Microbial metagenomes from three aquifers in the Fennoscandian shield terrestrial deep biosphere reveal metabolic partitioning among populations

Xiaofen Wu, Karin Holmfeldt, Valerie Hubalek, Daniel Lundin,

Mats Åström, Stefan Bertilsson & Mark Dopson

Supplementary Files

Supplementary File 1. Planktonic cells > 0.1 µm (16S rRNA gene tag sequencing) or 0.22 µm (16S rRNA gene tag sequencing and metagenomes) were collected by connecting a High-Pressure Stainless Steel Filter Holder (Millipore) with a downstream needle valve and pressure gauge directly to the borehole and the water allowed to flow under *in situ* pressure. Three borehole section volumes were allowed to pass before placing a mixed cellulose ester membrane filter (Merck Millipore) in the holder. After an appropriate volume of water had been filtered under *in situ* pressure the filter was placed in a sterile cryogenic tube (Thermo Scientific) and immediately frozen in liquid nitrogen. The tubes were transported to the laboratory (on the same day) and stored at -80°C until processing. DNA was directly extracted from the filters using the MO BIO PowerWater DNA isolation kit according to the manufacturer's instructions except that the final DNA was re-suspended in 50 - 60 µL of eluent. The quality and quantity of the extracted DNA were analyzed with a Thermo Scientific Nanodrop 2000 and Qubit 2.0 Fluorometer (Life Technologies), respectively. A negative control was carried out for the > 0.22 µm fraction by extracting DNA from a filter that that resulted in a DNA concentration below the Qubit 2.0 Fluorometer's detection limit (0.348 ng).

16S rRNA gene tag sequencing was carried out in duplicate by amplifying a portion of the 16S rRNA gene utilizing primers 341F and 805R (Herlemann et al 2011) according to published procedures (Hugerth et al 2014). Sequencing was carried out at the Science for Life Laboratory, Sweden (www.scilifelab.se) on the Illumina MiSeq platform as previously published (Lindh et al 2015).

Samples for the < 0.22 µm metagenome were collected in sterile containers before transport to the laboratory in < 1 h. Community DNA was prepared according to the iron chloride precipitation method (John et al 2011) with slight modification. Briefly, 0.22 µm filtrate (47 mm: cellulose nitrate, Sartorius; 142 mm: Isopore polycarbonate, Millipore) water

was mixed with $FeCl_3$ (1 mg/L final solution) and incubated 1 to 4 h in room temperature. The $FeCl_3$ aggregates were collected onto 0.8 µm filters (142 mm; Isopore, Millipore) and stored at 4°C in dark until use. One mL Ascorbate-EDTA buffer (0.2 M ascorbic acid, 0.1 M EDTA, 0.2 M $MgCl_2$, 0.125 M Tris, pH 6) was added per 1 L of $FeCl_3$-water solution on the filter. The solution was concentrated using Amicon Ultra-15 centrifugal device (50K, Millipore) at 2860 × g and treated with DNase I (50 U, Thermo Scientific) at 37°C for 30 min and EDTA (0.15 M) at 65°C for 15 min. Samples were treated overnight with Proteinase K (2 mg/mL) at 37°C and DNA was extracted with Wizard® PCR Preps DNA purification System (Promega).

Metagenome libraries were prepared using the ThruPlex DNA-seq Kit with 96 dual indexes (Rubricon, MI, USA) according to the manufacturer's instructions with the following changes. The protocols were automated using an Agilent NGS workstation (Agilent, CA, USA) using purification steps as previously described (Borgström et al 2011, Lundin et al 2010), clustered using onboard clustering. Metagenomes for the > 0.22 µm fraction was sequenced on two MiSeq flow cells (MCS 2.4.1.3/RTA 1.18.54) with a 2 × 301 setup in High Output mode (one library for each water type in each flow cell). Metagenomes for the < 0.22 µm fraction was sequenced on two HiSeq flow cells (HCS2.0.12.0/RTA 1.17.21.3) with a 2 x 101 setup in High Output mode (one library for each water type in each flow cell). Libraries MMS_B, UMS_A, and OSS_A were sequenced twice and all reads have been included in the analysis. Bcl to Fastq conversion was performed using bcl2Fastq v1.8.3 from the CASAVA software suite. The quality scale was Sanger / phred33 / Illumina 1.8+. All sequencing was carried out at the Science for Life Laboratory in Stockholm, Sweden.

Cells for counts were filtered through a 0.03 µm Whatman Nuclepore® Track-Etch membrane (Ø 13 mm) and the cells fixed with a final concentration of 1% (vol/vol) 0.02 µm sterile filtered formaline (37% formaldehyde solution) and stored in the dark at 4°C until

analysis. Cells on the filters were stained by placing the dry filter (sample side up) on a staining solution for 15 min in the dark. The staining solution was prepared by mixing 97.5 µl filtered deionized water with 2.5 µl of the 10% SYBR Green I working solution (the SYBR Green I stock was diluted 1:10 in 0.02 µm pre-filtered deionized water). After staining, the filter backside was carefully blotted on a Kimwipe, placed on a glass slide and 30 µl of anti-fade mounting solution was added (50% PBS/ 50% glycerol with 0.1% p-phenylenediamine), and overlaid with a cover slip. Microscope fields (> 30 fields) were randomly counted in an inverted Leica DMi8 fluorescence microscope at 1000 fold magnification.

Samples (7 µl) for electron micrographs were placed onto a formvar-coated 200-mesh grid and incubated for 2 min. The grid was dried by touching it transversely against a filter paper. 7 µl of 2% uranyl acetate was added onto the grid and incubated for 10 s before drying as describe above. The grid was further dried for > 15 min before placing in a FEI Tecnai G2 transmission electron microscope.

Bioinformatic analysis. 16S rRNA gene reads were de-multiplexed and analyzed via the UPARSE pipeline (Edgar 2013), species annotated by comparison to the SILVA database (SILVA 119) (Quast et al 2013) using SINA (Pruesse et al 2007), and final analysis in Explicet 2.10.5 (Robertson et al 2013).

Bioinformatic analysis of the metagenomes was carried out by first checking the quality of the metagenome paired reads using FastQC (version 0.11.2) before removing the adapters and low quality reads with SeqPrep and Sickle (version 1.210)(Joshi and Fass 2011) for planktonic cells > 0.22 µm and Trimmomatic(Bolger et al 2014) for cells < 0.22 µm. The trimmed reads were assembled using the *de novo* assembler Ray (version 2.3.1 & version 2.3.0) (Boisvert et al 2012) with k-mer sizes of 31, 41, 51, 61, 71, and 81. Newbler (version 2.6) was then used to merge all the contigs generated from different k-mers. The assembled contigs were binned using CONCOCT (version 0.3.0) (Alneberg et al 2014) which was

individually based on both composition and coverage. The assembled contigs >20 Kb were first cut up into 10 Kb segments using the script cut_up_fasta.py from the CONCOCT scripts package. All raw metagenome reads from cells < 0.22 µm or trimmed metagenome reads from cells > 0.22 µm were mapped back onto each assembled contigs, respectively using CONCOCT script map-bowtie2-markduplicates.sh which simultaneously removed the PCR duplicates and created the coverage file. A coverage table for each contig per sample was generated using the CONCOCT script gen_input_table.py. The coverage table and the cut-up contigs were used to run CONCOCT at with variable length thresholds (500, 700, 1000, 2000, 3000 nucleotides for cells < 0.22 µm and 1000, 2000, 3000, 4000, and 5000 nucleotides for cells > 0.22 µm). The clustering output by CONCOCT was evaluated using Prodigal (version 2.60) (Hyatt et al 2012), RPS-Blast (version 2.2.28+), and scripts from CONCOCT (COG_table.py and COGplot.R) (see figures below). The length threshold with the highest number of approved bins (which have ≤5 missing and ≤2 multiple copies of CONCOCT single copy genes) was chosen. The approved bins were then continuously extracted using the CONCOCT script extract_fasta_bins.py. Genome quality was assessed using GC content with checkM v0.9.7 (Parks et al 2014). Approved bins between planktonic bacterial and ultra-small bacteria were aligned using dnadiff from MUMmer (version 3.23) package (Kurtz et al 2004) (Supplementary File 6). All metagenomes were mapped back separately onto each approved bin using map-bowtie2-markduplicates.sh (Supplementary File 11).

Taxonomic analysis on approved bins was carried out using Phylosift v1.0.1 (Darling et al 2014). The most related species for each bin from the Phylosift taxa summary table were manually extracted from the NCBI genome database and PATRIC genome database. PhyloPhlan (version 0.99) (Segata et al 2013) was then used to make a phylogenetic tree for all the approved bins and the related species and visualized using GraPhlAn (Fig. 2) or

Archaeopteryx (Supplementary Fig. 2). Each extracted approved bin was annotated using Prokka v1.10 as a single genome (Seemann 2014). Functional genes and enzyme commission numbers were extracted and analyzed against the Kyoto Encyclopedia of Genes and Genomes (KEGG)(Kanehisa et al 2014) and MetaCyc databases (Caspi et al 2014).

Comparison of genome sizes with sequenced relatives. The estimated genome size of the approved bins were calculated by dividing the number of CONCOCT single copy genes by 36 (the total single copied genes) by the number of identified single copy genes and multiplying the number by the number of base pairs in the bin (Supplementary File 4). This value was compared to the size of its nearest sequenced neighbor based upon the phylogenetic tree in Supplementary File 7 and a ratio calculated by dividing the estimated genome size by the genome size of the sequenced closest relative from the NCBI database. A negative value means the metagenome bin estimated genome size was smaller than the reference and a positive number suggests a larger genome size.

Calculation of percentage fraction of small cell sizes. The percentage small cell size in the three waters was estimated by comparing the 16S rRNA gene sequences in the > 0.22 µm filters and > 0.1 µm filters. The calculation was based upon the assumption that an OTU less represented (% of reads) in the total (> 0.1 µm) size fraction was caused by dilution with cells passing the 0.22 µm filter. Hence, the difference in relative representation between the two fractions gives the percent small cells, i.e. if there is a 10-fold difference, small cells will make up 90 % of the total community. This calculation was carried out by first selecting OTUs represented by more than 0.5% in the > 0.1 µm fraction (giving at least 70 reads per OTU in the duplicates) before the OTU reads in the > 0.22 µm fraction was divided by the reads in the > 0.1 µm fraction (Supplementary File 9). Most estimated OTU enrichment factors converged at a value of 2, suggesting that overall 50% of all bacterial cells passed the 0.22 µm membrane filter (Supplementary File 9), but a conspicuous enrichment of a single

OTU in the undefined mixed water may also indicate preferential amplification of certain OTUs that may have influenced the results. This could also lead to an underestimate of the fraction of small cells in the undefined mixed water.

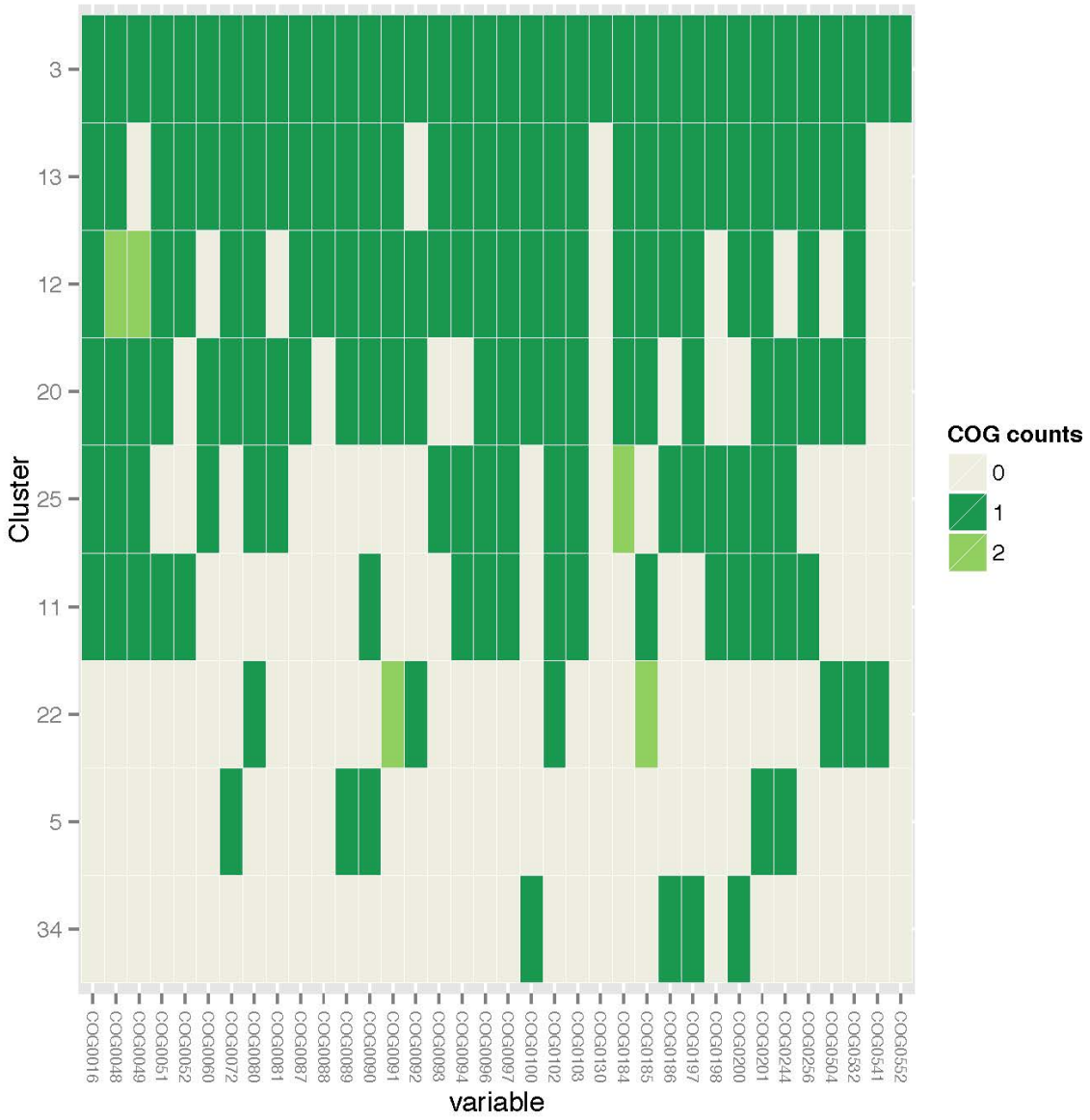
Supplementary File 2. Details of the sampling and number of sequenced reads for the duplicate metagenomes from the three water types.

