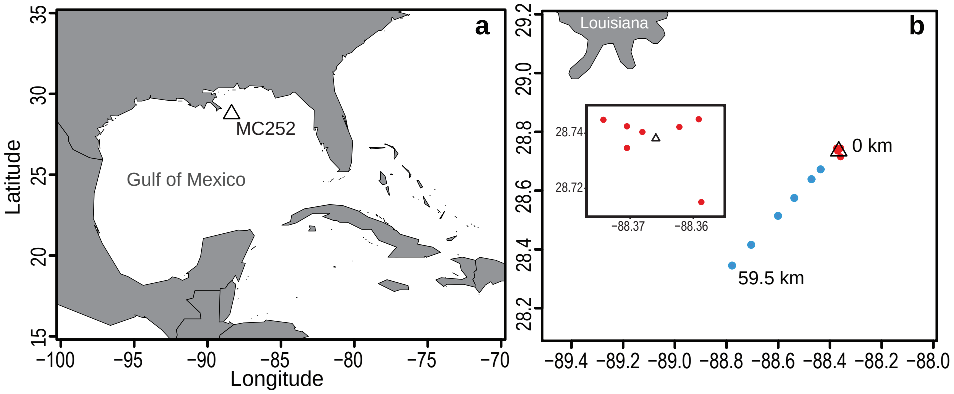

Figure S1. **(a)** Location of well MC252, and **(b)** sampling locations. Samples >3 km from MC252 are shown in blue, while near-well samples are shown in red (inset).

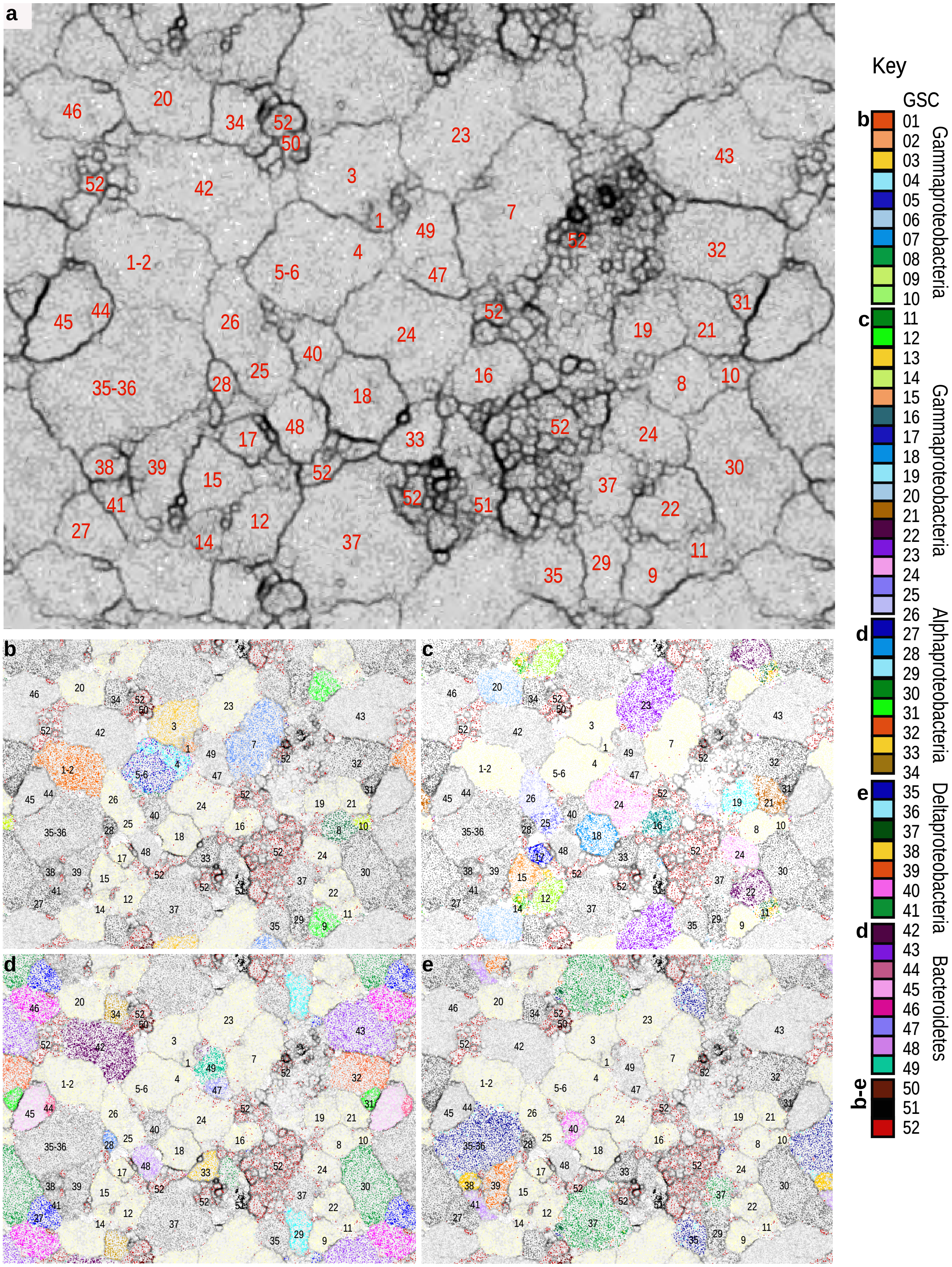

Figure S2. (a) ESOM of co-assembly constructed using tetranucleotide frequencies of 5 kbp long genomic fragments and differential coverage. Bin (GSC) numbers are shown in red. Dark lines demark cluster edges. Collections of small clusters are contigs containing virus-associated genes. **(b-e)** ESOM with data points representing 2 kbp long genomic fragments and colored by bin (see key).

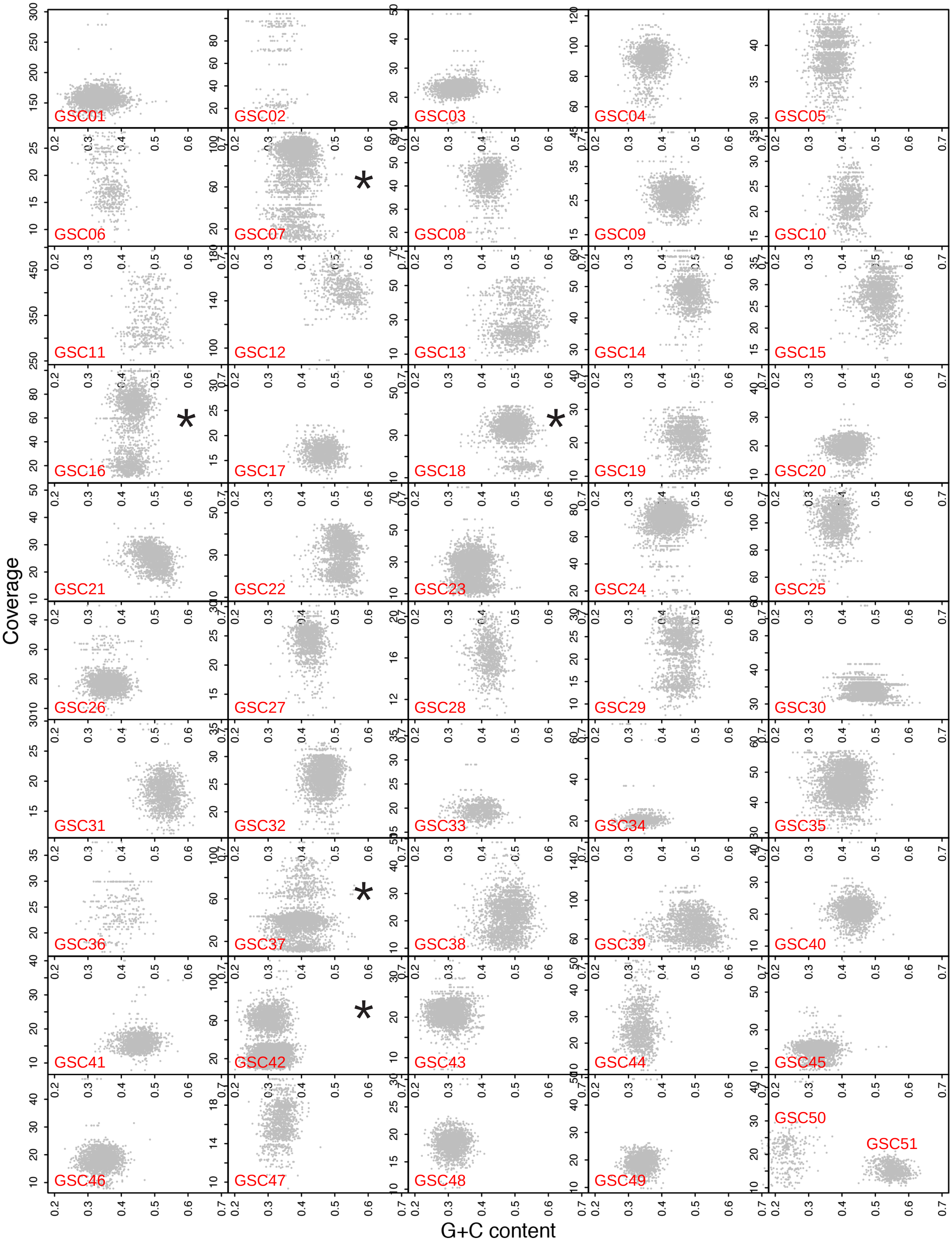

Figure S3. Coverage and GC content of contigs within each bin. Y-axes have different scales (LHS of each plot). *Compositionally similar co-binned genomes with imprecise coverage separation.

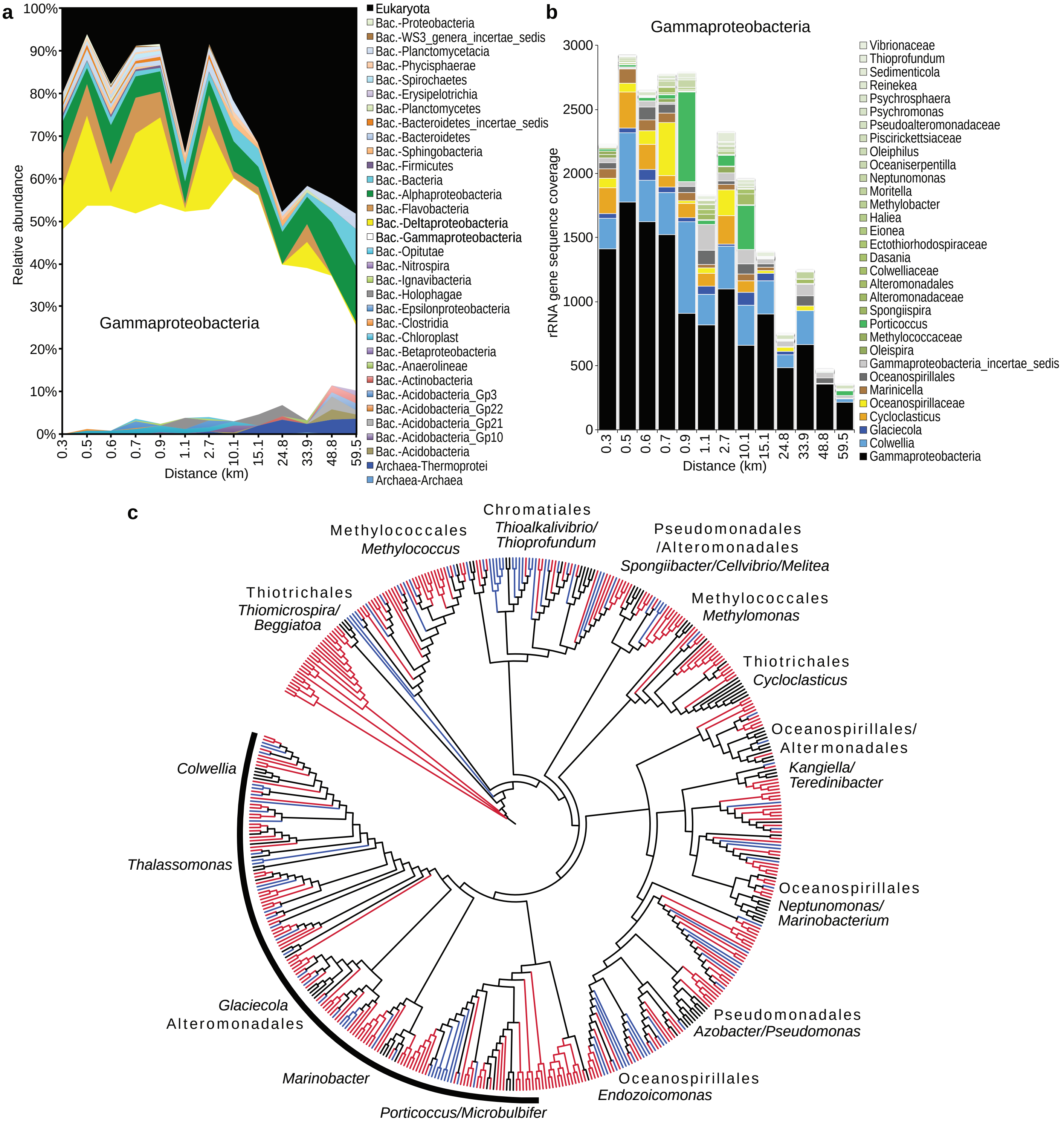

Figure S4. Phylogeny based on rRNA genes reconstructed from rarefied data, **(a)** Phylogeny per site. Eukaryota are relatively abundant in distal sites (23-52% versus 6-34% near-well). Bacteria, particularly Gammaproteobacteria increase in abundance relative to both Eukaryota and Archaea. **(b)** Stacked bar chart of Gammaproteobacteria rRNA gene sequence coverage per site based on RDP genus level designation or higher. The average number of unique designations is 13 ± 2 (1 standard deviation) for near-well sites, twice that for distal sites (7 ± 3). **(c)** Neighbor-joining tree depicting phylogenetic 16S rRNA gene diversity among the Gammaproteobacteria for near-well (red) and distal samples (blue), compared with reference sequences (black). Genus (italics) and order (bold) names are given for reference sequence clusters. Gammaproteobacterial richness at near-well sites was 317 (15,000x total 16S rRNA gene coverage), and the richness at distal sites was 152 (6,000x total 16S rRNA gene coverage).

