

Supplementary Information

5

Metagenomic analysis of size-fractionated picoplankton in a marine oxygen minimum zone

Sangita Ganesh, Darren J. Parris, Edward F. DeLong, Frank J. Stewart

10 **Table S1.** 16S rRNA gene amplicon sequencing statistics.

Sample ¹	Count	OTU ²
TOTAL sequences ³	742,553	
TOTAL following parsing ⁴	406,768	
Mean length ⁵	409 bp	
TOTAL unique OTUs		17,014
5p, <i>a</i>	18,837	1,298
5p, <i>b</i>	15,803	1,229
5s, <i>a</i>	20,506	1,225
5s, <i>b</i>	21,240	1,344
32p, <i>a</i>	13,540	1,818
32p, <i>b</i>	5,619	956
32s, <i>a</i>	18,681	2,354
32s, <i>b</i>	14,314	1,902
70p, <i>a</i>	6,281	760
70p, <i>b</i>	17,412	1,298
70s, <i>a</i>	15,929	1,991
70s, <i>b</i>	18,917	2,484
110p, <i>a</i>	11,336	1,717
110p, <i>b</i>	12,059	1,786
110s, <i>a</i>	19,057	2,066
110s, <i>b</i>	15,300	1,830
200p, <i>a</i>	12,026	1,903
200p, <i>b</i>	12,142	1,929
200s, <i>a</i>	17,436	1,564
200s, <i>b</i>	18,572	1,864
320p, <i>a</i>	9,647	1,493
320p, <i>b</i>	12,490	1,553
320s, <i>a</i>	19,410	2,163
320s, <i>b</i>	15,796	2,135
1000p, <i>a</i>	18,739	1,839
1000p, <i>b</i>	5,268	1,080
1000s, <i>a</i>	5,256	658
1000s, <i>b</i>	15,155	2,205

¹ Samples are labeled by depth and filter type, where p = prefilter (>1.6 μm), s = Sterivex (0.2-1.6 μm), and *a,b* indicate datasets generated by duplicate PCR reactions using DNA from the same sample

15 ² operational taxonomic unit (unique cluster of sequences sharing 97% nucleotide similarity)

³ generated on one full plate 454 run (Titanium chemistry); this plate also included 18S rRNA amplicon samples, which are not included in this analysis

20 ⁴ sequences sorted by barcode (sample ID) and parsed based on quality; deleted sequences included those with length <200 or >1000 bp (51,915), with 1 or more ambiguous bases (50,866), missing a quality score (0), with mean quality score <25 (741), with a homopolymer run exceeding 6 bp (4,864), or with mismatches in the primer sequence (177,586)

⁵ mean length of sequences remaining after parsing

25

50 **Table S3.** Bray-Curtis distances between metagenome samples based on sequence counts to SEED Subsystems.

		70		110		200		1000	
		p	s	p	s	p	s	p	s
70	p	**							
	s	0.09	**						
110	p	0.09	0.09	**					
	s	0.11	0.05	0.10	**				
200	p	0.09	0.07	0.08	0.08	**			
	s	0.12	0.06	0.12	0.04	0.09	**		
1000	p	0.11	0.11	0.11	0.13	0.10	0.14	**	
	s	0.10	0.07	0.11	0.08	0.09	0.09	0.12	**

55

p,s indicate prefilter and Sterivex samples, respectively

60 **Table S4.** Normalized¹ metagenome sequence counts per SEED Subsystem².

SEED Subsystem	70p	70s	110p	110s	200p	200s	1000p	1000s
Amino Acid Metabolism³	2125	2373	2320	2305	2142	2540	2460	2307
Alanine, serine, and glycine	319	360	321	322	222	316	238	323
Arginine; urea cycle, polyamines	275	319	337	400	382	451	442	331
Aromatic amino acids and derivatives	106	115	118	133	185	89	91	86
Branched-chain amino acids	363	463	502	421	428	458	590	506
Glutamine, glutamate, aspartate, asparagine; ammonia assimilation	106	42	63	26	70	20	102	101
Histidine Metabolism	213	205	212	177	127	176	136	128
Lysine, threonine, methionine, and cysteine	9	0	8	1	12	0	0	0
Lysine, threonine, methionine, and cysteine.1	602	767	666	733	629	906	714	720
Proline and 4-hydroxyproline	133	102	94	91	86	124	147	111
Carbohydrates	2878	3022	2876	3001	2919	3025	2710	2756
Aminosugars	35	26	47	21	53	22	147	27
Central carbohydrate metabolism	708	987	925	1056	970	1049	918	982
CO2 fixation	345	384	345	412	275	450	227	466
Di- and oligosaccharides	213	96	133	52	132	46	136	116
Fermentation	275	434	306	439	345	460	283	276
Glycoside hydrolases	106	63	39	90	49	70	34	69
Lacto-N-Biose I and Galacto-N-Biose Metabolic Pathway	0	0	0	0	0	1	0	0
Monosaccharides	390	311	408	269	354	300	340	212
One-carbon Metabolism	381	377	353	347	423	336	385	345
Organic acids	204	217	204	214	230	205	159	146
Polysaccharides	53	38	8	14	12	7	45	32
Sugar alcohols	168	89	110	87	74	80	34	84
Cell Division and Cell Cycle	912	1100	940	1066	929	1179	862	1049
Cell cycle in Prokaryota	629	641	408	620	551	738	510	686
Control of Macromoleculuar Synthesis	204	400	462	412	321	406	295	331
YgjD and YeaZ	80	59	71	35	58	35	57	32
Cell wall and Capsule	1718	1679	1520	1708	1694	1778	1417	2028
Capsular and extracellular polysacchrides	416	368	361	360	329	346	272	449
Cell wall of Mycobacteria	319	289	384	261	300	240	272	407
Gram-Negative cell wall components	257	386	313	431	399	462	317	407
Gram-Positive cell wall components	62	26	24	8	41	16	34	5
Peptidoglycan Biosynthesis	505	503	361	533	506	579	408	646
UDP-N-acetylmuramate from Fructose-6-phosphate Biosynthesis	89	69	47	80	62	95	68	69
YjeE	71	38	31	35	58	39	45	44
Clustering-based Subsystems	1284	1353	1285	1377	1377	1472	1224	1453
Bacterial Cell Division	531	580	415	529	539	656	340	713
Bacterial RNA-metabolizing Zn-dependent hydrolases	186	194	274	210	247	218	159	188
Carbohydrates.1	0	1	0	3	0	4	0	2