Sample ID	Sampling site	Water type	Water volume (L)	Total DNA amount (µg)	Number paired reads[a]	Number trimmed reads[a]	No. of Contigs (> 1000 bp)	Number of bins
> 0.22 µm								
MML_A	SA1229A	Modern marine	~50	1.75	8870782	8681684	7149	2
MML_B	SA1229A	Modern marine	~50	1.76	10338898	10192205	10115	2
UML_A	KA3105A_4	Undefined mixed	~140	2.09	8559387	8353204	9213	1
UML_B	KA3105A_4	Undefined mixed	~95	2.20	8180574	8059950	8910	2
OSL_A	KA3385A_1	Old saline	~280	0.17	8962216	8740784	23363	13
OSL_B	KA3385A_1	Old saline	~600	0.24	8111809	7966299	6576	4
< 0.22 µm								
MMS_A	SA1229A	Modern marine	~20	0.46	81090898	58524266	14824	8
MMS_B	SA1229A	Modern marine	~20	0.51	77980371	52314968	17500	6
UMS_A	KA3105A_4	Undefined mixed	~20	0.73	53690975	48997973	27168	6
UMS_B	KA3105A_4	Undefined mixed	~20	0.61	78483374	69598083	30202	9
OSS_A	KA3385A_1	Old saline	~30	0.35	68063273	60224718	26991	7
OSS_B	KA3385A_1	Old saline	~30	0.35	74998491	66488978	29093	9

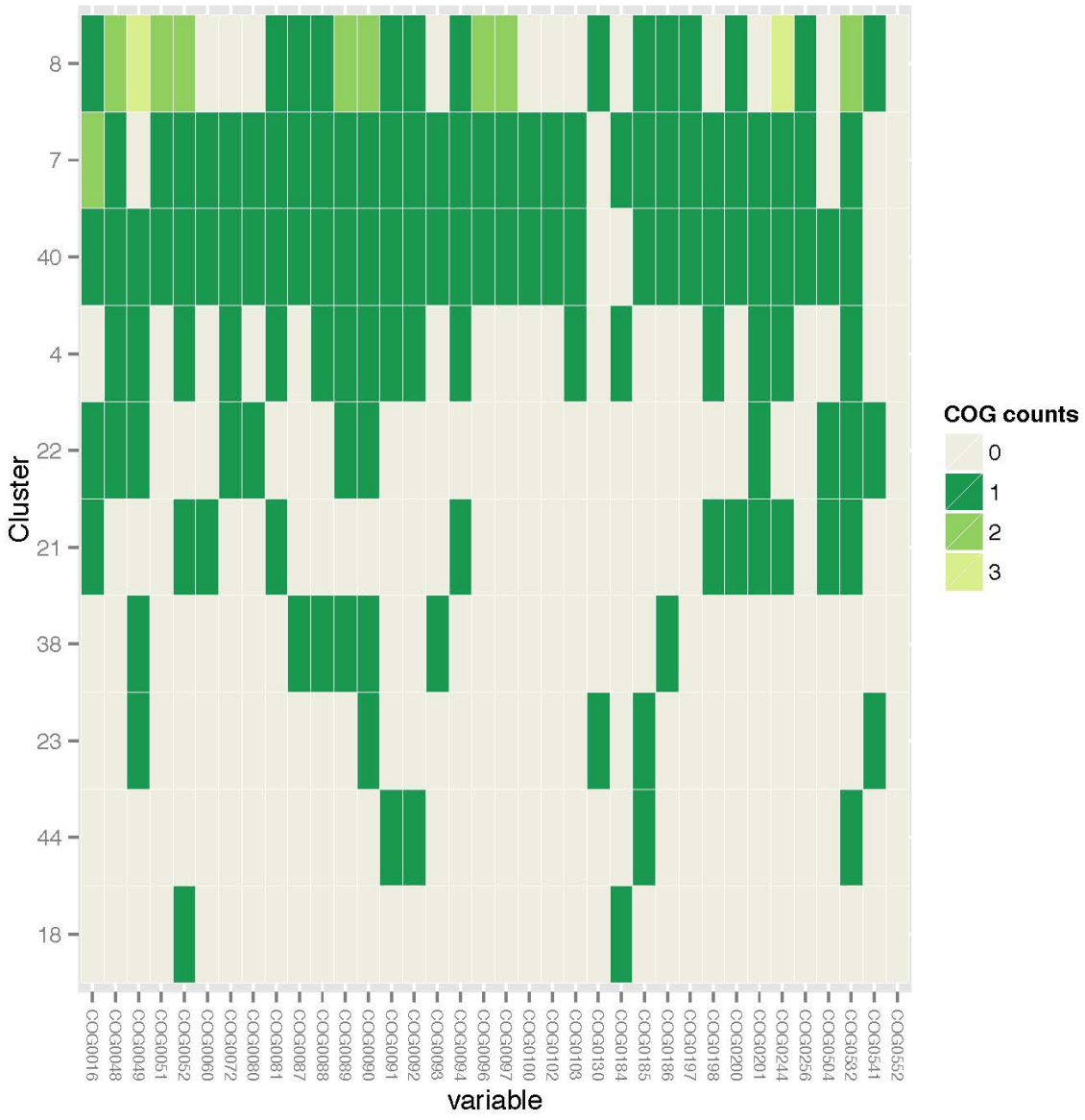
[a]Values for MMS_B, UMS_A, and OSS_A are totals for the duplicate sequencing.

Supplementary File 3. Evaluation of the near complete CONCOCT bins for the selected contig length cutoffs.

