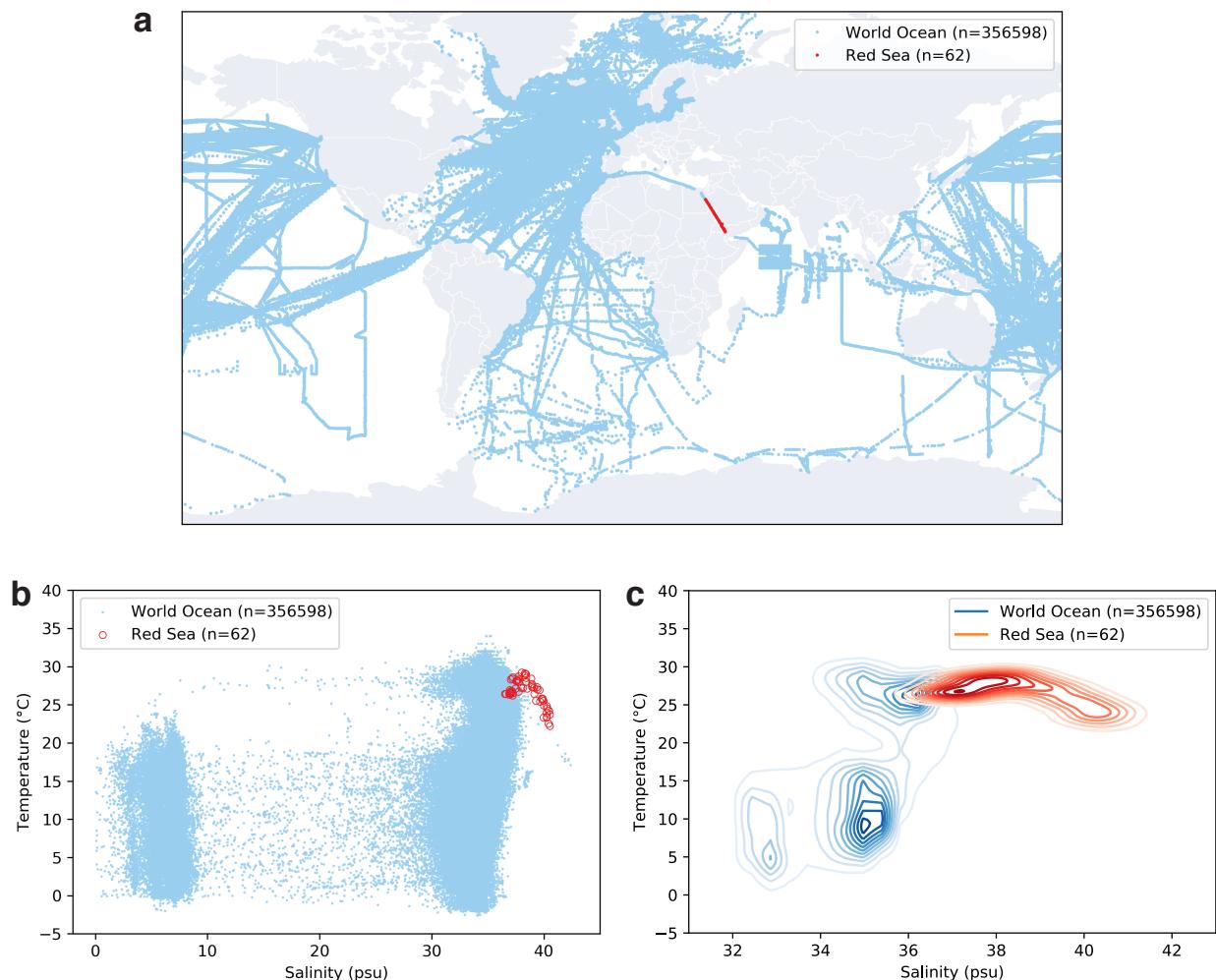
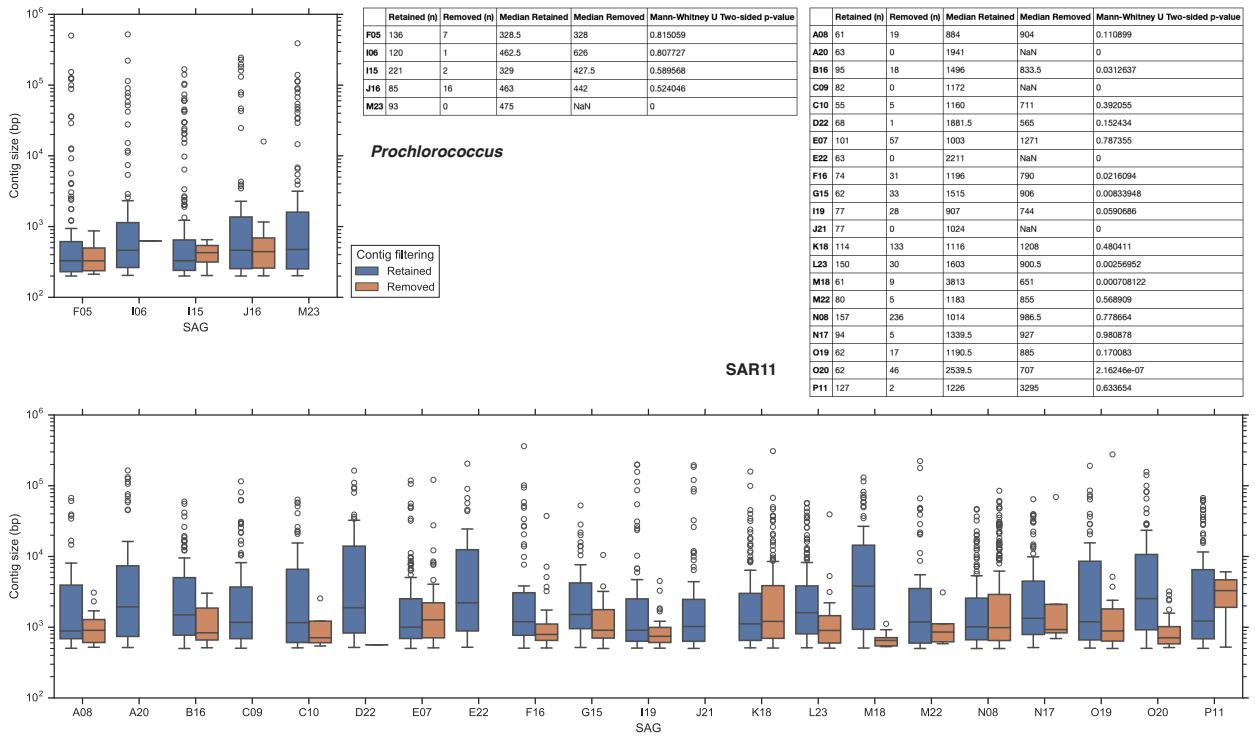