Catabolism of an unknown compound	0	0	0	1	0	1	11	2
CBSS-261594.1.peg.2640	0	0	0	1	0	0	0	0
CBSS-562.2.peg.5158 SK3 including	0	6	0	0	0	0	0	0
Clustering-based subsystems.1	106	84	86	157	95	108	125	74
Conserved gene cluster associated with Met-tRNA formyltransferase	239	216	227	200	247	209	204	192
LMPTP YfkJ cluster	9	9	16	6	8	8	0	0
LMPTP YwIE cluster	35	91	86	66	45	65	79	76
NusA-TFII Cluster	133	150	125	174	177	178	261	151
proteosome related	44	19	47	21	16	25	34	52
Putative hemin transporter	0	1	0	1	0	0	11	0
Putative sulfate assimilation cluster	0	0	0	1	0	0	0	0
Type III secretion system related	0	2	0	3	0	0	0	2
Type III secretion system, extended	0	0	8	3	4	0	0	0
Cofactors, Vitamins, Prosthetic Groups	1621	1654	1489	1435	1517	1557	1565	1532
Biotin	71	75	63	76	82	92	113	84
Coenzyme A	97	97	157	110	103	106	91	69
Coenzyme B	0	0	0	1	4	2	0	0
Coenzyme F420	44	21	24	15	12	8	23	17
Folate and pterines	390	359	219	340	304	309	329	358
Lipoic acid	53	51	110	55	86	55	68	32
NAD and NADP	133	162	172	133	177	125	147	116
Pyridoxine	27	67	47	70	41	63	34	30
Quinone cofactors	248	171	188	148	197	174	215	170
Riboflavin, FMN, FAD	97	78	94	58	62	51	79	69
Tetrapyrroles	381	461	337	337	317	406	385	456
Thiamin biosynthesis	80	112	78	91	132	165	79	131
DNA Metabolism	2754	2226	2132	2223	2257	2252	2268	2322
CRISPRs	0	10	31	1	4	4	11	42
DNA recombination	124	63	78	65	78	54	45	67
DNA repair	1293	1064	854	1033	1069	1073	1259	1091
DNA replication	1196	992	1050	1029	1015	1055	771	1059
DNA structural proteins, bacterial	27	19	24	36	21	27	57	15
Restriction-Modification System	97	35	71	35	58	22	23	32
YcfH	18	44	24	25	12	16	102	17
Dormancy and Sporulation	18	2	8	6	4	5	0	5
Spore Coat	18	0	0	1	0	0	0	0
Spore germination	0	1	0	0	0	0	0	0
Sporulation gene orphans	0	1	8	4	4	5	0	5
Fatty acids, Lipids, Isoprenoids	682	817	878	903	892	828	714	832
Fatty acids	204	287	329	373	333	301	340	336
Isoprenoids	248	237	290	220	271	213	125	249
Phospholipids	186	247	235	289	255	282	215	237
Polyhydroxybutyrate metabolism	44	44	24	22	33	31	11	5
Triacylglycerols	0	1	0	0	0	0	23	5
Membrane Transport	221	306	329	311	271	338	317	363

ABC transporters	62	122	63	112	70	115	45	111
<i>peptide_(TC_3.A.1.5.5)</i>	0	0	0	1	4	0	0	0
<i>branched-chain_amino_acid_(TC_3.A.1.4.1)</i>	26	77	41	95	40	89	30	64
<i>dipeptide_(TC_3.A.1.5.2)</i>	9	24	11	3	0	11	15	8
<i>alkylphosphonate_(TC_3.A.1.9.1)</i>	9	11	11	4	13	9	0	22
<i>tungstate_(TC_3.A.1.6.2)</i>	18	10	0	9	13	6	0	17
Choline Transport	0	0	8	0	0	0	0	0
ECF class transporters	18	16	0	17	0	13	0	15
Potassium homeostasis	80	112	196	139	127	143	125	143
Sugar Phosphotransferase Systems, PTS	0	2	0	0	4	0	34	7
Transport of Manganese	0	4	31	4	8	1	11	0
Transport of Molybdenum	0	0	0	0	4	0	0	0
Transport of Nickel and Cobalt	9	4	0	0	4	1	0	10
Transport of Zinc	35	35	31	22	41	48	45	54
Uni- Sym- and Antiporters	18	11	0	17	12	16	57	22
Metabolism of Aromatic Compounds	584	432	447	392	604	347	760	688
Anaerobic degradation of aromatic compounds	18	32	24	32	0	14	0	7
Aromatic Amin Catabolism	0	4	8	4	0	2	0	0
Benzoate transport and degradation cluster	89	69	94	88	107	84	159	76
carbazol degradation cluster	18	7	0	4	16	3	11	5
Gentisate degradation	53	32	8	44	29	45	57	54
Metabolism of central aromatic intermediates	159	116	133	64	197	74	215	155
p-cymene degradation	0	0	0	1	0	0	0	0
Peripheral pathways for catabolism of aromatic compounds	204	126	172	108	222	81	238	333
Salicylate and gentisate catabolism	44	46	8	47	33	43	79	57
Miscellaneous	221	189	243	213	197	194	283	205
Archease	0	4	0	0	4	1	0	0
Luciferases	0	0	0	1	0	1	0	0
Muconate lactonizing enzyme family	9	1	8	4	0	0	11	7
YaaA	18	1	8	1	8	1	23	12
YbbK	35	27	24	26	29	19	68	17
ZZ gjo need homes	159	155	204	180	156	172	181	168
Motility and Chemotaxis	558	324	682	249	493	219	703	375
Bacterial Chemotaxis	71	43	63	30	37	28	79	32
Flagellar motility in Prokaryota	487	281	619	218	456	190	624	343
Nitrogen Metabolism	576	817	846	1011	773	981	476	476
Allantoin Utilization	18	5	16	6	8	3	0	27
Amidase clustered with urea and nitrile hydratase functions	0	2	0	0	4	0	0	5
Ammonia assimilation	213	433	337	452	407	476	374	380
Denitrification	53	23	31	23	21	25	0	10
Dissimilatory nitrite reductase	0	4	24	4	8	7	0	0
Nitrate and nitrite ammonification	239	322	345	497	300	440	45	12
Nitric oxide synthase	18	12	24	4	8	2	23	35
Nitrogen fixation	9	7	8	6	4	3	23	0