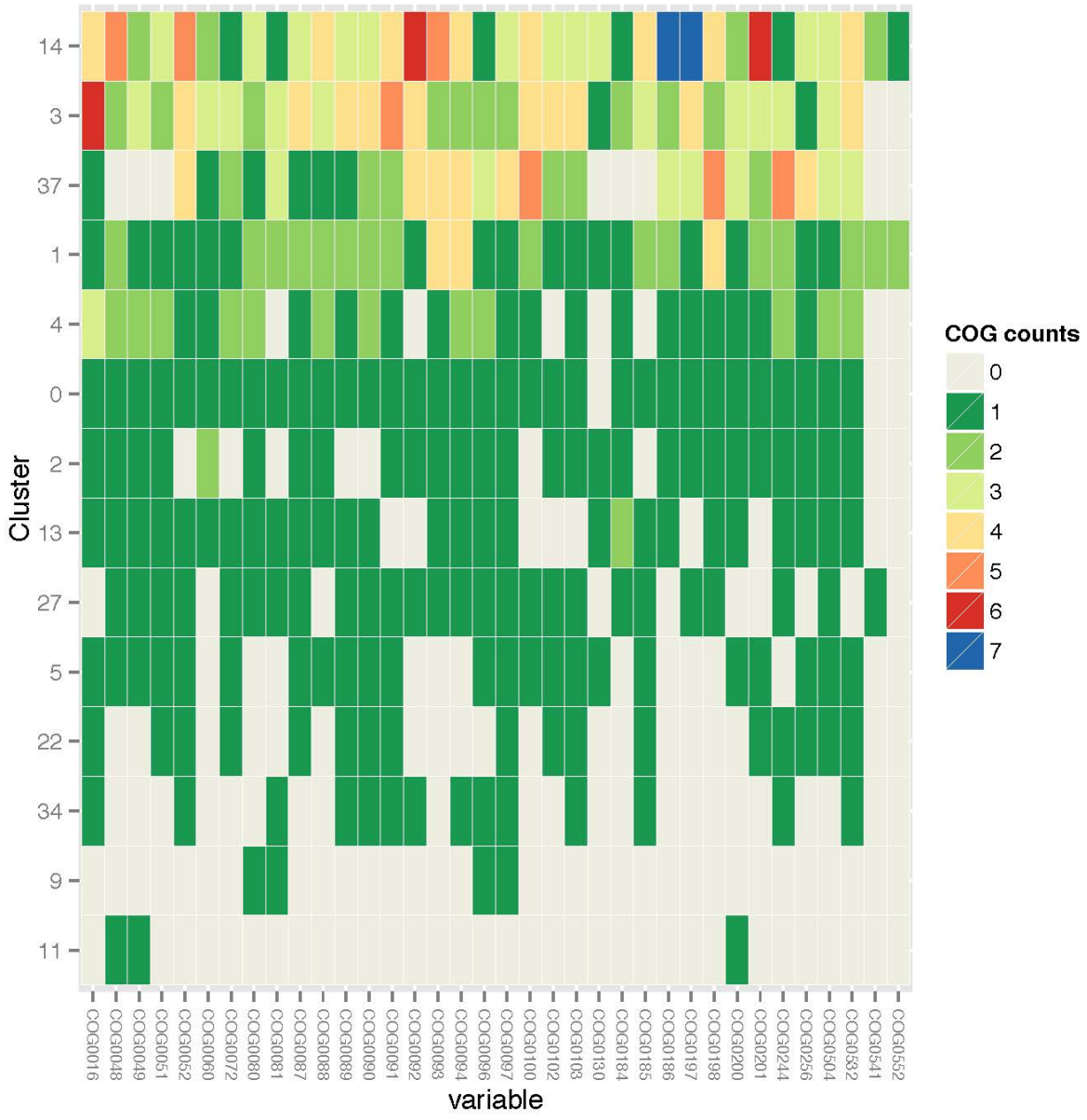
> 0.22 µm Modern marine: MML_A



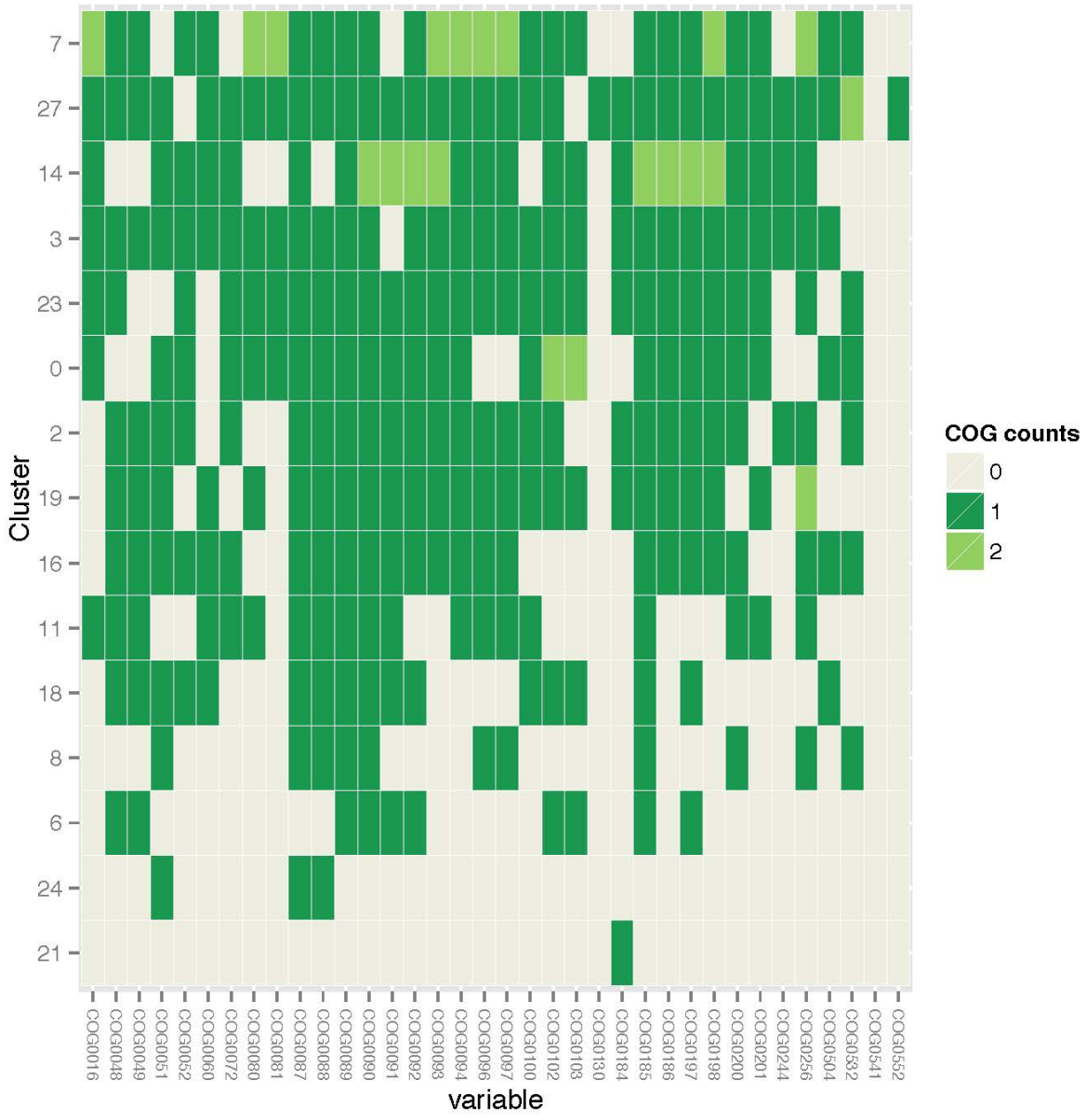
> 0.22 µm Modern marine: MML_B



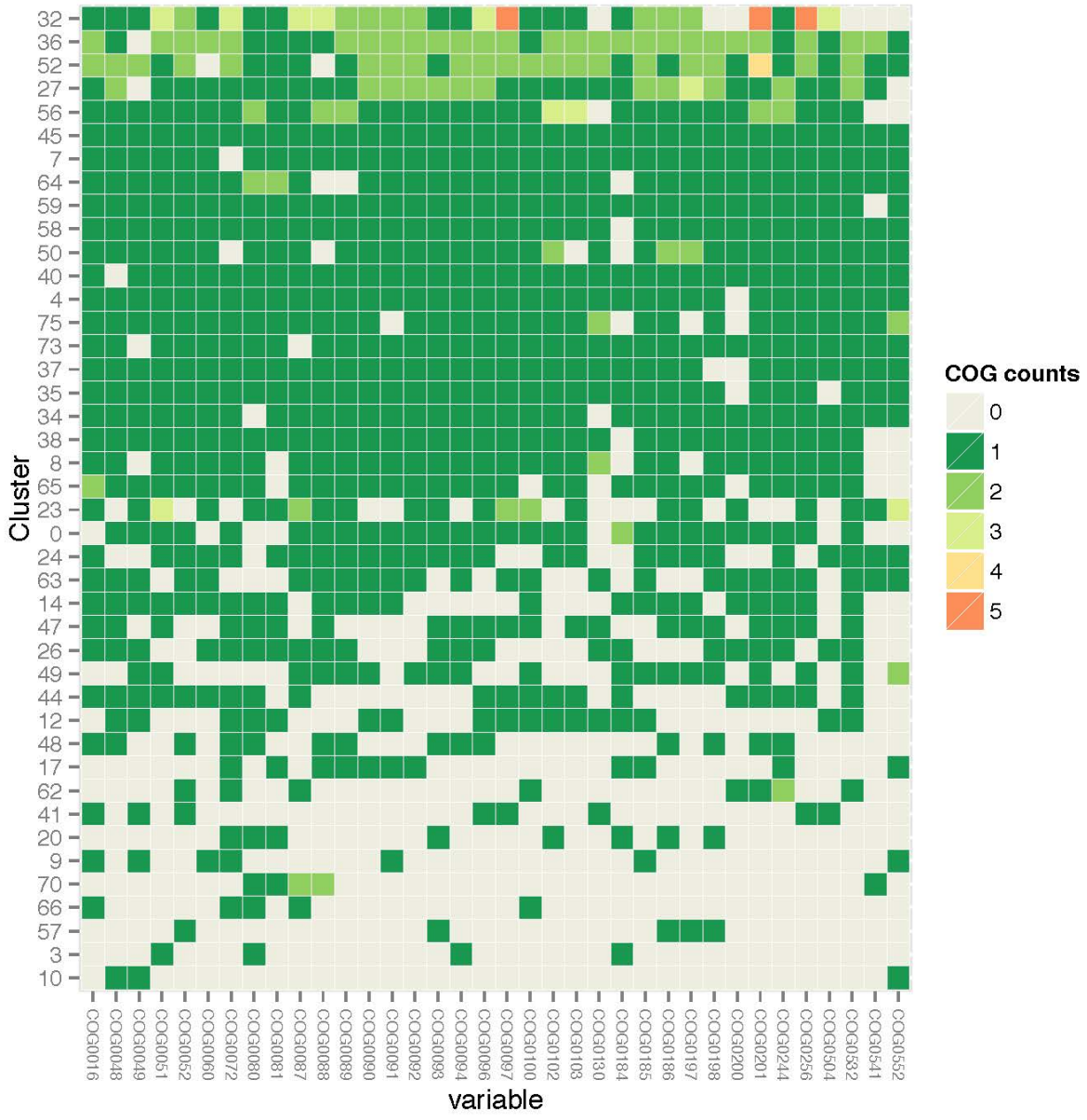
> 0.22 µm Undefined mixed: UML_A



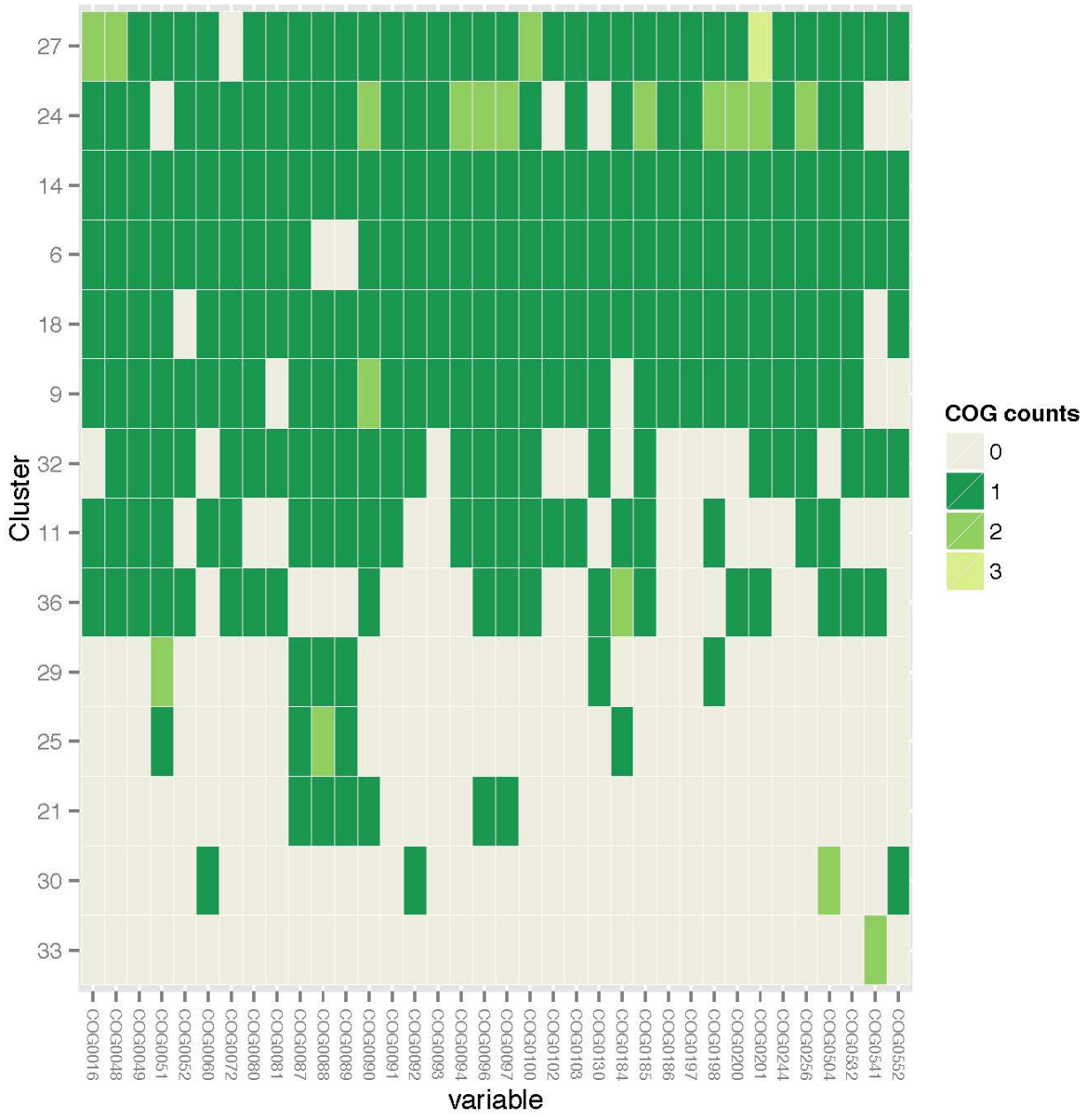
> 0.22 µm Undefined mixed: UML_B



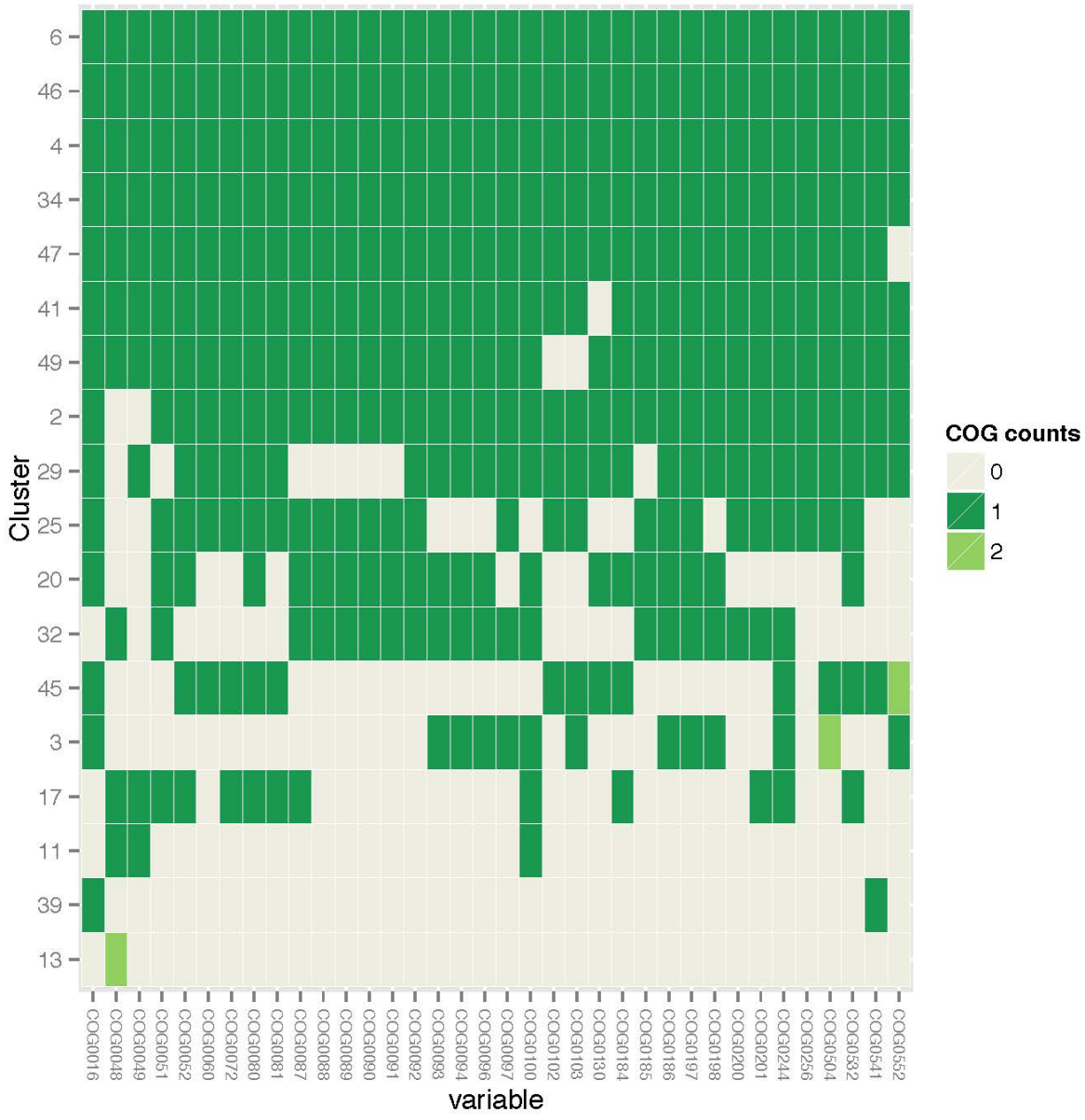
> 0.22 µm Old saline: OSL_A



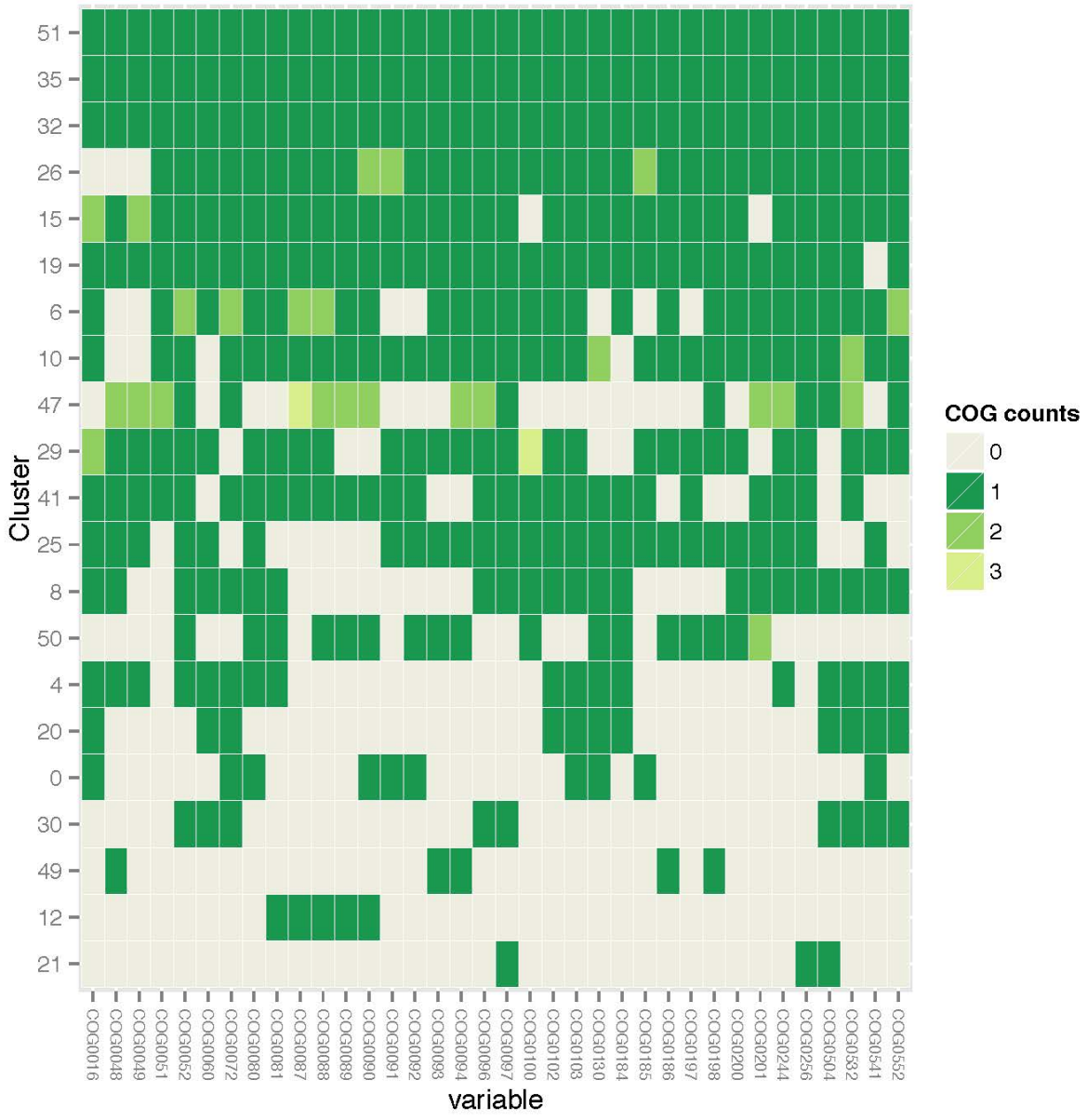
> 0.22 µm Old saline: OSL_B



