

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19

## Supplementary Information

### Salvaging high-quality genomes of microbial species from a meromictic lake using a hybrid sequencing approach

Yu-Hsiang Chen<sup>1,2,3</sup>, Pei-Wen Chiang<sup>3</sup>, Denis Yu Rogozin<sup>4,5</sup>, Andrey G.

Degermendzhy<sup>4</sup>, Hsiu-Hui Chiu<sup>3</sup>, Sen-Lin Tang<sup>2,3\*</sup>

<sup>1</sup>Bioinformatics Program, Taiwan International Graduate Program, National Taiwan University, Taipei, Taiwan

<sup>2</sup>Bioinformatics Program, Institute of Information Science, Taiwan International Graduate Program, Academia Sinica, Taipei, Taiwan

<sup>3</sup>Biodiversity Research Center, Academia Sinica, Taipei, Taiwan

<sup>4</sup>Institute of Biophysics, Siberian Division of Russian Academy of Sciences, Krasnoyarsk, Russia

<sup>5</sup>Siberian Federal University, Krasnoyarsk, Russia

\* Corresponding author

Biodiversity Research Center, Academia Sinica

E-mail: sltang@gate.sinica.edu.tw

Tel: +886-2-27893863

Fax: +886-2-27890844

20 **Supplementary Figures**

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

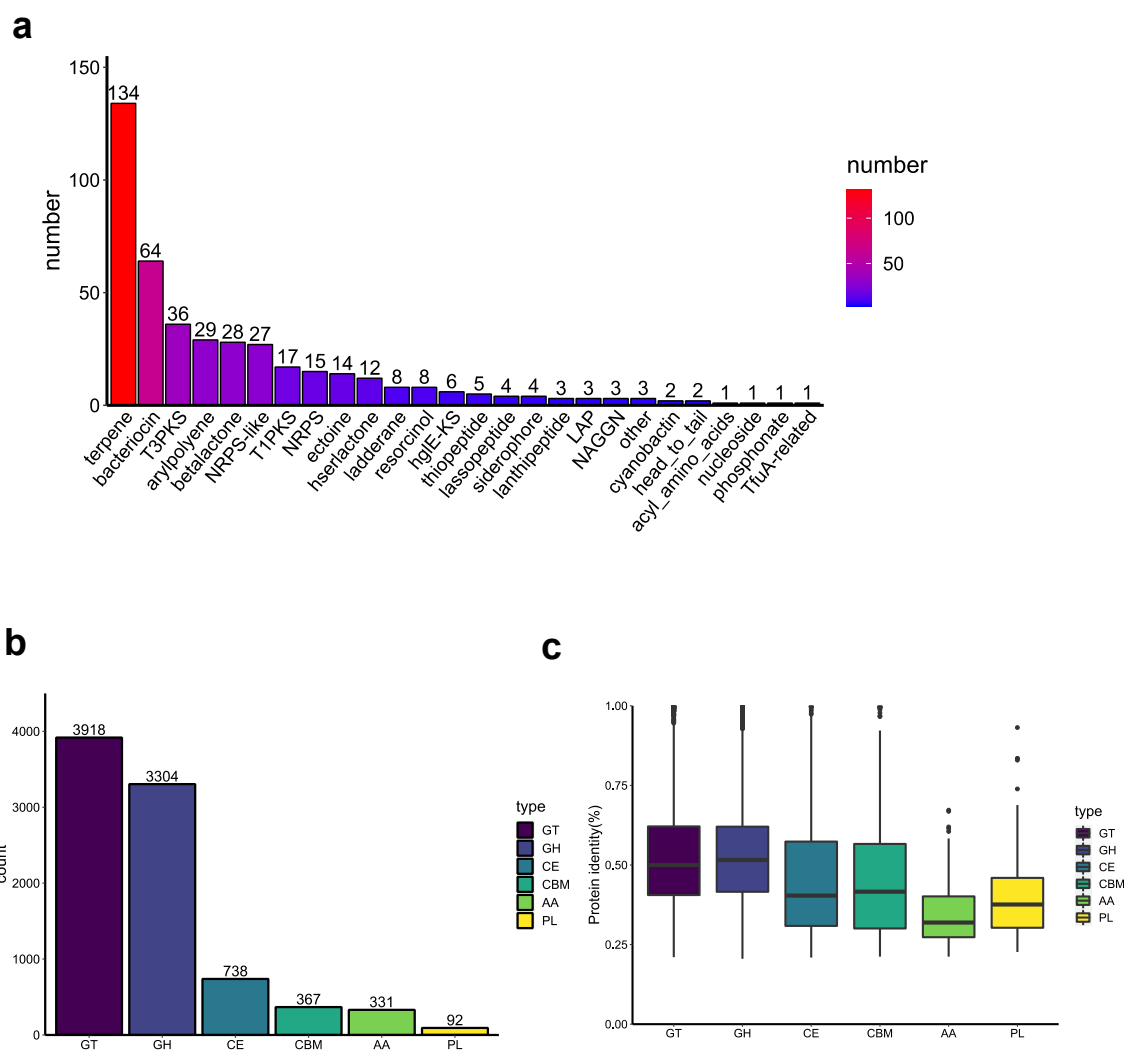
48

49

50

51

52



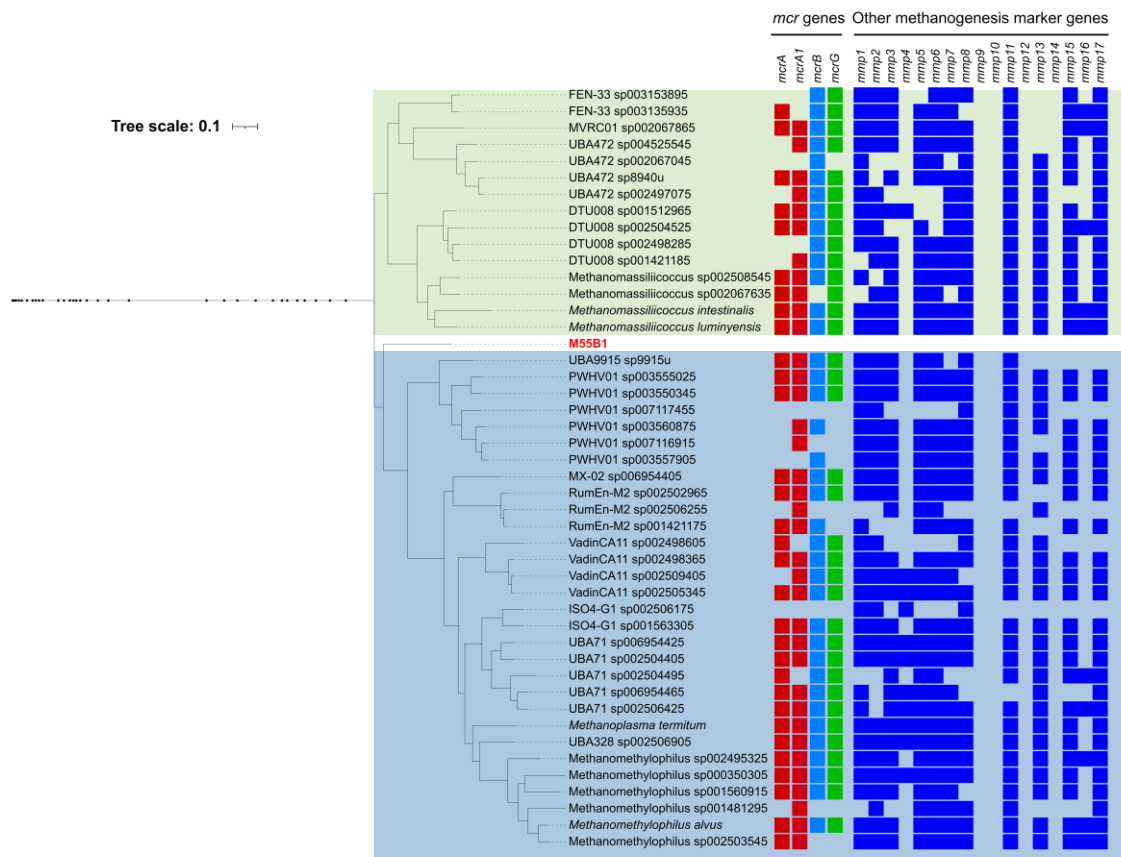
**Supplementary Figure 1. Predicted secondary metabolite biosynthetic clusters and carbohydrate active-enzymes from recovered MAGs.** **a.** The numbers of different types of BGCs identified from recovered MAGs by antiSMASH 5.0. The description of different types of secondary metabolite clusters can be found in the glossary from antiSMASH (<https://docs.antismash.secondarymetabolites.org/glossary/>). **b.** The numbers of different types of CAZymes identified from MAGs using dbCAN 5.0. **c.** Distribution of protein identities between newly identified CAZymes and their best hits in the NCBI nr database. The center line represents median, box limits represent upper and lower quartiles, whisker represents 1.5X interquartile range, and black dots represent outliers. Abbreviations in figure S1a: T3PKS, Type III Polyketide synthase