Nitrosative stress	27	7	63	19	12	24	11	7
Nucleosides and Nucleotides	1311	1331	1207	1323	1361	1239	850	1308
Adenosyl nucleosidases	9	4	8	3	16	2	23	7
AMP to 3-phosphoglycerate	0	2	0	0	0	0	0	0
De Novo Pyrimidine Synthesis	115	195	118	162	201	256	113	289
Detoxification.1	35	32	31	33	53	16	11	39
Hydantoin metabolism	213	117	259	167	247	112	45	106
Purine Utilization	213	148	219	135	156	108	34	99
Purines	602	667	415	627	514	571	499	624
Pyrimidines	35	48	16	22	41	16	45	35
Ribonucleotide reduction	89	117	141	174	132	157	79	109
Phages, Prophages, Transposable Elements	115	80	71	112	78	85	91	84
IbrA and IbrB: co-activators of prophage gene expression	0	0	0	3	4	0	0	0
Listeria phi-A118-like prophages	0	0	0	0	4	0	0	0
Staphylococcal phi-Mu50B-like prophages	115	74	55	108	70	85	79	76
Tn552	0	6	16	1	0	0	11	7
Phosphorous Metabolism	407	252	251	305	288	256	442	338
Alkylphosphonate utilization	35	12	8	14	4	8	68	25
High affinity phosphate transporter and control of PHO regulon	53	5	8	1	25	2	23	17
Phosphate metabolism	301	233	235	287	255	242	351	279
Phosphonate metabolism	18	1	0	3	4	4	0	17
Photosynthesis	115	25	31	3	0	3	0	2
Electron transport and photophosphorylation	106	21	31	3	0	3	0	2
Light-harvesting complexes	9	4	0	0	0	0	0	0
Potassium Metabolism	44	12	24	7	16	4	45	10
Glutathione-regulated potassium-efflux system and associated functions	44	12	24	7	16	4	45	10
Protein Metabolism	3232	3133	3135	3158	3100	2929	3413	3316
Protein biosynthesis	1877	1945	1920	1924	1813	1755	1995	2008
Protein degradation	514	429	400	412	423	366	385	533
Protein folding	602	548	596	539	555	560	692	528
Protein processing and modification	124	122	110	195	177	169	91	126
Secretion	62	16	47	28	21	28	113	44
Selenoproteins	53	73	63	62	111	51	136	76
Regulation and Cell Signaling	248	231	376	244	234	213	351	210
cAMP signaling in bacteria	142	68	102	55	62	41	113	67
CytR regulation	0	1	31	0	0	0	0	0
DNA-binding regulatory proteins, strays	0	10	0	8	0	4	11	7
Orphan regulatory proteins	18	16	78	12	45	12	34	0
Programmed Cell Death and Toxin-antitoxin Systems	0	7	0	6	16	12	11	2
Proteolytic pathway	0	0	24	0	0	0	0	5
Sex pheromones in Enterococcus faecalis and other Firmicutes	53	28	0	36	37	44	45	30
Stringent Response, (p)ppGpp metabolism	0	64	86	75	25	54	45	57
Two-component regulatory systems in Campylobacter	35	36	55	52	49	47	91	42

Respiration	1293	1916	1536	1975	1681	2021	1655	1520
ATP synthases	159	273	196	235	160	280	329	212
Biogenesis of c-type cytochromes	177	164	141	156	181	169	193	148
Biogenesis of cbb3-type cytochrome c oxidases	35	20	8	15	37	42	34	2
Biogenesis of cytochrome c oxidases	71	101	110	72	95	83	79	42
Carbon monoxide dehydrogenase maturation factors	9	9	39	6	8	11	34	22
Carbon monoxide induced hydrogenase	0	2	0	7	0	1	0	0
Cytochrome B6-F complex	0	1	0	1	4	0	23	0
Electron accepting reactions	257	345	212	391	411	442	317	249
Electron donating reactions	399	723	674	700	559	678	431	602
Formate hydrogenase	133	168	94	301	144	226	147	168
Reductive Dechlorination	0	0	0	1	0	0	0	0
Sodium Ion-Coupled Energetics	9	4	0	0	0	0	0	0
Soluble cytochromes and functionally related electron carriers	44	106	63	90	82	90	68	74
RNA Metabolism	1390	1454	1654	1502	1599	1592	1633	1453
Group II intron-associated genes	18	7	24	3	16	10	23	5
RNA processing and modification	629	596	713	586	711	600	522	590
Transcription	744	851	917	914	872	982	1088	859
Secondary Metabolism	142	234	227	209	177	209	91	128
Aromatic amino acids and derivatives.1	18	6	0	4	4	4	0	2
Bacterial cytostatics, differentiation factors and antibiotics	9	31	16	17	29	15	34	15
Biologically active compounds in metazoan cell defence and differentiation	35	36	71	26	16	30	11	10
Biosynthesis of phenylpropanoids	0	2	16	1	4	1	0	7
Plant Alkaloids	18	48	31	47	25	56	11	25
Plant Hormones	62	111	94	113	99	103	34	69
Stress Response	1417	1145	1168	1203	1081	1155	998	1162
Acid stress	0	1	8	0	8	0	0	0
Bacterial hemoglobins	0	6	8	1	8	0	0	10
Detoxification	142	25	0	15	29	11	11	25
Dimethylarginine metabolism	0	0	0	8	4	2	23	5
Flavohaemoglobin	0	0	8	0	0	0	0	0
Heat shock dnaK gene cluster extended	593	504	431	552	432	527	465	481
Hfl operon	80	144	110	122	107	158	79	67
Osmotic stress	186	178	180	207	164	227	147	212
Oxidative stress	390	266	368	291	304	229	215	350
SigmaB stress response regulation	27	17	31	4	12	1	11	12
Universal stress protein family	0	2	24	1	12	1	45	0
Sulfur Metabolism	248	174	306	149	230	195	204	234
Galactosylceramide and Sulfatide metabolism	27	10	39	8	12	1	11	0
Organic sulfur assimilation	106	65	157	47	95	53	91	151
Sulfate reduction-associated complexes	35	35	39	46	66	90	0	32
Sulfur oxidation	18	30	24	15	25	19	34	25
Thioredoxin-disulfide reductase	62	35	47	33	33	33	68	27