## Supplemental Information

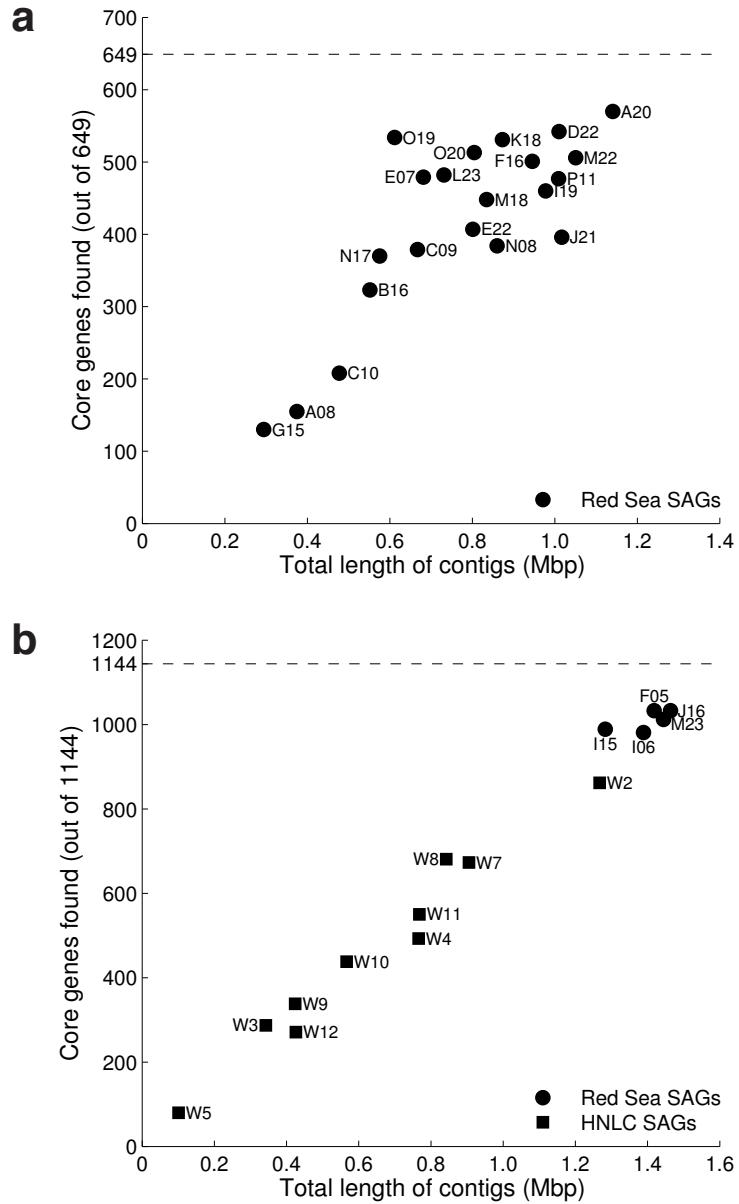
### SI Figures



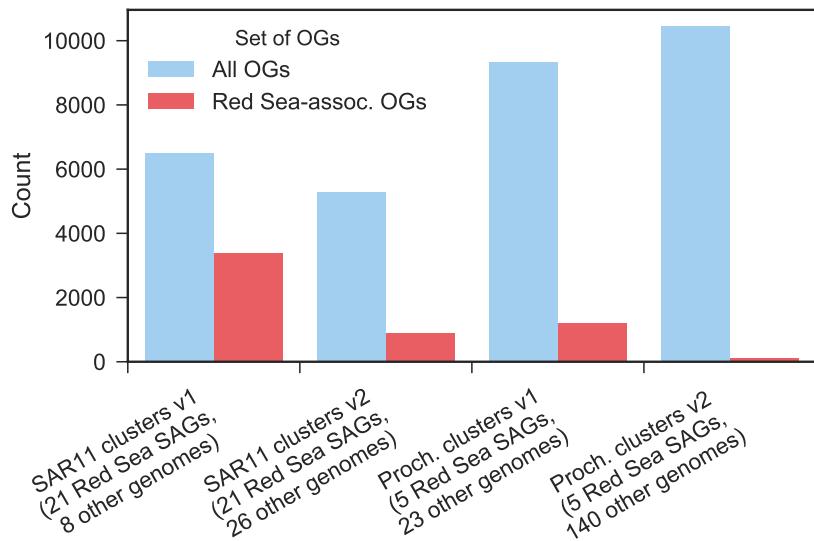
**Figure S1.** Temperature versus salinity relationships in the World Ocean Database 2013 (WOD13) colored by Red Sea or World Ocean excluding the Red Sea (World Ocean). (a) Map of samples in WOD13, with Red Sea samples ( $n=62$ ) in red. (b) Scatter plot of all points in WOD13, which spans from 1900 to 2010, and which includes a minority proportion of samples with salinities less than typical seawater (6% of samples  $<20$  psu). (c) Kernel density estimate (KDE) contour plot of those same samples, which shows the bivariate distribution of samples not visible in (b). Note that the majority of points have salinity  $>32$  psu and therefore empty sections of the KDE plot were cropped in (c) (note the  $x$ -scales). Both (b) and (c) show that the temperature–salinity distributions of Red Sea and World Ocean (non-Red Sea) samples are distinct, with Red Sea samples existing at the high end of global temperature distribution and extending beyond the non-Red Sea global salinity distribution.



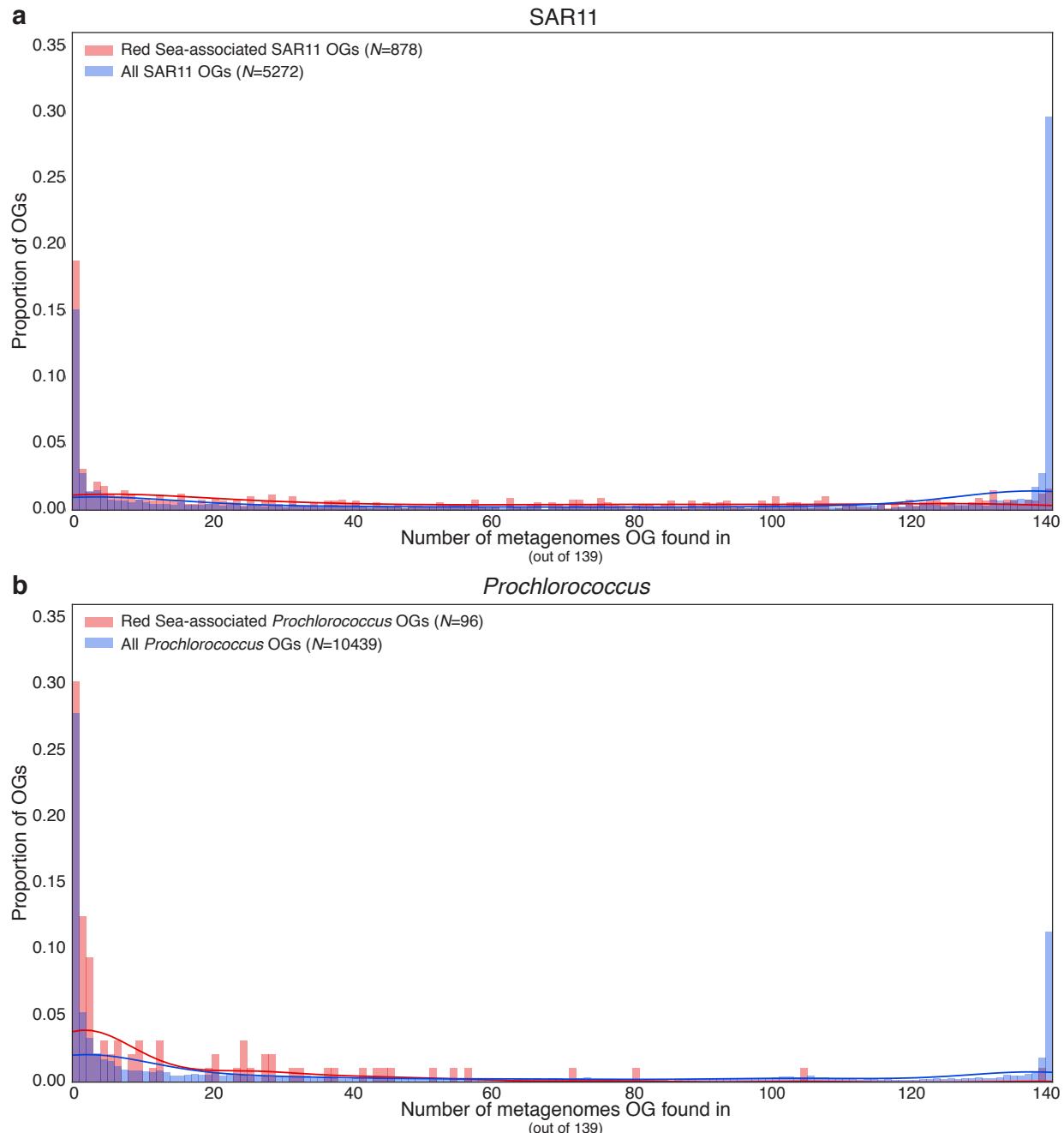
**Figure S2.** Contig size distributions of retained and removed contigs in the filtering of possible contamination from SAG assemblies. Boxplots show the median, interquartile range (IQR), and  $1.5 \times \text{IQR}$  (with outliers). Tables show, for each contig, the number  $n$  of contigs retained and removed, the median retained and removed contig size, and the Mann–Whitney U two-sided  $p$ -value comparing the distributions of retained and removed contig sizes.