53 (PKS); NPRS, Non-ribosomal peptide synthetase cluster; hglE-KS, heterocyst  
54 glycolipid synthase-like PKS; LAP, Linear azol(in)e-containing peptides; NAGGN, N-  
55 acetylglutaminyglutamine amide; head\_to\_tail, Head-to-tail cyclised cluster;  
56 acyl\_amino\_acids, N-acyl amino acid cluster; TfuA-related; TfuA-related ribosomally  
57 synthesized and post-translationally modified peptides. Abbreviation in figure S2b:  
58 c:GT, GlycosylTransferase; GH, Glycoside Hydrolase; CE, Carbohydrate esterase;  
59 CBM, Carbohydrate-Binding Module; AA, Auxiliary Activity; PL, Polysaccharide  
60 Lyase.

61

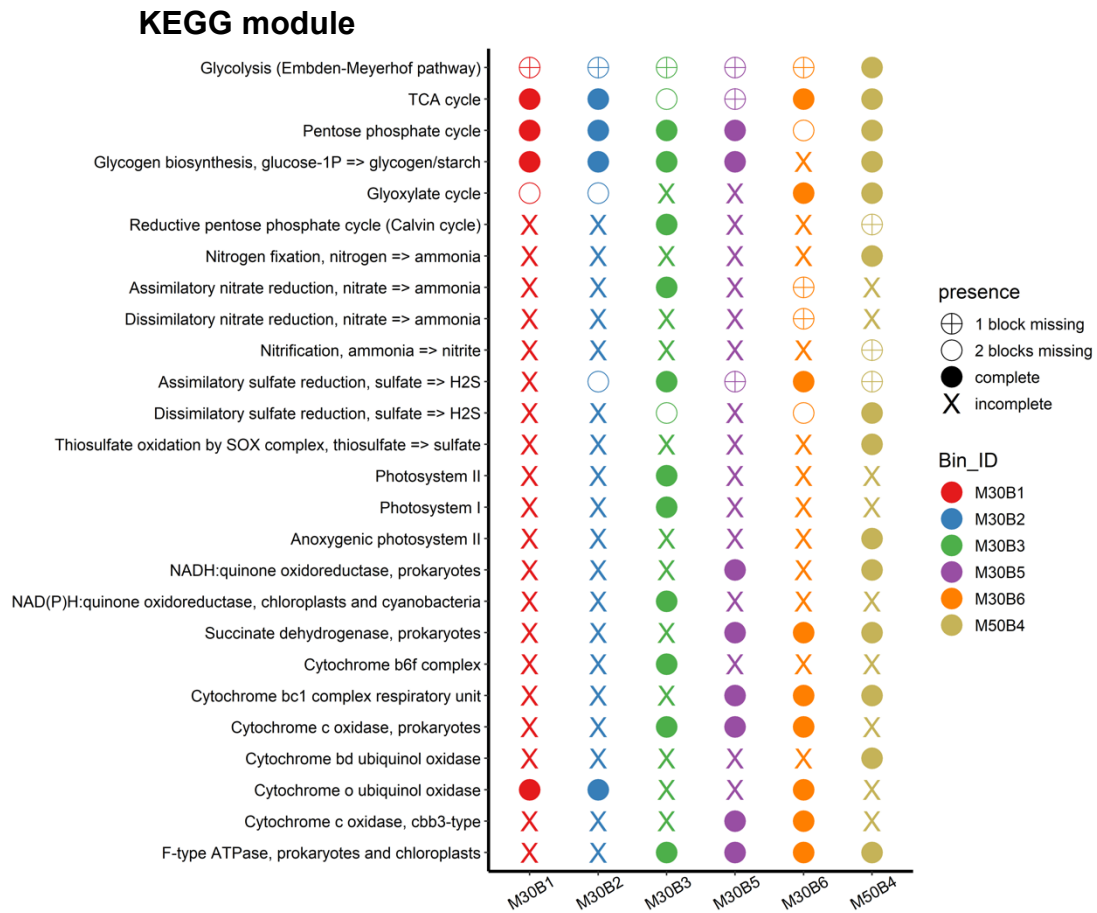
62

63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95