Virulence	2754	2093	2751	1924	2713	1708	3197	1902
Adhesion	133	67	78	97	132	57	136	67
Detection	44	32	39	41	45	41	79	5
Invasion and intracellular resistance	151	134	102	128	173	129	204	123
Iron Scavenging Mechanisms	496	386	470	334	419	228	488	338
Pathogenicity islands	124	130	118	101	107	117	181	91
Pseudomonas quinolone signal PQS	0	0	0	1	0	0	0	0
Quorum sensing and biofilm formation	9	11	8	11	12	4	11	2
Regulation of virulence	27	32	31	43	25	44	68	39
Resistance to antibiotics and toxic compounds	992	699	1074	631	917	626	1020	681
Streptococcus pyogenes Virulome	0	0	8	0	0	0	11	0
Ton and Tol transport systems	487	495	486	454	563	403	646	479
Toxins and superantigens	0	1	8	0	8	0	0	0
Type III, Type IV, Type VI, ESAT secretion systems	292	106	329	81	312	59	351	76
TOTAL non-normalized reads matching SEED²	2823	20268	3190	18104	6081	22837	2205	10132

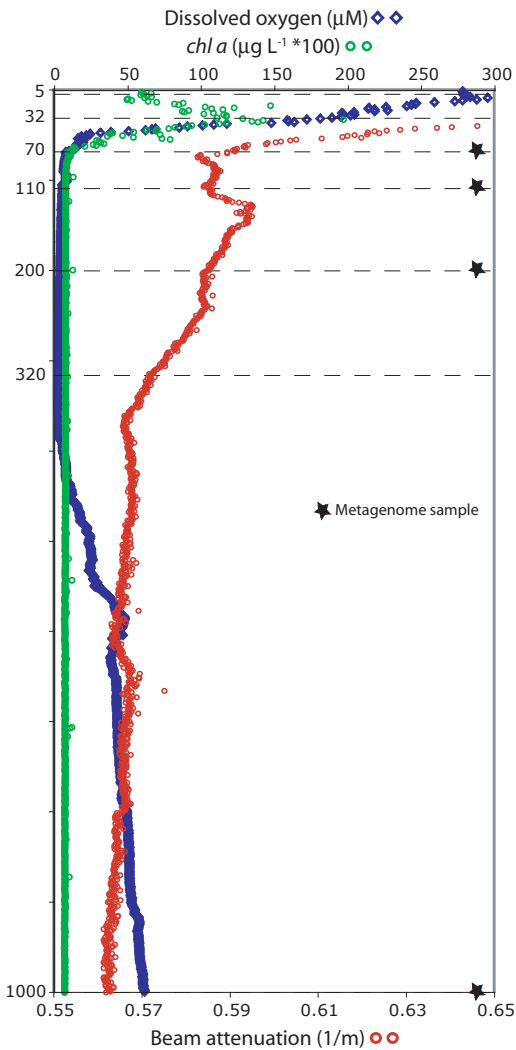
65 ¹ Counts are normalized to a standard dataset size (number of prokaryote reads matching SEED) of 25,000 to allow comparison across columns. To determine true read counts, divide values by the sample-specific scaling factor: 8.86 (70p), 1.23 (70s), 7.84 (110p), 1.38 (110s), 4.11 (200p), 1.09 (200s), 11.34 (1000p), 2.47 (1000s).

70 ² NOTE: Some gene sequences are classified into multiple Subsystems. The sum of read counts across all Subsystems exceeds the total number of SEED-assigned reads by 12-15%.

³ Bolded values are sums of Subsystem (Level 2) counts per broader functional category (Level 1).

