynthesis, animals and fungi, glycine => heme	deficient	deficient	deficient	deficient	deficient	deficient
M00873 Fatty acid biosynthesis in mitochondria, animals	deficient	deficient	deficient	deficient	deficient	deficient
M00874 Fatty acid biosynthesis in mitochondria, fungi	deficient	deficient	deficient	deficient	deficient	deficient
M00878 Phenylacetate degradation, phenylaxetate => acetyl-CoA/succinyl-CoA	deficient	none	none	none	deficient	none
M00880 Molybdenum cofactor biosynthesis, GTP => molybdenum cofactor	complete	deficient	complete	complete	complete	complete
M00881 Lipoic acid biosynthesis, plants and bacteria, octanoyl-ACP => dihydrolipoyl-E2/H	complete	complete	complete	complete	complete	complete
M00882 Lipoic acid biosynthesis, eukaryotes, octanoyl-ACP => dihydrolipoyl-H	deficient	deficient	deficient	deficient	deficient	deficient
M00883 Lipoic acid biosynthesis, animals and bacteria, octanoyl-ACP => dihydrolipoyl-H => dihydrolipoyl-E2	deficient	deficient	deficient	deficient	deficient	deficient
M00884 Lipoic acid biosynthesis, octanoyl-CoA => dihydrolipoyl-E2	deficient	deficient	deficient	deficient	deficient	deficient
M00892 UDP-N-acetyl-D-glucosamine biosynthesis, eukaryotes, glucose => UDP-GlcNAc	deficient	deficient	deficient	deficient	deficient	deficient
M00895 Thiamine biosynthesis, prokaryotes, AIR (+ DXP/glycine) => TMP/TPP	deficient	deficient	deficient	deficient	deficient	deficient
M00896 Thiamine biosynthesis, archaea, AIR (+ NAD+) => TMP/TPP	deficient	deficient	deficient	deficient	deficient	deficient
M00897 Thiamine biosynthesis, plants, AIR (+ NAD+) => TMP/thiamine/TPP	deficient	none	none	none	none	none
M00899 Thiamine salvage pathway, HMP/HET => TMP	deficient	deficient	deficient	deficient	deficient	deficient
M00909 UDP-N-acetyl-D-glucosamine biosynthesis, prokaryotes, glucose => UDP-GlcNAc	complete	complete	complete	complete	complete	complete
M00911 Riboflavin biosynthesis, fungi, GTP => riboflavin/FMN/FAD	deficient	deficient	deficient	deficient	deficient	deficient
M00912 NAD biosynthesis, tryptophan => quinolinate => NAD	deficient	deficient	deficient	deficient	deficient	deficient
M00913 Pantothenate biosynthesis, 2-oxoisovalerate/spermine => pantothenate	deficient	deficient	deficient	deficient	deficient	deficient
M00914 Coenzyme A biosynthesis, archaea, 2-oxoisovalerate => 4-phosphopantoate => CoA	deficient	deficient	deficient	deficient	deficient	deficient
M00925 Cobalamin biosynthesis, aerobic, uroporphyrinogen III => precorrin 2 => cobyrinate a,c-diamide	none	deficient	deficient	deficient	none	none
M00926 Heme biosynthesis, bacteria, glutamyl-tRNA => coproporphyrin III => heme	deficient	deficient	deficient	deficient	deficient	deficient
M00938 Pyrimidine deoxyribonuleotide biosynthesis, UDP => dTTP	deficient	deficient	deficient	deficient	deficient	complete
M00045 Histidine degradation, histidine => N-formiminoglutamate => glutamate	none	none	none	none	deficient	none
M00939 Pyrimidine degradation, uracil => 3-hydroxypropanoate	none	none	none	none	deficient	none
M00924 Cobalamin biosynthesis, anaerobic, uroporphyrinogen III => sirohydrochlorin => cobyrinate a,c-diamide	none	deficient	deficient	deficient	none	none
M00948 Hydroxyproline degradation, trans-4-hydroxy-L-proline => 2-oxoglutarate	none	deficient	none	deficient	none	deficient
Compatible Solute Transporters						
Betaine/carnitine transporter, BCCT family	1	1	1	1	1	1
Choline/glycine/proline betaine transport protein	0	1	1	1	1	1
Glycine betaine/proline transport system ATP-binding protein [EC:7.6.2.9]	1	2	2	2	2	2

Glycine betaine/proline transport system permease protein	1	2	2	2	2	2
Glycine betaine/proline transport system substrate-binding protein	3	2	2	4	2	3
Carbohydrate-Active Enzymes Classes						
Auxiliary activities	3	1	1	1	1	1
Carbohydrate-binding modules	5	5	3	4	4	2
Carbohydrate esterases	2	1	2	2	1	1
Glycoside hydrolases	11	7	8	7	7	7
Glycosyl transferases	28	22	19	26	16	21
Polysaccharide lyases	1	0	0	0	0	0

Table S6. The sequencing and rarefied data statistics of each sample.

Sample	Total reads	Filtered reads	Rarefied reads
S045_1	39460	38703	22031
S045_2	35598	34835	22031
S045_3	35906	35072	22031
S080_1	39164	38532	22031
S080_2	44984	43089	22031
S080_3	40641	39451	22031
S125_1	39004	31501	22031
S125_2	33540	27261	22031
S125_3	44873	35569	22031
S170_1	35273	33303	22031
S170_2	44862	42973	22031
S170_3	33487	32030	22031
S175_1	44221	36151	22031
S175_2	43916	36427	22031
S175_3	43718	31961	22031
S185_1	30641	28844	22031
S185_2	35248	29311	22031
S185_3	37997	35988	22031
S205_1	39406	34961	22031
S205_2	37001	33376	22031
S205_3	33164	30342	22031
S230_1	29148	28192	22031
S230_2	38850	36168	22031
S230_3	46375	44282	22031
S265_1	45779	42981	22031
S265_2	35453	33938	22031
S265_3	47809	45867	22031
S300_1	57808	54361	22031
S300_2	50148	48631	22031
S300_3	32621	31331	22031
S340_1	30349	28551	22031
S340_2	27999	26674	22031
S340_3	32837	29493	22031

Sample	Total reads	Filtered reads	Rarefied reads
W045_1	37573	33310	22031
W045_2	36310	33297	22031
W045_3	38489	35159	22031
W080_1	34197	31831	22031
W080_2	27170	25505	22031
W080_3	32981	31295	22031
W125_1	30512	30243	22031
W125_2	31082	30945	22031
W125_3	33956	33858	22031
W170_1	25309	25135	22031
W170_2	31761	31557	22031
W170_3	38476	38224	22031
W175_1	42009	41969	22031
W175_2	38346	38299	22031
W175_3	33180	33158	22031
W185_1	29826	29773	22031
W185_2	31730	31673	22031
W185_3	23774	23729	22031
W205_1	37806	37770	22031
W205_2	22101	22031	22031
W205_3	26386	26300	22031
W230_1	39212	38940	22031
W230_2	38296	38168	22031
W230_3	31824	31655	22031
W265_1	39047	36247	22031
W265_2	27307	25168	22031
W265_3	32658	29971	22031
W300_1	43860	43764	22031
W300_2	29241	28741	22031
W300_3	26387	26139	22031
W340_1	30458	29737	22031
W340_2	27595	26719	22031
W340_3	32902	32286	22031