Supplementary Figure 2. Pruned phylogenomic tree of *Methanomassiliococcales* and presence of methanogenesis marker genes in MAGs. The tree contains GTDB representative genomes of host-associated (highlighted in blue) and environmental clades (highlighted in green) in the *Methanomassiliococcales* order. The genomes are annotated with Pfam and KEGG orthologous genes by EnrichM. The methanogenesis markers were selected according to Borrel *et al* (2014)<sup>1,2</sup>. KO numbers and Pfam IDs are listed in Table S2. The colored tiles indicate that the gene is present. Scale bar represents changes per amino acid site. Abbreviation: mmp, methanogenesis marker protein.

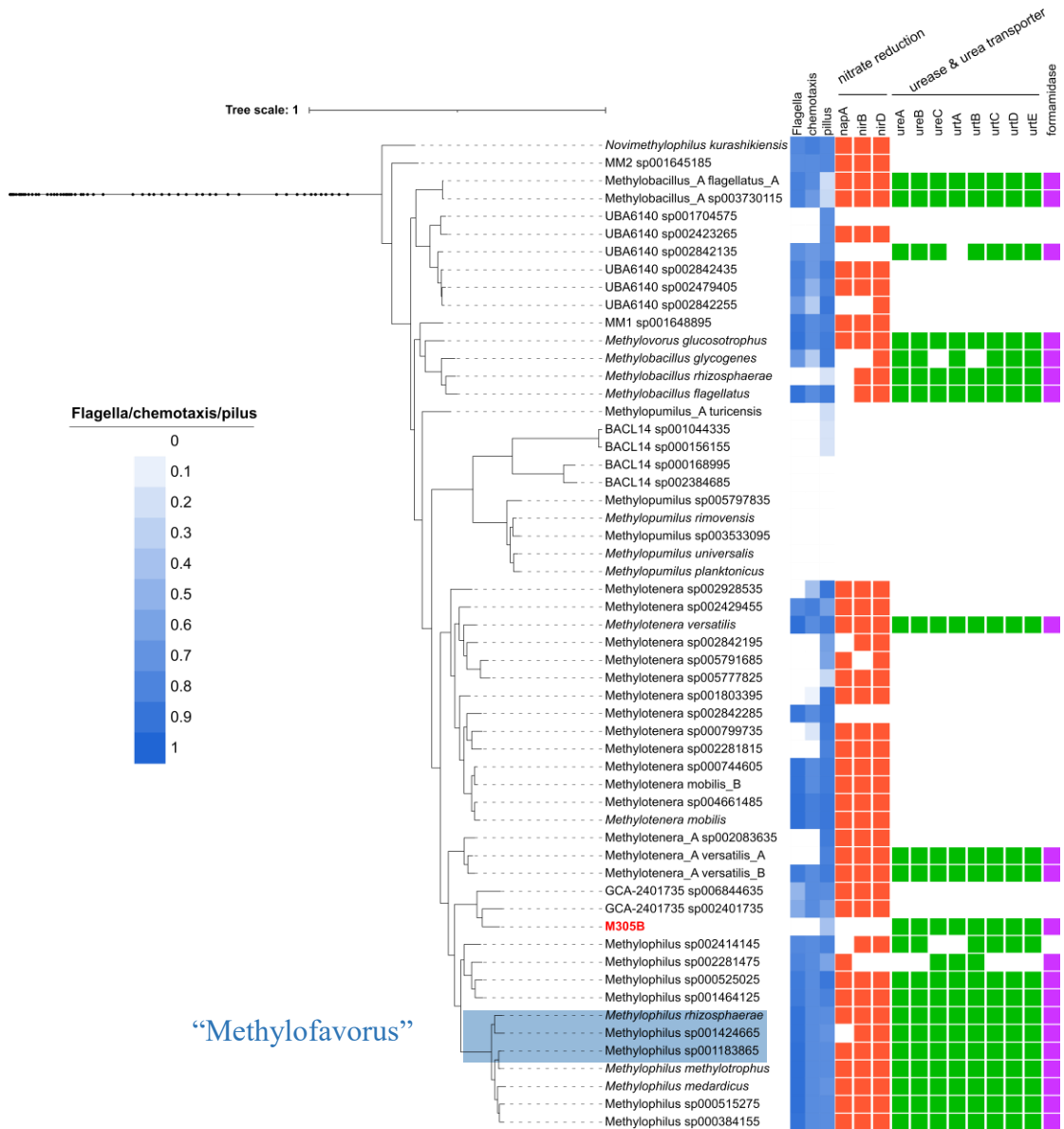
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122