Supplementary Figures

75

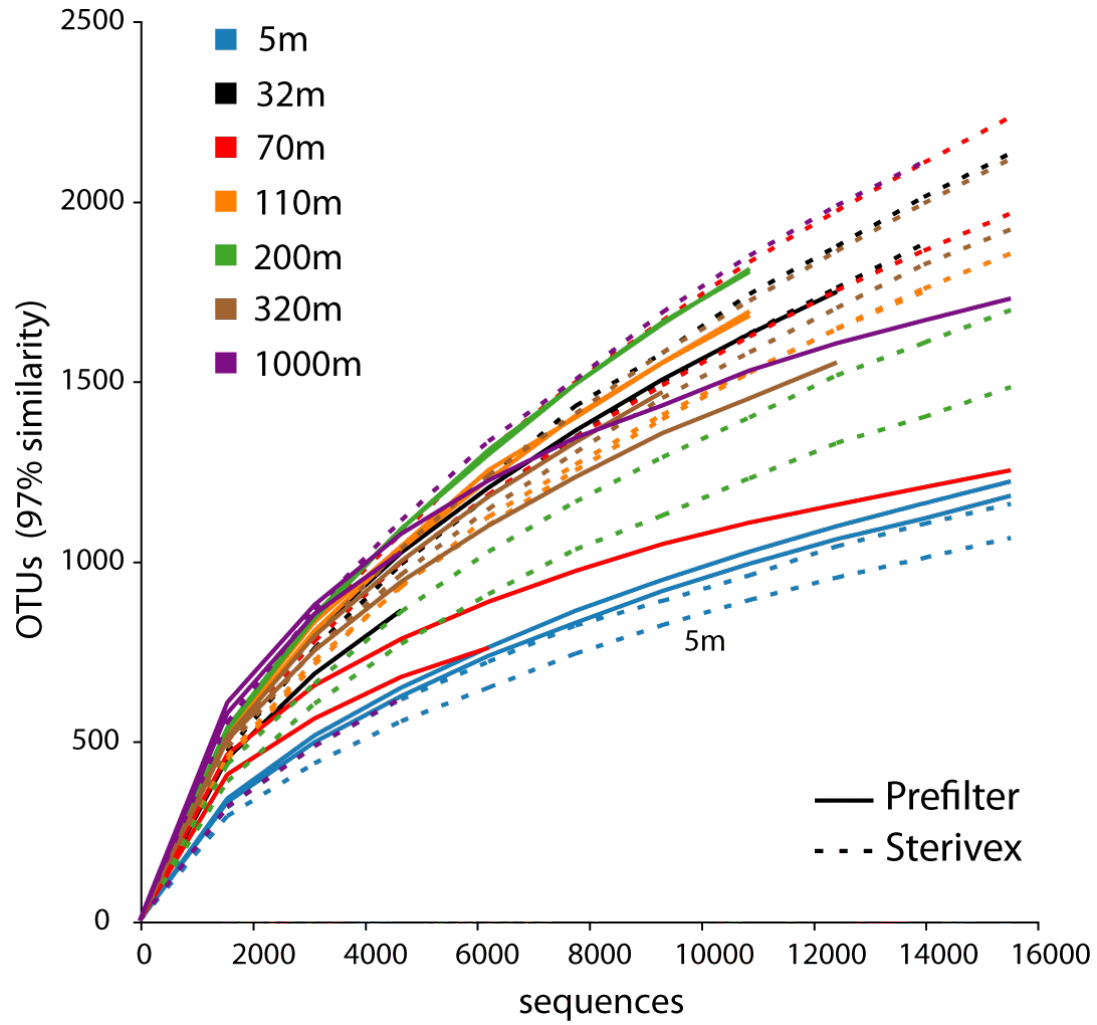


80 **Figure S1.** Dissolved oxygen, chlorophyll a, and particulate matter (beam attenuation
 85 coefficient) concentrations at Station 1 (20° 04.999S, 70° 48.001W). Data for DO and *chl a* are
 based on CTD-based DO sensor and fluorometer measurements from Cast #10 at ~2200 hrs
 on November 21, 2010. Data for beam attenuation are averages of transmissometer
 measurements from 20 casts at station #1 from November 19-24. Attenuation coefficients are
 90 not shown for depths less than 40 m, where values peaked at ~0.8 m⁻¹ close to the surface.
 16S rRNA gene amplicon diversity was analyzed for DNA collected at 7 depths (dashed lines).
 Metagenome analyses were conducted for 4 of these depths (stars). Note that *chl a* values (in
 µg L⁻¹) have been multiplied by 100 to match axis units. Oxygen profiles from other days during
 the 5-day sampling period (Nov. 19-23) closely resembled that shown in the Figure.

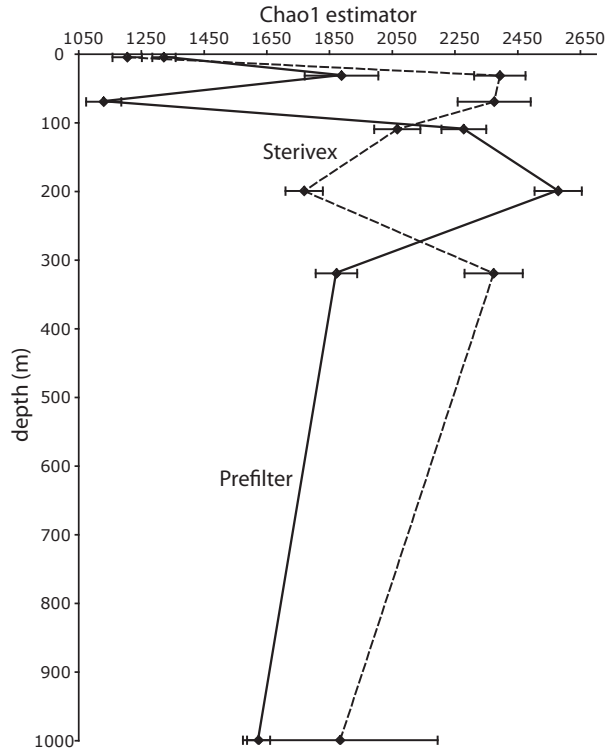
90

95

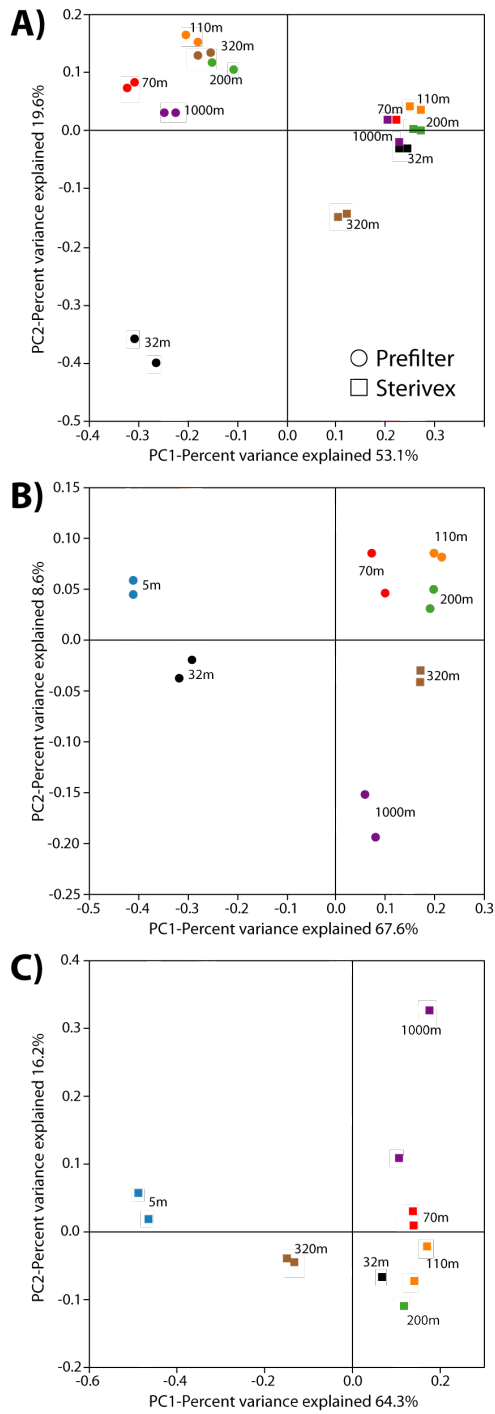
100



105 **Figure S2.** Number of observed OTUs (97% similarity clusters) as a function of sequencing depth, based on rarefaction of OTU counts. Duplicate samples reflect duplicate PCR reactions.