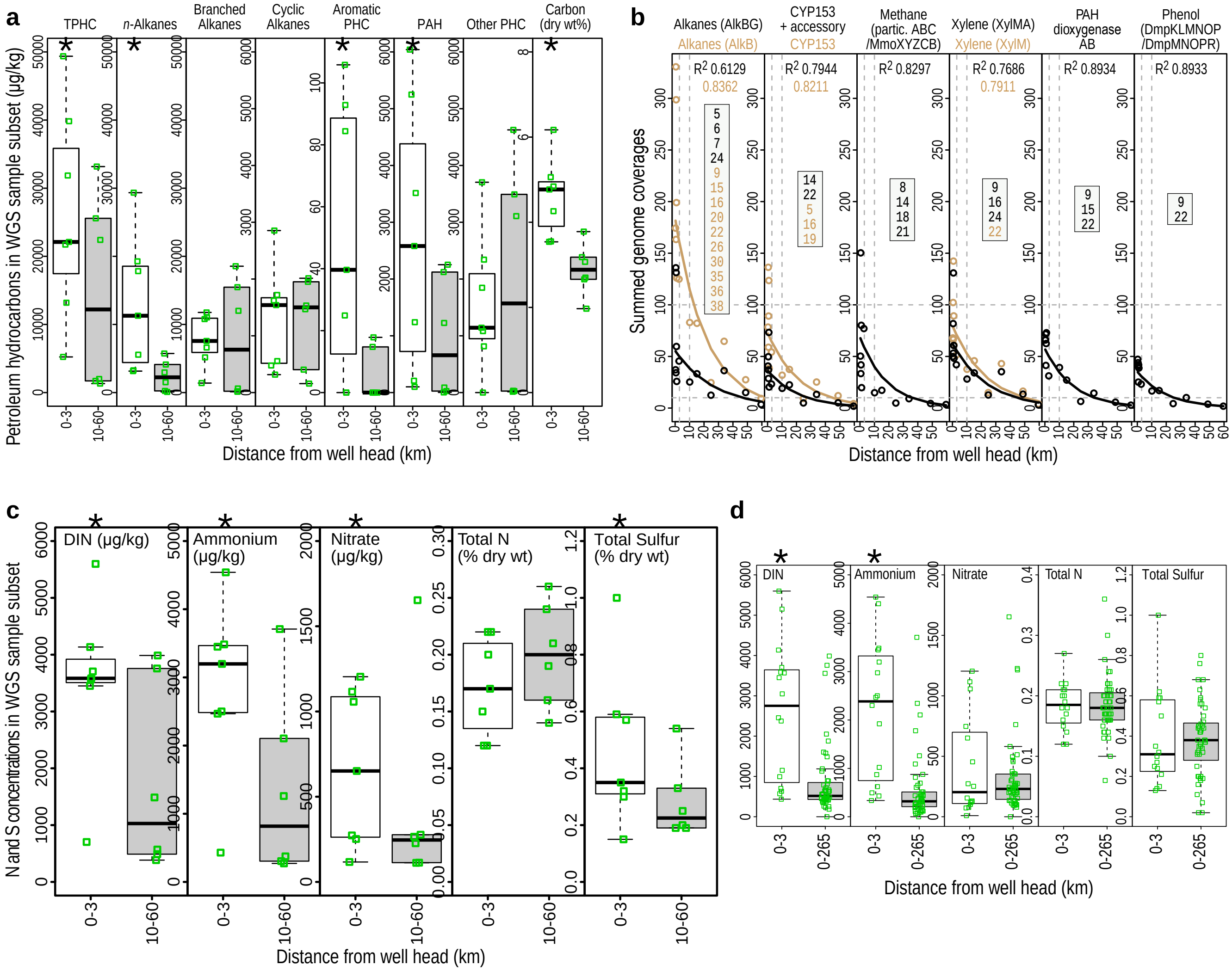

Figure S5. (a) Boxplots of petroleum hydrocarbon and carbon concentrations in the 13 metagenome samples 0-3 km (near-well) and 10-60 km (distal) from the MC252 wellhead. **(b)** Summed per site abundance of genome bins possessing genes associated with hydrocarbon degradation. Points are fitted with exponential curves. Bin numbers are listed in boxes central to each plot. Points in yellow represent the combined abundance of bin numbers given in black and yellow. **(c)** Boxplots of N and S concentrations in samples 0-3 km (near-well) and 10-60 km (distal) from the MC252 wellhead. DIN values comprised of ammonium + nitrate. **(d)** Boxplots of N and S concentrations in all 64 samples reported by Mason et al. (2014b) at 0-265 km distance. Abbreviations: (total) petroleum hydrocarbons, (T)PHC; weight, wt; dissolved inorganic nitrogen (DIN). *Concentrations higher near-well. The underlying geochemical data presented in boxplots are from Mason et al. (2014b supplementary online materials where total alkanes = total *n*-alkanes).

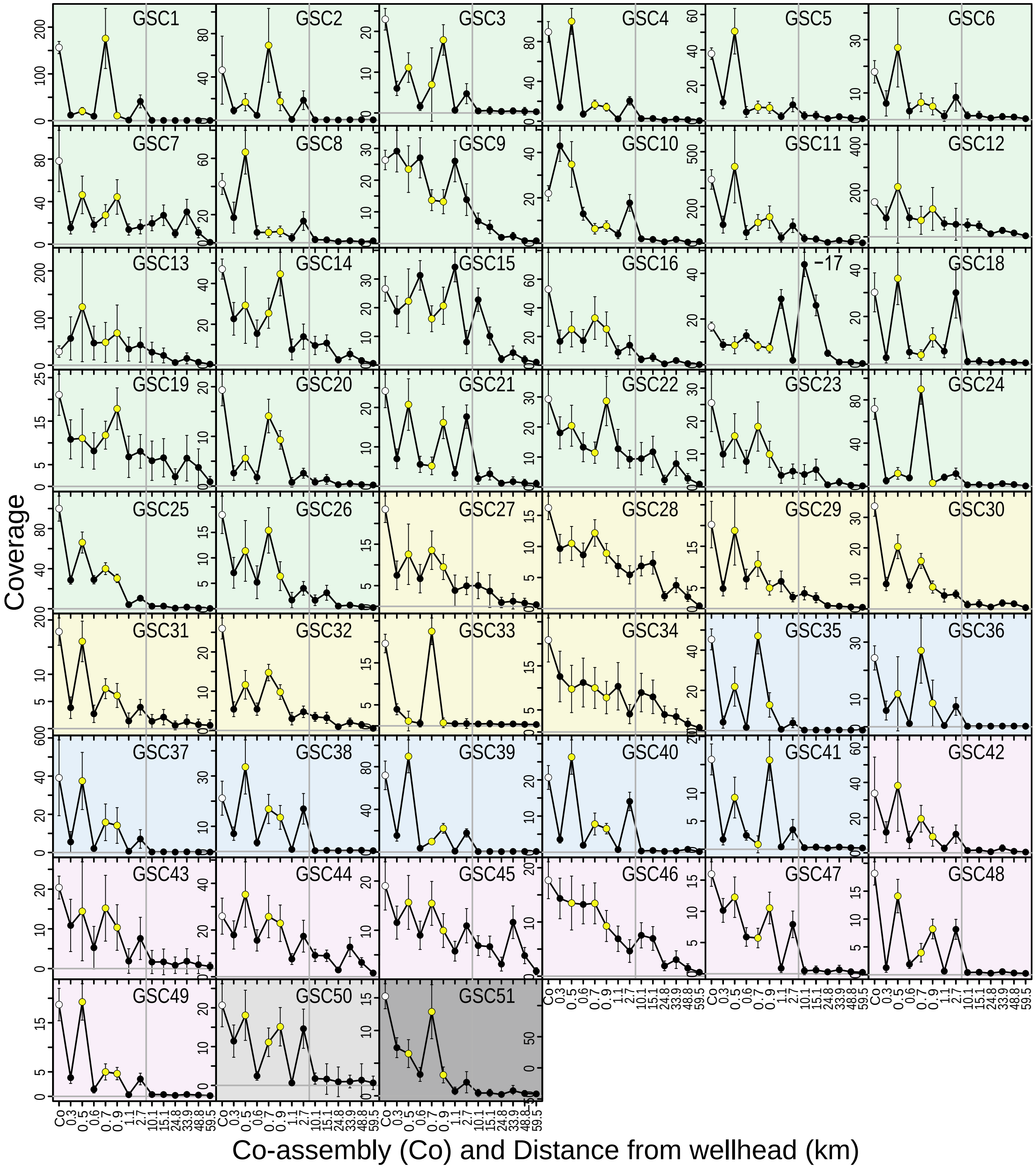

Figure S6. Average genome bin coverages per site determined by mapping reads to the co-assembly. Error bars = 1 standard deviation. Samples contributing to the co-assembly (white points) are bolded and points shown in yellow. Plots are organized by phylogenetic group: Gammaproteobacteria (green), Alphaproteobacteria (yellow), Deltaproteobacteria (blue), Cytophaga-Flavobacterium-Bacteroides group (CFB, pink), and unknown Bacteria (grey).

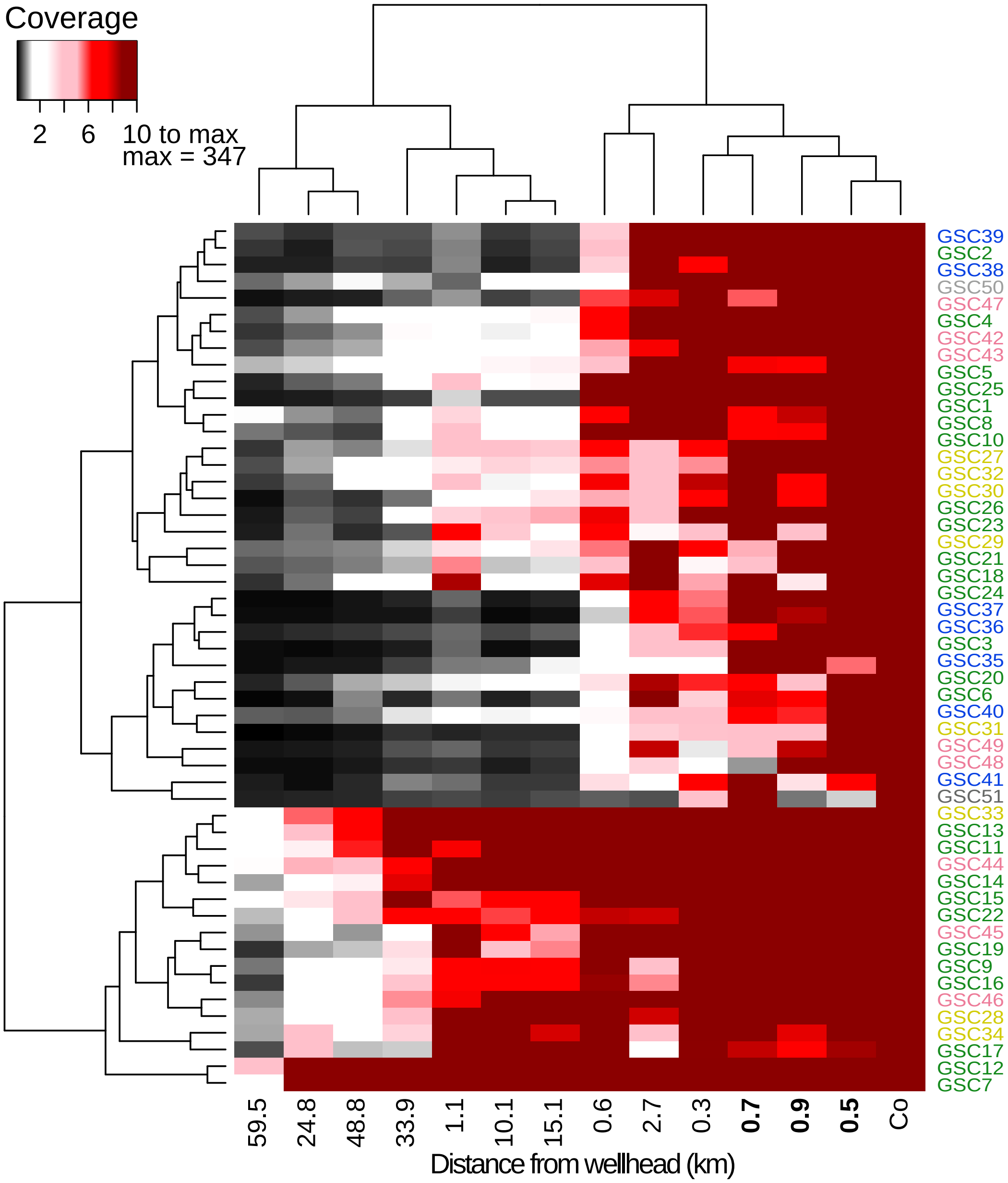

Figure S7. Heatmap and hierarchical clustering of genome bin abundance per site or Co-assembly (Co). Samples contributing to the co-assembly are bolded. Abundances ≥10 are assigned the same color (dark red). Bins (RHS) are colored by taxonomic group: Gammaproteobacteria (green), Alphaproteobacteria (yellow), Deltaproteobacteria (blue), Cytophaga-Flavobacterium-Bacteroides group (CFB, pink), and unknown Bacteria (grey).

