

(ST1)

Summary of the 137 marine genomes that were sequenced, assembled, and annotated as part of the Gordon and Betty Moore Foundation-funded marine microbial genome sequencing project.

Organism	Genome size (Mbp)	%GC	# proteins	ACC ID	NCBI PROJ ID	Kingdom	Phylum	Sampling Depth	Sampling Location
Aciduliprofundum boonei T469	2.97	39	2949	ABSD000000000	19483	Archaea	Euryarchaeota	1920 m	Lau Basin, Mariner vent field, 22S, 1800m
Alcanivorax sp. DG881	3.79	58	3337	ABRW000000000	19323	Bacteria	Proteobacteria	Surface waters	Paralytic shellfish producing dinoflagellate Gymnodinium catenatum isolated from the Derwent estuary, Tasmania, Australia. DG881 was isolated from a clonal culture of this dinoflagellate in Scotland 2001
Algoriphagus sp. PR1	4.78	38	4215	AAXU000000000	18947	Bacteria	Bacteroidetes	15 cm	Proterospongia culture
alpha proteobacterium BAL199	6.10	65	6128	ABHC000000000	19281	Bacteria	Proteobacteria	3 m	Baltic proper, St. BY31 (N 58° 35.90', E 18° 14.21')
Alteromonadales bacterium TW-7	4.10	39	3783	AAVS000000000	18465	Bacteria	Proteobacteria	surface waters	Coastal, off Scripps, detritus prepared from the diatom Talassiosira weissflogii
Alteromonas macleodii 'Deep ecotype'	4.41	44	4072	AAOD000000000	13374	Bacteria	Proteobacteria	3500 m	Urania Basin, Ionian Sea, Mediterranean
Aurantimonas sp. SI85-9A1	4.33	66	3650	AAPJ000000000	13607	Bacteria	Proteobacteria	Suboxic zone, ~100 m	Saanich Inlet, seasonally anoxic fjord in British Colombia, Canada, hydrothermal plumes, Black Sea, Alvinella worm surfaces
Bacillus sp. B14905	4.50	37	4624	AAXV000000000	18949	Bacteria	Firmicutes	Surface	Offshore Gulf of Mexico
Bacillus sp. NRRL B-14911	5.09	45	5691	AAOX000000000	13545	Bacteria	Firmicutes	10 m	Gulf of Mexico
Bacillus sp. SG-1	3.95	42	4337	ABCF000000000	19283	Bacteria	Firmicutes	10 m but also isolates from surface to 200 m	San Diego Bay surface sediments
beta proteobacterium KB13	1.33	35	1318	ABXG000000000	13602	Bacteria	Proteobacteria	Surface waters	Kaneohe Bay, Hawaii
Blastopirellula marina DSM 3645	6.65	57	6025	AANZ000000000	13488	Bacteria	Planctomycetes	Surface waters	Kiel Fjord, Baltic Sea, Germany
Brevundimonas sp. BAL3	3.64	67	3470	ABRU000000000	19287	Bacteria	Proteobacteria	4 m	Baltic Sea (63.30N, 19.48E)
Caldanaerobacter DSM 12653	2.32	36	1696	ABXP000000000	19289	Bacteria	Firmicutes	1500m	Okinawa Trough, 1500 m
Caminibacter mediatlanticus TB-2	1.66	27	1826	ABCJ000000000	19293	Bacteria	Proteobacteria	2300 m	Mid-Atlantic Ridge vent
Campylobacteriales bacterium GD 1	2.94	33	2745	ABXD000000000	19295	Bacteria	Proteobacteria	215 m	Gotland Deep, Baltic Sea, 215m, pelagic redoxcline of the Baltic Sea
Candidatus Pelagibacter sp. HTCC7211	1.46	29	1447	ABVS000000000	19339	Bacteria	Proteobacteria	N/A	BATS Bermuda-Atlantic-Time-Series study site
Candidatus Pelagibacter ubique HTCC1002	1.33	29	1393	AAPV000000000	13577	Bacteria	Proteobacteria	10 m	Coastal Pacific Ocean, Newport, Oregon
Carnobacterium sp. AT7	2.44	35	2388	ABHH000000000	19299	Bacteria	Firmicutes	2500 m	Aleutian Trench
Congregibacter litoralis KT71	4.33	57	3941	AAOA000000000	13585	Bacteria	Proteobacteria	Surface waters	North Sea near Helgoland, Germany
Croceibacter atlanticus HTCC2559	2.95	33	2719	AAMP000000000	13570	Bacteria	Bacteroidetes	Surface, 10 m	Sargasso Sea, Atlantic Ocean, BATS
Cyanobium sp. PCC 7001	2.83	68	2771	ABSE000000000	19301	Bacteria	Cyanobacteria	N/A	Long Island, NY, USA
Cyanothece sp. CCY0110	5.88	36	6475	AAXW000000000	18951	Bacteria	Cyanobacteria	10 m	Coastal seawaters off Zanzibar
Dokdonia donghaensis MED134	3.30	38	2944	AAMZ000000000	13544	Bacteria	Bacteroidetes	1 m	NW Mediterranean Sea

Erythrobacter litoralis HTCC2594	3.05	63	3011	CP000157	13480	Bacteria	Proteobacteria	Surface, 10 m	Sargasso Sea, Atlantic Ocean, BATS