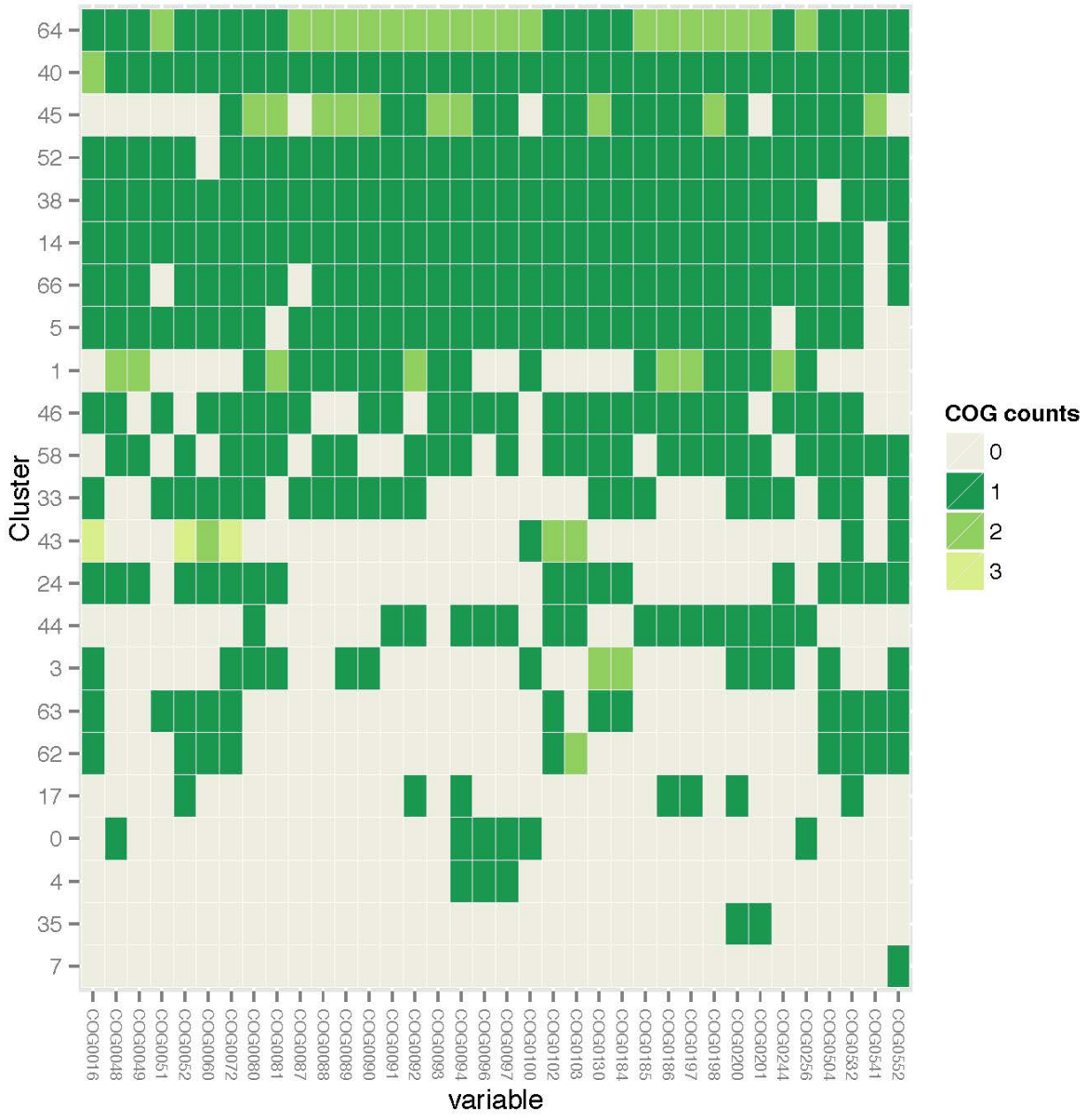
< 0.22 µm Modern marine: MMS_A



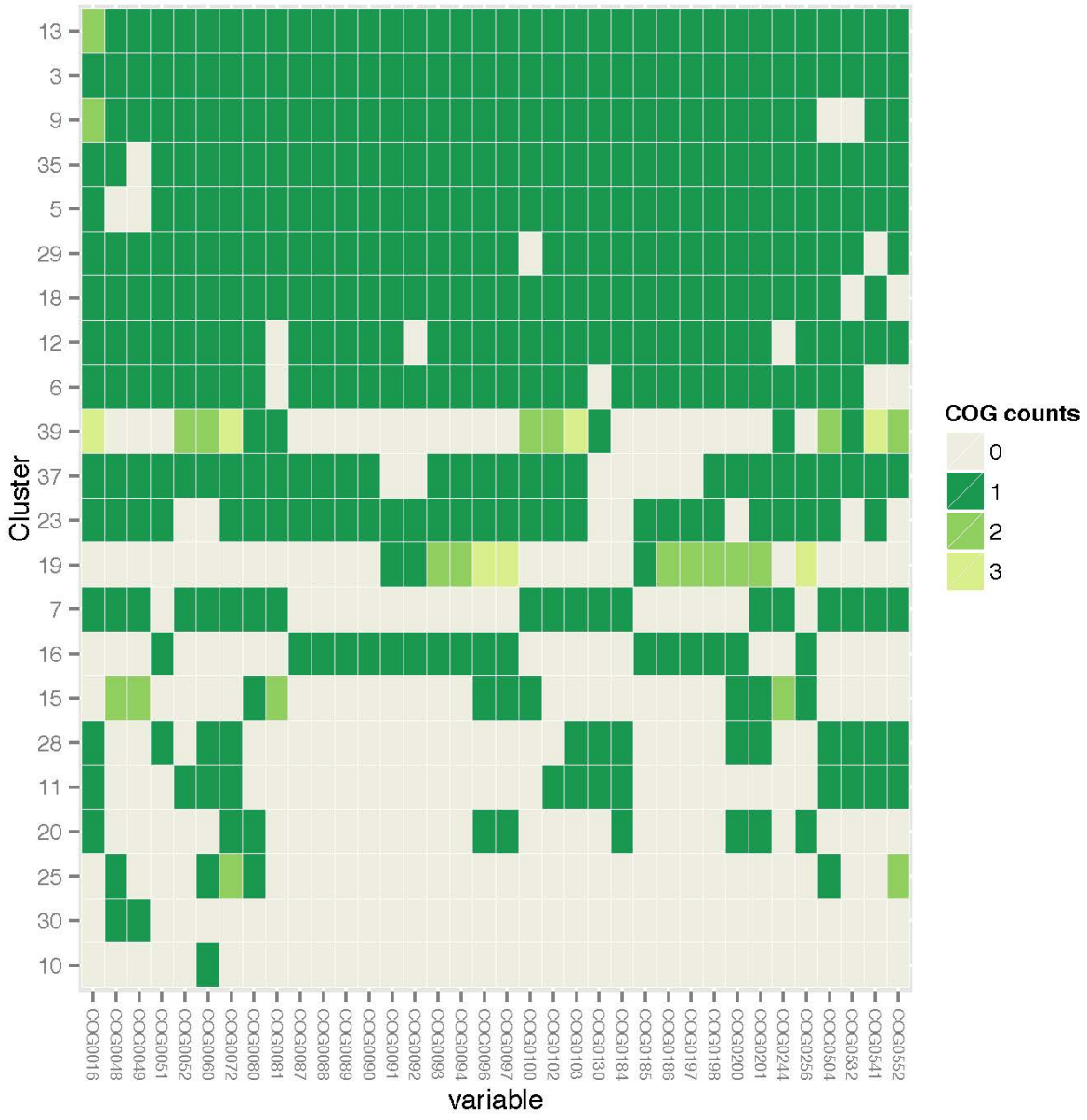
< 0.22 µm Modern marine: MMS_B



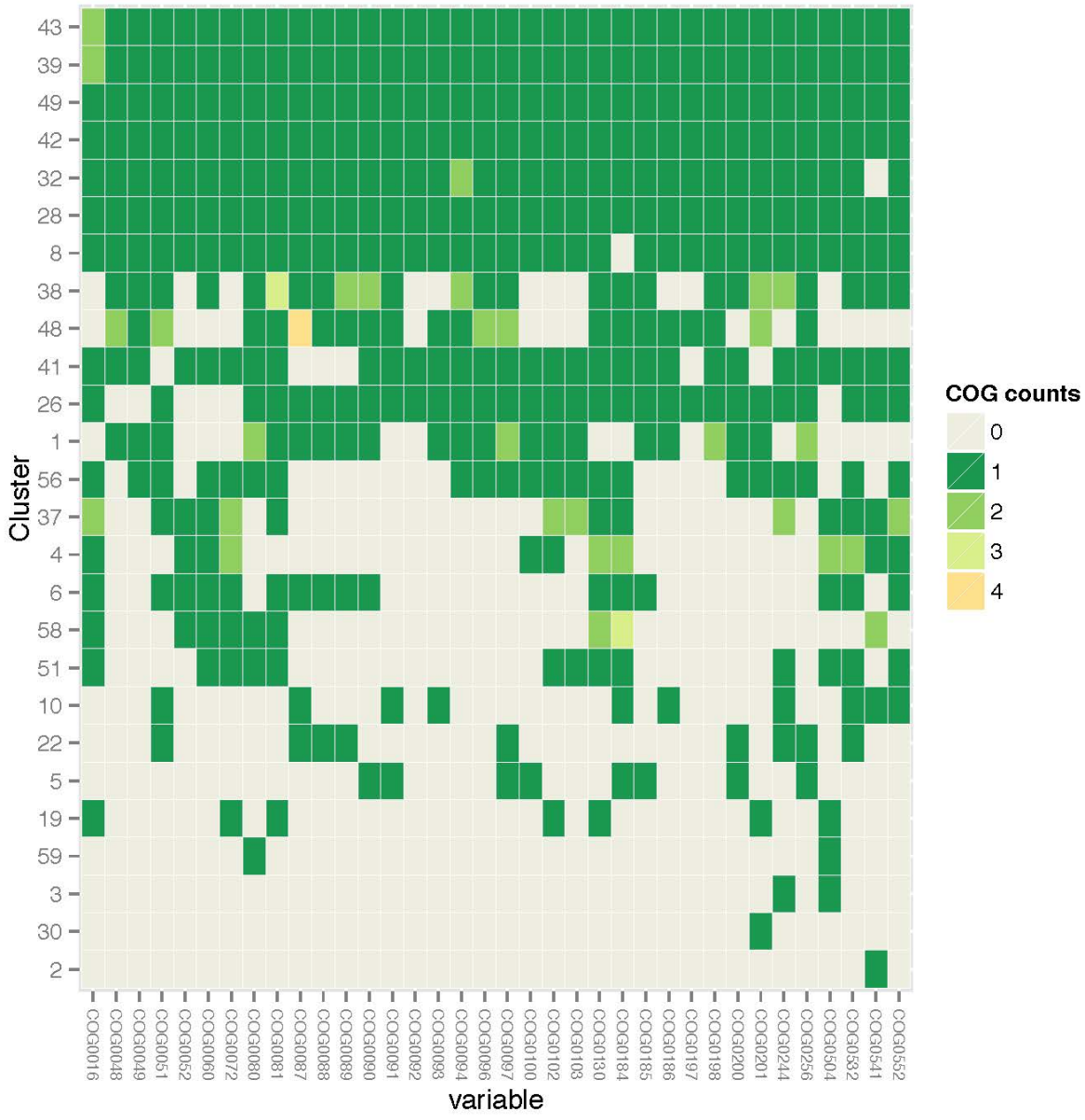
< 0.22 µm Undefined mixed: UMS_A



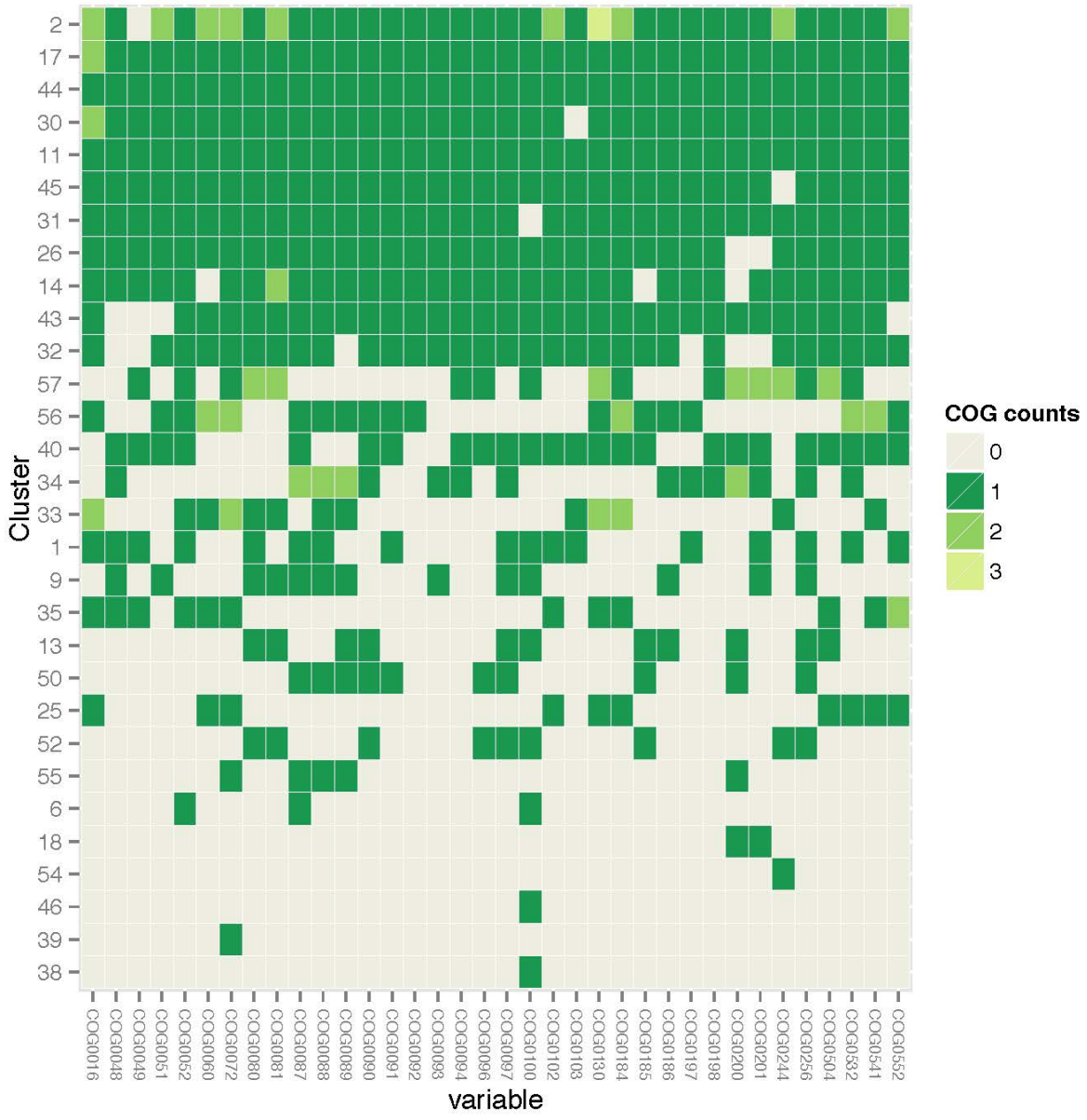
< 0.22 µm Undefined mixed: UMS_B



< 0.22 µm Old saline: OSS_A



< 0.22 µm Old saline: OSS_B



Supplementary File 4. Sequence information for each approved phylogenetic bin.