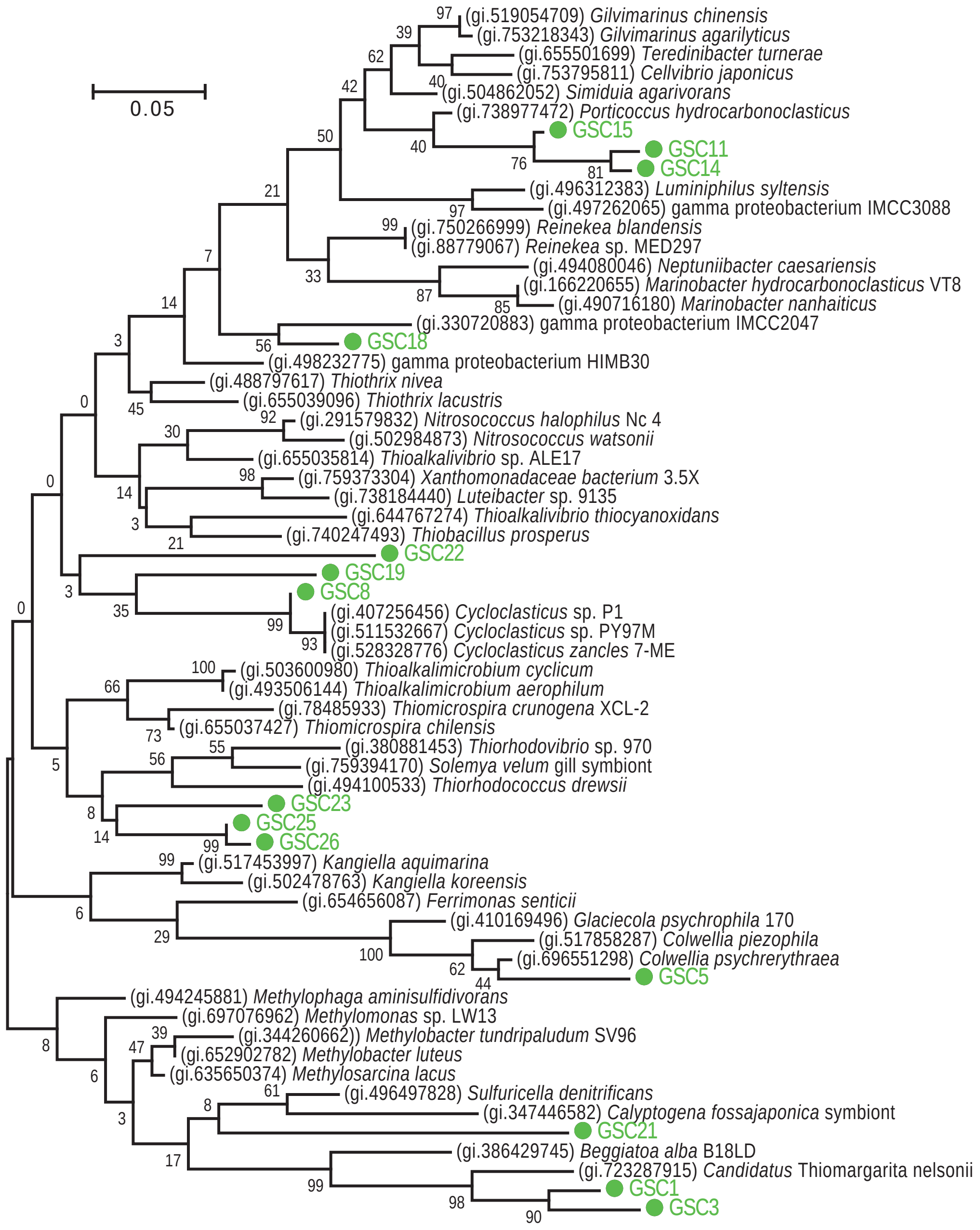

Figure S8. Maximum-Likelihood tree showing the phylogenetic relatedness of genome bins based on recovered predicted protein sequences of the highly conserved housekeeping gene recombinase A (*recA*). Sequences from our seafloor genome bins are shown in green, while those from reference organisms are in black. Reference sequence GenBank accession numbers are in parentheses. Tree construction employed 190 amino acid positions and 500 bootstrap replicates.

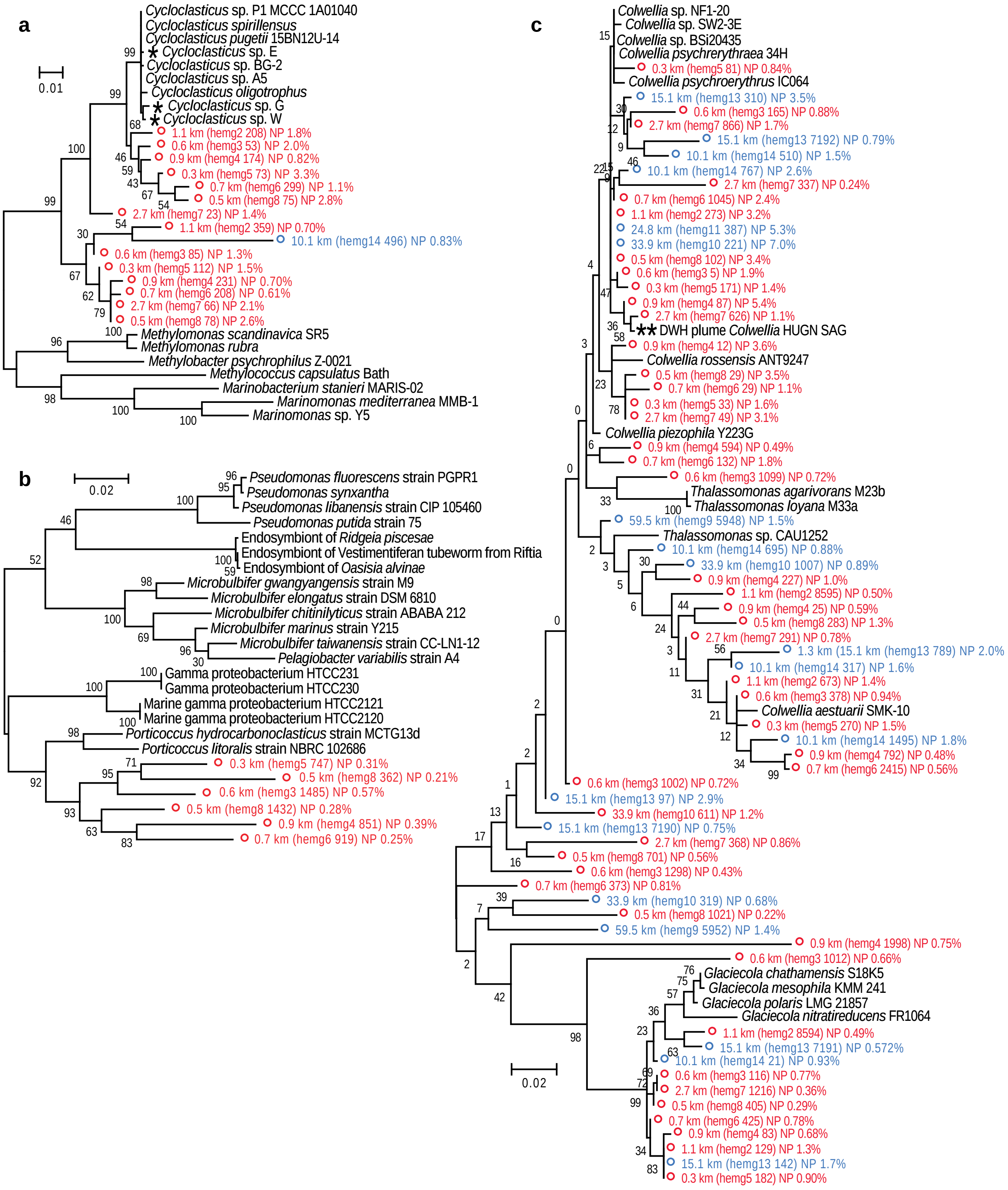

Figure S9. Maximum-Likelihood trees showing the phylogenetic relatedness of 16S rRNA genes sequences from near-well (red) and distal (blue) sites that are most closely related to **(a)** *Cycloclasticus*, **(b)** *Porticoccus*, and **(c)** *Colwellia* species. Sequence identifiers are in parentheses. Norm Priors (NP) denote abundances per sample. Tree construction employed near full-length 16S rRNA genes and 500 bootstrap replicates.16S sequences related to these hydrocarbonoclastic genera were recovered from rarified sequence data up to ~10 km from MC252 for *Cycloclasticus* and *Porticoccus*, or 33.9 km in the case of *Colwellia*. *Northern Gulf of Mexico *Cycloclasticus* strains (Geiselbrecht *et al*., 1998). **16S from a DWH plume *Colwellia* single amplified genome (SAG) sequence (Mason *et al*., 2014a).

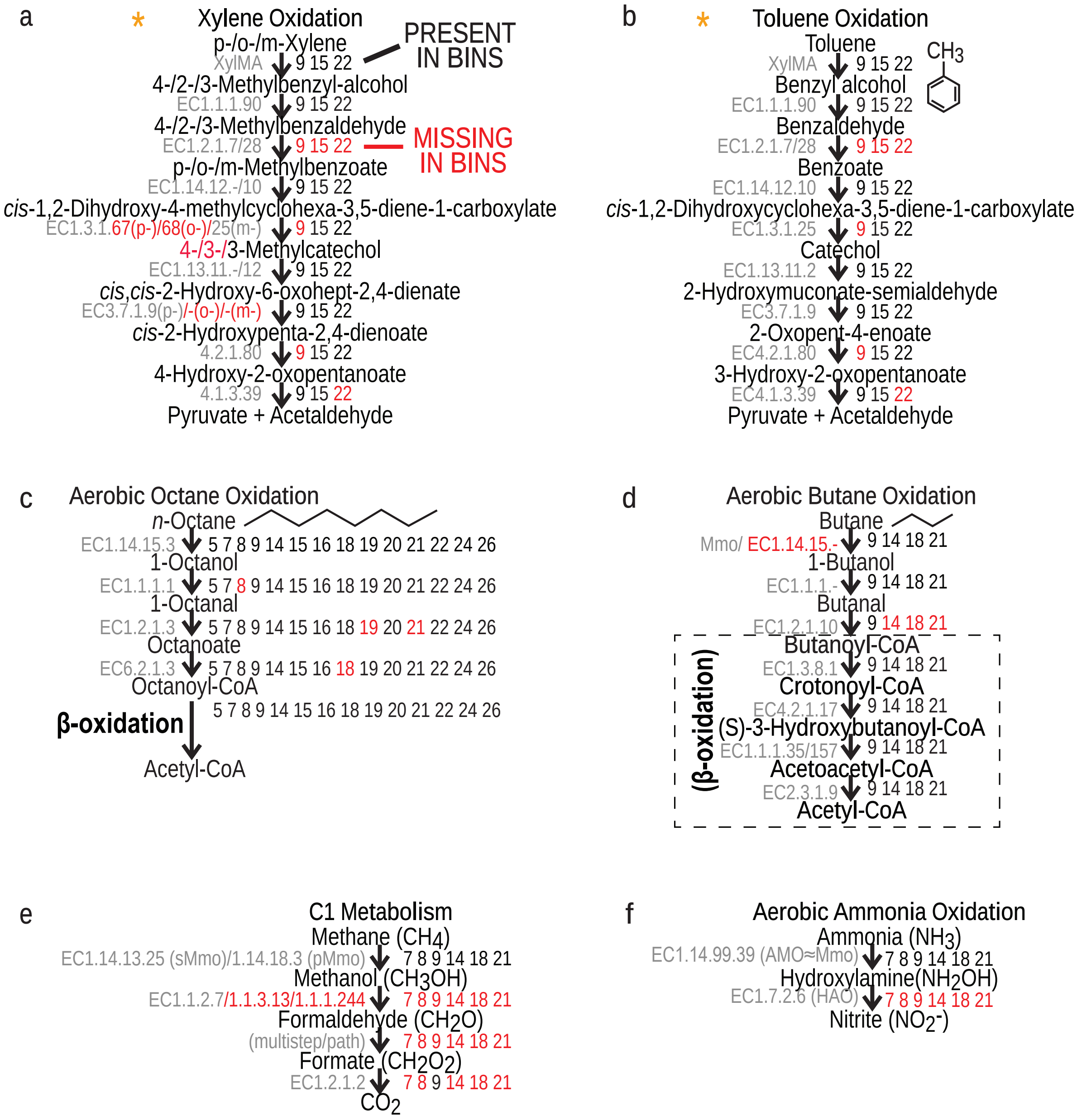

Figure S10. (a-b) Aromatic hydrocarbon oxidation pathways indicated by a yellow * in genome bin Group 4 (Fig. 4). **(c-d)** Alkane oxidation pathways using **(c)** alkane 1-monoxygenase and targeting C5-C10 alkanes, or **(d)** methane monooxygenase, Mmo, and targeting C1-C5 alkanes. **(e-f)** Incomplete methane and ammonium oxidation pathways using Mmo. **(a-f)** Genome bin numbers are shown to the right of pathway arrows. Genes present are shown in grey. Bins with some of the pathway genes present are shown in black. Genes missing and bins with genes unidentified for particular steps are shown in red.

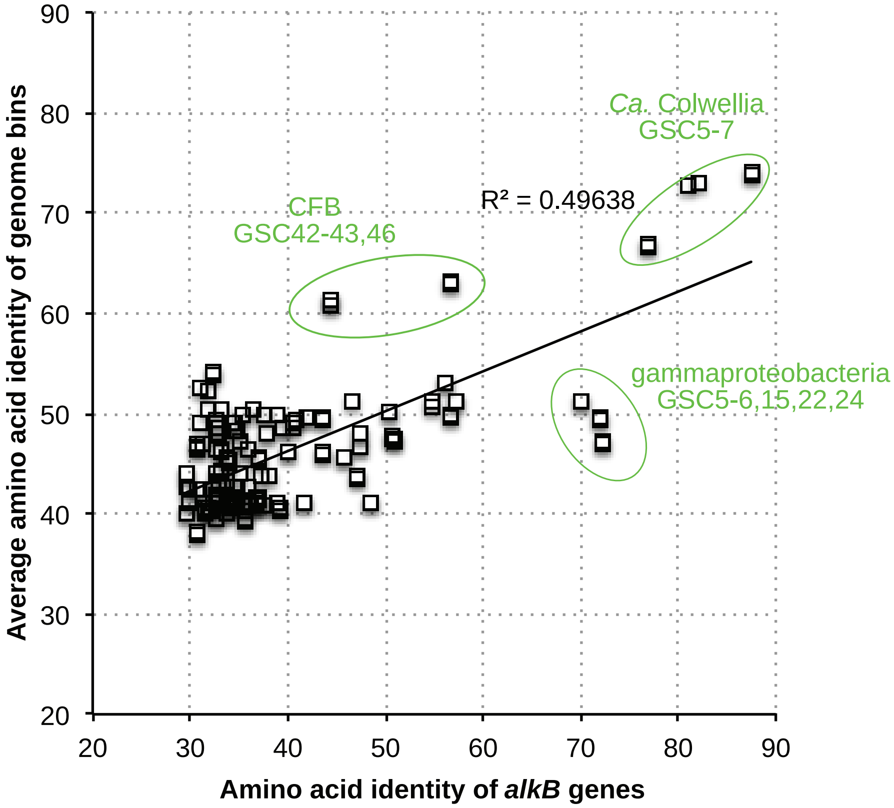

Figure S11. Comparisons of *alkB* gene and genome bin relatedness based on pairwise amino acid identities determined during reciprocal searches.

