Supplementary information for

Evolution in action: Habitat-transition from sediment to the pelagial leads to genome-

streamlining in Methylophilaceae

Michaela M. Salcher, Daniel Schaefle, Melissa Kaspar, Stefan M. Neuenschwander, Rohit

Ghai

Supplementary Tables S1-S5 and a set of HMMs of can be found as separate dataset (SalcherTablesS1-S5.xlsx; Salcher_methdbHMM.gz)

 Table S1: Characteristics of analysed genomes.

Table S2: Public metagenomes used for recruitment analyses and RPKG values (reads per kb of genome per Gb of metagenome) of all genomes.

Table S3: Membrane transporters annotated in the genomes.

Table S4: Selected metabolic modules present in the genomes.

Table S5: Complete list of proteins used for creating HMMs related to methylotrophicfunctions. The entire set of HMMs is available as supplementary data set (Salcher_meth-HMMs.txt)

Figure S1: Phylogeny of 39 strains of 'Ca. Methylopumilus spp.'.

Figure S2: Average nucleotide identity (ANI) of 39 strains of 'Ca. Methylopumilus sp.'.

Figure S3: Average amino acid identity (AAI) of Methylophilaceae.

Figure S4: Average nucleotide identity (ANI) of Methylophilaceae.

Figure S5: Genomic statistics of Methylophilaceae.

Figure S6: Amino acid usage of Methylophilaceae in relation to genome size.

Figure S7: Core- and pangenome analysis of Methylophilaceae.

Figure S8: Methylcitric acid (MCA) pathway.

Figure S9: Conserved methylotrophic pathways - H4F, methanol oxidation and RuMP.

Figure S10: Conserved methylotrophic pathways - methanol and formate oxidation.

Figure S11: XoxF and MxaF type methanol dehydrogenases.

Figure S12: Horizontal gene transfers of different rhodopsins.

Figure S13: Horizontal gene transfer of the NQR pathway in marine HTCC2181.

Figure S14: Horizontal gene transfer of an ectoine biosynthesis pathway and a sodium:proline symporter in marine OM43

Figure S15: Horizontal gene transfer an oligopeptide permease in marine OM43

Figure S16: Horizontal gene transfer of sodium transporters in marine OM43



Figure S1: Phylogeny of 39 strains of 'Ca. Methylopumilus spp.'.

Phylogenomic tree of 39 complete genomes of 'Ca. Methylopumilus spp' that form three distinct species ('Ca. M. rimovensis', 'Ca. M. planktonicus', 'Ca. M. universalis'). The tree is based on 983 common, concatenated protein sequences (337501 amino acid sites) with 'Ca. Methylosemipumilus turicensis' used as outgroup. Bootstraps values (100 bootstraps) are indicated by circles at individual nodes, the scale bar applies to 8% sequence divergence. The isolation source is indicated by different colours.



Figure S2: Average nucleotide identity (ANI) of 39 strains of 'Ca. Methylopumilus sp.'.

Values >95% identity (species level delineation) are marked by black boxes. The colour coding of genomes is according to isolation source as displayed in Fig. S1.



Figure S3: Average amino acid identity (AAI) of Methylophilaceae.

Values >70% identity (approx. genus level delineation) are marked by black boxes. The colour coding of genomes is according to Fig. 1.



Figure S4: Average nucleotide identity (ANI) of Methylophilaceae.

Values >95% identity (species level delineation) are marked by black boxes. The colour coding of genomes is according to Fig. 1.



Figure S5: Genomic statistics of Methylophilaceae.

Phylogenomic tree (for details see Fig. 1 and Fig. S1) with selected genomic features related to genome-streamlining in Methylophilaceae.



Figure S6: Amino acid usage of Methylophilaceae in relation to genome size.



b



Figure S7: Core- and pangenome analysis of Methylophilaceae.