Erythrobacter sp. NAP1	3.27	60	3177	AAMW00000000	13512	Bacteria	Proteobacteria	Surface waters	NW Atlantic Ocean
Erythrobacter sp. SD-21	2.97	62	2941	ABCG00000000	19305	Bacteria	Proteobacteria	10 m	San Diego Bay surface sediments
Flavobacteria bacterium BAL38	2.81	31	2612	AAXX00000000	18953	Bacteria	Bacteroidetes	4 m	Baltic Sea (63.30N, 19.48E)
Flavobacteria bacterium BBFL7	3.08	34	2592	AAPD00000000	13604	Bacteria	Bacteroidetes	Surface waters	Scripps Pier, La Jolla, CA
Flavobacteriales bacterium ALC-1	3.83	32	3445	ABHI00000000	19307	Bacteria	Bacteroidetes	surface waters	Scripps Pier, La Jolla, CA
Flavobacteriales bacterium HTCC2170	3.88	37	3478	AAOC00000000	13595	Bacteria	Bacteroidetes	10 m	Coastal Pacific Ocean, Newport, Oregon
Flavobacterium sp. MED217	4.24	39	3735	AANC00000000	13573	Bacteria	Bacteroidetes	1 m	NW Mediterranean Sea
Fulvimarina pelagi HTCC2506	3.80	61	3754	AATP00000000	13506	Bacteria	Proteobacteria	Surface, 10 m	Sargasso Sea, Atlantic Ocean, BATS
gamma proteobacterium HTCC5015	2.61	54	2459	ABSJ00000000	19309	Bacteria	Proteobacteria	Surface	Sargasso Sea, Atlantic Ocean, BATS
Gammaproteobacterium HTCC2148	4.33	52	3832	ABXQ00000000	30725	Bacteria	Proteobacteria	10 m	station NH15, Oregon Coast, USA
Gammaproteobacterium NOR5-3	4.21	56	3671	ACCX00000000	31353	Bacteria	Proteobacteria	Surface waters	top oxic layer of a sandy sediment core, List, Sylt (55°01'N, 8°26'E) Germany, 19.10.2005
Gammaproteobacterium NOR51-B	3.26	56	2931	ACCY00000000	31355	Bacteria	Proteobacteria	Surface waters	top oxic layer of a muddy sediment core, List, Sylt (55°02'N, 8°24'E) Germany, 19.10.2005
Glaciecola sp. HTCC2999	2.52	40	2267	ABST00000000	20121	Bacteria	Proteobacteria	10m	Coastal Pacific Ocean, Newport, Oregon
Hoeflea phototrophica DFL-43	4.46	59	4357	ABIA00000000	19311	Bacteria	Proteobacteria	N/A	Toxic dinoflagellates Alexandrium ostenfeldii and Prorocentrum lima
Hydrogenivirga sp. 128-5-R1-1	3.04	43	3808	ABHJ00000000	19313	Bacteria	Aquificae	2200 m	Lau Basin, 20S, 176W, 2200 m
Idiomarina baltica OS145	2.72	47	2608	AAMX00000000	13408	Bacteria	Proteobacteria	30 m	Baltic Sea
Janibacter sp. HTCC2649	4.23	68	4100	AAMN00000000	13546	Bacteria	Actinobacteria	10 m	Sargasso Sea, Atlantic Ocean, BATS
Kordia algicida OT-1	5.02	34	4514	ABIB00000000	19315	Bacteria	Bacteroidetes	1 m	Masan Bay, South Sea, Korea
Lentisphaera araneosa HTCC2155	6.02	40	5104	ABCK00000000	13437	Bacteria	Lentisphaerae	10 m	Coastal Pacific Ocean, Newport, Oregon
Limnobacter sp. MED105	3.39	52	3245	ABCT00000000	19317	Bacteria	Proteobacteria	Surface	E Mediterranean Sea (32.81N, 34.69E)
Loktanella vestfoldensis SKA53	3.06	59	3068	AAMS00000000	13444	Bacteria	Proteobacteria	2-5 m	North Atlantic
Lyngbya sp. PCC 8106	7.04	41	6142	AAVU00000000	13409	Bacteria	Cyanobacteria	Intertidal	Mellum, Germany
marine actinobacterium PHSC20C1	2.77	59	2668	AAOB00000000	13373	Bacteria	Actinobacteria	Surface waters	Coastal Antarctic Seawater - Western Antarctic Peninsula
marine gamma proteobacterium HTCC2080	3.58	51	3185	AAVV00000000	18487	Bacteria	Proteobacteria	surface 10m	Coastal Pacific Ocean, Newport, Oregon
marine gamma proteobacterium HTCC2143	3.93	47	3662	AAVT00000000	13605	Bacteria	Proteobacteria	10 m	Coastal Pacific Ocean, Newport, Oregon
marine gamma proteobacterium HTCC2207	2.62	49	2388	AAPI00000000	13606	Bacteria	Proteobacteria	10 m	Coastal Pacific Ocean, Newport, Oregon
Marinitoga piezophila KA3	1.82	45	1770	ABXR00000000	19319	Bacteria	Thermotogae	2630 m	East Pacific Rise (13degreesN), 2630 m
Marinobacter algicola DG893	4.41	56	4127	ABCP00000000	19321	Bacteria	Proteobacteria	surface waters	Isolated in Scotland from a clonal culture of the paralytic shellfish producing dinoflagellate Gymnodinium catenatum. G. catenatum was isolated from the Yellow Sea, Korea in 1998.

