

Supplemental Information

SI Figures

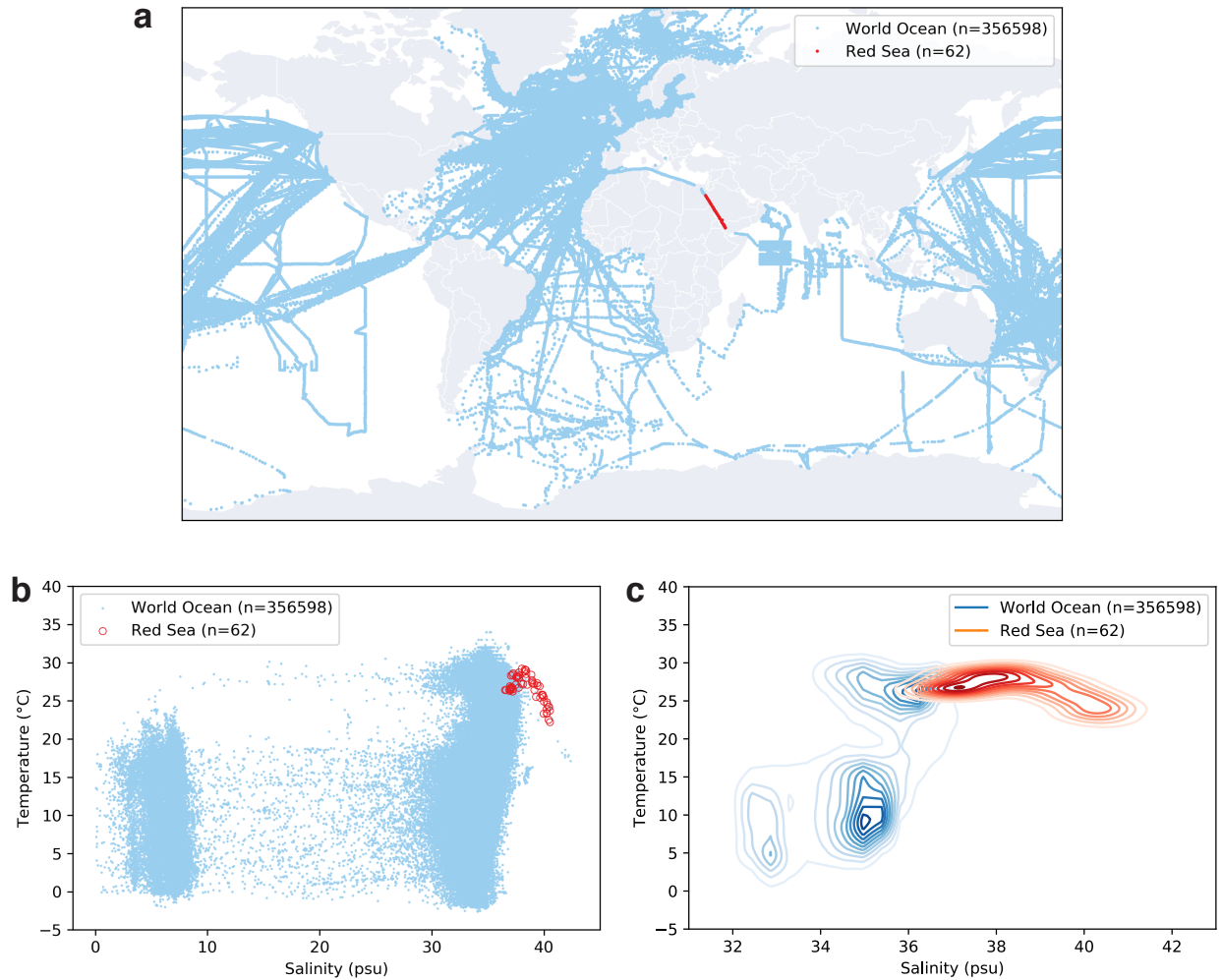


Figure S1. Temperature versus salinity relationships in the World Ocean Database 2013 (WOD13) colored by Red Sea or World Ocean excluding the Red Sea (World Ocean). (a) Map of samples in WOD13, with Red Sea samples ($n=62$) in red. (b) Scatter plot of all points in WOD13, which spans from 1900 to 2010, 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.