110 **Figure S3.** Chao1 estimator as a function of water column depth. Data points are mean values based on rarefaction at a standardized sequence count ($n = 4996$) per sample, with bars indicating 95% confidence intervals for the rarefied measurements. Data from both PCR duplicates are combined for averaging.



115

Figure S4. Principle component analysis of community relatedness based on relative bacterial taxon abundance in the 16S rRNA gene amplicon pool, as quantified by the weighted Unifrac metric. *A)* Excludes samples from the 5m depth. *B)* Prefilter only samples. *C)* Sterivex only samples.

120

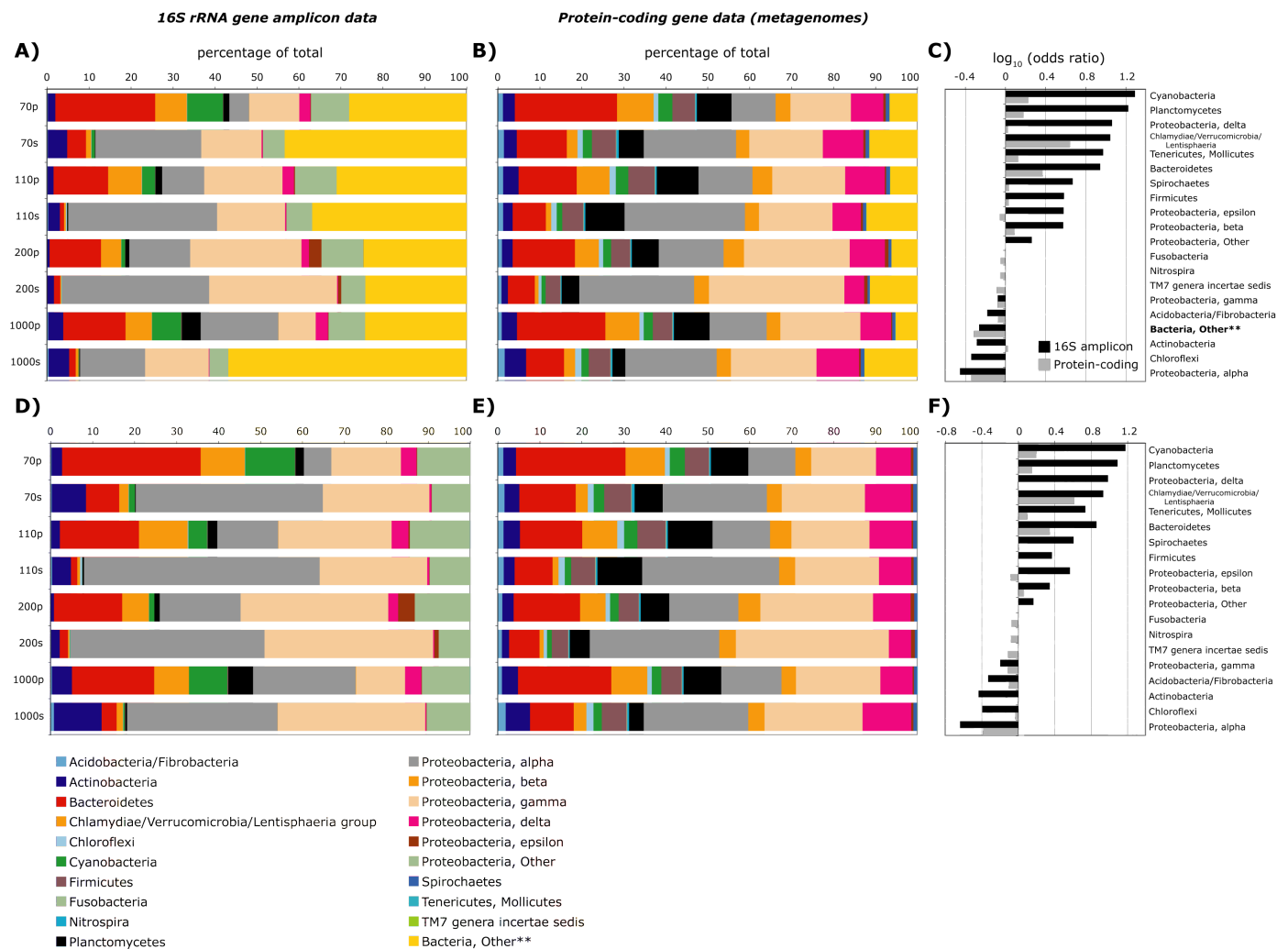
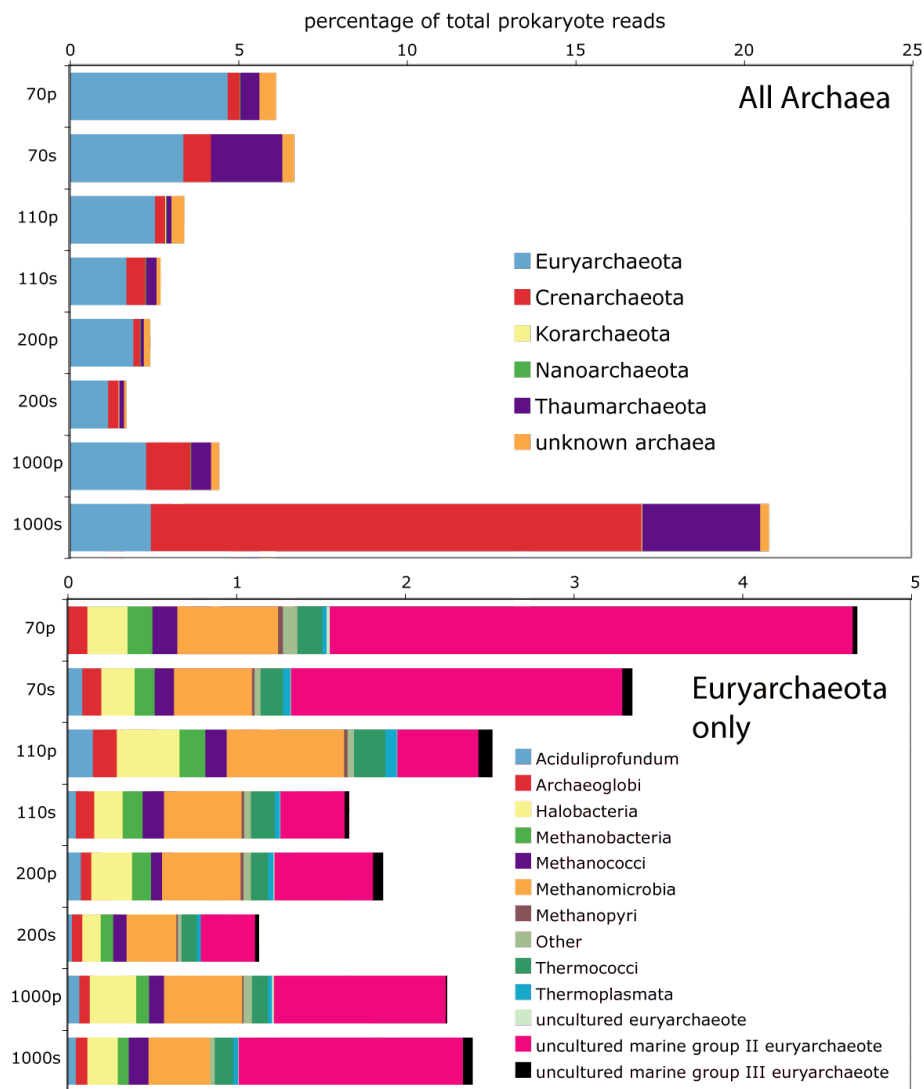