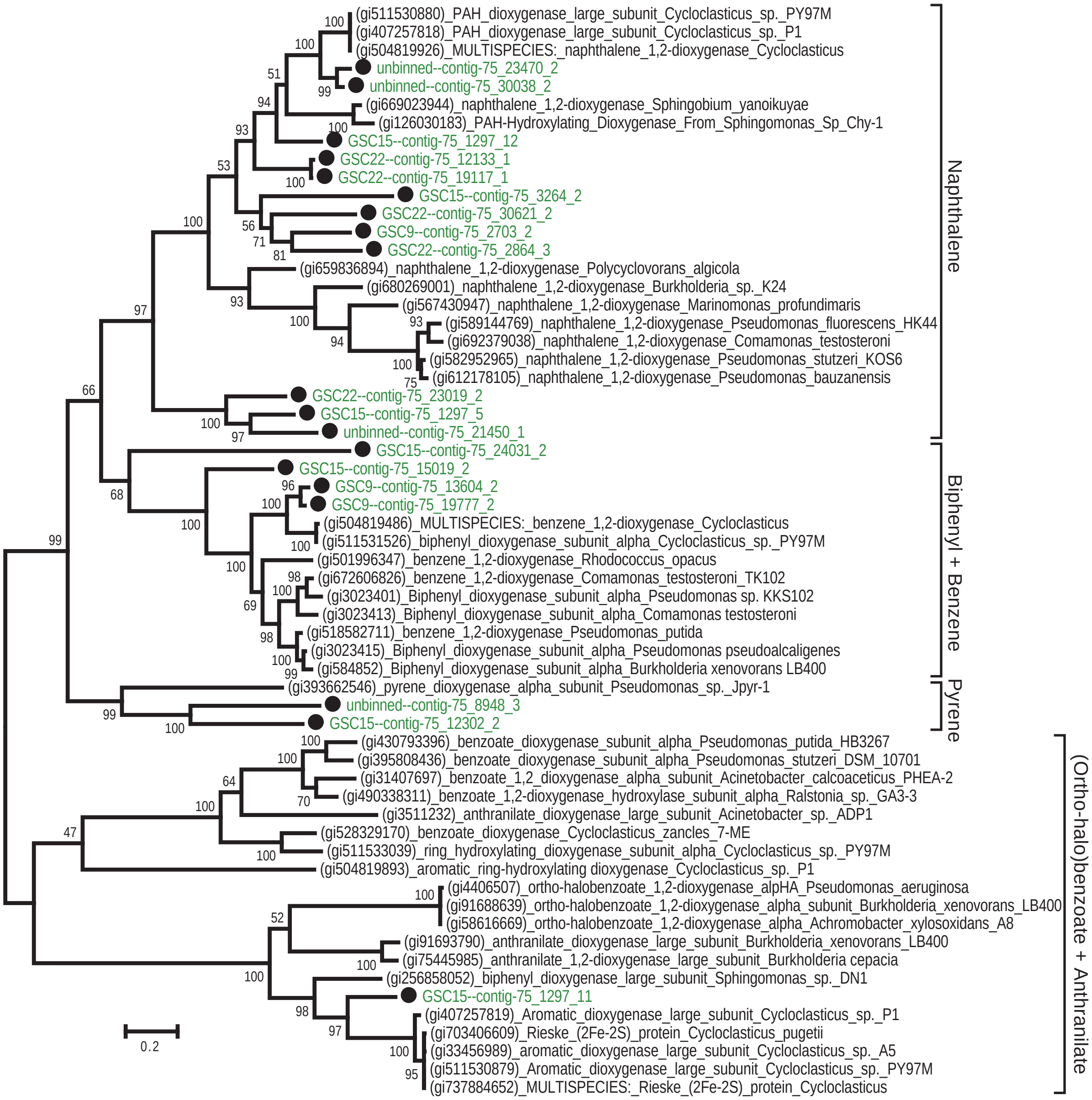

Figure S12. Maximum-Likelihood tree depicting the phylogenetic relatedness among predicted protein sequences of (near) full length candidate PAH dioxygenases alpha subunits from our genome bins (green) compared with reference naphthalene, biphenyl, pyrene and anthranilate PAH dioxygenases and benzene and (ortho-halo-)benzoate aromatic dioxygenases. GenBank accession numbers for reference sequences are in parentheses. Tree construction employed 340 amino acid positions and 500 bootstrap replicates.

Table S1. Sequence data input (trimmed), and IDBA-UD assembly summary.

Sample ID	Distance (km)	WGS PE reads (millions)[1]	WGS length (Gbp)	Assembled length (Mbp)[2]	N50	Max contig	Contig count*	WTS PE reads (millions)[1]	Reads mapped to rRNA	Reads mapped to CDS	rRNA (% of reads)	RNA types analyzed	RNA amplified
BP_331	0.3	166	17	28	3,417	73,930	8,010	48	6,161,432	714,996	89.6	rRNA+mRNA	Yes
BP_315	0.5	275	28	63	4,052	101,350	15,864	58	3,956,483	330,029	92.3	rRNA+mRNA	-
BP_350	0.6	153	15	26	3,672	74,260	7,077	52	3,243,259	185,051	94.6	rRNA+mRNA	-
BP_139	0.7	187	19	79	3,981	76,457	19,965	50	3,018,430	306,824	90.8	rRNA+mRNA	-
BP_143	0.9	204	21	58	4,101	51,853	14,517	58	3,854,101	353,822	91.6	rRNA+mRNA	-
BP_366	1.1	266	27	34	3,760	131,389	8,923	37	3,390,160	377,874	90.0	rRNA+mRNA	Yes
BP_120	1.3	-	-	-	-	-	-	37	1,808,118	244,029	88.1	rRNA	Yes
BP_278	2.7	144	15	28	3,384	73,930	7,896	53	3,293,179	291,043	91.9	rRNA+mRNA	-
Average	**1.0**	**199**	**20**	**45**	**3,767**	80,746	**11,750**	**49**	**3,590,645**	**350,459**	**91.1**	**-**	**-**
BP_101	10.1	285	29	18	3,648	78,608	4,640	32	3,326,370	522,218	86.4	rRNA+mRNA	Yes
BP_155	15.1	344	35	12	3,464	58,165	3,517	21	1,990,232	225,102	89.8	rRNA+mRNA	Yes
BP_444	24.8	369	37	2	3,182	26,366	557	-	-	-	-	-	-
BP_463	33.9	291	29	6	3,139	42,788	1,732	36	1,337,250	202,961	86.8	rRNA+mRNA	Yes
BP_501	48.8	355	36	10	4,238	26,854	2,481	-	-	-	-	-	-
BP_186	59.5	295	30	1	2,646	23,506	481	-	-	-	-	-	-
Average	**27.6**	**305**	**31**	**8**	**3,386**	42,715	**2,235**	**30**	**2,217,951**	**316,760**	**87.7**	**-**	**-**
Co-assembly	**0.5+0.7+0.9**	**666**	**68**	**181**	**4,748**	**134,227**	**41,000**	**-**	-	-	**-**	**-**	**-**

Note: Co-assembly and co-assembled samples are shown in blue.

[1] Trimmed read counts

[2] Contigs >2kbp long.

Table S2. Genomic bin characteristics.

Bin GSC	Nearest related genome (reference)	AAI[1] (%) to reference	Nearest RecA	RecA %ID	Average G+C	Average coverage	%CGs /107	%Surplus CGs /107	Genomes per bin	Phylog. cluster	AAI[2] (%) to bin	CDS count	Bin length (bp)	Contig count
1	*Candidatus* Thiomargarita nelsonii	66	*Ca.* Thiomargarita nelsonii	84	32	156	83	4	1	1	86 GSC3	4,088	4,888,626	787
2	*Candidatus* Thiomargarita nelsonii	65	-	-	32	46	10	0	1	1	80 GSC1	264	276,972	41
3	*Candidatus* Thiomargarita nelsonii	66	*Ca.* Thiomargarita nelsonii	86	32	23	76	1	1	1	86 GSC1	2,709	2,996,330	465
4	*Colwellia psychrerythraea* 34H	73	-	-	36	90	41	2	1	2	82 GSC6	1,929	1,928,427	406
5	*Colwellia psychrerythraea* 34H	73	*C. psychrerythraea*	92	36	38	53	3	1	2	81 GSC4	2,020	2,113,321	218
6	*Colwellia psychrerythraea* 34H	68	-	-	36	18	7	0	1	2	82 GSC4	643	575,276	175
7	*Colwellia psychrerythraea* 34H	90	-	-	38	90,22	57	2	2*	2	74 GSC4	4,925	5,186,372	909
8	*Cycloclasticus zancles*	73	*Cycloclasticus* sp. P1, *C.zancles*	94	42	42	50	0	1	3	86 GSC10	1,691	1,571,359	288
9	*Cycloclasticus* sp. P1	79	*Cycloclasticus* sp. P1, *C.zancles*	94	44	26	48	21	2**	3	73 GSC8	2,643	2,296,057	617
10	*Cycloclasticus* sp. P1	72	-	-	42	22	17	0	1	3	86 GSC8	1,070	889,100	249
11	*Porticoccus hydrocarbonoclasticus* MCTG13d	59	*P. hydrocarbonoclasticus* MCTG13d	86	49	347	30	0	1	4	83 GSC13	563	547,319	149
12	*Porticoccus hydrocarbonoclasticus* MCTG13d	58	-	-	51	150	13	0	1	4	82 GSC11	776	758,082	213
13	*Porticoccus hydrocarbonoclasticus* MCTG13d	56	-	-	50	29	14	1	1	4	82 GSC11	1,516	1,432,455	384
14	*Porticoccus hydrocarbonoclasticus* MCTG13d	61	*P. hydrocarbonoclasticus* MCTG13d	84	48	47	82	1	1	4	73 GSC11	1,541	1,513,903	287
15	*Porticoccus hydrocarbonoclasticus* MCTG13d	60	*P. hydrocarbonoclasticus* MCTG13d	88	50	27	80	0	1	4	64 GSC14	1,824	1,791,859	350
16	gammaproteobacterium HdN1	54	-	-	43	71,21	38	0	2*	-	54 GSC18	2,657	2,489,254	673
17	*Teredinibacter turnerae* T7901	58	-	-	46	17	31	2	1	-	55 GSC14	1,630	1,513,391	426
18	gammaproteobacterium IMCC2047	61	gammaproteobacterium IMCC2047	85	49	33,15	79	11	2*	-	55 GSC14	2,537	2,473,031	476
19	gammaproteobacterium IMCC2047	54	gammaproteobacterium HIMB30	76	47	21	48	0	1	-	55 GSC18	1,722	1,606,611	336
20	*Cycloclasticus* sp. P1	50	-	-	41	19	69	2	1	-	50 GSC23	2,790	2,689,849	549
21	*Candidatus* Thiomargarita nelsonii	53	*Thiothrix nivea*	79	49	24	80	1	1	-	53 GSC8	1,979	1,865,361	371
22	*Nitrosococcus watsonii*	51	*Ca.* Tenderia electrophaga	79	51	29	42	15	2**	-	50 GSC21	2,713	2,555,975	636
23	*Kangiella koreensis* SW-125	54	*Oleiagrimonas soli, K. aquimarina*	80	37	26	64	20	2**	-	51 GSC5	5,478	5,384,675	1,125
24	*Hahella chejuensis* KCTC 2396	57	-	-	41	72	76	0	1	-	55 GSC18	4,534	4,514,783	667
25	*Cycloclasticus* sp. P1	50	*Thiorhodovibrio* sp. 970	87	37	100	49	2	1	5	73 GSC26	1,633	1,525,989	378
26	*Kangiella koreensis* SW-125	48	Solemya velum gill symbiont, *Methylophaga frappieri*	81	35	18	34	2	1	5	72 GSC25	2,765	2,737,699	608
27	*Micavibrio aeruginosavorus*	57	*M. aeruginosavorus* EPB	79	42	23	87	22	2**	-	54 GSC29	1,757	1,606,244	268
28	*Micavibrio aeruginosavorus*	54	*O. indicum, O. pacidicum*	77	41	16	41	2	1	-	53 GSC27	1,259	1,046,950	296
29	*Micavibrio aeruginosavorus*	56	-	-	44	20	38	7	1	-	54 GSC27	2,107	1,839,492	441
30	Rhodobacterales bacterium HTCC2255	63	Rhodobacterales bacterium HTCC2255	83	47	34	97	2	1	6	60 GSC31	3,424	3,436,081	110
31	Rhodobacterales bacterium HTCC2255	59	*Sulfitobacter mediterraneus*	86	52	18	42	0	1	6	60 GSC30	1,756	1,405,749	386
32	*Oceanibaculum indicum* P24	53	*Sneathiella glossodoripedis*	90	46	26	89	3	1	-	51 GSC28	3,365	3,184,965	458
33	*Oceanibaculum indicum* P24	51	endosymbiont of *Acanthamoeba* sp. UWC36	76	39	20	66	1	1	-	51 GSC32	1,095	976,746	159
34	*Rickettsia bellii*	54	-	-	33	21	76	1	1	-	48 GSC33	1,078	999,820	178
35	*Polaribacter* sp. MED152	61	*Lutibacter* sp. LP1	94-96	30	64,22	67	74	3* **	11	66 GSC44	7,657	7,284,376	1,605
36	*Polaribacter* sp. MED152	74	*Polaribacter atrinae*	96	30	20	88	65	3**	11	63 GSC42	3,939	4,125,614	523
37	*Polaribacter* sp. MED152	60	*Tenacibaculum dicentrarchi*	93	33	26	13	0	1	11	65 GSC42	1,327	1,218,909	372
38	*Aequorivita sublithincola* 9-3	66	*Ulvibacter* sp. LPB0005	90	32	19	52	6	1	11	64 GSC46	2,770	2,718,766	495
39	*Formosa* sp. AK20	68	*Algibacter lectus*	92	34	18	54	2	1	11	64 GSC45	2,765	2,643,860	527
40	*Marinilabilia salmonicolor* JCM 21150	50	-	-	33	16	34	0	1	-	53 GSC49	1,288	1,278,012	314
41	*Anaerophaga* sp. HS1	52	*Marinifilum fragile*	84	30	18	46	2	1	-	53 GSC47	2,107	2,098,621	475
42	*Marinilabilia salmonicolor* JCM 21150	49	Spingobacterium sp. PM2-P1-29	82	33	19	32	0	1	-	54 GSC47	2,075	2,112,534	522
43	*Desulfobacterium autotrophicum* HRM2	53	*Desulfatirhabdium butyrativorans*	80	40	45	69	3	1	12	70 GSC37	5,893	6,773,533	593
44	*Desultotalea psychrophila* LSv54	47	-	-	39	24	0	22	2**	12	64 GSC35	421	475,316	65
45	*Desulfobacterium autotrophicum* HRM2	54	*D. autotrophicum* HRM2	88	39	76,38,16	93	7	3*	12	70 GSC35	5,659	6,042,683	861
46	*Desulfobacula toluolica* Tol2	65	-	-	47	21	27	3	1	-	52 GSC37	2,528	2,270,163	722
47	*Desulfobulbus propionicus* 1pr3	59	*Desulfobulbus japonicus*	83	48	72	68	1	1	-	57 GSC41	2,397	2,382,941	491
48	*Desulfotalea psychrophila* LSv54	60	-	-	43	21	59	1	1	-	58 GSC41	2,229	2,096,970	552
49	*Desulfocapsa sulfexigens* SB164P1	74	-	-	45	16	53	3	1	-	59 GSC40	1,775	1,684,524	476
50	uncultured bacterium (gcode 4)	46	-	-	23	21	10	0	1	-	47 GSC2	539	484,216	145
51	*Ignavibacterium album* Mat9-16	43	-	-	55	15	12	-	1	-	43 GSC33	1,125	1,037,969	322
52	(Virus-associated)	-	-	-	40	44	-	-	-	-	-	11,444	7,892,159	1,225
-	(Unbinned)	-	-	-	-	-	-	-	-	-	-	59,689	53,523,903	16,518