(a) Core- and pangenome analysis of all 72 Methylophilaceae strains in relation to 39 '*Ca*. Methylopumilus spp. strains. (b) Shared protein families of sediment *Methylophilus* and *Methylotenera*, '*Ca*. Methylosemipumilus turicensis', '*Ca*. Methylopumilus spp.' and marine OM43 at a 50% identity clustering.



Figure S8: Methylcitric acid (MCA) pathway

(a) Arrangement and protein similarities of genes coding for the methylcitric acid (MCA) pathway marked in dark yellow and flanking genes. (b) Similarities of *acnB* genes in marine OM43 and 'Ca. M. turicensis', located at different genomic regions. (c) RAxML tree (100 bootstraps) of concatenated MCA proteins.



Figure S9: XoxF and MxaF type methanol dehydrogenases

Phylogenetic tree (100 bootstraps) of XoxF and MxaF type methanol dehydrogenases in Methylophilaceae. Isolation sources of strains are indicated by different colours. Note that some strains have multiple copies of XoxF and MxaF. Several PQQ-dependent alcohol dehydrogenases (PQQ-ADHs type 9) and other distantly related dehydrogenases were used as outgroup.



Figure S10: Conserved methylotrophic pathways - H4F, methanol oxidation and RuMP

(a) Synteny of different methylotrophic pathways in Methylophilaceae. Genes involved in methanol oxidation, the H4F-pathway of formaldehyde oxidation, the RuMP cycle, and a complete MCA cycle are marked by different colours, as well as protein similarities (%) of individual genes. (b) Phylogenetic trees (100 bootstraps) of concatenated MxaRSACKL (left) and Zwf-GndA-PgI proteins (right). Isolation sources of strains are indicated by different colours.



Figure S11: Conserved methylotrophic pathways - methanol and formate oxidation

(a) Synteny of different methylotrophic pathways in Methylophilaceae. Genes involved in methanol oxidation, formate oxidation and synthesis of the cofactor molybdenum are marked by different colours, as well as protein similarities (%) of individual genes. (b) Phylogenetic trees (100 bootstraps) of concatenated XoxFJ (left), XoxG (middle) and FdhABCDE (right). Isolation sources of strains are indicated by different colours. Note that some strains have multiple copies of XoxG.



Figure S12: Horizontal gene transfers of different rhodopsins

RAxML trees (100 bootstraps) of (a) proteorohodopsins of '*Ca*. Methylopumilus spp.' and '*Ca*. Methylosemipumilus turicensis' and (b) xantho-like rhodopsins of '*Ca*. Methylopumilus spp.' and marine OM43. Protein sequence identities for the closest relatives are given below (a) or at the left side (b). Isolation sources of the closest relatives are indicated by different colours, actinorhodopsins were used as outgroup for both trees. (c) Amino acid alignment of xantho-like rhodopsins with residues involved in caroteinoid binding in *Salinibacter ruber* DSM13855 (Luecke et al. 2008, Imasheva et al. 2009). Letters c, g, k, and r stand for the chain, glucoside, keto group, and ring, respectively, and indicate part of the carotenoid in the vicinity of a residue. Note ring binding site at position 156 highlighted in red, where glycine is replaced with tryptophan in all strains indicating that the antenna might not bind.



Figure S13: Horizontal gene transfer of the NQR pathway in marine HTCC2181

(a) Arrangement and protein similarities of genes coding for the Na+ translocating NADHquinone reductase (NQR) marked in green colours and flanking genes. (b) RAxML tree (100 bootstraps) of concatenated NqrABCDEF proteins of the NQR pathway. Isolation sources of the strains are indicated by different colours.



Figure S14: Horizontal gene transfer of an ectoine biosynthesis pathway and a sodium:proline symporter in marine OM43

(a) Arrangement and protein similarities of genes coding for the ectoine biosynthesis pathway (*ectABC*) marked in green colours and flanking genes, including a sodium:proline symporter (marked in blue). Aspartate kinase *lysC* as well as aspartate-semialdehyde dehydrogenase *asd* are located elsewhere in the genomes of KB13 and MBRS-H7 and more similar to *lysC* and *asd* of other Methylophilaceae. RAxML trees (100 bootstraps) of (b) concatenated EctABC proteins of the ectoine biosynthesis pathway and (c) the sodium:proline symporter PutP. Isolation sources of the strains are indicated by different colours.