Figure S5. Relative abundance of major bacterial divisions at four depths in the OMZ based on the taxonomic identification of 16S rRNA gene fragments (A,D) and protein-coding genes identified in metagenomes (B,E). 16S rRNA data are those from Figure 1 (main text), with certain groups collapsed to higher taxonomic levels to match taxonomic groupings based on protein-coding genes. Panels A and B include sequences in the category “Bacteria, Other**”, which contains sequences unidentifiable to Phylum level, as well as sequences in bacterial divisions that were detected in one analysis (16S or metagenome) but not the other. These include 1) candidate divisions BRC1, OD1, SR1 and WS3, which were identifiable only in the 16S amplicon analysis, and 2) the Aquificae, Caldiserica, Deferribacteres, Deinococcus-Thermus, Elusimicrobia, Gemmatimonadetes, Synergistetes, Thermotogae, and candidate divisions NC10 and OP1, all of which were identifiable only based on protein-coding gene annotations. Protein-coding genes annotated as “unknown prokaryote” were also included in the “Bacteria, Other” category, although a proportion of these could belong to Archaea. Panels C and D exclude the “Bacteria, Other” category. The category “Proteobacteria, Other” was present only in the 16S amplicon analysis (based on the RDP classification in QIIME). Panels C and F show variation in the relative abundance of bacterial divisions between filter size fractions, as inferred from both 16S and protein-coding gene data. Values are the base-10 logarithm of the odds ratio: the ratio of the odds a taxon occurs in the prefilter fraction to the odds it occurs in the Sterivex fraction. Positive values indicate taxa that are more likely to occur in the prefilter. Values are based on counts pooled from only the four depths shown, with corrections for differences in dataset size. Relative abundances are calculated as a percentage of total identifiable bacterial sequences in C, and total identifiable bacterial sequences excluding “Bacteria, Other” in F. Samples are labeled by depth and filter type, where p = prefilter (>1.6 μm), s = Sterivex (0.2-1.6 μm).



150

Figure S6. Relative abundance and phylum-level taxonomic composition of protein-coding reads matching the domain Archaea. Relative abundance is expressed as a percentage of total prokaryote protein-coding reads. Phylum designations are based on annotations of NCBI-nr genes identified as top matches via BLASTX (bit score > 50). Samples are labeled by depth and filter type, where p = prefilter (>1.6 μm), s = Sterivex (0.2-1.6 μm).