Abbreviations: Phylogenetic (phylog.); average amino acid identity (AAI); ID (identity); core genes (CGs); coding DNA sequence (CDS). Bacteria bins are: Gammaproteobacteria (green), Alphaproteobacteria (yellow), Deltaproteobacteria (blue), Bacteroidetes (pink), and unclassified Bacteria (grey). AAI are given for comparisons between genome bins and: [1]the nearest related reference genome, or [2]the nearest related genome bin. Phylogenetically clustered genomes share AAIs of ≥60%, while un-clustered bins share <60% AAI with any other bin. Genome completion estimates are based on the percentage of 107 unique CGs present, without correction for co-binning. Bins estimated to be ≥50% complete are highlighted in purple. Co-binned genomes are clearly indicated by: *the detection of ≥2 compositionally similar genomes using ESOM with imprecise separation of contigs based on coverage; and are suggested by **the presence of surplus CG copies in a bin, which are typically present in a genome as single copy genes.

Table S3. Pearson's correlations between distance and (hydro)carbon concentrations or genome coverage.

	Distance		Distance	
	Correlation coefficient (linear)	P-value	Correlation coeff. (exponential*)	P-value
Total_petroleum_hydrocarbons	-0.17	1.00	0.03	1.00
Aromatics	-0.40	1.00	NaN	NaN
Alkanes	-0.39	1.00	-0.15	1.00
PAHs	-0.26	1.00	-0.10	1.00
Branched_alkanes	0.15	1.00	0.08	1.00
Cyclic_alkanes	0.07	1.00	0.16	1.00
Other_petroleum_hydrocarbons	0.22	1.00	NaN	NaN
Carbon	-0.61	0.17	-0.73	0.02
GSC1	-0.33	1.00	-0.82	0.01
GSC2	-0.41	1.00	-0.80	0.02
GSC3	-0.51	1.00	-0.82	0.01
GSC4	-0.37	1.00	-0.82	0.01
GSC5	-0.37	1.00	-0.80	0.02
GSC6	-0.44	1.00	-0.82	0.01
GSC7	-0.51	1.00	-0.52	0.14
GSC8	-0.40	1.00	-0.86	0.01
GSC9	-0.78	0.08	-0.95	0.00
GSC10	-0.52	1.00	-0.91	0.00
GSC11	-0.49	1.00	-0.87	0.00
GSC12	-0.65	0.62	-0.85	0.01
GSC13	-0.69	0.36	-0.89	0.00
GSC14	-0.68	0.44	-0.85	0.01
GSC15	-0.77	0.11	-0.84	0.01
GSC16	-0.73	0.22	-0.91	0.00
GSC17	-0.38	1.00	-0.51	0.14
GSC18	-0.44	1.00	-0.76	0.02
GSC19	-0.72	0.26	-0.77	0.02
GSC20	-0.49	1.00	-0.82	0.01
GSC21	-0.58	1.00	-0.84	0.01
GSC22	-0.74	0.18	-0.81	0.02
GSC3	-0.69	0.39	-0.88	0.00
GSC24	-0.32	1.00	-0.76	0.02
GSC25	-0.59	0.99	-0.94	0.00
GSC26	-0.63	0.74	-0.89	0.00
GSC27	-0.75	0.15	-0.89	0.00
GSC28	-0.87	0.01	-0.80	0.02
GSC29	-0.62	0.76	-0.91	0.00
GSC30	-0.59	1.00	-0.90	0.00
GSC31	-0.50	1.00	-0.82	0.01
GSC32	-0.65	0.62	-0.85	0.01
GSC33	-0.27	1.00	-0.63	0.07
GSC34	-0.85	0.01	-0.82	0.01
GSC35	-0.40	1.00	-0.82	0.01
GSC36	-0.46	1.00	-0.81	0.02
GSC37	-0.45	1.00	-0.84	0.01
GSC38	-0.51	1.00	-0.82	0.01
GSC39	-0.38	1.00	-0.81	0.02
GSC40	-0.44	1.00	-0.78	0.02
GSC41	-0.41	1.00	-0.75	0.02
GSC42	-0.51	1.00	-0.87	0.00
GSC43	-0.65	0.62	-0.89	0.00
GSC44	-0.66	0.56	-0.77	0.02
GSC45	-0.64	0.65	-0.70	0.03
GSC46	-0.81	0.04	-0.88	0.00
GSC47	-0.65	0.61	-0.90	0.00
GSC48	-0.48	1.00	-0.75	0.02
GSC49	-0.42	1.00	-0.82	0.01
GSC50	-0.58	1.00	-0.72	0.03
GSC51	-0.52	1.00	-0.88	0.00

Values with significant p-values (≤0.05) are highlighted in yellow.

Correlations were determined using the R psych package, and p-values were adjusted for multiple tests using the "hommel" method.

*Linear correlation following log transformation of data.

Table S4. Percentage of contigs with mapped reads at each site.

GSC	Co	0.3	0.5	0.6	0.7	0.9	1.1	2.7	10.1	15.1	24.8	33.9	48.8	59.5	Min*	Min*	Min*
1	100	100	100	100	100	100	100	100	93	92	90	91	92	90	90	100	90
2	100	100	100	100	100	100	100	100	93	93	83	88	83	90	83	100	83
3	100	100	100	100	100	100	99	100	96	94	93	91	93	89	89	99	89
4	100	100	100	100	100	100	100	100	100	99	98	100	99	93	93	100	93
5	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
6	100	100	100	100	100	100	98	100	98	95	95	96	97	87	87	98	87
7	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
8	100	100	100	100	100	100	100	100	99	99	97	98	94	98	94	100	94
9	100	100	100	100	100	100	100	100	100	100	100	100	99	99	99	100	99
10	100	100	100	100	100	100	100	100	100	100	98	100	91	98	91	100	91
11	100	100	100	100	100	100	100	100	100	100	99	100	98	95	95	100	95
12	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
13	100	100	100	100	100	100	100	100	99	99	98	99	98	94	94	100	94
14	100	100	100	100	100	100	100	100	100	100	100	100	100	96	96	100	96
15	100	100	100	100	100	100	100	100	100	100	100	100	100	97	97	100	97
16	100	100	100	100	100	100	100	100	100	100	97	100	97	85	85	100	85
17	100	100	100	100	100	100	100	100	100	100	100	99	100	90	90	100	90
18	100	98	100	100	99	100	100	100	93	92	87	90	89	83	83	98	83
19	100	100	100	100	100	100	100	100	100	100	100	100	100	96	96	100	96
20	100	100	100	100	100	100	96	100	98	100	75	81	73	65	65	96	65
21	100	100	100	100	100	100	100	100	100	100	97	94	90	85	85	100	85
22	100	100	100	100	100	100	100	100	100	100	100	100	100	97	97	100	97
23	100	100	100	100	100	100	100	100	100	100	95	98	89	77	77	100	77
24	100	100	100	100	100	100	100	100	99	99	98	100	100	86	86	100	86
25	100	100	100	100	100	100	100	100	100	100	98	100	98	77	77	100	77
26	100	100	100	100	100	100	100	100	100	100	97	96	92	76	76	100	76
27	100	100	100	100	100	100	100	100	100	100	100	99	99	84	84	100	84
28	100	100	100	100	100	100	100	100	100	100	100	100	100	92	92	100	92
29	100	100	100	100	100	100	100	100	100	100	100	95	88	82	82	100	82
30	100	100	100	100	100	100	100	100	100	100	100	100	100	99	99	100	99
31	100	100	100	100	100	100	100	100	100	100	96	100	96	86	86	100	86
32	100	100	100	100	100	100	100	100	100	100	100	100	100	98	98	100	98
33	100	100	98	98	100	97	96	96	94	94	89	89	86	84	84	96	84
34	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
35	100	100	100	100	100	100	100	100	89	89	88	91	89	88	88	100	88
36	100	100	100	100	100	100	100	100	89	94	92	92	89	89	89	100	89
37	100	100	100	100	100	100	99	100	79	83	73	82	76	68	68	99	68
38	100	100	100	100	100	100	99	100	72	81	65	80	82	63	63	99	63
39	100	100	100	100	100	100	99	100	85	89	84	85	86	83	83	99	83
40	100	100	100	100	100	100	100	100	87	93	75	79	99	54	54	100	54
41	100	100	100	100	94	100	92	100	74	79	61	78	66	59	59	92	59
42	100	100	100	100	100	100	100	100	99	99	98	99	98	95	95	100	95
43	100	100	100	100	100	100	100	100	99	100	98	100	98	96	96	100	96
44	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
45	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
46	100	100	100	100	100	100	100	100	100	100	100	100	100	99	99	100	99
47	100	100	100	100	100	100	100	100	94	95	90	94	88	81	81	100	81
48	100	100	100	100	100	100	100	100	96	95	95	96	93	89	89	100	89
49	100	100	100	100	100	100	95	100	94	91	86	86	86	72	72	95	72
50	100	100	100	99	100	100	98	100	99	100	98	97	98	93	93	98	93
51	100	100	100	100	100	100	97	100	85	82	73	94	82	72	72	97	72

*Minimum % of contigs detected across all, proximal or distal sites. Abbreviation: Co-assembly, Co.

Table S5. Average mapped-read genome bin coverages per site.