**Figure S3.** Completeness of (a) SAR11 and (b) *Prochlorococcus* single-cell amplified genomes (SAGs) based on the number of single-copy core genes. Completeness of *Prochlorococcus* SAGs was generally higher than SAR11 SAGs.



**Figure S4.** Bar chart showing number of the number of ortholog groups (OGs) identified in two versions of clustering and how many OGs were Red-Sea-associated (found only in the Red Sea genomes). As more genomes are added to the clustering, the number of OGs found only in Red Sea SAGs approaches zero.



**Figure S5.** Histograms (bars) and kernel density estimates (lines) showing the distribution of (a) SAR11 and (b) *Prochlorococcus* OGs (all and Red-Sea-associated) by the number of *Tara Oceans* metagenomes they were found in. Red-Sea-associated OGs were less likely to be found in a high proportion of metagenomes than all OGs and were 10–20% more likely to not be found in any metagenome.

## SI Tables

**Table S1.** Properties of reference genomes of *Prochlorococcus* and SAR11. These include genomes used for phylogenetic trees and PCA ordination. Additional genomes were included for OG clustering and are provided in supplementary file 1.

Genus	Strain	Location	Clade	Size (bp)	G+C (%)	Accession no.	Alternate accession no.	Completeness (CheckM, %)	Redundancy (CheckM, %)
<i>Prochlorococcus</i>	MED4	Mediterranean Sea, 5 m	HL I	1657990	30.8	NC_005072		99.5	0.27
<i>Prochlorococcus</i>	MIT9515	Equatorial Pacific, 15 m	HL I	1704176	30.8	NC_008817		100.0	0.00
<i>Prochlorococcus</i>	AS9601	Arabian Sea, 50 m	HL II	1669886	31.3	NC_008816		99.6	0.00
<i>Prochlorococcus</i>	MIT9202	South Pacific, 79 m	HL II	1691453	31.1	NZ_DS999537	NZ_ACDW00000000.scaffold	98.8	0.27
<i>Prochlorococcus</i>	MIT9215	Equatorial Pacific, surface	HL II	1738790	31.2	NC_009840		99.7	0.27
<i>Prochlorococcus</i>	MIT9301	Sargasso Sea, 90 m	HL II	1641879	31.3	NC_009091		99.5	0.00
<i>Prochlorococcus</i>	MIT9312	Gulf Stream, 135 m	HL II	1709204	31.2	NC_007577		99.7	0.00
<i>Prochlorococcus</i>	NATL1A	Northern Atlantic, 30 m	LL I	1864731	35.0	NC_008819		99.7	0.00
<i>Prochlorococcus</i>	NATL2A	Northern Atlantic, 10 m	LL I	1842899	35.1	NC_007335		99.5	0.00
<i>Prochlorococcus</i>	SS120	Sargasso Sea, 120 m	LL II	1751080	36.4	NC_005042		100.0	0.00
<i>Prochlorococcus</i>	MIT9211	Equatorial Pacific, 83 m	LL III	1688963	38.0	NC_009976		99.7	0.00
<i>Prochlorococcus</i>	MIT9303	Sargasso Sea, 100 m	LL IV	2682807	50.0	NC_008820		100.0	0.14
<i>Prochlorococcus</i>	MIT9313	Gulf Stream, 135 m	LL IV	2410873	50.7	NC_005071		99.5	0.14
SAR11	HTCC7211	Sargasso Sea, open-ocean	Ia	1456888	29.0	NZ_DS995298	NZ_ABVS00000000.scaffold	100.0	0.47
SAR11	HTCC9565	N. Pacific, Oregon, open-ocean	Ia	1279674	29.0	SRX037743	PRJNA278082	99.1	0.00
SAR11	HIMB5	N. Pacific, Hawaii, coastal	Ia	1343202	28.6	NC_018643		61.4	0.00
SAR11	HTCC1002	N. Pacific, Oregon, coastal	Ia	1322994	29.7	NZ_CH724130	NZ_AAPV00000000.scaffold	100.0	0.47
SAR11	HTCC1062	N. Pacific, Oregon, coastal	Ia	1308759	29.7	NC_007205		100.0	0.47
SAR11	HIMB114	N. Pacific, Hawaii, coastal	IIla	1237371	29.6	NZ_GG704918	NZ_ADAC00000000.scaffold	98.8	2.38
SAR11	IMCC9063	Arctic Ocean, Norway, coastal	IIla	1284727	31.7	NC_015380		98.8	0.00
SAR11	HIMB59	N. Pacific, Hawaii, coastal	Va	1410127	32.3	NC_018644		88.6	0.00

**Table S2.** OGs found only in Red Sea single-cell genomes that are annotated as something other than “hypothetical protein”. All non-hypothetical RS-OGs including singletons are provided in supplemental file 2. OG descriptions are provided in supplemental file 4.