**Supplementary Figure 3. Summary of metabolic features in the newly recovered complete MAGs.**

The genomes are annotated with KEGG orthologues, and the completeness of the metabolic pathways is determined by EnrichM. The KEGG module numbers of each metabolic pathway are listed in Table S2. Each KEGG module is made of multiple functional units (blocks). Missing blocks represents missing gene(s) that form a functional unit. Incomplete modules indicate that more than two blocks are missing.

123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155



**Supplementary Figure 4. Pruned phylogenomic tree of *Methylophilaceae*.**

The tree contains GTDB representative genomes in the family *Methylophilaceae*. The genomes are annotated with Pfam and KEGG orthologous genes by EnrichM. The genes are selected according to Salcher *et al.*<sup>3</sup>. KO numbers and Pfam IDs are listed in Table S3. The colored tiles indicate that the gene is present. The heatmap represents the proportions calculated by dividing the number of flagella, chemotaxis, and pilus-related genes found in the genome with related gene numbers in the KEGG database (list in Table S2). The scale bar represents changes per amino acid site.

156 **Supplementary Table 1. Statistics on Nanopore and Illumina reads**

	<b>3.0m</b>	<b>5.0m</b>	<b>5.5m</b>
<b>Nanopore reads</b>			
Mean read length	2,063	1,000	7,006
Mean read quality	10.6	11.1	11.6
Median read length	761	472	3,669
Median read quality	11.1	11.3	11.9
Number of reads	6,705,136	12,569,375	683,928
Read length N50	5,578	1,775	13,545
Total bases	13,83 Gbp	12,57 Gbp	4,79 Gbp
Top 5 longest reads :			
<b>1.</b>	150,131	133,860	133,934
<b>2.</b>	132,395	130,047	120,861
<b>3.</b>	128,047	128,699	114,105
<b>4.</b>	121,232	123,796	114,062
<b>5.</b>	120,736	123,662	112,369
<b>Illumina reads</b>			
Total reads (F+R)	486,655,848	579,511,508	509,594,810
Sequence length (bp)	45-100	45-100	45-100
Mean read length (F, R)	86, 91	91, 90	90, 93
Total bases	43.27 gbp	52.53 gbp	46.74 gbp
Q20 bases	42.12 gbp	52.53 gbp	45.82 gbp
Q30 bases	38.06 gbp	52.53 gbp	42.41 gbp