GSC	Co	0.3	0.5	0.6	0.7	0.9	1.1	2.7	10.1	15.1	24.8	33.9	48.8	59.5	Cnt*	Cnt*	Cnt*
1	156.34	11.91	20.29	9.42	175.76	10.64	1.19	41.16	0.53	0.53	0.30	0.46	0.37	0.28	7	7	0
2	46.28	9.04	16.73	4.62	69.15	17.44	0.79	18.54	0.38	0.48	0.30	0.52	0.57	0.41	6	6	0
3	22.97	6.02	11.13	1.60	6.96	17.91	0.67	4.74	0.49	0.61	0.37	0.50	0.41	0.31	6	6	0
4	89.61	14.24	100.23	7.16	16.64	14.11	2.32	20.47	2.60	2.74	0.90	2.11	1.50	0.53	11	7	4
5	37.88	10.30	50.57	4.94	7.61	7.23	2.30	8.99	2.78	2.89	1.16	2.24	1.45	1.04	13	7	6
6	17.92	6.09	26.97	3.21	6.44	4.97	1.34	8.40	1.42	1.51	0.60	1.12	0.98	0.34	10	7	3
7	78.19	15.45	46.32	18.25	27.30	44.37	13.77	16.34	19.61	27.18	9.99	30.35	10.87	1.76	13	7	6
8	41.80	17.88	64.39	7.34	7.37	8.13	3.44	15.43	2.20	2.02	0.86	1.46	0.69	1.37	11	7	4
9	26.37	29.16	23.48	27.07	13.70	13.20	26.02	13.85	7.08	5.24	1.94	2.30	0.89	0.87	11	7	4
10	22.00	42.91	34.75	12.97	6.34	7.54	3.88	17.77	1.94	1.62	0.57	1.63	0.46	0.72	10	7	3
11	348.42	101.31	418.61	57.84	112.94	143.09	30.56	95.28	25.54	21.87	4.37	15.02	7.35	1.48	13	7	6
12	150.44	81.35	216.21	82.18	71.33	120.43	57.47	53.86	51.28	47.49	12.92	27.63	16.36	4.29	13	7	6
13	28.87	56.78	123.45	47.00	48.42	68.10	34.11	42.93	27.57	20.71	5.68	14.44	6.40	2.21	13	7	6
14	47.07	22.66	29.28	15.42	25.43	44.62	7.62	13.94	9.60	10.78	2.58	5.39	2.19	0.83	12	7	5
15	26.60	18.67	22.29	31.28	16.14	20.67	34.17	8.02	22.77	10.13	2.11	4.32	1.86	0.98	12	7	5
16	52.94	16.59	25.04	17.13	32.89	25.22	9.01	13.94	4.08	5.45	0.96	3.27	1.10	0.40	11	7	4
17	16.66	8.74	8.51	12.72	8.15	7.25	28.79	1.96	43.91	25.97	5.01	1.13	1.09	0.52	12	7	5
18	30.11	2.78	35.94	5.07	3.87	11.24	5.45	29.97	1.10	1.24	0.64	1.03	0.77	0.58	10	7	3
19	21.06	10.81	11.06	8.15	11.77	17.84	6.80	8.05	5.90	6.58	2.23	6.49	4.36	1.07	13	7	6
20	19.34	2.60	5.62	1.74	14.10	9.28	0.75	2.58	0.77	1.34	0.27	0.48	0.28	0.21	7	6	1
21	24.20	6.97	20.74	5.56	5.18	16.14	3.24	17.67	1.94	3.15	0.75	1.17	0.80	0.66	10	7	3
22	29.30	18.05	20.42	13.26	11.45	28.64	12.74	9.30	9.46	11.73	2.36	7.82	2.89	0.95	12	7	5
23	25.50	9.92	15.45	7.68	18.31	9.89	3.48	4.74	3.76	5.20	0.61	1.41	0.47	0.28	10	7	3
24	71.72	5.24	12.02	7.82	89.89	3.05	8.40	11.78	1.40	1.65	0.71	2.54	1.71	0.39	11	7	4
25	99.93	28.70	66.12	29.01	40.05	30.34	4.15	10.52	2.58	2.69	0.60	1.68	0.75	0.33	10	7	3
26	18.46	7.06	11.36	5.20	15.40	6.41	1.71	4.01	1.66	3.14	0.53	0.71	0.39	0.22	9	7	2
27	23.41	7.49	12.56	6.67	13.54	9.50	3.83	4.95	5.02	3.67	0.93	1.24	0.79	0.41	10	7	3
28	16.21	9.65	10.49	8.64	12.17	8.90	6.84	5.44	6.86	7.36	2.00	3.76	1.92	0.43	12	7	5
29	20.22	4.82	18.78	7.10	10.75	4.96	6.54	2.75	3.67	2.59	0.72	0.56	0.37	0.30	9	7	2
30	33.65	8.19	20.39	7.60	15.73	7.18	4.40	4.84	1.34	1.62	0.66	1.94	1.72	0.44	11	7	4
31	17.84	3.84	16.05	2.72	7.35	6.09	1.41	3.94	1.34	2.11	0.59	1.26	0.74	0.61	10	7	3
32	25.97	5.40	11.63	5.42	14.74	9.79	3.00	4.75	3.46	3.22	0.96	2.07	1.43	0.53	11	7	4
33	19.60	3.94	1.15	0.60	22.51	0.73	0.50	0.54	0.45	0.53	0.33	0.46	0.36	0.32	3	3	0
34	20.86	12.59	9.76	11.21	9.98	7.80	10.36	4.05	8.94	7.97	3.96	3.49	1.84	0.97	12	7	5
35	45.36	4.20	21.86	1.61	47.07	12.78	0.65	3.84	0.22	0.27	0.19	0.30	0.23	0.22	6	6	0
36	24.40	5.76	11.69	1.15	26.99	8.37	0.46	7.21	0.19	0.24	0.22	0.25	0.25	0.21	6	6	0
37	39.09	5.56	37.45	2.10	15.81	14.15	0.65	7.09	0.27	0.34	0.20	0.34	0.25	0.20	6	6	0
38	21.17	7.06	33.58	3.53	16.89	13.55	0.81	16.96	0.32	0.45	0.31	0.45	0.47	0.31	6	6	0
39	71.88	15.51	89.69	3.60	9.83	22.23	0.85	17.91	0.44	0.53	0.39	0.56	0.56	0.53	6	6	0
40	20.67	3.55	26.31	1.92	7.80	6.51	0.72	14.04	0.32	0.48	0.23	0.35	0.80	0.17	6	6	0
41	15.91	1.75	9.16	2.46	0.88	15.79	0.43	3.48	0.29	0.39	0.21	0.39	0.27	0.22	5	5	0
42	33.78	11.72	38.18	7.34	19.38	9.20	2.55	10.61	1.33	1.58	0.63	2.67	0.85	0.42	10	7	3
43	20.40	10.86	14.42	5.25	15.20	10.34	1.89	7.59	1.64	1.69	0.84	1.84	0.98	0.53	10	7	3
44	25.95	17.90	35.23	15.57	25.75	22.91	7.58	17.37	9.16	8.97	2.90	12.76	6.13	1.60	13	7	6
45	19.00	11.55	15.66	8.97	15.46	9.94	5.75	10.92	6.88	6.68	3.12	11.62	4.86	1.68	13	7	6
46	17.69	14.33	13.49	13.24	13.44	9.26	6.90	4.63	7.54	6.96	1.91	3.05	1.48	0.72	12	7	5
47	15.96	10.15	12.26	5.89	5.73	10.53	0.88	7.91	0.47	0.58	0.30	0.63	0.32	0.23	6	6	0
48	18.18	1.28	14.12	1.88	3.95	8.21	0.64	8.15	0.41	0.46	0.28	0.55	0.32	0.25	6	6	0
49	18.70	3.81	19.24	1.42	4.99	4.64	0.34	3.54	0.38	0.37	0.20	0.39	0.26	0.15	6	6	0
50	20.66	11.41	18.08	2.44	11.12	15.13	0.64	14.61	1.76	1.64	0.92	1.00	1.36	0.66	10	6	4
51	15.24	7.37	6.46	3.28	12.88	3.20	0.70	2.08	0.43	0.44	0.21	0.79	0.36	0.29	6	6	0
Average	43	15	38	11	24	19	7	13	6	6	2	4	2	1			
Standard Deviation	53	19	65	15	31	26	11	16	11	9	2	6	3	1			
% present out of 51	100	100	100	98	98	98	69	98	69	71	31	65	41	18			

*Genome counts (Cnt) across all, proximal or distal sites based on an average coverage of >=1.

Abbreivation: Co-assembly, Co.

Table S6. Pairwise average amino acid identities (AAI) shared between genome bins.

GSC	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	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
1	100	80	86	49	49	45	49	53	51	51	49	49	49	52	52	50	49	52	51	50	52	49	49	52	49	47	45	45	44	44	44	45	44	43	43	40	43	42	44	44	44	41	42	41	42	41	41	41	41	41	40
2	79	100	67	48	49	43	48	51	50	47	50	48	50	53	51	45	50	52	53	49	52	49	48	53	48	44	42	44	46	44	45	45	46	42	44	32	44	40	46	41	44	41	41	39	38	38	40	39	39	47	39
3	86	68	100	49	49	46	49	53	52	52	50	49	49	52	52	51	49	52	52	50	52	50	49	51	49	48	44	44	43	44	44	45	44	43	42	41	44	43	44	44	45	41	42	42	41	41	42	41	41	40	40
4	49	48	49	100	81	82	74	50	49	49	48	49	48	50	49	50	50	50	50	48	50	46	51	51	46	46	43	42	43	43	43	44	44	42	42	41	42	41	42	42	42	41	41	41	41	41	40	39	41	38	39
5	49	49	49	80	100	74	73	50	48	50	49	48	48	50	49	50	49	50	50	48	50	46	51	51	46	47	43	43	42	43	43	44	44	42	42	44	42	41	42	43	42	41	41	41	41	40	40	40	41	39	39
6	45	43	46	82	74	100	67	48	49	47	47	47	46	47	46	48	48	48	46	46	47	43	47	47	43	44	38	39	39	41	39	41	42	40	40	41	41	40	40	41	41	39	39	38	38	38	39	39	38	40	35
7	49	49	49	74	73	67	100	50	49	50	49	48	48	50	50	49	50	50	50	48	50	46	51	51	46	46	43	43	42	43	43	44	43	42	42	40	42	41	41	43	41	41	41	40	41	42	40	39	40	39	39
8	53	51	53	50	50	48	50	100	73	86	50	49	49	52	52	51	50	52	51	50	53	50	49	52	49	48	44	45	43	44	43	44	44	42	43	41	43	42	43	43	43	41	41	42	41	41	42	41	40	39	39
9	51	50	52	49	48	48	48	73	100	73	50	48	49	50	51	49	49	50	50	49	51	50	48	50	48	46	43	43	42	43	42	43	43	41	42	40	42	41	43	42	42	40	41	41	41	40	40	40	39	39	39
10	51	48	52	49	49	47	50	86	73	100	49	47	47	50	50	49	49	51	50	49	52	49	49	50	47	47	42	43	43	42	42	43	43	41	42	47	42	40	42	42	42	40	40	40	40	40	41	40	41	39	39
11	50	50	50	48	49	47	49	50	50	49	100	82	82	73	64	50	53	53	53	48	51	49	49	53	47	45	43	43	43	44	43	44	44	41	41	39	41	41	41	43	41	40	41	41	40	40	39	39	40	38	39
12	49	48	49	49	48	47	48	49	48	48	82	100	77	70	60	47	52	52	51	47	50	46	47	51	45	45	40	40	40	40	40	42	41	38	40	37	40	40	42	42	42	39	39	38	39	39	37	38	39	36	38
13	49	50	49	48	48	46	48	49	48	47	83	77	100	68	60	47	50	51	49	45	48	46	47	49	44	44	42	43	40	43	42	43	42	42	40	38	41	40	41	41	40	40	39	40	40	39	39	40	38	36	38
14	51	53	52	50	50	47	50	52	50	50	73	70	68	100	64	52	55	55	54	50	53	49	50	54	48	47	45	44	44	44	44	45	44	42	43	40	43	43	45	45	44	41	42	40	41	41	41	41	40	39	40
15	51	52	52	49	49	46	49	52	51	50	63	60	60	64	100	53	54	55	54	50	52	49	51	54	48	47	45	44	43	44	44	45	43	42	42	37	43	42	45	44	44	41	42	40	42	42	41	42	40	41	40
16	50	45	50	50	50	47	49	50	49	49	50	47	48	52	52	100	50	54	52	49	50	47	48	53	48	46	44	43	43	44	44	44	45	44	41	37	42	41	45	43	43	41	42	40	40	40	40	41	40	39	41
17	49	50	49	50	49	47	50	50	49	49	53	51	50	55	54	49	100	53	52	48	50	47	49	53	47	46	43	43	42	43	44	43	42	41	42	37	42	39	42	41	42	41	41	40	41	41	42	39	41	42	39
18	52	52	52	51	50	47	50	52	51	51	53	52	51	56	55	54	53	100	55	49	52	50	50	55	48	47	44	45	44	44	44	45	45	43	42	38	42	42	44	44	43	41	42	42	41	42	41	41	40	40	40
19	51	52	52	50	49	46	50	51	50	50	53	51	50	54	54	52	53	55	100	49	52	50	51	54	49	47	44	44	44	44	44	45	45	42	43	41	43	41	44	44	43	41	41	40	41	41	41	41	40	40	40
20	50	48	50	48	48	45	48	50	49	49	48	47	45	50	50	49	47	49	49	100	50	47	51	50	50	50	44	44	43	44	44	44	44	42	42	41	43	42	43	44	43	41	41	41	42	41	42	41	41	38	40
21	52	53	53	50	50	47	49	54	52	52	51	50	48	53	52	50	50	53	52	50	100	50	49	52	49	48	44	44	44	44	44	45	45	43	43	42	43	43	45	45	43	41	42	42	42	41	40	41	40	40	40
22	49	49	49	46	46	43	46	50	50	49	49	46	46	49	50	47	47	49	50	47	50	100	47	48	47	45	43	43	43	43	43	44	44	42	42	39	42	41	43	42	42	40	41	40	40	40	41	41	40	39	39
23	49	48	49	51	51	47	50	50	49	49	49	48	47	50	50	48	49	50	51	50	49	47	100	50	49	49	44	44	43	43	44	45	44	42	42	40	42	41	43	43	42	42	41	41	42	42	41	41	41	40	40
24	51	52	51	52	51	47	51	52	50	50	53	52	49	54	54	53	53	54	54	50	52	48	51	100	48	47	45	44	44	44	44	45	45	42	43	39	43	42	44	43	43	41	41	40	41	41	40	41	40	39	40
25	48	48	49	46	46	44	47	49	48	47	46	44	44	48	48	48	48	48	49	50	49	47	49	48	100	72	45	45	44	44	44	45	44	43	42	35	43	41	42	43	41	41	41	42	41	41	41	42	40	37	40
26	47	44	48	46	47	44	46	48	46	47	46	45	44	47	47	46	46	47	47	50	48	45	49	47	73	100	43	42	43	42	43	43	43	41	41	42	41	41	42	41	42	41	41	42	41	42	40	41	40	42	40
27	45	42	44	43	43	38	43	44	43	42	43	40	42	45	45	44	43	44	44	44	44	43	43	44	45	43	100	53	54	48	49	50	49	46	42	37	42	41	44	44	43	41	42	40	41	41	41	41	40	39	41
28	45	44	44	42	43	39	43	44	43	43	43	40	42	44	45	44	42	45	44	44	44	43	44	44	45	42	53	100	53	49	49	51	50	48	42	37	42	40	43	42	41	42	42	41	42	41	42	41	40	38	40
29	44	46	43	43	43	39	42	43	42	43	43	40	41	44	43	44	42	44	44	43	44	43	43	44	44	42	54	53	100	48	48	50	49	45	41	36	42	40	42	43	42	41	41	41	40	41	40	40	40	39	40
30	44	44	44	43	43	41	43	44	43	42	44	41	43	44	44	44	43	44	44	44	45	43	44	44	43	42	49	49	48	100	60	50	48	46	41	38	42	41	43	43	43	40	41	41	41	41	40	40	40	39	40
31	44	46	44	43	43	40	43	43	42	43	43	40	42	44	44	44	44	44	44	44	44	43	44	44	44	43	49	49	48	60	100	51	48	46	41	36	42	40	44	42	43	40	41	39	40	40	40	40	39	43	39
32	45	45	45	44	44	41	44	44	44	43	44	42	43	45	45	44	43	45	45	44	45	44	45	45	45	43	50	51	50	50	51	100	51	48	42	38	42	41	43	43	43	41	42	41	42	42	41	41	40	39	39
33	44	46	44	44	44	42	43	45	44	43	44	41	42	44	44	45	42	45	45	44	45	44	44	45	45	43	49	50	49	48	48	51	100	48	42	42	43	42	43	43	43	41	42	41	42	40	41	41	41	38	43
34	43	42	43	42	42	40	42	41	41	41	41	39	42	42	42	44	41	43	42	42	43	42	42	42	43	41	46	47	46	46	46	48	48	100	40	39	42	40	43	42	42	40	42	41	41	41	39	42	40	38	39
35	43	45	42	42	42	40	42	43	42	42	41	40	40	43	43	42	41	42	43	42	43	42	42	43	42	41	42	42	41	41	41	42	42	40	100	64	70	52	46	47	47	41	41	41	41	41	41	41	41	41	41
36	40	32	42	41	43	41	40	41	40	47	39	37	37	40	37	37	37	38	41	41	41	39	41	39	34	41	38	38	36	38	35	38	40	39	64	100	59	44	46	42	46	39	41	40	42	38	38	36	41	47	37
37	43	44	44	42	42	42	42	43	42	42	42	40	41	43	43	42	42	43	43	43	43	42	42	43	43	41	42	42	42	42	42	42	43	42	70	59	100	52	47	49	47	41	42	41	41	41	41	42	41	41	40
38	42	40	42	41	41	40	41	42	41	40	41	40	40	43	42	41	39	42	41	42	42	41	42	42	42	41	41	40	40	41	40	41	42	40	51	44	52	100	46	47	45	40	40	40	41	41	41	41	40	37	39
39	44	46	44	42	43	41	42	43	43	42	41	43	41	46	45	45	42	45	44	43	44	44	43	44	43	42	44	43	42	43	44	43	43	43	46	46	47	46	100	56	57	42	43	40	42	42	41	42	42	40	40
40	44	41	44	42	43	41	43	43	42	42	43	42	41	45	44	43	42	44	44	44	45	42	43	44	43	41	44	42	43	43	42	43	43	42	47	43	49	47	56	100	59	42	43	40	42	43	41	42	41	39	42
41	44	44	45	42	42	40	42	43	42	42	41	42	40	44	44	43	42	43	44	43	43	42	42	43	41	42	43	41	42	43	42	43	43	42	47	49	47	45	57	58	100	42	42	43	41	42	41	41	40	38	41
42	41	41	41	40	41	39	41	41	40	40	40	39	40	41	41	41	41	41	41	41	41	40	41	41	41	41	41	42	40	40	40	41	41	40	41	39	41	40	41	42	41	100	63	65	60	61	51	50	50	40	40
43	42	40	42	41	41	39	41	42	41	40	41	39	39	42	42	42	41	41	42	41	42	41	41	42	41	42	42	42	41	41	41	42	42	42	41	41	42	40	43	43	42	63	100	62	59	61	49	50	50	41	40
44	41	39	41	41	41	38	41	41	41	39	41	38	40	41	40	40	40	42	40	41	42	39	41	40	42	42	41	41	41	41	39	41	41	41	41	39	41	40	40	41	43	66	62	100	59	60	49	49	49	41	40
45	41	39	42	41	41	37	41	41	41	40	40	39	40	42	42	40	41	41	41	42	42	40	42	41	41	41	41	41	40	41	40	42	41	41	41	41	41	40	42	42	41	60	59	59	100	64	48	49	50	39	40
46	41	38	41	41	40	38	42	41	40	40	40	39	39	42	42	40	41	42	41	41	42	40	42	41	42	42	41	41	41	41	40	42	40	41	40	37	41	41	42	43	42	61	61	60	64	100	49	49	50	40	41
47	41	40	41	40	40	38	41	42	41	41	40	38	39	41	41	40	42	41	41	42	40	41	41	40	41	40	41	41	40	40	40	41	41	40	41	37	41	41	42	41	41	51	49	49	49	49	100	53	54	36	40
48	41	39	41	39	40	39	39	41	40	40	40	37	39	41	42	41	39	41	41	41	41	41	41	41	42	41	42	42	40	40	40	41	41	42	41	36	42	41	42	42	41	50	50	49	49	48	53	100	52	39	40
49	41	39	41	41	41	38	40	40	39	41	40	39	38	40	40	40	41	40	40	41	40	40	41	40	40	40	40	40	40	40	39	40	41	40	41	42	41	40	42	41	40	50	50	48	50	49	53	52	100	38	40
50	41	48	40	38	40	40	39	39	39	41	38	36	36	39	41	39	43	40	40	38	40	39	40	39	38	40	39	38	39	39	41	39	39	38	41	47	41	37	40	39	38	40	41	40	39	40	35	39	38	100	36
51	40	38	41	39	39	36	39	40	39	39	38	38	38	40	41	40	38	39	40	40	40	39	40	40	41	40	41	40	40	40	39	39	42	39	41	37	40	39	40	42	40	41	41	40	40	41	40	40	40	35	100