Genus	OG number	OG size	Example SAG	Example description	Example protein
<i>Prochlorococcus</i>	proch20358	1	PI06	Methionine import ATP-binding protein MetN	Proch_AAA795-I06_00177
<i>Prochlorococcus</i>	proch20359	1	PI06	Ubiquinone/menaquinone biosynthesis C-methyltransferase UbiE	Proch_AAA795-I06_00179
<i>Prochlorococcus</i>	proch20363	1	PI06	Ubiquinone/menaquinone biosynthesis C-methyltransferase UbiE	Proch_AAA795-I06_00210
<i>Prochlorococcus</i>	proch20373	1	PI15	GDP-perosamine synthase	Proch_AAA795-I15_00082
<i>Prochlorococcus</i>	proch20377	1	PI15	tRNA-2-methylthio-N(6)-dimethylallyl adenosine synthase	Proch_AAA795-I15_00146
<i>Prochlorococcus</i>	proch20394	1	PJ16	Inner membrane protein Yhal	Proch_AAA795-J16_00432
<i>Prochlorococcus</i>	proch20409	1	PJ16	UDP-4-amino-4-deoxy-L-arabinose--oxoglutarate aminotransferase	Proch_AAA795-J16_01544
<i>Prochlorococcus</i>	proch20421	1	PM23	Glucosylglycerol-phosphate synthase	Proch_AAA795-M23_01055
<i>Prochlorococcus</i>	proch20425	1	PM23	UvrABC system protein B	Proch_AAA795-M23_01207
SAR11	pelag11768	5	SB16	N-acylneuraminate cytidylyltransferase	Pelagi_AAA795-B16_00606
SAR11	pelag11869	4	SA20	Guanidinopropionase	Pelagi_AAA795-A20_00614
SAR11	pelag11963	4	SB16	Alkyl hydroperoxide reductase AhpD	Pelagi_AAA795-B16_00388
SAR11	pelag11968	3	SF16	Formate dehydrogenase	Pelagi_AAA795-F16_00091
SAR11	pelag12057	3	SG15	sn-glycerol-3-phosphate import ATP-binding protein UgpC	Pelagi_AAA795-G15_00156
SAR11	pelag12189	2	SF16	Fimbrial protein	Pelagi_AAA795-F16_00208
SAR11	pelag12217	2	SA08	N-acylneuraminate cytidylyltransferase	Pelagi_AAA795-A08_00237
SAR11	pelag12258	2	SL23	ATP synthase subunit alpha	Pelagi_AAA795-L23_00301
SAR11	pelag12412	2	SB16	4-methylaminobutanoate oxidase (formaldehyde-forming)	Pelagi_AAA795-B16_00337
SAR11	pelag12623	2	SA08	GDP-perosamine synthase	Pelagi_AAA795-A08_00275
SAR11	pelag12658	2	SI19	Alcohol dehydrogenase	Pelagi_AAA797-I19_00217
SAR11	pelag12665	2	SM18	Bifunctional homocysteine S-methyltransferase/5,10-methylenetetrahydrofolate reductase	Pelagi_AAA795-M18_01040
SAR11	pelag12668	2	SC09	Inositol 2-dehydrogenase/D-chiro-inositol 3-dehydrogenase	Pelagi_AAA795-C09_00458
SAR11	pelag12669	2	SC09	3-oxoacyl-[acyl-carrier-protein] reductase FabG	Pelagi_AAA795-C09_00465
SAR11	pelag12679	2	SD22	DNA polymerase I	Pelagi_AAA795-D22_00674
SAR11	pelag12688	2	SE07	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase (ferredoxin)	Pelagi_AAA795-E07_00143
SAR11	pelag12700	2	SM18	Heparin-sulfate lyase	Pelagi_AAA795-M18_00113
SAR11	pelag12701	2	SI19	4-hydroxy-tetrahydrodipicolinate synthase	Pelagi_AAA797-I19_00542
SAR11	pelag12703	2	SK18	Inositol 2-dehydrogenase	Pelagi_AAA795-K18_00566
SAR11	pelag12707	2	SM18	putative GTP-binding protein YjiA	Pelagi_AAA795-M18_00527
SAR11	pelag12710	2	SM22	Phosphoribosylglycinamide formyltransferase	Pelagi_AAA795-M22_00022
...	...	...	...	...	...

**Table S3.** Best-case scenarios of OG recruitment by *Tara* metagenomes. These are the number of OGs found (and fraction of total OGs found) if each new metagenome added adds the most new OGs. The full version of this table is provided in supplemental file 3.