Marinobacter sp. ELB17	4.89	54	4850	AAXY00000000	18955	Bacteria	Proteobacteria	17 m	East Lobe Lake Bonney, Antarctica, 17m
Marinomonas sp. MED121	5.12	40	4820	AAANE00000000	13587	Bacteria	Proteobacteria	1 m	NW Mediterranean Sea
Mariprofundus ferrooxydans PV-1	2.87	54	2866	AATS00000000	13615	Bacteria	Proteobacteria	1100 m	Loihi Seamount, Pacific Ocean
Methylophilales bacterium HTCC2181	1.30	37	1338	AAUX00000000	18317	Bacteria	Proteobacteria	10 m	Coastal Pacific Ocean, Newport, Oregon
Microcoleus chthonoplastes PCC 7420	8.65	45	8294	ABRS0000	19325	Bacteria	Cyanobacteria	N/A	Salt Marsh, Woods Hole, Massachusetts
Microscilla marina ATCC 23134	9.77	40	8319	AAWS00000000	13411	Bacteria	Bacteroidetes	surface waters	Scripps Inst. Oceanogr., Marine aquarium outflow; San Diego, USA
Moritella sp. PE36	5.22	40	4781	ABCQ00000000	13535	Bacteria	Proteobacteria	~3600 m	Pacific Ocean, 288 km from San Diego, USA
Nitrobacter sp. Nb-311A	4.10	60	4256	AAAY00000000	13538	Bacteria	Proteobacteria	Surface waters	Tropical Eastern Atlantic
Nitrococcus mobilis Nb-231	3.62	59	3503	AAOF00000000	13475	Bacteria	Proteobacteria	Surface waters	Equatorial Pacific
Nitrosococcus oceani C-27	3.47	50	3396	ABSG00000000	13547	Bacteria	Proteobacteria	Surface waters	Pacific Ocean
Nodularia spumigena CCY9414	5.32	41	4860	AAVW00000000	13447	Bacteria	Cyanobacteria	Surface waters	Bornholm Sea, Baltic Sea
Oceanibulbus indolifex HEL-45	4.11	59	4153	ABID00000000	19327	Bacteria	Proteobacteria	10 m	3 km off the coast of Helgoland in the North Sea, 10 m
Oceanicaulis alexandrii HTCC2633	3.17	63	3029	AAMQ00000000	13448	Bacteria	Proteobacteria	Surface, 10 m	Sargasso Sea, Atlantic Ocean, BATS
Oceanicola batsensis HTCC2597	4.44	66	4212	AAMO00000000	13449	Bacteria	Proteobacteria	Surface, 10 m	Sargasso Sea, Atlantic Ocean, BATS
Oceanicola granulosus HTCC2516	4.04	70	3792	AAOT00000000	13450	Bacteria	Proteobacteria	Surface, 10 m	Sargasso Sea, Atlantic Ocean, BATS
Oceanobacter sp. RED65	3.53	43	3314	AAQH00000000	13562	Bacteria	Proteobacteria	1 m	Red Sea
Oceanospirillum sp. MED92	3.87	46	3688	AAOW00000000	13561	Bacteria	Proteobacteria	1 m	E Mediterranean Sea
Octadecabacter antarcticus 238	5.37	55	5834	ABSK00000000	19331	Bacteria	Proteobacteria	N/A	350 km offshore off Deadhorse, Alaska (73° 01' N, 148° 31' W)
Octadecabacter antarcticus 307	4.89	54	5495	ABSH00000000	19333	Bacteria	Proteobacteria	N/A	McMurdo Sound, Antarctica (77° 53' S, 166° 35' E)
Parvularcula bermudensis HTCC2503	2.91	60	2733	AAMU00000000	13509	Bacteria	Proteobacteria	Surface, 10 m	Sargasso Sea, Atlantic Ocean, BATS
Pedobacter sp. BAL39	5.79	44	5101	ABCM00000000	19337	Bacteria	Bacteroidetes	4 m	Baltic Sea (63.30N, 19.48E)
Phaeobacter gallaeciensis 2.10	4.16	59	3948	ABIE00000000	19353	Bacteria	Proteobacteria	Surface waters	surface of the marine algae Ulva lactuca (Australia)
Phaeobacter gallaeciensis BS107	4.23	59	4059	ABIF00000000	19355	Bacteria	Proteobacteria	N/A	Seawater from larval cultures of the scallop Pecten maximus, in A Coruna, Galicia, Spain
Photobacterium profundum 3TCK	6.11	41	5549	AAPH00000000	13563	Bacteria	Proteobacteria	Surface waters	San Diego
Photobacterium sp. SKA34	4.95	39	4567	AAOU00000000	13566	Bacteria	Proteobacteria	2-5 m	North Atlantic
Planctomyces maris DSM 8797	7.78	50	6480	ABCE00000000	15662	Bacteria	Planctomycetes	Neretic water	Coastal Bacterioplankton, Puget Sound, Washington, USA
Plesiocystis pacifica SIR-1	10.59	70	8450	ABCS00000000	19341	Bacteria	Proteobacteria	N/A	Japan