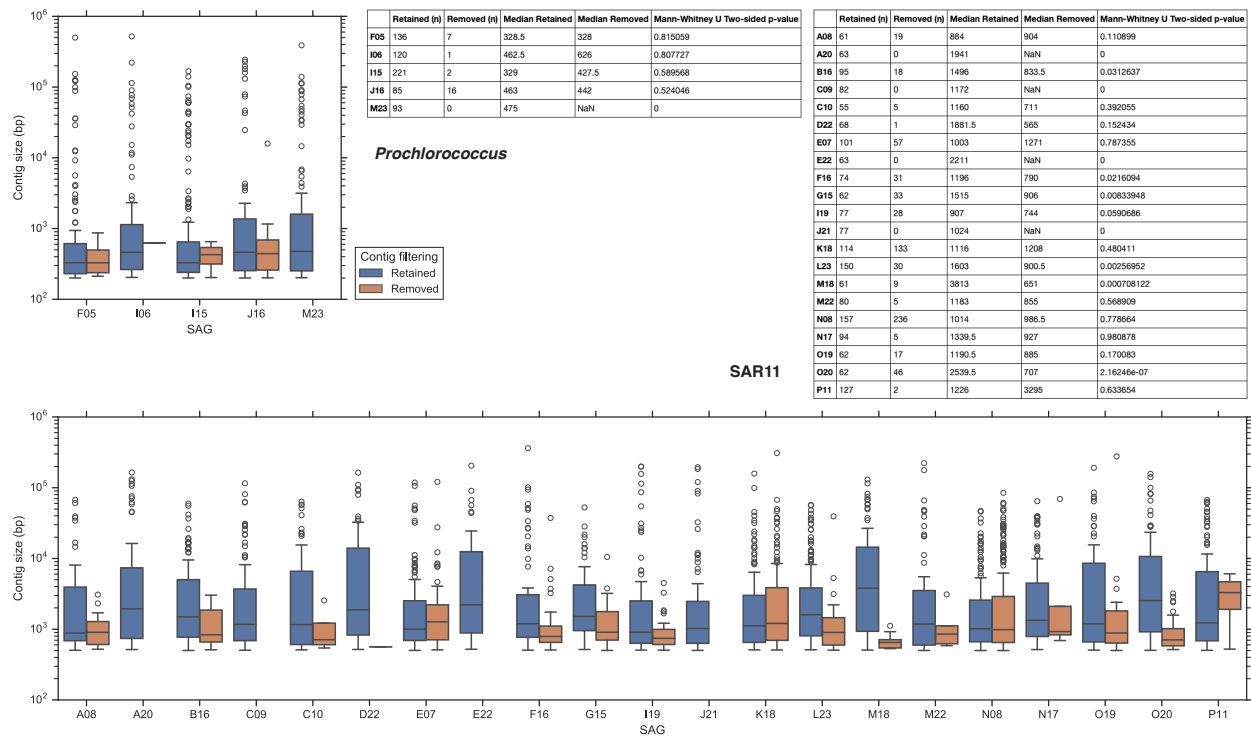


Figure S2. Contig size distributions of retained and removed contigs in the filtering of possible contamination from SAG assemblies. Boxplots show the median, interquartile range (IQR), and $1.5 \times \text{IQR}$ (with outliers). Tables show, for each contig, the number n of contigs retained and removed, 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.