Order	All SAR11 OGs best-case scenario			Red Sea-assoc. SAR11 OGs best-case scenario			All <i>Prochlorococcus</i> OGs best-case scenario			Red Sea-assoc. <i>Prochlorococcus</i> OGs best-case scenario		
	Metagenome	OGs found	Fraction of total OGs found	Metagenome	OGs found	Fraction of total OGs found	Metagenome	OGs found	Fraction of total OGs found	Metagenome	OGs found	Fraction of total OGs found
1	pelag_133_SRF_0.22-3_1e-5	3495	0.7810	pelag_094_SRF_0.22-3_1e-5	459	0.6438	proch_052_SRF_0.22-1.6_1e-5	4840	0.6422	proch_031_SRF_0.22-1.6_1e-5	35	0.5224
2	pelag_030_SRF_0.22-1.6_1e-5	3744	0.8366	pelag_124_SRF_0.22-3_1e-5	529	0.7419	proch_025_DCM_0.22-1.6_1e-5	5471	0.7259	proch_048_SRF_0.22-1.6_1e-5	44	0.6567
3	pelag_123_MIX_0.22-3_1e-5	3888	0.8688	pelag_100_MES_0.22-3_1e-5	565	0.7924	proch_111_DCM_0.22-3_1e-5	5818	0.7719	proch_140_SRF_0.22-3_1e-5	50	0.7463
4	pelag_094_SRF_0.22-3_1e-5	3968	0.8867	pelag_030_SRF_0.22-1.6_1e-5	590	0.8275	proch_032_SRF_0.22-1.6_1e-5	6093	0.8084	proch_045_SRF_0.22-1.6_1e-5	55	0.8209
5	pelag_141_SRF_0.22-3_1e-5	4031	0.9008	pelag_140_SRF_0.22-3_1e-5	609	0.8541	proch_137_DCM_0.22-3_1e-5	6311	0.8373	proch_052_SRF_0.22-1.6_1e-5	58	0.8657
6	pelag_100_MES_0.22-3_1e-5	4081	0.9120	pelag_123_MIX_0.22-3_1e-5	625	0.8766	proch_099_SRF_0.22-3_1e-5	6444	0.8550	proch_096_SRF_0.22-3_1e-5	60	0.8955
7	pelag_085_DCM_0.22-3_1e-5	4122	0.9211	pelag_133_SRF_0.22-3_1e-5	635	0.8906	proch_034_DCM_0.22-1.6_1e-5	6564	0.8709	proch_004_DCM_0.22-1.6_1e-5	61	0.9104
8	pelag_030_DCM_0.22-1.6_1e-5	4152	0.9278	pelag_141_SRF_0.22-3_1e-5	645	0.9046	proch_045_SRF_0.22-1.6_1e-5	6656	0.8831	proch_032_DCM_0.22-1.6_1e-5	62	0.9254
9	pelag_124_SRF_0.22-3_1e-5	4178	0.9336	pelag_150_SRF_0.22-3_1e-5	653	0.9158	proch_030_DCM_0.22-1.6_1e-5	6726	0.8924	proch_032_SRF_0.22-1.6_1e-5	63	0.9403
10	pelag_125_MIX_0.22-3_1e-5	4200	0.9385	pelag_057_SRF_0.22-3_1e-5	659	0.9243	proch_138_SRF_0.22-3_1e-5	6789	0.9008	proch_038_DCM_0.22-1.6_1e-5	64	0.9552
11	pelag_084_SRF_0.22-3_1e-5	4219	0.9428	pelag_110_DCM_0.22-3_1e-5	665	0.9327	proch_124_SRF_0.22-3_1e-5	6842	0.9078	proch_072_DCM_0.22-3_1e-5	65	0.9701
12	pelag_041_SRF_0.22-3_1e-5	4236	0.9466	pelag_111_DCM_0.22-3_1e-5	670	0.9397	proch_132_DCM_0.22-3_1e-5	6886	0.9136	proch_111_DCM_0.22-3_1e-5	66	0.9851
13	pelag_025_SRF_0.22-1.6_1e-5	4250	0.9497	pelag_138_MES_0.22-3_1e-5	674	0.9453	proch_048_SRF_0.22-1.6_1e-5	6927	0.9191	proch_136_SRF_0.22-3_1e-5	67	1.0000
14	pelag_000_DCM_0.22-1.6_1e-5	4262	0.9524	pelag_004_DCM_0.22-1.6_1e-5	677	0.9495	proch_018_DCM_0.22-1.6_1e-5	6964	0.9240	proch_004_SRF_0.22-1.6_1e-5	67	1.0000
15	pelag_140_SRF_0.22-3_1e-5	4274	0.9551	pelag_112_DCM_0.22-3_1e-5	680	0.9537	proch_031_SRF_0.22-1.6_1e-5	6995	0.9281	proch_007_DCM_0.22-1.6_1e-5	67	1.0000
16	pelag_140_SRF_0.22-1.6_1e-5	4284	0.9573	pelag_132_SRF_0.22-3_1e-5	683	0.9579	proch_032_DCM_0.22-1.6_1e-5	7023	0.9318	proch_007_SRF_0.22-1.6_1e-5	67	1.0000
17	pelag_018_SRF_0.22-3_1e-5	4294	0.9596	pelag_009_SRF_0.22-1.6_1e-5	685	0.9607	proch_064_SRF_0.22-3_1e-5	7051	0.9355	proch_009_DCM_0.22-1.6_1e-5	67	1.0000
18	pelag_078_MES_0.22-3_1e-5	4303	0.9616	pelag_036_SRF_0.22-1.6_1e-5	687	0.9635	proch_100_DCM_0.22-3_1e-5	7074	0.9386	proch_009_SRF_0.22-1.6_1e-5	67	1.0000
19	pelag_102_DCM_0.22-3_1e-5	4312	0.9636	pelag_048_SRF_0.22-1.6_1e-5	689	0.9663	proch_125_SRF_0.22-3_1e-5	7096	0.9415	proch_018_DCM_0.22-1.6_1e-5	67	1.0000