Polaribacter dokdonensis MED152	2.97	30	2679	AANA00000000	13543	Bacteria	Bacteroidetes	1 m	NW Mediterranean Sea
Polaribacter irgensii 23-P	2.75	34	2557	AAOG00000000	13451	Bacteria	Bacteroidetes	Surface 10-25 m	Penola Strait, Western Antarctic Peninsula

Prochlorococcus marinus str. AS9601	1.67	31	1921	CP000551	13548	Bacteria	Cyanobacteria	50 m	Arabian Sea
Prochlorococcus marinus str. MIT 9202	1.69	31	1891	ACDW00000000	31353	Bacteria	Cyanobacteria	79 m	South Pacific
Prochlorococcus marinus str. MIT 9211	1.69	38	1855	AALP00000000	13551	Bacteria	Cyanobacteria	83 m	Equatorial Pacific (0°/140° W)
Prochlorococcus marinus str. MIT 9303	2.68	50	2997	CP000554	13496	Bacteria	Cyanobacteria	100 m	Sargasso Sea
Prochlorococcus marinus str. MIT 9515	1.70	30	1906	CP000552	13617	Bacteria	Cyanobacteria	Surface waters	Equatorial Pacific
Prochlorococcus marinus str. NATL1A	1.86	34	2193	CP000553	15660	Bacteria	Cyanobacteria	30 m	North Atlantic 37.4°N; 40°W
Pseudoalteromonas tunicata D2	4.98	39	4504	AAOH00000000	13456	Bacteria	Proteobacteria	N/A	Tunicate surface (Ciona intestinalis), Sweden
Pseudovibrio sp. JE062	5.73	52	5238	ABXL00000000	19345	Bacteria	Proteobacteria	20 m	Key Largo, Florida
Psychroflexus torquis ATCC 700755	4.32	34	4038	AAPR00000000	13542	Bacteria	Bacteroidetes	1.5m	Prydz Bay, Antarctica
Psychromonas sp. CNPT3	2.95	38	2727	AAPG00000000	13583	Bacteria	Proteobacteria	~5800 m	Central North Pacific Ocean
Reinekea sp. MED297	4.51	52	4238	AAOE00000000	13534	Bacteria	Proteobacteria	1 m	NW Mediterranean Sea
Rhodobacterales bacterium HTCC2083	4.02	53	4179	ABXE00000000	13508	Bacteria	Proteobacteria	10 m	Coastal Pacific Ocean, Newport, Oregon
Rhodobacterales bacterium HTCC2150	3.58	49	3667	AAXZ00000000	18973	Bacteria	Proteobacteria	Surface	Coastal Pacific Ocean, Newport, Oregon
Rhodobacterales bacterium HTCC2255	2.30	36	2240	AATR00000000	17677	Bacteria	Proteobacteria	10 m	Coastal Pacific Ocean, Newport, Oregon
Rhodobacterales bacterium HTCC2654	4.53	64	4712	AAMT00000000	13614	Bacteria	Proteobacteria	Surface, 10 m	Sargasso Sea, Atlantic Ocean, BATS
Rhodobacterales bacterium Y4I	4.33	64	4133	ABXF00000000	19349	Bacteria	Proteobacteria	N/A	Coastal seawater near St. Mary's Georgia
Rhodobacterales KLH11	4.49	57	4277	ACCW00000000	31351	Bacteria	Proteobacteria	20 m	Key Largo, Florida
Robiginitalea biformata HTCC2501	3.53	55	3228	AAOI00000000	13461	Bacteria	Bacteroidetes	Surface, 10 m	Sargasso Sea, Atlantic Ocean, BATS
Roseobacter GAI101	4.52	58	4203	ABXS00000000	13560	Bacteria	Proteobacteria	N/A	Coastal seawater collected off Georgia in December 1997
Roseobacter litoralis Och 149	4.68	57	4746	ABIG00000000	19357	Bacteria	Proteobacteria	N/A	From seaweed
Roseobacter sp. AzwK-3b	4.18	61	4145	ABCR00000000	19361	Bacteria	Proteobacteria	surface waters	Elkhorn Slough, estuary of Monterey Bay.
Roseobacter sp. CCS2	3.50	55	3660	AAYB00000000	18995	Bacteria	Proteobacteria	surface water, 0.5 m	near Bodega Head County park, California, lat/long data are:123 degrees 3.97 west and 38 degrees 18.30 north
Roseobacter sp. MED193	4.65	57	4535	AANB00000000	13590	Bacteria	Proteobacteria	1 m	123 degrees 3.97 west and 38 degrees 18.30 north
Roseobacter sp. SK209-2-6	4.56	57	4537	AAYC00000000	18997	Bacteria	Proteobacteria	total depth >2500 m, 267 m in the oxygen minimum zone	NW Mediterranean Sea
									267 m in the oxygen minimum zone of the Arabian Sea, one of the three major denitrification zones in the open ocean.