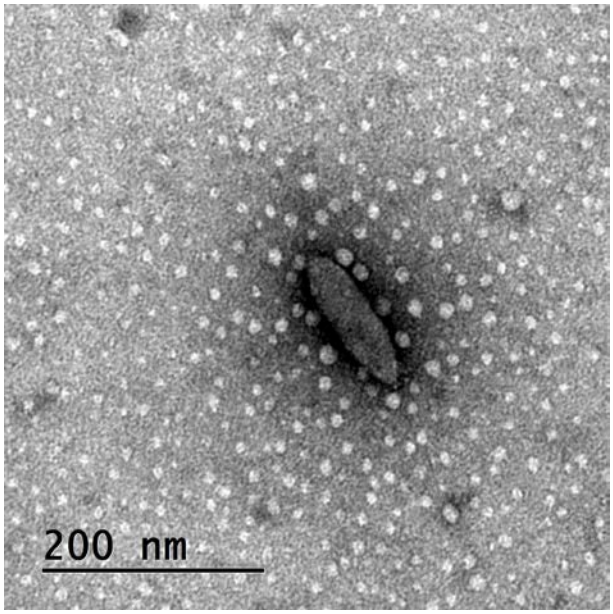
Bins	Estimated genome size (bp)	No. contigs	n50[a]	GC%	No. single copy genes	No. duplicated single copy genes	Contig length cutoff	Mapped reads (%)[b]	Suggested phylogeny (NCBI)
> 0.22 µm									
Modern marine									
MML_A1	1184545	322	90	31.0	31/36	0/36	1000	2.23	Candidate OD1
MML_B1	768338	185	50	29.7	32/36	0/36	1000	1.31	Candidate OD1
MML_A2	4809034	447	164	31.3	36/36	0/36	1000	5.43	*Chlorobi/Ignavibacteriae*
MML_B2	1815935	594	166	48.5	31/36	1/36	1000	0.50	Unclassified bacteria
Undefined mixed									
UML_A1	1069396	299	79	30.8	33/36	0/36	1000	1.42	Candidate OD1
UML_B1	3047695	462	141	57.7	33/36	1/36	3000	2.47	δ-*Proteobacteria*
UML_B2	807120	158	62	42.8	31/36	0/36	3000	0.28	Unclassified bacteria
Old saline									
OSL_A1	4676176	441	208	60.9	35/36	0/36	1000	4.85	β-*Proteobacteria*
OSL_B1	4527143	433	204	60.9	36/36	0/36	1000	57.73	β-*Proteobacteria*
OSL_A2	4709108	817	186	36.6	33/36	2/36	1000	6.19	*Chlorobi/Ignavibacteriae*
OSL_B2	3945601	643	157	36.6	34/36	0/36	1000	2.95	*Chlorobi/Ignavibacteriae*
OSL_A3	1977375	236	68	44.9	34/36	0/36	1000	1.08	*Planctomycetes*
OSL_B3	3586970	329	123	63.5	34/36	0/36	1000	2.72	β-*Proteobacteria*
OSL_A4	854997	90	29	30.7	34/36	0/36	1000	0.77	Unclassified archaea
OSL_B4	922260	201	50	35.0	32/36	1/36	1000	0.90	Unclassified bacteria
OSL_A5	1476292	199	55	35.1	34/36	0/36	1000	0.74	Candidate OD1
OSL_A6	1225838	121	42	46.8	33/36	0/36	1000	0.77	Candidate OP11
OSL_A7	3755959	489	139	50.8	35/36	0/36	1000	1.85	Unclassified bacteria

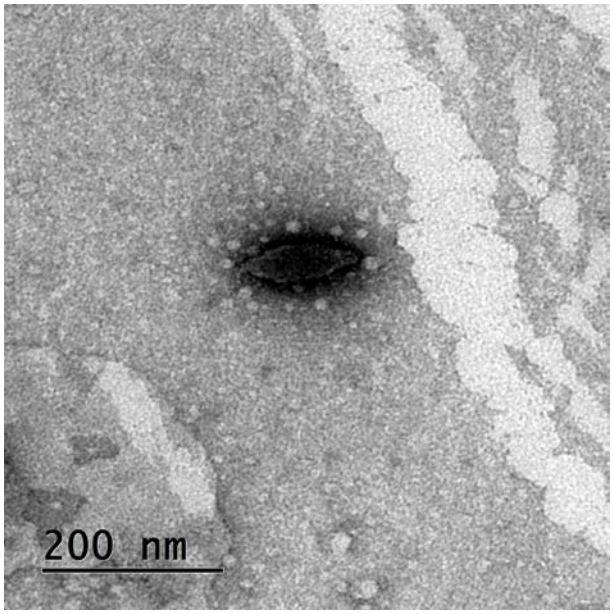
OSL_A8	2602682	262	90	50.3	36/36	0/36	1000	1.31	*Chloroflexi*
OSL_A9	1533041	172	52	34.4	35/36	0/36	1000	1.01	Unclassified archaea
OSL_A10	4149137	562	146	63.5	35/36	0/36	1000	2.48	δ-*Proteobacteria*
OSL_A11	2899667	482	130	57.6	34/36	0/36	1000	1.22	δ-*Proteobacteria*
OSL_A12	640835	173	40	37.5	32/36	2/36	1000	0.25	Unclassified archaea
OSL_A13	2536971	257	82	52.6	35/36	0/36	1000	1.26	Unclassified bacteria
< 0.22 µm									
Modern marine									
MMS_A1	4817225	625	174	63.3	34/36	0/36	2000	1.13	β-*Proteobacteria*
MMS_B1	5088473	1065	219	63.3	34/36	2/36	1000	0.97	β-*Proteobacteria*
MMS_A2	3263998	318	155	52.9	36/36	0/36	2000	14.80	β-*Proteobacteria*
MMS_B2	3858532	380	185	52.1	36/36	0/36	1000	14.43	β-*Proteobacteria*
MMS_A3	4950005	486	235	64.5	36/36	0/36	2000	1.96	α-*Proteobacteria*
MMS_B3	5047710	475	222	64.4	35/36	0/36	1000	1.44	α-*Proteobacteria*
MMS_A4	7192220	679	309	59.0	36/36	0/36	2000	24.11	γ-*Proteobacteria*
MMS_B4	7321248	703	333	59.0	36/36	0/36	1000	22.78	γ-*Proteobacteria*
MMS_A5	3883770	373	174	60.4	36/36	0/36	2000	0.90	β-*Proteobacteria*
MMS_B5	5961411	617	245	59.4	36/36	0/36	1000	1.34	β-*Proteobacteria*
MMS_A6	5300198	484	177	61.1	34/36	0/36	2000	0.79	α-*Proteobacteria*
MMS_A7	4129394	374	155	66.3	35/36	0/36	2000	1.24	β-*Proteobacteria*
MMS_A8	3841005	420	135	64.4	35/36	0/36	2000	0.57	α-*Proteobacteria*
MMS_B6	2750588	1136	331	65.3	32/36	2/36	1000	0.36	γ-*Proteobacteria*
Undefined mixed									
UMS_A1	5387357	510	245	63.1	35/36	0/36	1000	3.34	β-*Proteobacteria*
UMS_B1	5352342	521	253	63.1	36/36	0/36	3000	3.57	β-*Proteobacteria*
UMS_A2	3260094	441	119	52.9	35/36	0/36	1000	0.60	β-*Proteobacteria*
UMS_B2	3317403	290	128	52.9	34/36	0/36	3000	0.59	β-*Proteobacteria*
UMS_A3	4644512	426	177	42.5	36/36	1/36	1000	2.63	γ-*Proteobacteria*

UMS_B3	4658983	417	170	42.6	36/36	1/36	3000	2.63	γ-*Proteobacteria*
UMS_A4	772466	70	35	47.8	32/36	0/36	1000	0.40	Candidate OD1
UMS_B4	765744	66	32	47.8	32/36	0/36	3000	0.38	Candidate OD1
UMS_A5	4719207	453	190	60.8	35/36	0/36	1000	1.39	α-*Proteobacteria*
UMS_B5	4261742	383	179	60.1	34/36	0/36	3000	1.27	α-*Proteobacteria*
UMS_A6	3645268	930	209	52.7	33/36	0/36	1000	0.74	γ-*Proteobacteria*
UMS_B6	3652465	462	132	52.7	34/36	1/36	3000	0.73	γ-*Proteobacteria*
UMS_B7	5807134	495	195	68.8	33/36	0/36	3000	1.37	α-*Proteobacteria*
UMS_B8	3342490	287	114	67.5	34/36	0/36	3000	1.59	α-*Proteobacteria*
UMS_B9	4082676	359	145	57.9	35/36	0/36	3000	0.63	α-*Proteobacteria*
Old saline									
OSS_A1	5335906	522	253	63.1	36/36	0/36	1000	2.35	β-*Proteobacteria*
OSS_B1	5438778	532	257	63.1	36/36	0/36	1000	2.60	β-*Proteobacteria*
OSS_A2	4693420	441	182	42.6	36/36	1/36	1000	1.89	γ-*Proteobacteria*
OSS_B2	4816940	456	177	42.5	35/36	1/36	1000	1.64	γ-*Proteobacteria*
OSS_A3	3273504	309	130	67.6	36/36	0/36	1000	3.51	α-*Proteobacteria*
OSS_B3	3359936	318	122	67.5	35/36	0/36	1000	3.88	α-*Proteobacteria*
OSS_A4	5404527	519	245	62.8	36/36	0/36	1000	2.33	α-*Proteobacteria*
OSS_B4	5561662	543	259	62.7	36/36	0/36	1000	3.35	α-*Proteobacteria*
OSS_A5	2648571	285	98	69.5	36/36	1/36	1000	0.76	*Actinobacteria*
OSS_B5	2663712	246	106	69.6	36/36	1/36	1000	0.91	*Actinobacteria*
OSS_A6	5013912	779	194	46.1	35/36	1/36	1000	0.69	γ-*Proteobacteria*
OSS_B6	4961327	643	178	46.2	35/36	0/36	1000	0.67	γ-*Proteobacteria*
OSS_A7	3225612	1141	324	34.6	35/36	0/36	1000	0.35	*Bacteroidetes*
OSS_B7	3094981	612	135	65.9	33/36	1/36	1000	0.81	α-*Proteobacteria*
OSS_B8	5171545	464	176	64.5	34/36	0/36	1000	1.08	α-*Proteobacteria*
OSS_B9	4971160	422	189	63.8	32/36	0/36	1000	1.35	α-*Proteobacteria*

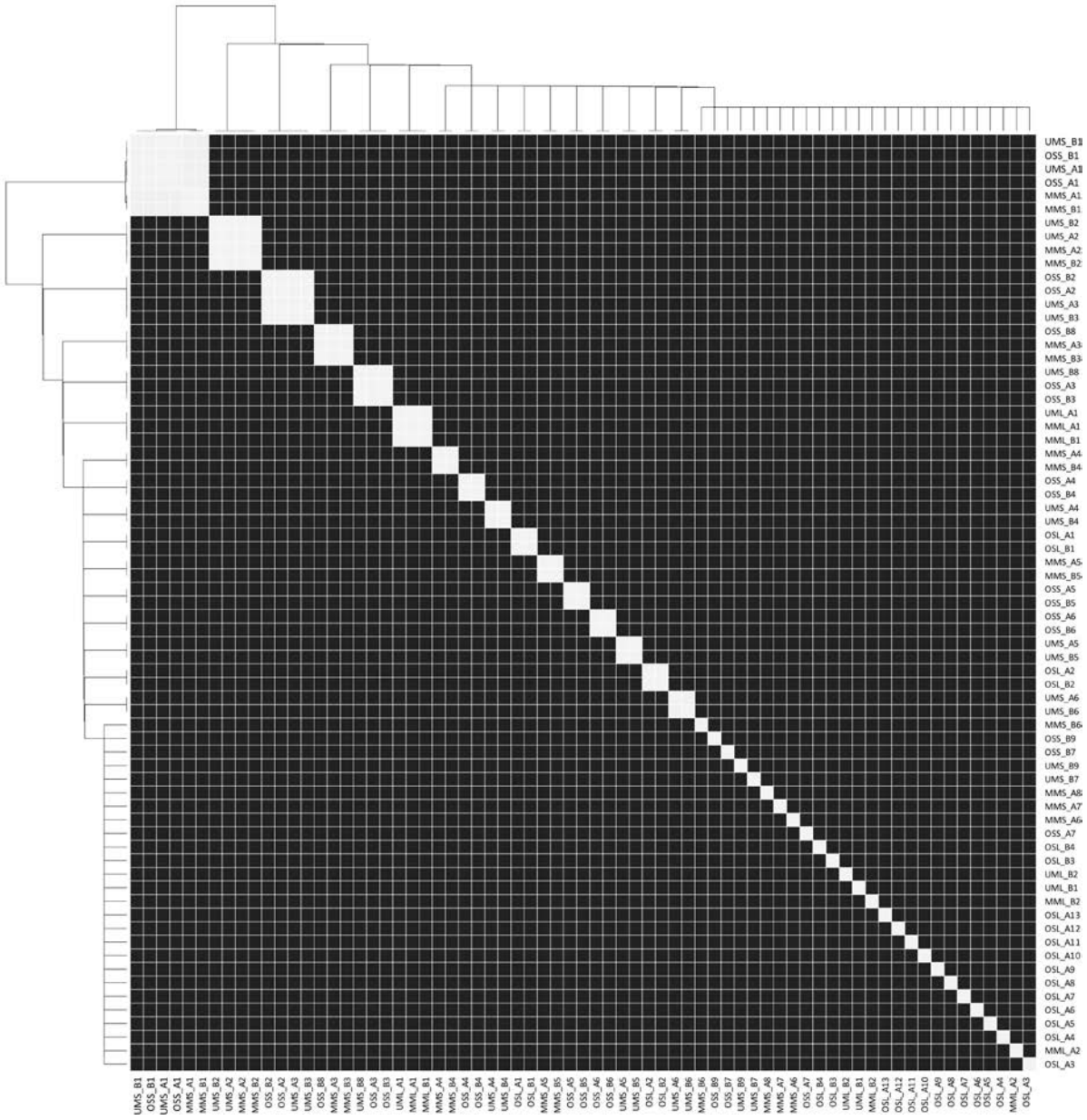
[a] n50: number of the largest contigs that sum up to 50% of the total sum of bases
[b]Mapped reads (%) from respective metagenomes. Percentages for MMS_B, UMS_A, and OSS_A are average values for the duplicate sequencing.

Supplementary File 5. Transmission electron micrographs of modern marine water showing cell sizes < 0.22 µm. The water samples were fixed with a final concentration of 1% formalin and stained with 2% uranyl acetate.