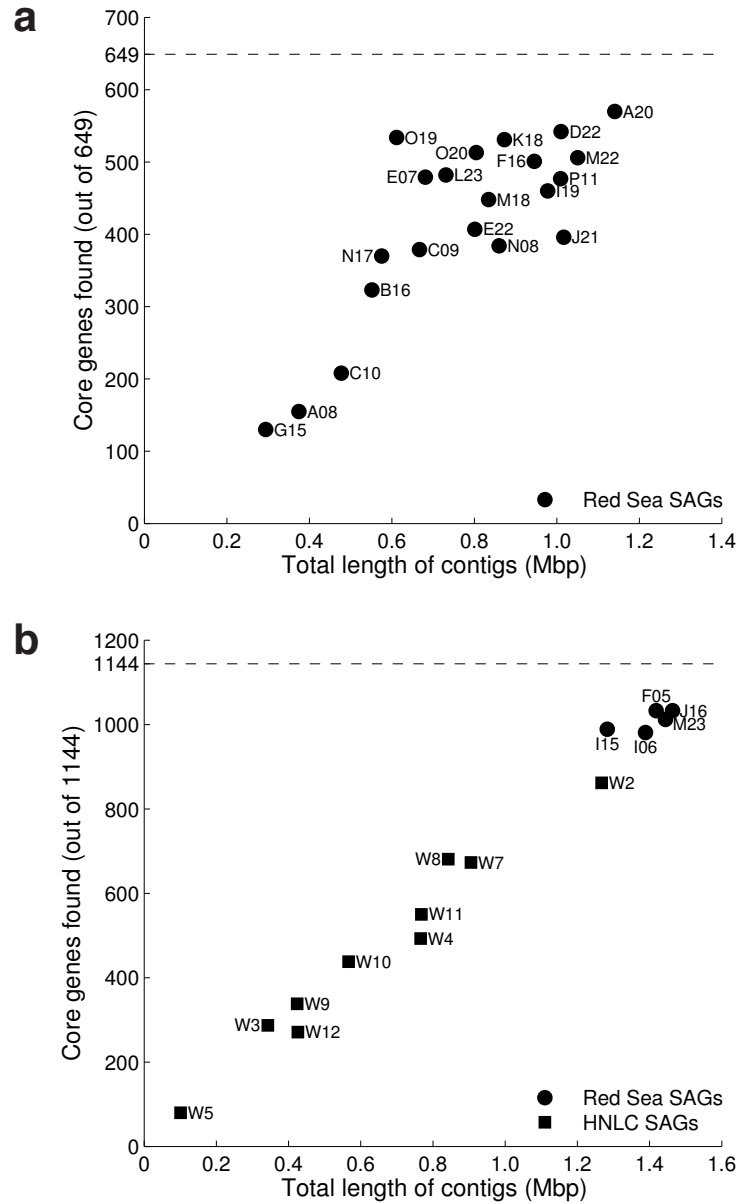


Figure S3. Completeness of (a) SAR11 and (b) *Prochlorococcus* single-cell amplified genomes (SAGs) based on the number of single-copy core genes. Completeness of *Prochlorococcus* SAGs was generally higher than SAR11 SAGs.

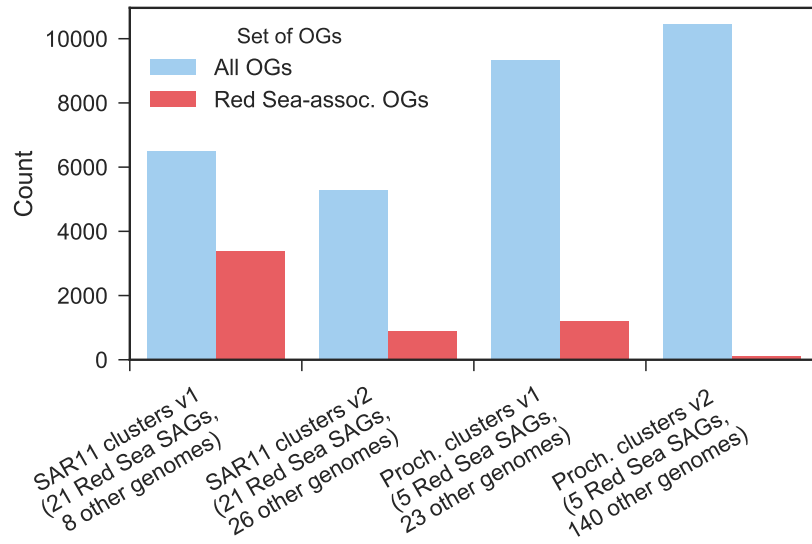


Figure S4. Bar chart showing number of the number of ortholog groups (OGs) identified in two versions of clustering and how many OGs were Red-Sea-associated (found only in the Red Sea genomes). As more genomes are added to the clustering, the number of OGs found only in Red Sea SAGs approaches zero.