Roseovarius nubinhibens ISM	3.67	63	3547	AALY00000000	13464	Bacteria	Proteobacteria	Surface waters	Caribbean Sea (22° 4 N/74° 4 W)
Roseovarius sp. 217	4.76	60	4772	AAMV00000000	13576	Bacteria	Proteobacteria	Surface waters	Coastal seawater, station L4, off the coast of Plymouth, England, UK
Roseovarius sp. HTCC2601	5.43	66	5452	AATQ00000000	13574	Bacteria	Proteobacteria	Surface, 10 m	Sargasso Sea, Atlantic Ocean, BATS
Roseovarius sp. TM1035	4.21	60	4102	ABCL00000000	19363	Bacteria	Proteobacteria	ca. 5-7 mm	Dinoflagellate culture
Ruegeria sp. R11	3.82	59	3657	ABXM00000000	30725	Bacteria	Proteobacteria	2-4m	Bare Island, Botany Bay, Sydney, Australia (33°59'38" S, 151°16'00"E)
Sagittula stellata E-37	5.26	65	5067	AAYA00000000	18971	Bacteria	Proteobacteria	surface water, 0.25 m	Coastal Georgia, USA
Shewanella benthica KT99	4.35	45	4235	ABIC00000000	13387	Bacteria	Proteobacteria	~9000 m	Tonga-Kermadec Trench
Sphingomonas sp. SKA58	3.95	63	3914	AAQG00000000	13584	Bacteria	Proteobacteria	2-5 m	North Atlantic
Stappia aggregata IAM 12614	6.56	59	6314	AAUW00000000	18315	Bacteria	Proteobacteria	N/A	Damariscotta River, Maine
Stenotrophomonas sp. SKA14	5.02	66	4469	ACDV00000000	19369	Bacteria	Proteobacteria	4 m	Skagerrak (57.57N, 10.44E), 4 m
Sulfitobacter sp. EE-36	3.55	60	3474	AALV00000000	13501	Bacteria	Proteobacteria	Surface waters	Coastal Georgia, USA
Sulfitobacter sp. NAS-14.1	4.00	60	3962	AALZ00000000	13591	Bacteria	Proteobacteria	Surface waters	North Atlantic Ocean (33° 3 N/74° 3 W)
Synechococcus sp. BL107	2.28	54	2507	AATZ00000000	13559	Bacteria	Cyanobacteria	100+ m	Blanes Bay, Mediterranean Sea
Synechococcus sp. PCC 7335	5.96	48	5586	ABRV00000000	19377	Bacteria	Cyanobacteria	N/A	Snail shell, intertidal zone, Puerto Penasco, Mexico
Synechococcus sp. RS9916	2.66	59	2961	AAUA00000000	13557	Bacteria	Cyanobacteria	10 m	Station A, Eilat, Red Sea (29 N/34 E)
Synechococcus sp. RS9917	2.58	64	2770	AANP00000000	13555	Bacteria	Cyanobacteria	10 m	Station A, Eilat, Red Sea (29 N/34 E)
Synechococcus sp. WH 5701	3.04	65	3346	AANO00000000	13554	Bacteria	Cyanobacteria	N/A	Long Island Sound
Synechococcus sp. WH 7805	2.62	57	2883	AAOK00000000	13553	Bacteria	Cyanobacteria	N/A	33 N/67 W
Thermococcus barophilus MP	2.06	41	2146	ABSF00000000	19379	Archaea	Euryarchaeota	3550 m	Chimney Wall, Mid-Atlantic Ridge, Snakepit (23°22'41.18N, 44°56'49.84W)
Thermococcus sp. AM4	2.09	54	2168	ABXN00000000	19381	Archaea	Euryarchaeota	2600 m	East Pacific Rise (13degreesN), 2600 m
unidentified eubacterium SCB49	3.21	34	2948	ABCO00000000	19389	Bacteria	Bacteroidetes	0.5 m	Southern California Bight (32.53N, 117.15W)
Verrucomicrobiae bacterium DG1235	5.77	54	4909	ABSI00000000	19391	Bacteria	Verrucomicrobia	N/A	Isolated in Scotland from a clonal culture of the dinoflagellate Scrippsiella trochoidea. S. trochoidea was isolated in Plymouth harbour, United Kingdom.
Vibrio alginolyticus 12G01	5.16	44	4732	AAPS00000000	13571	Bacteria	Proteobacteria	Surface waters	Plum Island Ecosystem-LTER, USA
Vibrio angustum S14	5.10	39	4558	AAOJ00000000	13567	Bacteria	Proteobacteria	Surface waters	Botany Bay, Sydney, Australia
Vibrio campbellii AND4	4.26	44	3935	ABGR00000000	13564	Bacteria	Proteobacteria	2-5 m	Andaman Sea
Vibrio fischeri MJ11	4.50	38	4039	ABIH00000000	19393	Bacteria	Proteobacteria	20m	Squid light organ, Japan
Vibrio parahaemolyticus 16	4.53	45	4036	ACCV00000000	19395	Bacteria	Proteobacteria	N/A	Tampa Bay; also detected in Hawaii
Vibrio shilonii AK1	5.70	43	5360	ABCH00000000	19397	Bacteria	Proteobacteria	3 m	Bleached coral Mediterranean, 3m
Vibrio sp. MED222	4.89	43	4590	AAND00000000	13616	Bacteria	Proteobacteria	1 m	NW Mediterranean Sea
Vibrio splendidus 12B01	5.60	44	5231	AAMR00000000	13568	Bacteria	Proteobacteria	Surface waters	Plum Island Ecosystem-LTER, USA
Vibrionales bacterium SWAT-3	5.84	44	5408	AAZW00000000	19399	Bacteria	Proteobacteria	surface water	Scripps Pier, La Jolla, CA

(ST2)

Statistics for recruitment of GOS reads to the sequenced marine genomes at various recruitment thresholds. For recruitment threshold of X% match identity, fogX and docX denote fraction of genome covered and depth of coverage, respectively. Genomes have been reverse sorted by doc50. Each genome is categorized as being in the high recruiting (HG) or low recruiting (LG) set; this information is subsequently used in the genomic and functional characterization.