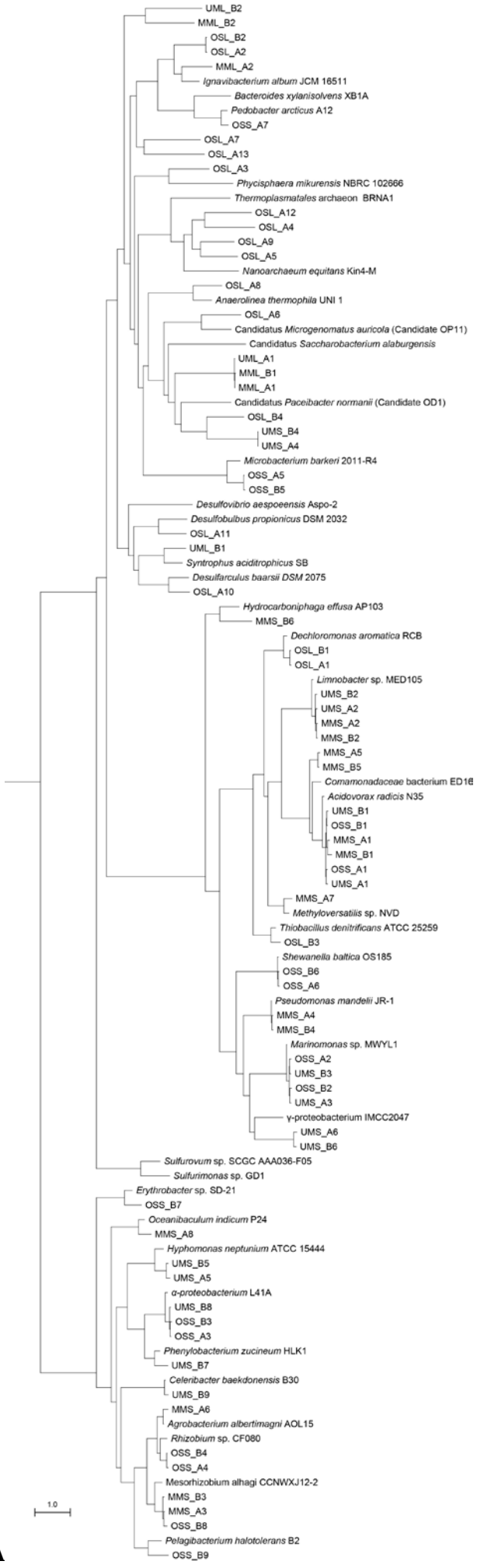




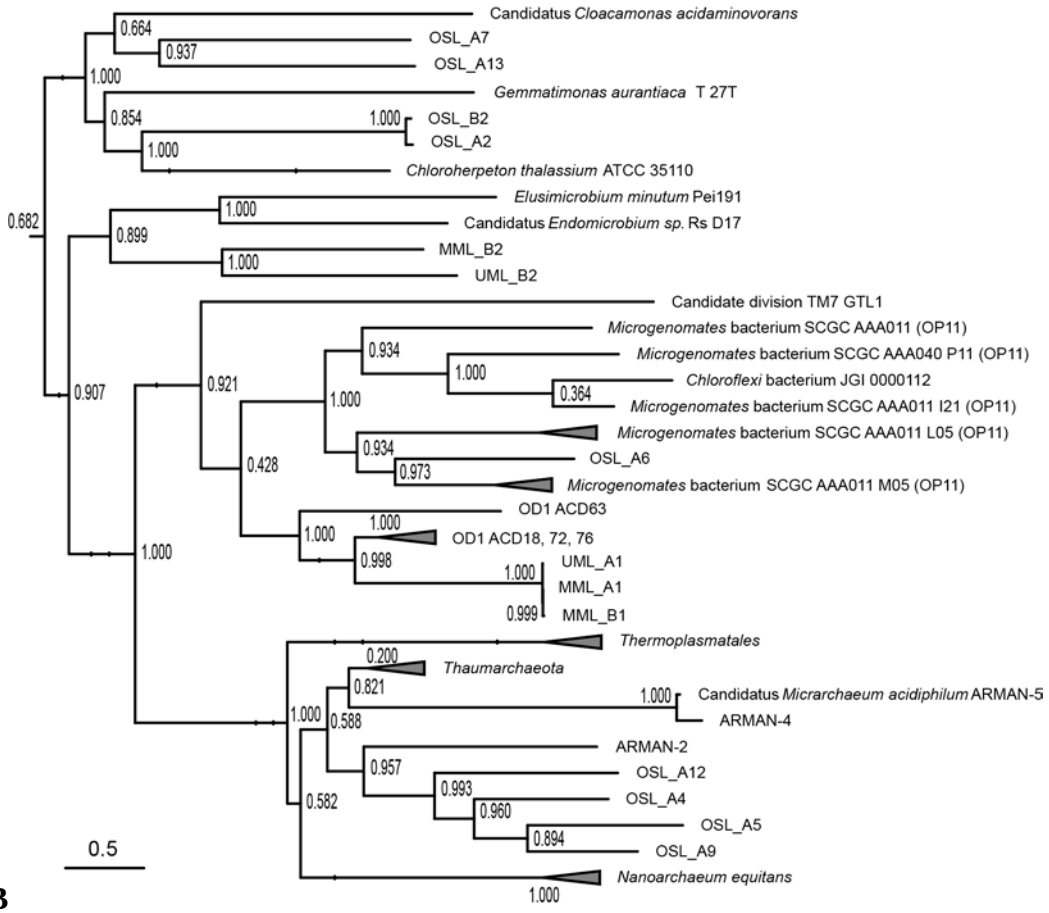
Supplementary File 6. Alignment from all metagenomes showing total nucleotide similarity when contigs from the CONCOCT bins were aligned to each other. Boxes show bins with > 50% nucleotide similarity (white) and < 50% similarity (black).



Supplementary File 7. Whole genome phylogenetic tree of the relationship between the CONCOCT bins visualized by Archaeopteryx (A) and a detailed tree of the poorly classified species (B). The scale bar equals 1.0% in (A) and number of nucleotide substitutions per site (B).


UML_B2
MML_B2
OSL_B2
OSL_A2
MML_A2
Ignavibacterium album JCM 16511
Bacteroides xylanisolvens XB1A
Pedobacter arcticus A12
OSS_A7
OSL_A7
OSL_A13
OSL_A3
Phycisphaera mikurensis NBRC 102666
Thermoplasmatales archaeon BRNA1
OSL_A12
OSL_A4
OSL_A9
OSL_A5
Nanoarchaeum equitans Kin4-M
OSL_A8
Anaerolinea thermophila UNI 1
OSL_A6
Candidatus Microgenomatus aurícola (Candidate OP11)
Candidatus Saccharobacterium alaburgensis
UML_A1
MML_B1
MML_A1
Candidatus Paceibacter normanii (Candidate OD1)
OSL_B4
UMS_B4
UMS_A4
Microbacterium barkeri 2011-R4
OSS_A5
OSS_B5
Desulfovibrio aespoeensis Aspo-2
Desulfobulbus propionicus DSM 2032
OSL_A11
UML_B1
Syntrophus aciditrophicus SB
Desulfarculus baarsii DSM 2075
OSL_A10
Hydrocarboniphaga effusa AP103
MMS_B6
Dechloromonas aromatica RCB
OSL_B1
OSL_A1
Limnobacter sp. MED105
UMS_B2
UMS_A2
MMS_A2
MMS_B2
MMS_A5
MMS_B5
Comamonadaceae bacterium ED16
Acidovorax radicis N35
UMS_B1
OSS_B1
MMS_A1
MMS_B1
OSS_A1
UMS_A1
MMS_A7
Methyloversatilis sp. NVD
Thiobacillus denitrificans ATCC 25259
OSL_B3
Shewanella baltica OS185
OSS_B6
OSS_A6
Pseudomonas mandelii JR-1
MMS_A4
MMS_B4
Marinomonas sp. MWYL1
OSS_A2
UMS_B3
OSS_B2
UMS_A3
γ-proteobacterium IMCC2047
UMS_A6
UMS_B6
Sulfurovum sp. SCGC AAA036-F05
Sulfurimonas sp. GD1
Erythrobacter sp. SD-21
OSS_B7
Oceanibaculum indicum P24
MMS_A8
Hyphomonas neptunium ATCC 15444
UMS_B5
UMS_A5
α-proteobacterium L41A
UMS_B8
OSS_B3
OSS_A3
Phenylobacterium zucineum HLK1
UMS_B7
Celeribacter baekdonensis B30
UMS_B9
MMS_A6
Agrobacterium albertimagni AOL15
Rhizobium sp. CF080
OSS_B4
OSS_A4
Mesorhizobium alhagi CCNWXJ12-2
MMS_B3
MMS_A3
OSS_B8
Pelagibacterium halotolerans B2
OSS_B9
1.0

A


Candidatus Cloacamonas acidaminovorans
0.664
OSL_A7
0.937
OSL_A13
1.000
Gemmatimonas aurantiaca T 27T
0.854
1.000
OSL_B2
OSL_A2
1.000
Chloroherpeton thalassium ATCC 35110
Elusimicrobium minutum Pei191
1.000
0.682
Candidatus Endomicrobium sp. Rs D17
0.899
MML_B2
1.000
UML_B2
Candidate division TM7 GTL1
Microgenomates bacterium SCGC AAA011 (OP11)
0.934
Microgenomates bacterium SCGC AAA040 P11 (OP11)
1.000
Chloroflexi bacterium JGI 0000112
0.907
0.921
0.364
Microgenomates bacterium SCGC AAA011 I21 (OP11)
1.000
Microgenomates bacterium SCGC AAA011 L05 (OP11)
0.934
OSL_A6
0.428
0.973
Microgenomates bacterium SCGC AAA011 M05 (OP11)
OD1 ACD63
1.000
1.000
OD1 ACD18, 72, 76
1.000
0.998
UML_A1
1.000
MML_A1
0.999
MML_B1
Thermoplasmatales
0.200
Thaumarchaeota
0.821
1.000
Candidatus Micrarchaeum acidiphilum ARMAN-5
1.000
0.588
ARMAN-4
ARMAN-2
0.957
OSL_A12
0.993
OSL_A4
0.582
0.960
OSL_A5
0.894
OSL_A9
0.5
Nanoarchaeum equitans
1.000

B

Supplementary File 8. Metabolic characteristics identified in the metagenomic bins from the three water types. The listed pathways are based upon BioCyc (http://biocyc.org/) and KEGG (http://www.genome.jp/kegg/). Additional pathways that were searched for but are not listed as they were negative in all cases include: ferric reduction as a terminal electron acceptor; methanogenesis; anaerobic ammonia oxidation; and the reductive TCA cycle, incomplete TCA cycle, 3-hydroxypropanoate cycle, and reductive acetyl CoA pathway for CO_2 fixation.

		Metabolic pathways								Respiration & electron transport					Nutrient fixation		Nutrients
Group	**Bins**	**Pyruvate[a]**	**Acetyl-CoA to butyrate II[b]**	**H_2 prod[c]**	**Ethanol[d]**	**Formate to CO_2[e]**	**Methane[f]**	**H_2[g]**	**Sulfur oxidation[h]**	**Sulfate reduction[i]**	**Sulfur Reduction[j]**	**Nitrate reduction[k]**	**Denitrifi-cation[l]**	**Rnf Complex[m]**	**CBB cycle[n]**	**N_2 Fixation[o]**	**Uptake[p]**
Modern marine > 0.22 µm																	
Group II	MML_A2	Acetate	-[q]	II/ IV	II (Pu)/ III (Pu)[r]	-	-	III	Tet (Pr)[s] / thio	-	-	-	-	+[t]	-	-	Phosphate (Pr)
	MML_B2	Acetate	-	IV	-	-	-	III	Tet (Pr)	-	-	-	-	+	-	-	Sulfate, phosphate, zinc
Ungrouped	MML_A1	-	-	II / III	-	-	-	-	-	-	-	-	-	-	-	-	-
	MML_B1	-	-	II /III	-	-	-	-	-	-	-	-	-	-	-	-	-
Modern marine < 0.22 µm																	
Group I	MMS_A2	Ethanol 2/3	-	-	I	+	-	-	Thio	-	-	Nitrite	-	-	+	-	Phosphate, sulfate, nitrate (Pr), taurine (Pr)
	MMS_B2	Ethanol 2/3	-	-	I	+	-	-	Thio	-	-	Nitrite	-	-	+	-	Sulfate, phosphate, nitrate (Pr), taurine
	MMS_A5	Propionate 5/7	-	II	-	+	-	-	Thio	-	-	Nitrite	-	-	+	-	Phosphate, sulfate, zinc, nitrate (Pr), taurine
	MMS_B5	Ethanol 2/3 & propionate 5/7	-	II	I	+	-	-	Thio	-	-	Nitrite	-	-	+	-	Phosphate, sulfate, nitrate (Pr), taurine
Group II	MMS_A6	Lactate	-	IV	II / III / IV	Pu	-	III	Thio	-	-	-	-	-	-	+	Phosphate, sulfate, iron, zinc, manganese, nitrate (Pr), nickel (Pr), taurine (Pr)
	MMS_A8	Propionate 5/7	-	IV	-	+	-	III	-	-	-	-	-	-	-	-	Phosphate, sulfate, iron, zinc, nitrate (Pr)
Group III	MMS_A3	Propionate 5/7	–	-	-	Pu	+	-	Tet (Pr) / thio	-	-	Nitrite	-	-	-	-	Sulfate, phosphate, manganese, taurine

	MMS_B3	Propionate 5/7	-	-	II (Pu) / III (Pu) / IV (Pu)	Pu	+	-	Tet (Pr) / thio	-	-	Nitrite	-	-	-	-	Sulfate, phosphate, manganese, iron, taurine
Group IV	MMS_A4	Lactate	-	IV	II (Pu) / III (Pu) / IV (Pu)	+	-	III	Thio	-	-	Ammonia	N_2 (Pr)	+	-	-	Phosphate, sulfate, iron, zinc, manganese, nitrate (Pr), taurine
	MMS_B4	Lactate	-	IV	II (Pu) / III (Pu) / IV (Pu)	+	-	III	Thio	-	-	Ammonia	N_2 Pr	+	-	-	Phosphate, sulfate, iron, zinc, manganese, nitrate (Pr), taurine
Group V	MMS_A7	Ethanol 2/3	-	II / IV	I / II / III / IV	+	-	III	Tet (Pr) / thio	-	-	Nitrite	-	+	+	+	Phosphate, sulfate, iron, nitrate (Pr), taurine
Ungrouped	MMS_A1	Propionate 5/7	-	-	II (Pu) / III (Pu) / IV (Pu)	+	-	-	Thio	-	-	Ammonia	-	-	-	-	Phosphate, sulfate, nickel, nitrate (Pr), taurine
	MMS_B1	Propionate 5/7	-	-	II (Pu) / III (Pu) / IV (Pu)	+	-	-	Thio	-	-	Ammonia	-	-	-	-	Phosphate, sulfate, nickel, nitrate (Pr)
	MMS_B6	-	-	-	III	-	-	-	Thio	-	-	-	-	+	-	-	Phosphate, nitrate (Pr)