Gammaproteobacteria: GSC1-26; Alphaproteobacteria: GSC27-34; Deltaproteobacteria: GSC35-41; Bacteroidetes: GSC42-49; Other: 50-51.

Table S7. Average amino acid identities (AAI) shared between genome bins (left) and reference genomes (top).

GSC	Beggiatoa_alba_B18LD	Beggiatoa_sp._PS	Candidatus_Thiomargarita_nelsonii	Thiothrix_nivea_DSM_5205	Colwellia_piezophila	Colwellia_psychrerythraea_34H	Glaciecola_arctica_BSs20135	Cycloclasticus_pugetii	Cycloclasticus_sp._strain_P1	Cycloclasticus_zancles	Methylomicrobium_alcaliphilum_20Z	marine_gamma_proteobacterium_HTCC2143	Porticoccus_hydrocarbonoclasticus_MCTG13d	gamma_proteobacterium_HTCC2207	Simiduia_agarivorans_SA1	Teredinibacter_turnerae_T7901	Saccharophagus_degradans_2-40	gamma_proteobacterium_IMCC2047	gamma_proteobacterium_HdN1	Nitrosococcus_halophilus_Nc4	Nitrosococcus_watsonii	Thioalkalivibrio_sp._HL-EbGR7	Neptuniibacter_caesariensis	Kangiella_koreensis_SW-125	Hahella_chejuensis_strain_KCTC_2396	Micavibrio_aeruginosavorus	Oceanibaculum_indicum_P24	Rhodobacteraceae_bacterium_HTCC2150	Rhodobacterales_bacterium_HTCC2255	Dinoroseobacter_shibae_DFL_12	Thalassospira_profundimaris_WP0211	Rickettsia_bellii	Desulfobacterium_autotrophicum_strain_HRM2	Desulfatibacillum_alkenivorans_strain_AK-01	Desulfobacula_toluolica_strain_Tol2	Desulfococcus_oleovorans_Hxd3	Desulfobulbus_propionicus_1pr3	Desulfotalea_psychrophila_LSv54	Desulfocapsa_sulfexigens_SB164P1	Polaribacter_sp._MED152	Formosa_sp._AK20	Aequorivita_sublithincola_strain_9-3	Flavobacteriales_bacterium_ALC-1	Anaerophaga_sp._HS1	Marinilabilia_salmonicolor_JCM_21150	uncultured_bacterium_gcode_4	Caldithrix_abyssi_DSM_13497	Ignavibacterium_album_Mat9-16	
GSC01	59	64	66	53	49	49	42	50	53	53	53	51	52	51	50	50	50	52	51	52	55	54	51	52	51	46	45	43	44	43	41	41	43	43	43	44	44	44	44	40	40	41	40	40	40	40	43	43	Gammaproteobacteria
GSC02	58	64	65	53	50	49	42	50	53	53	54	52	54	51	50	51	51	52	52	52	54	55	52	50	50	45	45	43	44	42	40	40	44	44	44	46	45	45	44	40	39	40	39	40	39	40	43	44	
GSC03	59	63	66	52	50	49	42	50	53	53	52	51	52	52	51	50	50	52	51	52	55	54	51	51	51	45	45	43	44	43	40	41	44	44	43	44	44	44	44	40	40	41	40	40	40	40	43	43	
GSC04	49	49	50	50	73	73	48	48	50	50	49	50	50	50	50	49	50	50	49	47	49	49	51	51	50	43	43	43	43	42	40	40	42	41	42	42	42	42	42	39	40	40	40	39	39	39	41	41	
GSC05	49	49	50	49	73	73	48	47	50	50	49	50	50	50	50	49	49	51	50	48	49	49	51	52	50	44	43	43	43	42	40	40	42	42	41	42	42	42	43	40	40	40	40	39	39	39	41	41	
GSC06	47	47	45	47	66	68	46	45	48	48	46	46	47	46	47	46	46	48	46	46	47	47	48	47	46	39	40	39	42	40	40	39	41	41	41	41	42	42	43	38	39	39	39	39	39	37	39	39	
GSC07	49	49	50	49	81	90	49	47	50	50	48	50	50	50	49	49	49	51	49	47	49	49	51	51	49	44	43	43	43	43	41	40	42	41	41	41	42	42	42	40	40	41	40	39	40	39	41	41	
GSC08	53	52	54	52	50	50	43	68	73	73	55	50	51	51	51	51	51	51	51	51	53	53	51	51	51	44	44	43	44	43	41	41	43	43	43	43	43	43	43	40	40	41	40	40	40	39	42	42	
GSC09	52	51	52	51	49	48	43	75	79	79	54	49	50	50	49	50	50	51	49	51	52	52	50	50	50	44	43	42	42	42	40	40	42	41	41	42	42	43	43	39	39	40	39	39	39	38	40	41	
GSC10	52	50	52	51	49	49	44	67	72	72	54	49	50	50	49	49	50	50	49	50	52	52	50	49	50	42	43	41	43	42	40	40	42	41	41	42	42	42	42	39	39	40	39	40	40	37	41	41	
GSC11	50	49	50	50	50	49	44	46	51	50	50	53	59	57	55	54	54	54	52	50	51	50	53	51	52	45	44	43	44	43	41	41	42	43	41	42	43	43	42	39	39	40	40	39	39	39	42	41	
GSC12	50	50	49	49	48	48	43	46	50	50	48	52	58	55	53	52	52	52	50	47	49	50	51	48	50	41	42	41	41	41	40	37	41	40	40	40	42	42	42	38	39	39	38	38	38	38	39	39	
GSC13	49	48	49	48	47	47	42	46	49	50	48	52	56	54	52	51	52	52	49	47	49	49	50	50	50	44	43	43	43	42	40	39	41	41	39	41	42	42	42	38	38	40	38	38	39	38	40	40	
GSC14	52	52	52	52	51	50	43	48	52	52	51	55	61	58	56	55	56	55	53	50	52	52	54	51	54	45	44	43	44	43	41	40	43	43	43	43	44	44	44	40	40	41	40	39	39	41	42	43	
GSC15	52	52	52	52	51	50	43	50	53	53	51	54	60	58	56	55	55	55	53	50	52	52	53	52	54	45	45	43	44	43	40	40	44	43	42	43	44	44	44	40	40	41	40	40	40	40	42	42	
GSC16	51	50	50	50	49	50	42	48	51	51	50	52	53	52	52	51	51	52	54	48	51	50	52	51	51	44	44	42	44	43	41	40	43	42	42	44	43	44	44	40	39	40	39	40	40	41	41	42	
GSC17	50	50	50	50	49	49	43	47	51	51	50	53	55	54	55	58	58	53	52	48	50	50	52	50	52	43	44	42	43	43	40	39	42	42	41	42	43	42	43	40	39	40	40	39	40	40	40	41	
GSC18	52	52	53	52	50	50	44	48	52	52	51	54	55	54	54	53	54	61	54	51	53	53	54	52	54	45	45	43	44	44	41	40	43	43	43	43	44	44	44	40	40	40	40	40	40	40	42	42	
GSC19	52	52	52	52	50	50	44	48	52	52	51	54	54	53	53	52	53	54	52	50	52	52	53	51	53	45	45	43	44	43	41	40	43	43	42	43	43	44	44	40	40	41	40	39	39	40	42	42	
GSC20	50	50	50	50	48	48	44	47	50	50	49	49	50	49	48	48	49	49	49	48	50	50	48	50	49	45	44	43	43	42	41	40	43	42	42	42	43	44	43	40	40	41	40	40	40	39	42	42	
GSC21	53	53	53	52	50	50	43	49	53	53	53	50	52	51	50	50	51	52	51	51	53	52	51	52	51	45	45	43	44	43	41	40	44	43	43	43	44	45	44	40	40	41	40	40	40	40	42	42	
GSC22	50	49	50	49	46	46	42	48	50	50	49	48	49	49	48	48	48	49	48	49	51	50	48	48	48	43	44	43	43	42	40	39	42	42	42	43	43	43	43	39	39	40	39	39	39	39	41	41	
GSC23	49	49	49	49	51	51	46	47	50	50	49	49	50	50	49	48	49	50	49	47	50	49	49	54	49	44	44	43	43	42	41	40	42	42	42	42	42	43	43	41	41	41	41	39	40	39	41	42	
GSC24	52	51	52	51	51	51	44	49	52	52	51	54	54	53	53	52	53	55	52	49	52	51	55	52	57	45	45	43	44	43	42	41	43	42	42	43	43	44	43	40	40	41	40	39	40	39	42	42	
GSC25	49	49	49	49	47	46	42	47	50	50	47	47	47	48	47	47	47	47	48	46	49	49	48	49	48	45	44	43	43	42	40	39	43	43	42	42	43	44	43	40	40	41	40	40	40	40	41	42	
GSC26	47	47	48	47	47	46	43	44	48	48	46	46	46	46	46	46	46	47	46	46	47	47	46	48	47	43	42	42	42	41	40	39	42	41	41	41	42	42	41	41	40	41	40	39	39	38	40	41	
GSC27	45	44	46	44	44	43	39	42	44	44	44	44	44	44	44	43	43	44	44	42	44	44	44	45	44	57	52	48	48	47	45	44	43	43	42	43	43	43	43	39	39	41	40	39	39	40	42	43	Alphaproteobacteria
GSC28	45	44	45	45	43	42	38	42	45	44	44	43	44	44	43	43	43	44	44	42	44	44	44	44	44	54	53	48	50	47	45	44	42	42	42	42	43	42	42	40	39	41	39	40	40	40	42	42	
GSC29	44	44	45	44	42	42	39	42	44	44	43	43	44	43	43	43	43	43	43	42	43	44	44	44	43	56	51	47	48	46	45	43	41	42	41	41	42	41	42	39	39	41	39	39	39	39	41	41	
GSC30	44	44	47	44	43	43	40	43	45	45	43	44	44	44	44	43	43	44	44	42	44	44	44	44	44	50	50	60	63	58	45	43	42	42	41	42	43	42	42	40	39	40	39	39	39	39	41	42	
GSC31	45	44	46	45	43	44	40	43	44	44	44	44	44	45	44	43	43	44	44	42	44	45	44	45	44	50	50	58	59	58	46	43	42	41	41	42	42	42	42	39	38	39	39	38	38	40	42	41	
GSC32	46	44	46	45	44	44	41	44	46	46	44	45	45	45	45	44	44	44	45	43	45	45	45	45	45	51	53	50	50	49	47	44	42	43	42	43	43	43	43	40	40	41	40	40	40	40	42	42	
GSC33	45	44	46	45	43	43	39	42	45	45	44	44	45	44	44	44	44	44	44	43	44	45	45	45	45	50	51	48	48	47	45	46	43	43	42	43	43	43	43	39	40	41	39	40	40	39	42	43	
GSC34	44	43	43	43	42	42	37	40	43	43	43	42	43	42	42	41	41	43	43	41	43	43	42	44	43	47	48	44	45	44	41	54	41	42	41	42	42	42	42	39	38	41	39	39	39	40	43	43	
GSC35	42	42	43	43	42	42	39	41	43	43	42	42	42	42	42	42	42	42	42	41	42	42	42	42	42	42	41	41	41	40	39	39	53	51	53	52	47	47	46	40	40	41	40	41	41	40	43	42	Deltaproteobacteria
GSC36	40	41	43	42	40	41	40	41	39	42	44	36	39	41	38	40	38	38	38	42	41	38	39	39	38	36	40	36	38	39	39	37	45	44	46	44	44	47	46	40	41	42	42	39	40	41	40	38	
GSC37	43	43	43	44	43	42	39	42	43	43	43	43	43	43	42	42	42	42	43	41	43	42	43	43	43	42	42	41	42	41	40	40	54	51	54	53	47	48	47	41	40	42	40	41	41	40	43	43	
GSC38	42	41	41	42	41	41	38	42	42	42	41	41	42	42	42	41	41	42	41	41	42	42	42	42	42	41	41	40	40	40	38	39	57	50	65	51	47	46	46	40	40	41	40	40	40	38	42	42	
GSC39	44	44	45	44	43	42	41	43	44	45	44	44	44	44	43	43	44	44	44	43	44	43	43	44	43	44	44	42	43	41	41	40	47	47	47	47	59	54	56	40	40	41	41	41	40	40	44	44	
GSC40	44	44	43	44	43	43	40	42	44	44	43	43	44	43	43	43	43	43	43	42	44	44	44	44	44	43	43	42	42	42	40	39	49	48	48	48	56	60	59	40	40	41	40	41	41	40	43	43	
GSC41	43	44	44	44	42	42	39	43	44	44	43	43	43	43	43	42	43	43	43	42	45	44	43	43	43	42	43	42	41	41	39	39	47	48	47	48	57	58	74	40	40	41	41	40	40	40	43	43	
GSC42	41	41	41	41	41	41	40	40	41	42	41	41	41	41	41	41	41	42	41	40	41	41	41	41	41	42	40	40	40	40	38	39	41	41	41	41	42	42	42	61	59	58	58	48	48	39	44	44	Bacteroidetes
GSC43	42	42	42	42	42	42	40	41	42	42	42	42	42	42	42	41	41	41	42	41	42	41	41	43	42	42	41	41	41	41	38	39	42	42	41	42	42	42	42	74	59	58	59	48	47	41	44	45	
GSC44	40	39	41	41	41	41	40	40	41	42	41	41	41	41	41	41	41	41	41	39	41	41	41	41	40	42	40	41	40	40	39	39	41	41	41	40	41	42	42	60	58	58	58	47	47	40	43	44	
GSC45	41	41	42	41	42	41	39	40	42	42	41	41	42	41	41	41	41	41	41	40	41	41	41	42	41	42	40	40	40	40	38	39	42	41	41	41	42	41	42	57	60	66	60	47	47	40	43	44	
GSC46	41	41	41	41	42	42	39	41	42	42	41	41	42	42	42	41	41	41	41	40	41	41	41	42	41	42	40	41	41	40	38	39	41	42	41	41	41	42	42	59	68	62	67	47	47	39	43	44	
GSC47	41	41	41	41	40	40	39	40	41	41	41	41	41	40	40	40	41	40	40	40	41	41	40	41	41	42	40	40	40	40	38	38	42	42	42	42	42	41	42	48	47	49	47	50	50	38	45	46	
GSC48	41	41	41	41	40	40	39	39	41	41	41	40	41	41	41	40	40	40	41	40	41	40	40	41	41	42	41	40	40	39	38	39	42	42	42	42	41	42	42	48	48	49	48	52	51	40	44	45	
GSC49	40	40	40	41	41	41	39	39	41	41	41	41	40	40	40	40	40	40	40	39	40	40	40	41	41	40	39	39	39	39	37	39	41	41	40	41	42	41	42	47	48	49	47	49	49	40	44	44	
GSC50	40	39	40	39	40	39	38	38	39	39	39	39	39	40	40	39	40	40	39	37	38	38	38	40	41	40	39	38	40	39	37	39	40	41	39	41	39	40	40	40	40	40	40	39	38	46	40	39	Bacteria
GSC51	40	40	40	40	39	39	38	38	40	40	39	40	40	40	40	39	39	40	40	39	40	40	40	40	41	39	40	38	40	39	37	38	43	42	40	41	41	41	42	39	39	41	39	40	40	37	43	43	
	Gammaproteobacteria																									Alphaproteobacteria							Deltaproteobacteria							Bacteroidetes						Bacteria	Ignavibacteriae		