Genomes	HG/LG	fog50	doc50	fog60	doc60	fog70	doc70	fog80	doc80	fog90	doc90
Candidatus Pelagibacter sp. HTCC7211	HG	97.94	531.95	97.83	523.50	97.22	449.50	95.20	205.53	89.89	13.55
Candidatus Pelagibacter ubique HTCC1062	HG	96.91	149.32	96.63	144.57	95.16	107.35	93.30	38.57	89.55	13.21
Candidatus Pelagibacter ubique HTCC1002	HG	94.57	142.69	94.29	138.18	93.17	103.86	91.08	35.71	87.35	12.36
Prochlorococcus marinus str. AS9601	HG	96.49	81.71	96.30	81.65	95.17	81.35	94.26	80.68	92.35	73.23
Prochlorococcus marinus MIT 9301	HG	97.13	60.69	96.98	60.65	95.92	60.42	95.26	59.83	94.18	53.78
Prochlorococcus marinus MIT 9312	HG	94.96	35.56	94.80	35.50	93.71	35.17	91.82	34.07	80.43	19.69
Prochlorococcus marinus str. MIT 9202	HG	93.89	26.53	93.70	26.47	92.82	26.24	91.28	25.68	85.76	20.78
Synechococcus CC9605	HG	86.87	15.03	86.70	15.01	85.78	14.79	83.80	14.03	76.56	10.96
Prochlorococcus marinus MED4	HG	96.27	8.64	96.15	8.58	94.93	8.21	93.66	7.42	90.89	5.44
Prochlorococcus marinus MIT 9215	HG	80.11	5.93	79.93	5.92	78.76	5.83	76.92	5.63	66.89	4.41
Nitrosopumilus maritimus SCM1	HG	64.71	5.74	63.74	5.55	57.35	3.82	27.97	0.92	2.15	0.05
Polaribacter dokdonensis MED152	HG	45.01	5.37	42.27	4.85	24.47	1.92	4.82	0.19	0.63	0.02
Rhodobacterales bacterium HTCC2255	HG	40.43	4.84	39.04	4.32	32.34	1.56	22.93	0.47	9.00	0.13
Methylophilales bacterium HTCC2181	HG	81.67	4.21	79.69	3.83	61.78	2.04	31.58	0.54	27.81	0.43
Flavobacteriales bacterium ALC-1	HG	35.52	3.57	32.73	3.19	16.05	1.28	1.67	0.11	0.25	0.05
Synechococcus sp. WH8102	HG	83.33	3.38	82.99	3.34	81.07	2.95	74.20	2.01	56.94	1.04
Synechococcus WH 7803	HG	66.32	2.94	65.29	2.87	52.23	2.11	15.48	0.49	1.69	0.06
beta proteobacterium KB13	HG	73.70	2.64	71.65	2.32	60.96	1.31	45.76	0.72	29.63	0.40
Rhodobacterales bacterium HTCC2150	HG	35.37	2.50	30.75	2.12	13.38	0.69	2.31	0.06	0.26	0.00
Prochlorococcus marinus str. MIT 9515	HG	68.93	2.37	68.05	2.31	63.49	2.01	52.52	1.37	10.30	0.24
Flavobacteria bacterium BAL38	HG	34.14	2.30	30.68	2.00	13.76	0.63	2.29	0.04	0.27	0.01
marine gamma proteobacterium HTCC2080	HG	56.07	2.22	53.25	2.01	38.90	1.00	15.95	0.24	1.84	0.04
Gammaproteobacterium NOR51-B	HG	35.86	2.18	32.32	1.93	13.61	0.79	1.09	0.04	0.15	0.01
marine gamma proteobacterium HTCC2207	HG	52.29	2.05	49.80	1.87	31.44	0.91	7.30	0.14	1.10	0.03
Rhodobacterales bacterium HTCC2083	HG	36.46	1.91	31.73	1.59	15.83	0.50	3.34	0.05	0.54	0.01
Flavobacteriales bacterium HTCC2170	HG	30.29	1.70	26.25	1.45	9.42	0.39	0.61	0.01	0.26	0.00
Croceibacter atlanticus HTCC2559	HG	30.78	1.67	26.25	1.41	9.52	0.41	0.69	0.01	0.10	0.00
Synechococcus sp. WH 7805	HG	46.41	1.43	45.13	1.38	33.65	0.96	8.32	0.20	1.34	0.03
unidentified eubacterium SCB49	HG	28.03	1.31	24.75	1.14	10.25	0.44	1.39	0.06	0.32	0.01
Kordia algicida OT-1	HG	20.34	1.26	17.98	1.10	7.37	0.42	0.71	0.05	0.18	0.02
Prochlorococcus marinus str. NATL1A	HG	58.29	1.13	57.66	1.10	55.49	0.99	52.22	0.88	32.96	0.46
Prochlorococcus marinus NATL2A	HG	57.24	1.11	56.80	1.09	55.06	1.00	51.56	0.88	31.67	0.43
Roseobacter sp. CCS2	HG	27.59	1.05	23.32	0.85	9.42	0.27	1.22	0.03	0.04	0.00