Figure S15: Horizontal gene transfer an oligopeptide permease in marine OM43

RAxML trees (100 bootstraps) of the dipeptide/tripeptide permease DtpD unique for the marine OM43 cluster (in bold). Isolation sources of the closest relatives are indicated in different colours.



c) GItT (sodium:dicarboxylate symporter) Nitrosomonas ureae (4) de la Nitrosomonas ursupi W/P_002294018
 Nitrosomonas p. ALI2[2]WP_013648258
 Nitrosomonas p. ALI2[2]WP_013648258
 Nitrosomonas p. Nitrosomonas p. Superior and the second program of the second progra freshwater marine hypersaline δ soil, plant associa human / animal a sludge, contamina mixed / unknown Nitrosomonas cryotolerans/WF_U2040000 Nitrosomonas communis (3) Nitrosomonas sp. Nm34/WF_090690249 Nitrosomonas sp. Nm38/WF_090578389 Nitrosomonas sp. Nm58/WF_090578389 human / animal associated, sludge, contaminated sources Nitrosospira sp. Nsp13/WP_0905
 Nitrosospira sp. Nsp13/WP_090906704
 Nitrosospira lacus/WP_004176361
 Nitrosospira sp. Nsp13/WP_004176361
 Nitrosospira sp. Nsp13/WP_0041512209
 Nitrosospira sp. Nsp13/WP_0041512209 ○ 95-100
○ 85-94
○ 65-84
● 45-64
● <45 bootstraps -\$ 'Ca. M. rimovensis' MMS-RI-1 beta proteobacterium KB13|WP 009851906 Methylophilales bacterium MBRS-H7|WP_048357097 'Ca. M. rimovensis' MMS-RI-1 bbrouseWP_107864229 aWP 009764229 osospira briensis|WP_074795828 Nitrosospira sp. Nsp44|WP_121080673 • Crenothrx polysporsfWP_08746050 Samethanical WP_018746050 Samethanical WP_018818742 Methylophilus sp. 11WP_19883683 Methylophilus sp. 21WP_01883683 Methylophilus sp. 21WP_024930899 Methylophilus sp. 21WP_024930899 Methylophilus sp. 21WP_024930899 Methylophilus sp. 21WP_024917412 Methylophilus sp. 21WP_024917412 Methylophilus sp. 21WP_024917412 Methylophilus mediardicus MMS-M-37 Methylophilus mediardicus MMS-M-3 Methylomon 091471901 8 Methylophius sp. Leaf41 (1)(VP _0552280) Methylophius sp. Leaf41 (1)(VP _0552280) Methylophius sp. Leaf41 (1)(VP _0552280) Methylophius sp. VPC2](VP _04535572 Methylophius sp. VPC2](VP _04535573 Methylophius sp. VPC2](VP _047545573 Methylophius sp. VPC2](VP _047545574 Methylophius sp. VPC2](VP _047545556 Methylophius sp. VPC2](VP _047545576 Methylophius sp. VPC2](VP _047454576 Methylophius sp. VPC2](VP _047454577 Methylophius sp. VPC2](VP _047454577 Methylophius sp. VPC2](VPC2)(VPC2) Methylophius sp. VPC2](VPC2)(VPC2) Methylophius sp. VPC2](VPC2)(VPC2) Methylophius sp. VPC2)(VPC2)(VPC2) Methylophius sp. VPC2)(VPC2)(VPC2) Methylophius sp. VPC2)(VPC2)(VPC2) Methylophius sp. VPC2)(VPC2)(VPC2) Me biovar El Tor str. N16961 NP_230813.2 5 -6 Methylobacillus rhizosphaerae Ca-68|WP 089 Methylobacillus flagellatus KT|WP_011478577 Ð Aburyos and a second se Pse cholerae 01 20 Pseudomonas pelagia (2)
 Pseudomonas selegais (2)
 Pseudomonas selegais (2)
 Pseudomonas selegais (2)
 Celivibrio - Celivibrio - SPSB006/WP, 08736856
 Celivibrio - SPSB006/WP, 087468656
 Celivibrio - SPSB006/WP, 087468656
 Celivibrio - matusiWP_094985080 Pseudomo Pseudomo Vibrio -6 0.4 Cytophagaceee (11)