20	pelag_124_MIX_0.22-3_1e-5	4320	0.9654	pelag_068_MES_0.22-3_1e-5	691	0.9691	proch_132_SRF_0.22-3_1e-5	7118	0.9444	proch_018_SRF_0.22-1.6_1e-5	67	1.0000
21	pelag_099_SRF_0.22-3_1e-5	4327	0.9669	pelag_099_SRF_0.22-3_1e-5	693	0.9719	proch_041_DCM_0.22-1.6_1e-5	7138	0.9471	proch_023_DCM_0.22-1.6_1e-5	67	1.0000
22	pelag_137_DCM_0.22-3_1e-5	4334	0.9685	pelag_110_SRF_0.22-3_1e-5	695	0.9748	proch_122_DCM_0.22-3_1e-5	7157	0.9496	proch_023_SRF_0.22-1.6_1e-5	67	1.0000
23	pelag_145_MIX_0.22-3_1e-5	4341	0.9701	pelag_132_DCM_0.22-3_1e-5	697	0.9776	proch_141_SRF_0.22-3_1e-5	7176	0.9521	proch_025_DCM_0.22-1.6_1e-5	67	1.0000
24	pelag_145_SRF_0.22-3_1e-5	4348	0.9716	pelag_137_DCM_0.22-3_1e-5	699	0.9804	proch_004_DCM_0.22-1.6_1e-5	7193	0.9544	proch_025_SRF_0.22-1.6_1e-5	67	1.0000
25	pelag_025_DCM_0.22-1.6_1e-5	4353	0.9727	pelag_009_DCM_0.22-1.6_1e-5	700	0.9818	proch_100_MES_0.22-3_1e-5	7210	0.9566	proch_030_DCM_0.22-1.6_1e-5	67	1.0000
26	pelag_033_SRF_0.22-1.6_1e-5	4358	0.9739	pelag_018_DCM_0.22-1.6_1e-5	701	0.9832	proch_072_SRF_0.22-3_1e-5	7225	0.9586	proch_030_SRF_0.22-1.6_1e-5	67	1.0000
27	pelag_085_SRF_0.22-3_1e-5	4363	0.9750	pelag_030_DCM_0.22-1.6_1e-5	702	0.9846	proch_112_DCM_0.22-3_1e-5	7240	0.9606	proch_033_SRF_0.22-1.6_1e-5	67	1.0000
28	pelag_009_SRF_0.22-1.6_1e-5	4367	0.9759	pelag_036_DCM_0.22-1.6_1e-5	703	0.9860	proch_094_SRF_0.22-3_1e-5	7254	0.9625	proch_034_DCM_0.22-1.6_1e-5	67	1.0000
29	pelag_018_DCM_0.22-1.6_1e-5	4371	0.9768	pelag_041_SRF_0.22-1.6_1e-5	704	0.9874	proch_102_SRF_0.22-3_1e-5	7268	0.9643	proch_034_SRF_0.22-1.6_1e-5	67	1.0000
30	pelag_038_MES_0.22-1.6_1e-5	4375	0.9777	pelag_056_SRF_0.22-3_1e-5	705	0.9888	proch_111_SRF_0.22-3_1e-5	7280	0.9659	proch_036_DCM_0.22-1.6_1e-5	67	1.0000
31	pelag_039_MES_0.22-1.6_1e-5	4379	0.9785	pelag_064_SRF_0.22-3_1e-5	706	0.9902	proch_096_SRF_0.22-3_1e-5	7291	0.9674	proch_036_SRF_0.22-1.6_1e-5	67	1.0000
32	pelag_109_SRF_0.22-3_1e-5	4383	0.9794	pelag_076_MES_0.22-3_1e-5	707	0.9916	proch_140_SRF_0.22-3_1e-5	7302	0.9688	proch_037_MES_0.22-1.6_1e-5	67	1.0000
33	pelag_111_SRF_0.22-3_1e-5	4387	0.9803	pelag_084_SRF_0.22-3_1e-5	708	0.9930	proch_149_SRF_0.22-3_1e-5	7313	0.9703	proch_038_MES_0.22-1.6_1e-5	67	1.0000
34	pelag_132_DCM_0.22-3_1e-5	4391	0.9812	pelag_096_SRF_0.22-3_1e-5	709	0.9944	proch_052_DCM_0.22-1.6_1e-5	7322	0.9715	proch_038_SRF_0.22-1.6_1e-5	67	1.0000
35	pelag_009_DCM_0.22-1.6_1e-5	4394	0.9819	pelag_102_DCM_0.22-3_1e-5	710	0.9958	proch_137_SRF_0.22-3_1e-5	7331	0.9727	proch_039_DCM_0.22-1.6_1e-5	67	1.0000
36	pelag_034_SRF_0.22-1.6_1e-5	4397	0.9826	pelag_102_SRF_0.22-3_1e-5	711	0.9972	proch_038_DCM_0.22-1.6_1e-5	7339	0.9737	proch_039_MES_0.22-1.6_1e-5	67	1.0000
37	pelag_042_SRF_0.22-1.6_1e-5	4400	0.9832	pelag_110_MES_0.22-3_1e-5	712	0.9986	proch_093_SRF_0.22-3_1e-5	7347	0.9748	proch_041_DCM_0.22-1.6_1e-5	67	1.0000
38	pelag_048_SRF_0.22-1.6_1e-5	4403	0.9839	pelag_138_SRF_0.22-3_1e-5	713	1.0000	proch_098_DCM_0.22-3_1e-5	7354	0.9757	proch_041_SRF_0.22-1.6_1e-5	67	1.0000
39	pelag_052_DCM_0.22-1.6_1e-5	4406	0.9846	pelag_004_SRF_0.22-1.6_1e-5	713	1.0000	proch_128_SRF_0.22-3_1e-5	7361	0.9766	proch_042_DCM_0.22-1.6_1e-5	67	1.0000
40	pelag_052_SRF_0.22-1.6_1e-5	4409	0.9853	pelag_007_DCM_0.22-1.6_1e-5	713	1.0000	proch_142_SRF_0.22-3_1e-5	7368	0.9776	proch_042_SRF_0.22-1.6_1e-5	67	1.0000