157

158

159

160

161

162

163

164

165

**Supplementary Table 2. KEGG Orthology and Pfam families used in this study**

Genes	Proteins/ functions	KEGG/TIGRFAMs #
<b>Methanogenesis</b>		
<i>mcrA</i>	Methyl-CoM reductase subunit A	K00399
<i>mcrA2</i>	Methyl-CoM reductase subunit A2	K00400
<i>mcrB</i>	Methyl-CoM reductase subunit beta	K00401
<i>mcrG</i>	Methyl-CoM reductase subunit G	K00402
<i>mmp1</i>	Methanogenesis marker protein 1	TIGR03266
<i>mmp2</i>	Methanogenesis marker protein 2	TIGR03267
<i>mmp3</i>	Methanogenesis marker protein 3	TIGR03268
<i>mmp4</i>	Methanogenesis marker protein 4	TIGR03270
<i>mmp5</i>	Methanogenesis marker protein 5	TIGR03271
<i>mmp6</i>	Methanogenesis marker protein 6	TIGR03272
<i>mmp7</i>	Methanogenesis marker protein 7	TIGR03274
<i>mmp8</i>	Methanogenesis marker protein 8	TIGR03275
<i>mmp9</i>	Methanogenesis marker protein 9	TIGR03277
<i>mmp10</i>	Methanogenesis marker protein 10 (putative methyl coenzyme M reductase-arginine methyltransferase)	TIGR03278
<i>mmp11</i>	Methanogenesis marker protein 11	TIGR03280
<i>mmp12</i>	Methanogenesis marker protein 12	TIGR03281
<i>mmp13</i>	Methanogenesis marker protein 13	TIGR03282
<i>mmp14</i>	Methanogenesis marker protein 14	TIGR03285
<i>mmp15</i>	Methanogenesis marker protein 15	TIGR03286
<i>mmp16</i>	Methanogenesis marker protein 16	TIGR03287
<i>mmp17</i>	Methanogenesis marker protein 17	TIGR03291
<b>Metabolic pathways</b>		
	Glycolysis (Embden-Meyerhof pathway)	M00001
	TCA cycle	M00009
	Pentose phosphate cycle	M00004
	Glycogen biosynthesis	M00854
	Glyoxylate cycle	M00012



	Reductive pentose phosphate cycle	M00165
	Nitrogen fixation	M00175
	Assimilatory nitrate reduction	M00531
	Dissimilatory nitrate reduction	M00530
	Nitrification, ammonia => nitrite	M00528
	Assimilatory sulfate reduction, sulfate => H2S	M00176
	Dissimilatory sulfate reduction, sulfate => H2S	M00596
	Thiosulfate oxidation by SOX complex	M00595
	Photosystem II	M00161
	Photosystem I	M00161
	Anoxygenic photosystem II	M00597
	NADH:quinone oxidoreductase, prokaryotes	M00144
	NAD(P)H:quinone oxidoreductase, chloroplasts and cyanobacteria	M00145
	Succinate dehydrogenase, prokaryotes	M00149
	Cytochrome b6f complex	M00162
	Cytochrome bc1 complex respiratory unit	M00151
	Cytochrome c oxidase, prokaryotes	M00155
	Cytochrome bd ubiquinol oxidase	M00153
	Cytochrome o ubiquinol oxidase	M00417
	Cytochrome c oxidase, cbb3-type	M00156
<b>Flagellum</b>		
<i>flhG, fleN</i>	flagellar biosynthesis protein FlhG	K04562
<i>flhB2</i>	flagellar biosynthesis protein	K04061
<i>flgA</i>	flagella basal body P-ring formation protein	K02386
<i>flgB</i>	flagellar basal-body rod protein FlgB	K02387
<i>flgC</i>	flagellar basal-body rod protein FlgC	K02388
<i>flgD</i>	flagellar basal-body rod modification protein	K02389
<i>flgE</i>	flagellar hook protein FlgE	K02390
<i>flgF</i>	flagellar basal-body rod protein FlgF	K02391
<i>flgG</i>	flagellar basal-body rod protein FlgG	K02392
<i>flgH</i>	flagellar L-ring protein precursor FlgH	K02393