Scale (% identity):

34 🟦 🟦 🟦 🟦 ⬜ 🟥 🟥 🟥 90

Table S8. Summary of candidate hydrocarbon degradation genes distributed among 25 bacterial genome bins, GSC52 and the unbinned fraction.

Predicted protein	Substrate	GSC4	GSC5	GSC6	GSC7	GSC8	GSC9	GSC10	GSC14	GSC15	GSC16	GSC18	GSC19	GSC20	GSC21	GSC22	GSC24	GSC26	GSC30	GSC35	GSC37	GSC42	GSC43	GSC46	GSC48	GSC49	GSC52	unbinned	Sum
Transmembrane alkane 1-monooxygenase AlkB	*n*-alkanes	0	1	1	1	0	4	0	0	2	2	0	0	1	0	2	2	1	2	0	0	1	1	1	0	0	0	6	
Alkane-1 monooxygenase accessory--Rubredoxin-2 AlkG	*n*-alkanes	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	2	
Alkane-1 monooxygenase accessory--Rubredoxin-2 RubB/RubA2	*n*-alkanes	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
Alkane-1 monooxygenase accessory--Rubredoxin-NAD(+) reductase AlkT	*n*-alkanes	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	38
Cytochrome P450 accessory:2Fe-2S ferredoxin	*n*-alkanes	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
Cytochrome P450 accessory:Ferredoxin reductase--Rubredoxin-NAD(+) reductase	*n*-alkanes	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Cytochrome P450--CYPX_CD_Family--CYP153_gene	*n*-alkanes	1	3	0	0	0	0	0	1	0	3	0	9	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	20
Soluble methane monooxygenase hydroxylase component A alpha-subunit MmoX	methane, ammoniumn, *n*-alkanes	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	
Soluble methane monooxygenase hydroxylase component A beta-subunit MmoY	methane, ammoniumn, *n*-alkanes	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	
Soluble methane monooxygenase hydroxylase component A gamma-subunit MmoZ	methane, ammoniumn, *n*-alkanes	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	
Soluble methane monooxygenase reductase component C MmoC	methane, ammoniumn, *n*-alkanes	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	
Soluble methane monooxygenase regulatory protein B MmoB	methane, ammoniumn, *n*-alkanes	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	6
Particulate methane monooxygenase subunit A	methane, ammoniumn, *n*-alkanes	0	0	0	0	1	0	0	1	0	0	2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	3	
Particulate methane monooxygenase subunit B	methane, ammoniumn, *n*-alkanes	0	0	0	0	1	1	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	6	
Particulate methane monooxygenase subunit C	methane, ammoniumn, *n*-alkanes	0	0	0	3	2	2	0	1	0	0	2	0	0	2	0	0	0	0	0	0	0	0	0	0	0	2	2	35
Aromatic dioxygenase L	PAHs	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
PAH dioxygenase L--Benzene/Biphenyl*	PAHs	0	0	0	0	0	2	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
PAH dioxygenase L--Benzene/Biphenyl/Pyrene/Naphthalene*	PAHs	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
PAH dioxygenase L--Naphthalene*	PAHs	0	0	0	0	0	4	0	0	8	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	5	
PAH dioxygenase L--Ortho-halobenzoate/Benzoate/Anthranilate*	PAHs	0	0	0	0	0	3	0	0	1	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	2	
PAH dioxygenase L--Pyrene*	PAHs	0	0	0	0	0	0	0	0	1	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	1	
Aromatic dioxygenase S	PAHs	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
PAH dioxygenase S--Benzene/Biphenyl*	PAHs	0	0	0	0	0	2	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
PAH dioxygenase S--Naphthalene*	PAHs	0	0	0	0	0	3	0	0	5	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	4	
PAH dioxygenase S--Ortho-halobenzoate/Benzoate/Anthranilate*	PAHs	0	0	0	0	0	2	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
PAH dioxygenase S--Pyrene*	PAHs	0	0	0	0	0	0	0	0	1	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	1	
PAH dioxygenase electron transfer component--Ferredoxin--NAD(+) reductase	PAHs	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
PAH dioxygenase electron transfer component--Rieske (2Fe-2S) ferredoxin	PAHs	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	74
Phenol 2-monooxygenase--CD=FAD-dependent Phenol hydoxylase (PHOX) family	phenol, benzene, toluene	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Phenol hydroxylase, P0 assembly protein DmpK/LapK	phenol, benzene, toluene	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	
Phenol hydroxylase, P1 oxygenase component DmpL	phenol, benzene, toluene	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	
Phenol hydroxylase, P2 regulatory component DmpM	phenol, benzene, toluene	0	0	0	0	0	2	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	
Phenol hydroxylase, P3 oxygenase component DmpN	phenol, benzene, toluene	0	0	0	0	0	2	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	
Phenol hydroxylase, P4 oxygenase component DmpO	phenol, benzene, toluene	0	0	0	0	0	2	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
Phenol hydroxylase, FAD- and [2Fe-2S]-containing reductase P5 component DmpP	phenol, benzene, toluene	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
Positive regulator of phenol hydroxylase, DmpR	phenol, benzene, toluene	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	23
Membrane protein involved in aromatic hydrocarbon degradation--CD=Toluene_X	toluene	0	1	0	0	2	1	0	0	0	6	0	0	0	0	1	1	0	0	6	3	3	1	0	1	1	0	4	
Toluene-4-monooxygenase protein_A TmoA	toluene	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
Toluene-4-monooxygenase protein_B TmoB	toluene	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
Toluene-4-monooxygenase protein_C TmoC	toluene	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
Toluene-4-monooxygenase protein_D TmoD	toluene	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
Toluene-4-monooxygenase protein_E TmoE	toluene	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
Toluene-4-monoxygenase electron transfer component TmoF	toluene	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	39
Xylene monooxygenase electron transfer component XylA	xylenes, toluene	0	0	0	0	0	2	0	0	0	3	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	2	
Xylene monooxygenase hydroxylase subunit XylM	xylenes, toluene	0	0	0	0	0	2	0	0	0	2	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	5	20
Ethylbenzene dehydrogenase alpha subunit EbdA	toluene (anaerobic)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	
Ethylbenzene dehydrogenase beta subunit EbdB	toluene (anaerobic)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
Ethylbenzene dehydrogenase gamma subunit EbdC	toluene (anaerobic)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	4
Benzylsuccinate synthase alpha subunit BssA	toluene (anaerobic)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	
Benzylsuccinate synthase beta subunit BssB	toluene (anaerobic)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
Benzylsuccinate synthase gamma subunit BssC	toluene (anaerobic)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	5
Sum		1	6	2	7	6	41	2	5	21	16	5	9	1	10	35	8	1	2	6	3	4	2	1	1	1	2	67	264
Genes present (1) or absent (0)	Alkanes	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	0	0	1	1	22
Genes present (1) or absent (0)	PAHs	0	0	0	0	0	1	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	5
Genes present (1) or absent (0)	Aromatic hydrocarbons	0	1	0	1	1	1	1	0	0	1	0	0	0	0	1	1	0	0	1	1	1	1	0	1	1	0	1	15

Bacteria bins are: Gammaproteobacteria (green), Alphaproteobacteria (yellow), Deltaproteobacteria (blue), Bacteroidetes (pink), and virus/unbinned (grey).

Abbreviations: PAH dioxygenase L (large subunit); PAH dioxygenase S (small subunit).

*PAH and aromatic dioxygenases with highest sequence identities.