**Table S4.** OGs driving separation in principal components analysis (PCA) of *Tara* Oceans surface samples based on *Prochlorococcus* and SAR11 OG abundances. These OGs were among the top 50 variables driving PCA separation of samples. Hypothetical proteins have been omitted from this list. OG descriptions are provided in supplemental file 4.

Genus	OG number	Description
<i>Prochlorococcus</i>	proch10265	LPS-assembly protein LptD
<i>Prochlorococcus</i>	proch11818	Intimin
<i>Prochlorococcus</i>	proch10782	LexA repressor
<i>Prochlorococcus</i>	proch12059	Outer membrane protein PagN
<i>Prochlorococcus</i>	proch11343	Undecaprenyl phosphate-alpha-4-amino-4-deoxy-L-arabinose arabinosyl transferase
<i>Prochlorococcus</i>	proch11663	Tyrosine-protein kinase ptk
<i>Prochlorococcus</i>	proch12291	Beta-barrel assembly-enhancing protease
<i>Prochlorococcus</i>	proch11496	Lipopolysaccharide assembly protein A
<i>Prochlorococcus</i>	proch11401	Lipopolysaccharide export system protein LptC
<i>Prochlorococcus</i>	proch12161	Phycoerythrin class 2 subunit gamma, linker polypeptide
<i>Prochlorococcus</i>	proch12010	C-phycoerythrin class 1 subunit alpha
SAR11	pelag10168	4-hydroxythreonine-4-phosphate dehydrogenase 2
SAR11	pelag10710	Lipopolysaccharide export system permease protein LptG
SAR11	pelag11313	2-O-methyltransferase Noel
SAR11	pelag11257	putative undecaprenyl-phosphate N-acetylglucosaminy1 1-phosphate transferase
SAR11	pelag11434	UDP-N,N'-diacetylbacillosamine 2-epimerase (hydrolyzing)
SAR11	pelag11679	Putative polysaccharide biosynthesis protein with aminopeptidase-like domain protein
SAR11	pelag10704	Heterocyst differentiation ATP-binding protein HepA
SAR11	pelag11235	CDP-glucose 4,6-dehydratase
SAR11	pelag11593	putative adenylyl-sulfate kinase
SAR11	pelag11308	23S rRNA (guanine(745)-N(1))-methyltransferase
SAR11	pelag11653	Phosphoribosylglycinamide formyltransferase
SAR11	pelag11002	CMP-N,N'-diacetyllegionaminic acid synthase
SAR11	pelag11422	dTDP-glucose 4,6-dehydratase
SAR11	pelag11743	GDP-6-deoxy-D-talose 4-dehydrogenase
SAR11	pelag11212	Undecaprenyl-phosphate alpha-N-acetylglucosaminy1 1-phosphate transferase
SAR11	pelag11890	Peptidoglycan-N-acetylglucosamine deacetylase
SAR11	pelag11566	Inner membrane ABC transporter permease protein YcjO
SAR11	pelag11963	Alkyl hydroperoxide reductase AhpD
SAR11	pelag11958	GDP-mannose-dependent alpha-mannosyltransferase
SAR11	pelag14916	Chitinase A

## Supplemental Material

Supplemental file 1 – Metadata of genomes used for analyses and OG clustering: strain, genus, code (for OG clustering), clade, Red Sea or not (True/False), latitude, longitude, depth in meters where isolated, type (Cultured/SAG), strain short name (without “Pro\_” or “SAR11\_”), and reference. genome\_metadata.xlsx

Supplemental file 2 – Full version of Table S2. tableS2\_red\_sea\_assoc\_non\_hypothetical.xlsx

Supplemental file 3 – Full version of Table S3. tableS3\_tara\_best\_case\_scenarios.xlsx

Supplemental file 4 – Metadata of OGs: OG number, number of genomes found in, example accession number, description of function, Red-Sea-associated (True/False), and genus. og\_metadata.xlsx

Supplemental file 5 – Genome OG counts: counts per genome of SAR11 OGs. og\_table\_sar.xlsx

Supplemental file 6 – Genome OG counts: counts per genome of *Prochlorococcus* OGs. og\_table\_pro.xlsx

Supplemental file 7 – *Tara* OG counts: counts per *Tara* metagenome of SAR11 OGs. tara\_counts\_sar.xlsx

Supplemental file 8 – *Tara* OG counts: counts per *Tara* metagenome of *Prochlorococcus* OGs. tara\_counts\_pro.xlsx

Protein sequence identifiers and FASTA sequences for each OG have been archived at <https://zenodo.org> with DOI 10.5281/zenodo.2634561.