Hallscomenbacter hydrossiWP_013767316
Gladecola sp. THG-37(WP_124746036
GyzaelWP_081477250
Gladecola sp. THG-37(WP_013020582
Gallionella capafernformans/WP_013020582
Gallionella capafernformans/WP_013020582
Comamonadaceae bacterium //NGI Gd282/WP_110289521
Comamonadaceae bacterium //NGI Gd282/WP_110289521
Methylophilales bacterium //NGI Gd28408
Methylophilales bacterium //NGI Gd2851865
orans hindinis/WP_08148327
Dysgonomonas ag/Inati/Inca/WP_10131284
Dysgonomonas ag/Inati/Inca/WP_10131284
Dysgonomonas sp. HGC4/WP_003046
Elusimicobum minitum/WP_07403346 Cytophagaceae (11) d) NhaP2 (sodium:proton antiporter) source marine hypersaline soil, plant associated human / animal associated sludge, contaminated sites mixed / unknown Mucinivorane sd ○ 95-100 ○ 85-94 ○ 65-84 ● 45-64 ● <45 De Imonas formicexedens/WP_076003246 Elusimicrobium minutum/WP_012415469 robium sp. An273/WP_087286733 Enhygromyxa salina б Elusim Enhygromyza salinalWP_108395005 Methanomassilicoccus luminyensik3WP_001579759 Prostheocchiots (4) Chiorobium phaeobacteroides(WP_012474328 Chiorobrepton thalassium)WP_012501177 Nitrosopumilacee (12) us sp. PAO1WPU 05732787
 Noviherbaspillum bag. Not180/WP_057202827

 Noviherbaspillum bag. Not180/WP_057202827

 Herbaspillum bag. Not180/WP_057202827

 Methylophilus bg. Ledf450/WP_05830205

 Methylophilus bg. Ledf450/WP_08178522

 Methylophilus bg. Ledf450/WP_081785260

 Methylophilus bg. Ledf450/WP_049637675

 Thiobacillus dentificians/WP_09753183

 Thiobacillus dentificians/WP_125180274

 Marinnicobum agarifyticum/WP_036187000

 Acararus comunityWP_10848014
 other Methylophilaceae osa PAO1|NP_252576.1 б cus communis/WP_108948614 iferula thiophila/WP_124951849 ferax ferrireducens/WP_011463698 rax sp. OTU1/WP_114973947 itrificans/WP_009207846 Sulfuricella denitrificans/WP Sulfuriferula multivorans/WP_12470 pphilus medardicus MMS-M-51 ophilus medardicus MMS_M-34 ophilus medardicus MMS-M-33 omonas clara/WP_124552630 Sulfu **Methy lophilaceae** 0H31|WP_024930378 Q8|WP_047515293 5|WP_029147858 sphaerae CBMB127|WP_091472049 other k Methy Methy Burkh L Bu



Figure S16: Horizontal gene transfer of sodium transporters in marine OM43

RAxML trees (100 bootstraps) of (a) the sodium:alanine symporter AlsT, the sodium:acetate symporter ActP, (c) the sodium:dicarboxylate symporter GltT, (d) the sodium:proton antiporter NhaP2, and (e) another putative sodium:proton antiporter (NhaE-like). Marine OM43 and '*Ca*. Methylopumilus' strains are highlighted in bold, isolation sources of the closest relatives are indicated in different colours.