		Metabolic pathways								Respiration & electron transport					Nutrient fixation		Nutrients
Group	**Bins**	**Pyruvate**	**Acetyl-CoA to butyrate II**	**H_2 prod**	**Ethanol**	**Formate to CO_2**	**Methane**	**H_2**	**Sulfur oxidation**	**Sulfate reduction**	**Sulfur reduction**	**Nitrate reduction**	**Denitrifi-cation**	**Rnf complex[I]**	**CBB cycle**	**N_2 fixation**	**Uptake**
Undefined mixed > 0.22 µm																	
Group V	UML_B1	Propionate 6/7	-	IV	-	Pu	-	III	Tet (Pr)	+	-	Nitrite	-	-	-	+	Sulfate, phosphate, manganese, taurine
Ungrouped	UML_A1	-	-	II / III	-	-	-	-	-	-	-	-	-	-	-	-	-
	UML_B2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Undefined mixed < 0.22 µm																	
Group I	UMS_A2	Ethanol 2/3 & propionate 5/7	-	-	I	-	-	-	Thio	-	-	Nitrite	-	-	+	-	Sulfate, phosphate, taurine, nitrate (Pr)
	UMS_B2	-	-	-	-	+	-	-	Thio	-	-	Nitrite	-	-	+	-	Sulfate, phosphate, taurine, nitrate (Pr)
Group II	UMS_A1	Lactate	-	-	II / III / IV	+	-	-	Thio	-	-	Ammonia	-	-	-	-	Sulfate, phosphate, taurine
	UMS_B1	Propionate 6/7	-	IV	II / III / IV	+	-	III	Thio	-	-	Ammonia	-	-	-	-	Sulfate, phosphate, taurine
Group III	UMS_A3	Ethanol 2/3 & lactate	+	-	I / II / III	-	-	-	Thio	-	-	-	-	+	-	-	Sulfate, phosphate, iron, zinc, nitrate (Pr)
	UMS_B3	Ethanol 2/3 & lactate	+	-	I	-	-	-	Thio	-	-	-	-	+	-	-	Sulfate, phosphate, iron, zinc, nitrate (Pr), taurine
Group IV	UMS_B9	Propionate 5/7 & lactate	+	IV	II / III / IV	Pu	-	III	Thio	-	-	-	-	-	+	-	Sulfate, phosphate, zinc, nitrate (Pr), taurine
Ungrouped	UMS_A4	-	-	-	-	-	-	-	Thio	-	-	-	-	-	-	-	-
	UMS_B4	-	-	-	-	-	-	-	Thio	-	-	-	-	-	-	-	-
	UMS_A5	Propionate 5/7 & lactate	-	-	-	-	-	-	Thio	-	-	-	-	-	-	-	Sulfate, phosphate, iron
	UMS_B5	Propionate 5/7	-	-	-	-	-	-	Thio	-	-	Nitrite	-	-	-	-	Sulfate, phosphate, Iron

	UMS_A6	-	-	-	-	-	-	-	Thio	-	-	-	-	-	-	-	Sulfate, phosphate, iron, taurine, zinc
Ungrouped	UMS_B6	-	-	-	-	-	-	-	Thio	-	-	-	-	-	-	-	Sulfate, phosphate, zinc
	UMS_B7	-	-	-	II (Pu)/III (Pu)/IV (Pu)	-	-	-	Thio	-	-	-	-	-	-	-	Sulfate, phosphate, iron, taurine
	UMS_B8	Lactate	-	-	II / III	-	-	-	-	-	-	-	-	-	-	-	Phosphate

		Metabolic pathways								Respiration & electron transport					Nutrient fixation		Nutrients
Group Number	**Bins**	**Pyruvate**	**Acetyl-CoA to butyrate II**	**H_2 prod**	**Ethanol**	**Formate to CO_2**	**Methane**	**H_2 oxidation**	**Sulfur oxidation**	**Sulfate reduction**	**Sulfur reduction**	**Nitrate reduction**	**Denitrifi-cation**	**Rnf complex[I]**	**CBB cycle**	**N_2 fixation**	**Uptake**
Old saline > 0.22 µm																	
Group I	OSL_A1	Propionate 6/7	+	II	III	+	+	-	Thio	-	-	Nitrite / Ammonia (Pr)	N_2 (Pr)	+	+	+	Sulfate, phosphate, Nitrate, iron
	OSL_B1	Propionate 6/7	+	II	III	+	-	-	Thio	-	-	Nitrite / Ammonia (Pr)	N_2 (Pr)	+	+	+	Sulfate, phosphate, nitrate, iron
	OSL_B3	-	-	II	-	Pr	+	-	Tet / thio	+	-	Nitrite	N_2O (Pr)	+	+	+	Sulfate, phosphate, nitrate, Zinc (Pr), taurine
Group II	OSL_A4	Acetate	-	-	-	-	-	-	-	-	-	-	-	-	+	-	Phosphate
	OSL_A12	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-
Group III	OSL_A5	Acetate	-	II / IV	-	-	-	III	-	-	+	-	-	-	-	-	-
	OSL_A9	Acetate	-	II / III / IV	-	-	-	III	-	-	+	-	-	-	+	+	Sulfate, phosphate
Group IV	OSL_A2	-	-	II / IV	-	-	-	III	Tet (Pr) / thio	-	Pr	Nitrite / Ammonia (Pr)	-	-	-	-	Phosphate
	OSL_B2	Propionate 6/7	-	II / III / IV	-	-	-	III	Tet (Pr) / thio	-	-	Nitrite / Ammonia (Pr)	-	-	-	-	Phosphate
Group V	OSL_A7	Acetate	-	II / IV	-	Pu	-	III	Tet (Pr) / thio	+	-	Nitrite	-	-	-	-	Phosphate, manganese, zinc
	OSL_A10	Propionate 6/7	-	II / IV	-	Pu	-	III	Tet (Pr) / thio	+	Pr	-	-	+	-	-	Sulfate, phosphate, nickel, zinc
Ungrouped	OSL_A3	-	+	II	-	-	-	-	-	-	-	-	-	+	-	-	-
	OSL_A6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Zinc
	OSL_A8	-	-	II / IV	-	-	-	III	Thio	-	-	-	-	-	-	-	Phosphate, manganese, zinc, iron
	OSL_A11	-	-	II	-	Pu	-	-	Tet (Pr) / thio	+	Pr	-	-	-	-	+	Phosphate, iron, zinc
	OSL_A13	Acetate	-	II / IV / V	-	Pu	-	III	-	-	-	-	-	-	-	-	Sulfate, phosphate, zinc

Ungrouped	OSL_B4	Lactate	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Old saline < 0.22 µm																	
Group IV	OSS_A4	Lactate	-	IV	II (Pu) / III / IV (Pu)	-	-	III	Thio	-	-	Nitrite	N_2O (Pr)	-	-	-	Phosphate, sulfate, nitrate (Pr), zinc, iron, nickel, manganese, taurine
	OSS_B4	Lactate	-	IV	II / III / IV	Pu	-	III	Thio	-	-	Nitrite	N_2O (Pr)	-	-	-	Phosphate, sulfate, nitrate (Pr), zinc, iron, nickel, manganese
Ungrouped	OSS_A6	Ethanol 2/3, propionate 5/7, lactate	-	-	-	+	-	-	Tet / thio	-	Pr	Nitrite / Ammonia (Pr)	-	+	-	-	Phosphate, sulfate, iron, manganese
	OSS_B6	Ethanol 2/3, propionate 5/7, lactate	-	-	-	+	-	-	Tet / thio	-	Pr	Nitrite / Ammonia (Pr)	-	+	-	-	Phosphate, sulfate, iron
	OSS_A1	-	-	-	III / IV	+	-	-	Thio	-	-	Nitrite	-	-	-	-	Phosphate, sulfate, nitrate (Pr), nickel, taurine
	OSS_B1	Propionate 6/7	-	-	II / III / IV	+	-	-	Thio	-	-	Ammonia	-	-	-	-	Phosphate, sulfate, nitrate (Pr), nickel
	OSS_A2	Ethanol 2/3, lactate	+	-	I	-	-	-	Thio	-	-	-	-	+	-	-	Phosphate, sulfate, nitrate (Pr), zinc, iron, taurine
	OSS_B2	Ethanol 2/3, lactate	+	-	I	-	-	-	Thio	-	-	-	-	-	-	-	Phosphate, sulfate, nitrate (Pr), zinc, iron, taurine
	OSS_A3	Propionate 5/7, lactate	-	-	II / III	-	-	-	-	-	-	-	-	-	-	-	Phosphate
	OSS_B3	Propionate 5/7, lactate	-	-	II / III	-	-	-	-	-	-	-	-	-	-	-	Phosphate
	OSS_A5	Lactate	-	-	III / IV (Pu)	+	-	-	Thio	-	-	-	-	-	-	-	Phosphate, sulfate, iron, zinc, manganese
	OSS_B5	Lactate	-	-	III (Pu) / IV (Pu)	+	-	-	Thio	-	-	-	-	-	-	-	Phosphate, sulfate, zinc, iron, manganese
	OSS_A7	Lactate	-	-	-	-	-	-	Tet (pr) / thio	-	-	-	-	-	-	-	Sulfate
	OSS_B7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Phosphate
	OSS_B8	Propionate 5/7	-	-	-	-	+	-	Tet (pr) / thio	-	-	Nitrite	-	-	-	-	Phosphate, sulfate, zinc, iron, taurine
	OSS_B9	-	-	-	II (Pu) / III (Pu)	-	-	-	Thio	-	-	-	-	-	-	+	Phosphate, sulfate, nitrate (Pr), zinc, iron, nickel (Pr), manganese, taurine

[a] Fermentation of pyruvate to either acetate, lactate, ethanol, or propionate. Putative products have been shown where homologs of all genes in the pathway to acetate have been identified (e.g. MML_B2), all genes in the pathway to lactate have been identified (e.g. MMS_A6), two out of three genes for the pathway to ethanol (e.g. MML_B2), and five out of seven genes for the pathway to propionate (e.g. MMS_A5).
[b] All acetyl-CoA fermentation to butyrate II genes present or absent.
[c] Hydrogen production via pathways I to VI.
[d] Ethanol degradation pathways I to IV.
[e] Formate oxidation to CO_2.
[f] Anaerobic methane oxidation coupled to nitrate reduction.
[g] Hydrogen oxidation pathway. The Roman numeral 'III' denotes the presence of genes for the anaerobic, NADP dependent hydrogen oxidation. This enzyme is reversible and it is possible that it produces hydrogen.
[h] Oxidation of the inorganic sulfur compounds tetrathionate (tet) or thiosulfate (thio).
[i] Presence of the key genes in sulfate reduction, *dsr* and *dsv* encoding dissimilatory sulfite reductase.
[j] Sulfur reduction to hydrogen sulfide by enzyme commission number 1.12.98.4 and *psrAB*
[k] Dissimilatory nitrate reduction is described as the potential end product of either nitrite or ammonia.
[l] Presence of gene homologs for nitrate reduction to nitrous oxide or all steps of nitrate reduction to nitrogen gas (denitrification).
[m] Rnf complex based on the presence of at least four of the *rnfABCDEG* genes.
[n] Presence of *cbbLMS* encoding the key enzyme, ribulose bisphosphate carboxylase (RubisCO).
[o] Nitrogen fixation.
[p] Gene homologs for nutrient uptake systems.
[q] '-'designates that gene homologs for the pathway/system were not identified.
[r] Pu, putative enzyme.
[s] Pr, precursor enzyme.
[t] '+'designates that gene homologs for the pathway/system were identified.

Supplementary File 9. Calculation of the median percentage small cells in each water type.

OTU	Number reads > 0.1 µm	% of total reads > 0.1 µm	Number reads > 0.22 µm	% of total reads > 0.22 µm	Enrichment (> 0.22 µm % OTUs/> 0.1 µm % OTUs)	Median (> 0.22 µm enriched OTUs)	min/max	% median cells < 0.22 µm	min/max
Modern marine									
OTU_000234	6515	14.47	19998	27.61	1.91	1.9	1.7/1.9	47.1	40.8/47.6
OTU_000280	232	0.52	712	0.98	1.91				
OTU_000258	3591	7.98	10917	15.07	1.89				
OTU_000001	6398	14.21	18634	25.73	1.81				
OTU_000033	509	1.13	1385	1.91	1.69				
OTU_000012	351	0.78	507	0.70	0.90				
OTU_000007	487	1.08	654	0.90	0.83				
OTU_000006	837	1.86	374	0.52	0.28				
OTU_000008	1074	2.39	449	0.62	0.26				
OTU_000003	1941	4.31	607	0.84	0.19				
OTU_000009	2045	4.54	421	0.58	0.13				
OTU_000004	1325	2.94	214	0.30	0.10				
OTU_000005	1538	3.42	194	0.27	0.08				
OTU_000018	485	1.08	59	0.08	0.08				
OTU_000002	11610	25.78	1	0.00	0.00				
Undefined mixed									
OTU_000001	120	0.81	11001	29.46	36.29	2.2	1.07/36.29	53.6	6.5/97.2
OTU_000024	78	0.53	1272	3.41	6.46				
OTU_000003	408	2.76	2865	7.67	2.78				

OTU_000014	110	0.74	426	1.14	1.53				
OTU_000007	670	4.53	1856	4.97	1.10				
OTU_000013	109	0.74	294	0.79	1.07				
OTU_000023	76	0.51	191	0.51	0.99				
OTU_000016	96	0.65	208	0.56	0.86				
OTU_000225	232	1.57	429	1.15	0.73				
OTU_000019	360	2.44	552	1.48	0.61				
OTU_000241	100	0.68	151	0.40	0.60				
OTU_000012	385	2.61	467	1.25	0.48				
OTU_000044	174	1.18	178	0.48	0.40				
OTU_000015	142	0.96	109	0.29	0.30				
OTU_000017	218	1.48	163	0.44	0.30				
OTU_000020	266	1.80	197	0.53	0.29				
OTU_000009	1041	7.04	605	1.62	0.23				
OTU_000005	922	6.24	502	1.34	0.22				
OTU_000034	90	0.61	45	0.12	0.20				
OTU_000029	178	1.20	59	0.16	0.33				
OTU_000028	515	3.48	35	0.09	0.07				
OTU_000004	3337	22.58	190	0.51	0.06				
OTU_000082	94	0.64	2	0.01	0.02				
OTU_000063	274	1.85	2	0.01	0.01				
OTU_000002	2058	13.93	7	0.02	0.00				
Old saline									
OTU_000006	341	0.75	1932	1.97	2.63	2.0	1.2/2.6	50.1	16.0/62.0
OTU_000012	242	0.53	1052	1.07	2.02				
OTU_000003	1351	2.96	5783	5.88	1.99				
OTU_000008	360	0.79	919	0.93	1.19				
OTU_000019	245	0.54	495	0.50	0.94				
OTU_000011	343	0.75	675	0.69	0.91				

OTU_000009	781	1.71	921	0.94	0.55
OTU_000023	421	0.92	246	0.25	0.27
OTU_000015	780	1.71	332	0.34	0.20
OTU_000007	1750	3.83	182	0.19	0.05
OTU_000139	1058	2.32	33	0.03	0.01
OTU_000005	3162	6.92	66	0.07	0.01
OTU_000084	4652	10.19	91	0.09	0.01
OTU_000034	1532	3.36	29	0.03	0.01
OTU_000140	1937	4.24	34	0.03	0.01
OTU_000002	9223	20.20	151	0.15	0.01
OTU_000030	3043	6.66	24	0.02	0.00
OTU_000025	1204	2.64	9	0.01	0.00
OTU_000024	1678	3.67	11	0.01	0.00
OTU_000004	7003	15.34	19	0.02	0.00

Supplementary File 10. Percentage reads of the CONCOCT bin in in the twelve metagenomes. None of the bins in this study showed signs of being the result of contamination as they had different relative abundances between the water types when scaled to the DNA concentrations (Supplementary File 2). Additionally, only one phylogenetic clade was present in all similarly processed metagenomes (Supplementary File 11) and the pattern of recruited reads did not increase with lower DNA concentrations. Values for MMS_B, UMS_A and OSS_A are averages of the duplicate sequencing.