<i>flgI</i>	flagellar P-ring protein precursor FlgI	K02394
<i>flgJ</i>	flagellar protein FlgJ	K02395
<i>flgK</i>	flagellar hook-associated protein 1 FlgK	K02396
<i>flgL</i>	flagellar hook-associated protein 3 FlgL	K02397
<i>flgM</i>	negative regulator of flagellin synthesis FlgM	K02398
<i>flgN</i>	flagella synthesis protein FlgN	K02399
<i>flhA</i>	flagellar biosynthesis protein FlhA	K02400
<i>flhB</i>	flagellar biosynthetic protein FlhB	K02401
<i>flhC</i>	flagellar transcriptional activator FlhC	K02402
<i>flhD</i>	flagellar transcriptional activator FlhD	K02403
<i>flhF</i>	flagellar biosynthesis protein FlhF	K02404
<i>fliA</i>	RNA polymerase sigma factor for flagellar operon	K02405
<i>fliC</i>	flagellin	K02406
<i>fliD</i>	flagellar hook-associated protein 2	K02407
<i>fliE</i>	flagellar hook-basal body complex protein FliE	K02408
<i>fliF</i>	flagellar M-ring protein FliF	K02409
<i>fliG</i>	flagellar motor switch protein FliG	K02410
<i>fliH</i>	flagellar assembly protein FliH	K02411
<i>fliI</i>	flagellum-specific ATP synthase [EC:3.6.3.14]	K02412
<i>fliJ</i>	flagellar FliJ protein	K02413
<i>fliK</i>	flagellar hook-length control protein FliK	K02414
<i>fliL</i>	flagellar FliL protein	K02415
<i>fliM</i>	flagellar motor switch protein FliM	K02416
<i>fliNY, fliN</i>	flagellar motor switch protein FliN/FliY	K02417
<i>fliOZ, fliO</i>	flagellar protein FliO/FliZ	K02418
<i>fliP</i>	flagellar biosynthetic protein FliP	K02419
<i>fliQ</i>	flagellar biosynthetic protein FliQ	K02420
<i>fliR</i>	flagellar biosynthetic protein FliR	K02421
<i>fliS</i>	flagellar protein FliS	K02422
<i>fliT</i>	flagellar protein FliT	K02423
<i>flbT</i>	flagellar protein FlbT	K06601
<i>flaF</i>	flagellar protein FlaF	K06602

<i>flaG</i>	flagellar protein FlaG	K06603
<i>flaI</i>	flagellar rod protein FlaI	K06604
<b>Chemotaxis</b>		
<i>motA</i>	chemotaxis protein MotA	K02556
<i>motB</i>	chemotaxis protein MotB	K02557
<i>mcp</i>	methyl-accepting chemotaxis protein	K03406
<i>cheA</i>	two-component system, chemotaxis family, CheA [EC:2.7.13.3]	K03407
<i>cheW</i>	purine-binding chemotaxis protein CheW	K03408
<i>cheX</i>	chemotaxis protein CheX	K03409
<i>cheC</i>	chemotaxis protein CheC	K03410
<i>cheD</i>	chemotaxis protein CheD [EC:3.5.1.44]	K03411
<i>cheB</i>	two-component system, chemotaxis family, response regulator CheB [EC:3.1.1.61]	K03412
<i>cheY</i>	two-component system, chemotaxis family, response regulator CheY	K03413
<i>cheZ</i>	chemotaxis protein CheZ	K03414
<i>cheV</i>	two-component system, chemotaxis family, response regulator CheV	K03415
<b>Pilus</b>		
<i>pilA</i>	type IV pilus assembly protein PilA	K02650
<i>flp, pilA</i>	pilus assembly protein Flp/PilA	K02651
<i>pilB</i>	type IV pilus assembly protein PilB	K02652
<i>pilC</i>	type IV pilus assembly protein PilC	K02653
<i>pilD, pppA</i>	leader peptidase (prepilin peptidase) / N- methyltransferase [EC:3.4.23.43 2.1.1.-]	K02654
<i>pilE</i>	type IV pilus assembly protein PilE	K02655
<i>pilF</i>	type IV pilus assembly protein PilF	K02656
<i>pilG</i>	twitching motility two-component system response regulator PilG	K02657
<i>pilH</i>	twitching motility two-component system response regulator PilH	K02658