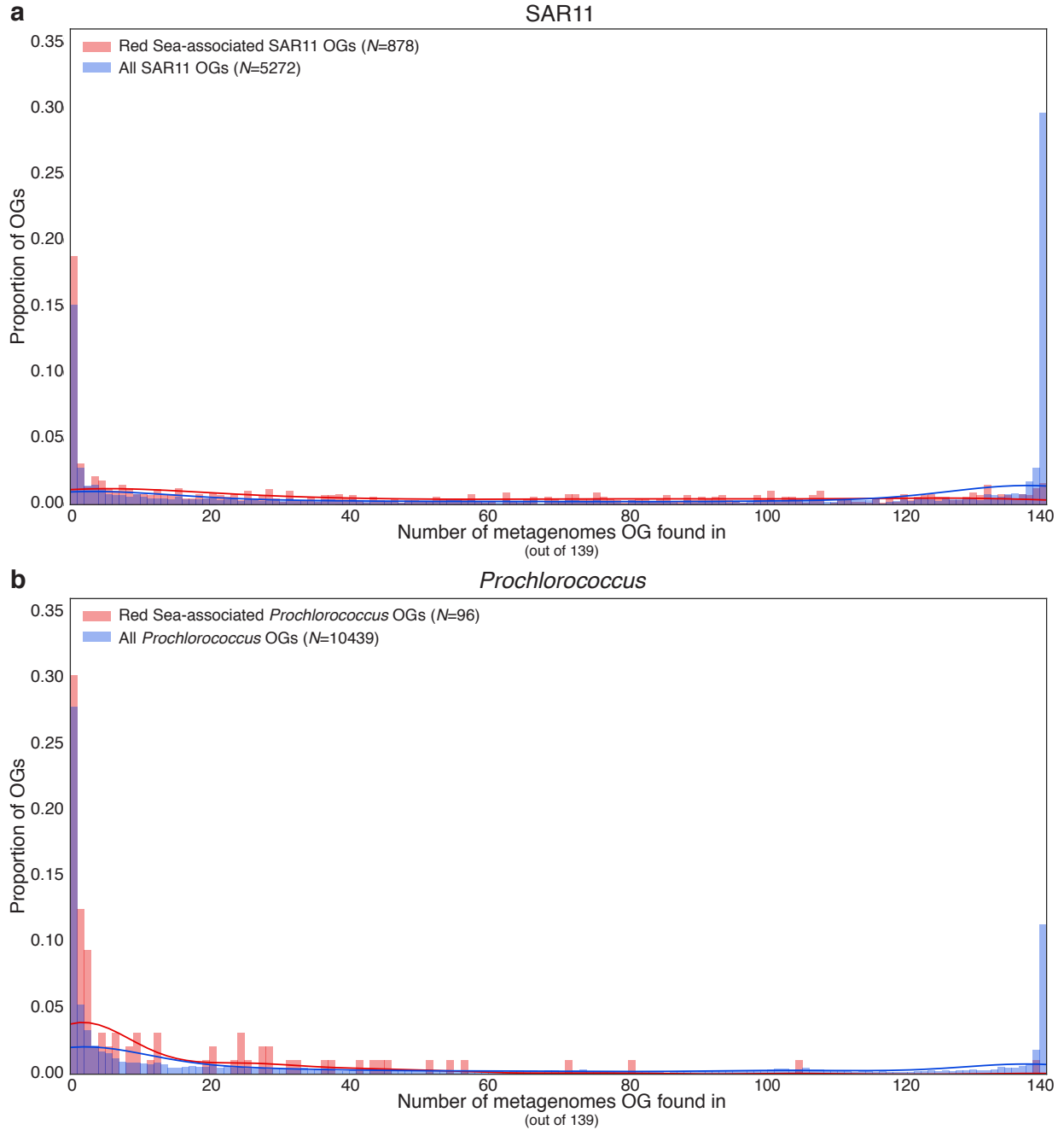


Figure S5. Histograms (bars) and kernel density estimates (lines) showing the distribution of (a) SAR11 and (b) *Prochlorococcus* OGs (all and Red-Sea-associated) by the number of *Tara* Oceans metagenomes they were found in. Red-Sea-associated OGs were less likely to be found in a high proportion of 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	IIIa	1237371	29.6	NZ_GG704918	NZ_ADAC00000000.scaffold	98.8	2.38
SAR11	IMCC9063	Arctic Ocean, Norway, coastal	IIIa	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)-dimethylallyladenine 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 cytidyltransferase	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 cytidyltransferase	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 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-acetylglucosaminyl 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-acetylglucosaminyl 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.