	Percentage reads in the metagenomes											
bin	**MML_A**	**MML_B**	**MMS_A**	**MMS_B**	**UML_A**	**UML_B**	**UMS_A**	**UMS_B**	**OSL_A**	**OSL_B**	**OSS_A**	**OSS_B**
UMS_A4	0.00	0.00	0.00	0.00	0.00	0.00	0.40	0.38	0.00	0.00	0.05	0.04
UMS_A1	0.00	0.00	1.17	1.00	0.00	0.00	3.34	3.50	0.00	0.01	2.31	2.51
UMS_A5	0.00	0.00	0.17	0.14	0.06	0.00	1.39	1.45	0.00	0.02	0.34	0.37
UMS_A3	0.00	0.00	0.27	0.31	0.00	0.00	2.63	2.61	0.00	0.00	1.87	1.64
UMS_A2	0.00	0.00	14.41	13.13	0.00	0.00	0.60	0.60	0.36	0.02	0.17	0.17
UMS_A6	0.00	0.00	0.05	0.08	0.00	0.00	0.74	0.71	0.00	0.00	0.07	0.06
UMS_B1	0.00	0.00	1.20	1.02	0.00	0.00	3.41	3.57	0.00	0.01	2.36	2.56
UMS_B5	0.00	0.00	0.17	0.14	0.00	0.00	1.23	1.27	0.06	0.02	0.21	0.23
UMS_B4	0.00	0.00	0.00	0.00	0.00	0.00	0.40	0.38	0.00	0.00	0.05	0.04
UMS_B6	0.00	0.00	0.05	0.08	0.00	0.00	0.40	0.73	0.00	0.00	0.41	0.06
UMS_B7	0.00	0.00	0.02	0.01	0.00	0.00	1.27	1.37	0.01	0.00	1.26	1.42
UMS_B3	0.00	0.00	0.27	0.31	0.00	0.00	2.64	2.63	0.00	0.00	1.88	1.65
UMS_B2	0.00	0.00	14.25	12.99	0.00	0.00	0.59	0.59	0.36	0.02	0.17	0.17
UMS_B8	0.00	0.00	0.10	0.08	0.00	0.00	1.49	1.59	0.19	0.02	3.36	3.73
UMS_B9	0.00	0.00	0.20	0.18	0.00	0.00	0.60	0.63	0.00	0.00	0.20	0.23
OSS_A7	0.00	0.00	0.09	0.12	0.00	0.00	0.10	0.09	0.00	0.00	0.35	0.30

	Percentage reads in the metagenomes											
bin	**MML_A**	**MML_B**	**MMS_A**	**MMS_B**	**UML_A**	**UML_B**	**UMS_A**	**UMS_B**	**OSL_A**	**OSL_B**	**OSS_A**	**OSS_B**
OSS_A1	0.00	0.00	1.20	1.02	0.00	0.00	3.40	3.57	0.00	0.01	2.35	2.55
OSS_A6	0.00	0.00	0.12	0.12	0.00	0.00	1.90	1.81	0.00	0.00	0.69	0.67
OSS_A2	0.00	0.00	0.27	0.31	0.00	0.00	2.65	2.63	0.00	0.00	1.89	1.65
OSS_A3	0.00	0.00	0.11	0.45	0.00	0.00	1.53	1.63	0.20	0.02	3.52	3.90
OSS_A5	0.00	0.00	0.04	0.03	0.00	0.00	0.02	0.02	0.00	0.00	0.76	0.92
OSS_A4	0.00	0.00	0.04	0.04	0.00	0.00	0.03	0.04	0.03	0.01	2.33	3.24
OSS_B4	0.00	0.00	0.05	0.04	0.00	0.00	0.03	0.04	0.03	0.01	2.42	3.35
OSS_B7	0.00	0.00	0.15	0.12	0.00	0.00	0.12	0.13	0.01	0.00	0.70	0.81
OSS_B5	0.00	0.00	0.04	0.03	0.00	0.00	0.02	0.02	0.00	0.00	0.75	0.91
OSS_B8	0.00	0.00	1.92	1.43	0.00	0.00	0.73	0.78	0.01	0.00	0.93	1.08
OSS_B2	0.00	0.00	0.27	0.31	0.00	0.00	2.63	2.61	0.00	0.00	1.87	1.64
OSS_B6	0.00	0.00	0.12	0.12	0.00	0.00	1.90	1.80	0.00	0.00	0.68	0.67
OSS_B9	0.00	0.00	0.20	0.01	0.00	0.00	0.01	0.01	0.00	0.00	0.98	1.35
OSS_B1	0.00	0.00	1.22	1.04	0.00	0.00	3.47	3.64	0.00	0.01	2.40	2.60
OSS_B3	0.00	0.00	0.10	0.09	0.00	0.00	1.52	1.62	0.20	0.02	3.49	3.88
MMS_A6	0.00	0.00	0.79	0.67	0.00	0.00	0.04	0.04	0.04	0.01	0.12	0.15
MMS_A2	0.00	0.00	14.80	13.49	0.00	0.00	0.62	0.62	0.37	0.02	0.17	0.18
MMS_A3	0.00	0.00	1.96	1.46	0.00	0.00	0.75	0.81	0.02	0.02	0.96	1.12
MMS_A4	0.00	0.00	24.11	22.11	0.00	0.00	0.09	0.11	0.01	0.01	0.13	0.15
MMS_A7	0.00	0.00	1.24	1.08	0.00	0.00	0.01	0.01	0.02	0.25	0.01	0.01
MMS_A5	0.00	0.00	0.90	0.79	0.00	0.00	0.02	0.02	0.00	0.01	0.02	0.02
MMS_A8	0.00	0.00	0.57	0.45	0.00	0.00	0.16	0.17	0.00	0.01	0.10	0.11
MMS_A1	0.00	0.00	1.13	0.96	0.00	0.00	2.72	2.85	0.00	0.01	1.86	2.02
MMS_B6	0.00	0.00	0.33	0.36	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.01
MMS_B1	0.00	0.00	1.14	0.97	0.00	0.00	2.75	2.89	0.00	0.01	1.88	2.05
MMS_B3	0.00	0.00	1.93	1.44	0.00	0.00	0.73	0.79	0.03	0.01	0.94	1.09
MMS_B2	0.00	0.00	15.74	14.43	0.00	0.00	0.75	0.76	0.39	0.12	0.23	0.23
MMS_B4	0.00	0.00	24.85	22.78	0.00	0.00	0.10	0.11	0.01	0.01	0.14	0.15
MMS_B5	0.00	0.00	1.51	1.34	0.00	0.00	0.04	0.05	0.05	0.63	0.03	0.04

	Percentage reads in the metagenomes											
bin	**MML_A**	**MML_B**	**MMS_A**	**MMS_B**	**UML_A**	**UML_B**	**UMS_A**	**UMS_B**	**OSL_A**	**OSL_B**	**OSS_A**	**OSS_B**
MML_A1	2.23	2.01	0.02	0.03	1.43	1.51	0.05	0.05	0.12	0.04	0.03	0.02
MML_A2	5.43	0.57	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MML_B1	1.48	1.31	0.00	0.00	0.95	1.02	0.00	0.00	0.03	0.01	0.00	0.00
MML_B2	0.51	0.50	0.01	0.01	0.12	0.13	0.01	0.01	0.00	0.00	0.00	0.00
UML_A1	2.06	1.85	0.00	0.00	1.42	1.50	0.01	0.01	0.07	0.03	0.01	0.01
UML_B1	0.15	0.09	0.01	0.01	2.26	2.47	0.06	0.06	0.14	0.06	0.04	0.04
UML_B2	0.02	0.02	0.00	0.00	0.24	0.28	0.00	0.00	0.01	0.00	0.00	0.00
OSL_A3	0.00	0.00	0.00	0.00	0.06	0.07	0.00	0.00	1.08	0.23	0.03	0.03
OSL_A4	0.01	0.01	0.00	0.00	0.03	0.03	0.00	0.00	0.77	0.31	0.00	0.00
OSL_A5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.74	0.21	0.00	0.00
OSL_A6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.77	0.19	0.01	0.01
OSL_A7	0.00	0.00	0.01	0.01	0.06	0.06	0.01	0.01	1.85	0.80	0.01	0.01
OSL_A8	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	1.31	0.26	0.01	0.01
OSL_A9	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	1.01	0.34	0.00	0.00
OSL_A10	0.04	0.06	0.00	0.00	0.03	0.02	0.00	0.00	2.48	0.56	0.03	0.03
OSL_A1	0.00	0.01	0.07	0.06	0.00	0.01	0.03	0.03	4.85	58.18	0.04	0.04
OSL_A2	0.00	0.00	0.00	0.00	0.02	0.02	0.00	0.00	6.19	3.14	0.01	0.01
OSL_A11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.22	0.25	0.01	0.01
OSL_A12	0.00	0.00	0.00	0.00	0.04	0.04	0.00	0.00	0.25	0.12	0.00	0.00
OSL_A13	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	1.26	0.33	0.00	0.00
OSL_B1	0.00	0.01	0.03	0.03	0.00	0.01	0.00	0.01	4.81	57.73	0.01	0.01
OSL_B3	0.00	0.00	0.01	0.01	0.02	0.06	0.01	0.02	0.03	2.72	0.02	0.02
OSL_B2	0.00	0.00	0.00	0.00	0.02	0.02	0.00	0.00	5.80	2.95	0.00	0.00
OSL_B4	0.00	0.00	0.00	0.00	0.30	0.29	0.00	0.00	3.37	0.90	0.00	0.00

Supplementary File 11. Analysis of cross-sample abundances onto the contig bins from all twelve replicates onto the contig bins from all twelve replicates.

References

Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ *et al* (2014). Binning metagenomic contigs by coverage and composition. *Nat Methods* **11:** 1144-1146.

Boisvert S, Raymond F, Godzaridis E, Laviolette F, Corbeil J (2012). Ray Meta: scalable de novo metagenome assembly and profiling. *Genome Biol* **13:** R122.

Bolger AM, Lohse M, Usadel B (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30:** 2114-2120.

Borgström E, Lundin S, Lundeberg J (2011). Large scale library generation for high throughput sequencing. *PLoS ONE* **6:** e19119.

Caspi R, Altman T, Billington R, Dreher K, Foerster H, Fulcher CA *et al* (2014). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Research* **42:** D459-D471.

Darling AE, Jospin G, Lowe E, Matsen FAt, Bik HM, Eisen JA (2014). PhyloSift: phylogenetic analysis of genomes and metagenomes. *PeerJ* **2:** e243.

Edgar RC (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods* **10:** 996-998.

Herlemann DP, Labrenz M, Jurgens K, Bertilsson S, Waniek JJ, Andersson AF (2011). Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *ISME J* **5:** 1571-1579.

Hugerth LW, Wefer HA, Lundin S, Jakobsson HE, Lindberg M, Rodin S *et al* (2014). DegePrime, a program for degenerate primer design for broad-taxonomic-range PCR in microbial ecology studies. *Appl Environ Microbiol* **80:** 5116-5123.

Hyatt D, LoCascio PF, Hauser LJ, Uberbacher EC (2012). Gene and translation initiation site prediction in metagenomic sequences. *Bioinformatics* **28:** 2223-2230.

John SG, Mendez CB, Deng L, Poulos B, Kauffman AK, Kern S *et al* (2011). A simple and efficient method for concentration of ocean viruses by chemical flocculation. *Environmen Microbiol Reports* **3:** 195-202.

Joshi NA, Fass JN (2011). Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files, Version 1.33 edn: https://github.com/najoshi/sickle.

Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M (2014). Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res* **42:** D199-205.

Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C *et al* (2004). Versatile and open software for comparing large genomes. *Genome Biol* **5:** R12.

Lindh MV, Figueroa D, Sjöstedt J, Baltar F, Lundin D, Andersson A *et al* (2015). Transplant experiments uncover Baltic Sea basin-specific responses in bacterioplankton community composition and metabolic activities. *Fron Microbiol* **6:** 10.3389/fmicb.2015.00223.

Lundin S, Stranneheim H, Pettersson E, Klevebring D, Lundeberg J (2010). Increased Throughput by Parallelization of Library Preparation for Massive Sequencing. *PLoS ONE* **5:** e10029.

Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW (2014). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *PeerJ PrePrints* **2:e554v1** https://dx.doi.org/10.7287/peerj.preprints.7554v7281.

Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig WG, Peplies J *et al* (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* **35:** 7188-7196.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P *et al* (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41:** D590-596.

Robertson CE, Harris JK, Wagner BD, Granger D, Browne K, Tatem B *et al* (2013). Explicet: graphical user interface software for metadata-driven management, analysis and visualization of microbiome data. *Bioinformatics* **29:** 3100-3101.

Seemann T (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30:** 2068-2069.

Segata N, Bornigen D, Morgan XC, Huttenhower C (2013). PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nature communications* **4:** 2304.