Alteromonas macleodii 'Deep ecotype'	HG	46.27	1.02	45.51	0.99	41.63	0.88	30.02	0.58	18.19	0.29
Polaribacter irgensii 23-P	LG	28.74	0.96	24.98	0.81	10.65	0.27	1.38	0.05	0.30	0.01
Gramella forsetii KT0803	LG	22.52	0.96	19.02	0.80	5.76	0.18	0.56	0.01	0.06	0.00
Synechococcus RCC307	LG	56.05	0.92	55.78	0.91	54.50	0.87	48.89	0.74	27.31	0.36
alpha proteobacterium BAL199	LG	16.58	0.88	14.12	0.78	5.46	0.30	0.69	0.08	0.12	0.02
Psychroflexus torquis ATCC 700755	LG	18.11	0.84	14.99	0.69	4.88	0.18	0.32	0.01	0.02	0.00
Silicibacter pomeroyi DSS-3	LG	22.94	0.73	20.61	0.65	10.64	0.28	1.34	0.03	0.23	0.01
Roseovarius nubinhibens ISM	LG	21.79	0.72	19.36	0.65	9.80	0.34	2.06	0.09	1.32	0.06
Synechococcus sp. RS9916	LG	27.15	0.71	26.25	0.70	21.34	0.59	11.53	0.29	4.20	0.07
Alteromonadales bacterium TW-7	LG	45.34	0.70	44.53	0.67	41.60	0.59	38.11	0.51	35.69	0.47
Roseobacter sp. SK209-2-6	LG	19.64	0.69	16.35	0.57	6.83	0.19	0.93	0.03	0.31	0.02
Roseobacter litoralis Och 149	LG	20.83	0.68	17.65	0.56	7.12	0.17	0.50	0.01	0.01	0.00
Roseobacter sp. MED193	LG	19.38	0.67	16.42	0.57	7.08	0.21	0.95	0.02	0.12	0.00
Algoriphagus sp. PR1	LG	15.85	0.66	13.51	0.56	3.99	0.11	0.32	0.01	0.00	0.00
Rhodobacterales KLH11	LG	17.81	0.64	15.72	0.54	6.80	0.20	0.88	0.02	0.14	0.01
Flavobacteria bacterium BBFL7	LG	18.07	0.62	14.89	0.51	4.34	0.12	0.26	0.00	0.00	0.00
Caminibacter mediatlanticus TB-2	LG	10.28	0.61	8.08	0.49	1.58	0.05	0.20	0.00	0.00	0.00
Ruegeria sp. R11	LG	20.72	0.61	18.17	0.53	8.71	0.26	1.95	0.06	0.82	0.02
Loktanella vestfoldensis SKA53	LG	23.67	0.60	20.18	0.52	8.63	0.20	0.74	0.01	0.05	0.00
marine gamma proteobacterium HTCC2143	LG	17.17	0.58	13.94	0.46	3.15	0.08	0.39	0.01	0.12	0.00
Roseobacter denitrificans Och 114	LG	20.81	0.57	17.95	0.48	7.70	0.17	0.69	0.01	0.10	0.00
Erythrobacter sp. SD-21	LG	35.11	0.57	34.23	0.55	28.59	0.45	12.39	0.18	2.13	0.03
Synechococcus sp. RS9917	LG	23.18	0.57	22.46	0.55	17.13	0.43	6.52	0.15	0.39	0.01
Roseovarius sp. TM1035	LG	20.41	0.56	17.62	0.48	7.32	0.17	0.88	0.02	0.36	0.01
Roseobacter sp. AzwK-3b	LG	20.58	0.55	18.02	0.46	7.67	0.18	1.01	0.02	0.31	0.01
Gammaproteobacterium HTCC2148	LG	19.32	0.52	16.09	0.43	5.40	0.12	0.97	0.02	0.26	0.00
Limnobacter sp. MED105	LG	33.19	0.50	32.65	0.49	31.60	0.46	29.51	0.42	9.00	0.12
Alcanivorax sp. DG881	LG	29.46	0.50	28.53	0.47	21.32	0.33	7.92	0.12	0.76	0.01
Roseobacter GAI101	LG	18.53	0.48	15.76	0.40	6.77	0.16	0.73	0.01	0.11	0.00
Octadecabacter antarcticus 307	LG	16.04	0.44	13.10	0.35	5.11	0.10	0.73	0.01	0.23	0.01
Silicibacter TM1040	LG	17.33	0.43	14.80	0.37	7.40	0.15	2.20	0.04	0.57	0.01
Synechococcus CC9311	LG	26.51	0.43	25.64	0.41	22.87	0.35	18.70	0.26	7.26	0.10
Phaeobacter gallaeciensis BS107	LG	16.32	0.42	14.09	0.36	6.58	0.15	1.25	0.03	0.50	0.01
Phaeobacter gallaeciensis 2.10	LG	15.77	0.41	13.80	0.35	6.58	0.16	1.01	0.03	0.18	0.01
Pseudovibrio sp. JE062	LG	8.05	0.40	6.22	0.35	1.78	0.21	0.87	0.19	0.38	0.01
Oceanibulbus indolifex HEL-45	LG	17.13	0.40	14.64	0.33	6.08	0.12	1.16	0.02	0.25	0.00

Octadecabacter antarcticus 238	LG	14.73	0.40	12.26	0.33	5.22	0.11	1.00	0.01	0.13	0.00
Roseovarius sp. 217	LG	16.49	0.39	13.90	0.32	5.66	0.11	0.63	0.01	0.12	0.00