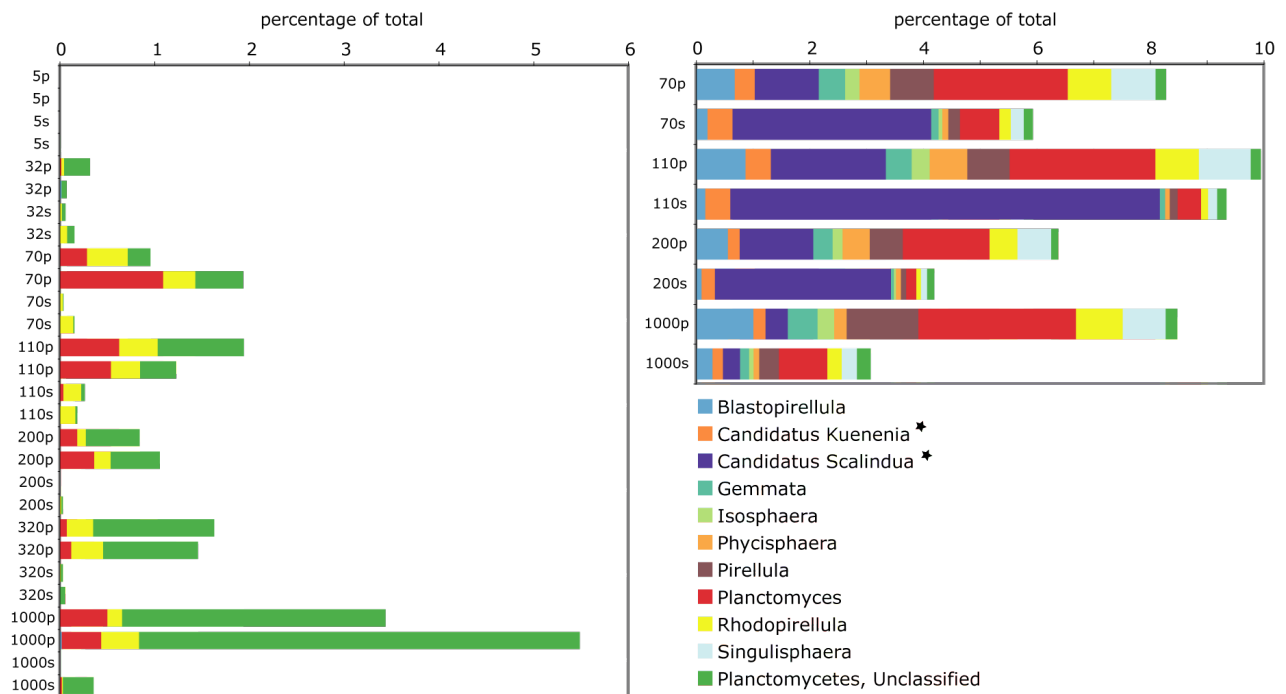
155

160

165

16S rRNA gene amplicon data

Protein-coding gene data (metagenomes)



170

Figure S7. Relative abundance and genus-level taxonomic composition of 16S rRNA (left) and protein-coding (right) reads matching Planctomycetes. Relative abundance is expressed as a percentage of total bacteria protein-coding reads (Protein-coding genes annotated as “unknown prokaryote” were also included in the total bacteria count, although a proportion of these could belong to Archaea). Taxon designations are based on RDP classifications of 16S rRNA gene fragments (left) and the annotations of NCBI-nr genes identified as top matches via BLASTX (bit score > 50; right). Samples are labeled by depth and filter type, where p = prefilter (>1.6 μm), s = Sterivex (0.2-1.6 μm). Stars mark taxa known to conduct anammox.

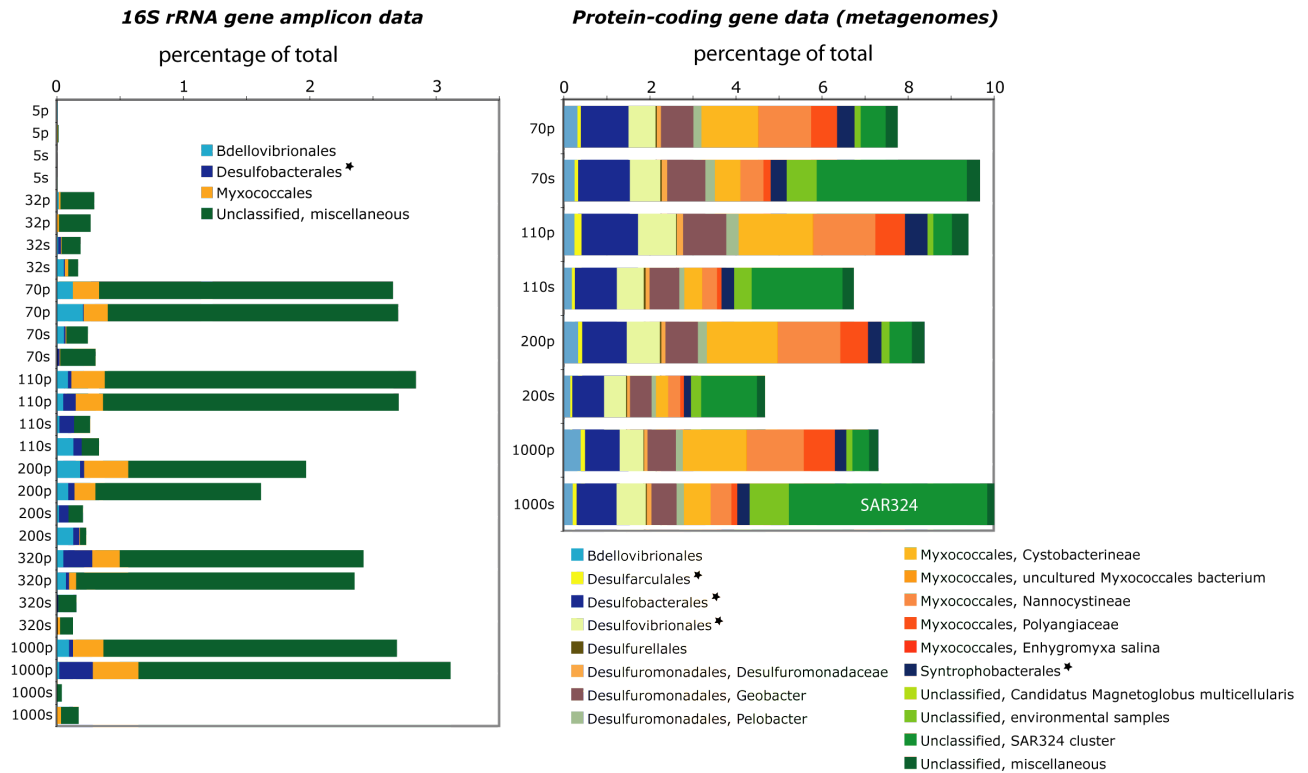
175

180

185

190

195



200 **Figure S8.** Relative abundance and taxonomic composition of 16S rRNA (left) and protein-
 coding (right) reads matching Deltaproteobacteria. Relative abundance is expressed as a
 percentage of total bacterial reads (Protein-coding genes annotated as “unknown prokaryote”
 were also included in the total bacteria count, although a proportion of these could belong to
 Archaea). Taxon designations are based on RDP classifications of 16S rRNA gene fragments
 205 (left) and the annotations of NCBI-nr genes identified as top matches via BLASTX (bit score >
 50; right). Samples are labeled by depth and filter type, where p = prefilter (>1.6 μm), s =
 Sterivex (0.2-1.6 μm). Stars mark groups known to contain sulfate-reducing members.