<i>pilI</i>	twitching motility protein PilI	K02659
<i>pilJ</i>	twitching motility protein PilJ	K02660
<i>pilK</i>	type IV pilus assembly protein PilK	K02661
<i>pilM</i>	type IV pilus assembly protein PilM	K02662
<i>pilN</i>	type IV pilus assembly protein PilN	K02663
<i>pilO</i>	type IV pilus assembly protein PilO	K02664
<i>pilP</i>	type IV pilus assembly protein PilP	K02665
<i>pilQ</i>	type IV pilus assembly protein PilQ	K02666
<i>pilR, pehR</i>	two-component system, NtrC family, response regulator PilR	K02667
<i>pilS, pehS</i>	two-component system, NtrC family, sensor histidine kinase PilS [EC:2.7.13.3]	K02668
<i>pilT</i>	twitching motility protein PilT	K02669
<i>pilU</i>	twitching motility protein PilU	K02670
<i>pilV</i>	type IV pilus assembly protein PilV	K02671
<i>pilW</i>	type IV pilus assembly protein PilW	K02672
<i>pilX</i>	type IV pilus assembly protein PilX	K02673
<i>pilY1</i>	type IV pilus assembly protein PilY1	K02674
<i>pilY2</i>	type IV pilus assembly protein PilY2	K02675
<i>pilZ</i>	type IV pilus assembly protein PilZ	K02676
<b>nitrate reductase</b>		
<i>napA</i>	nitrate reductase	K02567
<i>nirB</i>	nitrite reductase	K00362
<i>nirD</i>	nitrite reductase (NADH) small subunit	K00363
<b>Urease and urea transporter</b>		
<i>ureA</i>	urease subunit gamma	K01430
<i>ureB</i>	urease subunit beta	K01429
<i>ureC</i>	urease subunit alpha	K01428
<i>urtA</i>	urea transport system substrate-binding protein	K11959
<i>urtB</i>	urea transport system permease protein	K11960
<i>urtC</i>	urea transport system permease protein	K11961
<i>urtD</i>	urea transport system ATP-binding protein	K11962

<i>urtE</i>	urea transport system ATP-binding protein	K11963
<b>Others</b>		
formamidase	formamidase	K01455

167  
168

169

## 170 **Supplementary References**

- 171 1 Borrel, G. *et al.* Comparative genomics highlights the unique biology of  
172  
173 Methanomassiliicoccales, a Thermoplasmatales-related seventh order of  
174 methanogenic archaea that encodes pyrrolysine. *Bmc Genomics* **15**, doi:Artn  
175 679  
176 10.1186/1471-2164-15-679 (2014).
- 177 2 Zinke, L. A. *et al.* Evidence for non-methanogenic metabolisms in globally  
178 distributed archaeal clades basal to the Methanomassiliicoccales. *Environ*  
179 *Microbiol* **23**, doi:10.1111/1462-2920.15316 (2021).
- 180 3 Salcher, M. M., Schaeffle, D., Kaspar, M., Neuenschwander, S. M. & Ghai, R.  
181 Evolution in action: habitat transition from sediment to the pelagial leads to  
182 genome streamlining in Methylophilaceae. *Isme Journal* **13**, 2764-2777,  
183 doi:10.1038/s41396-019-0471-3 (2019).
- 184