Sphingomonas sp. SKA58	LG	26.87	0.38	26.15	0.37	22.99	0.31	17.12	0.22	8.92	0.11
Flavobacterium sp. MED217	LG	14.97	0.37	12.11	0.30	3.95	0.10	0.72	0.03	0.26	0.01
Synechococcus CC9902	LG	23.66	0.37	22.84	0.35	19.50	0.29	15.97	0.22	13.66	0.18
Alcanivorax borkumensis SK2	LG	27.90	0.36	26.51	0.34	22.98	0.28	21.08	0.25	20.21	0.23
Congregibacter litoralis KT71	LG	15.12	0.35	13.13	0.30	5.40	0.12	0.94	0.02	0.59	0.01
Dokdonia donghaensis MED134	LG	15.43	0.32	12.20	0.25	3.72	0.07	0.70	0.01	0.36	0.01
Pseudoalteromonas haloplanktis TAC125	LG	17.98	0.32	16.56	0.28	11.62	0.18	4.47	0.09	1.31	0.05
Hoeflea phototrophica DFL-43	LG	10.45	0.31	8.87	0.27	3.39	0.08	0.46	0.01	0.08	0.00
Sulfitobacter sp. EE-36	LG	15.15	0.31	12.98	0.26	5.07	0.10	0.66	0.01	0.23	0.00
Sulfitobacter sp. NAS-14.1	LG	14.80	0.30	12.72	0.25	5.39	0.11	0.97	0.01	0.29	0.00
Marinobacter aquaeolei VT8	LG	20.63	0.30	19.66	0.27	16.16	0.20	11.07	0.13	6.88	0.08
Gammaproteobacterium NOR5-3	LG	13.16	0.29	11.33	0.25	4.20	0.09	0.81	0.01	0.15	0.00
Sphingopyxis alaskensis RB2256	LG	20.05	0.28	19.39	0.27	16.09	0.21	9.47	0.12	2.11	0.03
Synechococcus sp. BL107	LG	15.80	0.28	14.78	0.26	11.69	0.18	7.13	0.09	4.75	0.06
Rhodobacterales bacterium Y4I	LG	12.55	0.27	11.01	0.24	4.95	0.10	0.97	0.02	0.42	0.01
Pseudoalteromonas tunicata D2	LG	6.86	0.27	5.54	0.24	1.72	0.16	0.44	0.10	0.34	0.04
Robiginitalea biformata HTCC2501	LG	10.73	0.26	8.97	0.23	2.74	0.06	0.17	0.00	0.04	0.00
Dinoroseobacter shibae DFL 12	LG	11.93	0.25	10.17	0.22	4.51	0.09	0.61	0.02	0.12	0.01
Marinomonas MWYL1	LG	8.50	0.25	6.65	0.17	1.45	0.02	0.45	0.01	0.39	0.00
Oceanospirillum sp. MED92	LG	9.39	0.22	6.91	0.15	1.44	0.02	0.20	0.00	0.06	0.00
Glaciecola sp. HTCC2999	LG	12.15	0.21	9.84	0.17	4.64	0.06	2.17	0.03	1.72	0.02
Marinomonas sp. MED121	LG	8.05	0.21	6.04	0.15	1.00	0.02	0.31	0.01	0.24	0.00
Erythrobacter litoralis HTCC2594	LG	14.52	0.21	13.78	0.20	9.60	0.13	4.25	0.06	2.04	0.03
Jannaschia CCS1	LG	9.41	0.21	7.69	0.16	2.68	0.04	0.22	0.00	0.02	0.00
Thermosipho africanus TCF52B	LG	5.17	0.20	3.58	0.15	0.40	0.01	0.00	0.00	0.00	0.00
Brevundimonas sp. BAL3	LG	15.21	0.20	14.64	0.19	12.52	0.16	7.32	0.09	2.16	0.03
Roseovarius sp. HTCC2601	LG	9.30	0.19	7.89	0.16	3.48	0.09	1.05	0.05	0.63	0.04
Prochlorococcus marinus CCMP1375	LG	8.82	0.19	7.63	0.16	3.25	0.05	0.73	0.01	0.07	0.00
Prochlorococcus marinus str. MIT 9211	LG	5.92	0.18	4.87	0.16	2.00	0.05	0.33	0.00	0.06	0.00
Rhodobacterales bacterium HTCC2654	LG	8.11	0.17	6.39	0.14	2.43	0.05	0.68	0.01	0.41	0.01
Sagittula stellata E-37	LG	7.49	0.16	6.26	0.14	2.32	0.05	0.34	0.01	0.07	0.00
Stappia aggregata IAM 12614	LG	7.01	0.15	5.72	0.13	1.86	0.04	0.31	0.01	0.02	0.00
Oceanicola batsensis HTCC2597	LG	8.17	0.15	7.01	0.12	3.01	0.04	0.90	0.01	0.28	0.00
Erythrobacter sp. NAP1	LG	10.83	0.15	9.88	0.13	6.02	0.08	1.67	0.02	0.30	0.00
Oceanobacter sp. RED65	LG	7.92	0.14	5.78	0.10	1.09	0.01	0.20	0.00	0.07	0.00
Oceanicola granulosus HTCC2516	LG	7.37	0.13	6.39	0.11	2.47	0.04	0.73	0.01	0.19	0.00
Maricaulis maris MCS10	LG	6.78	0.13	5.68	0.10	2.74	0.05	0.90	0.02	0.08	0.00
Saccharophagus degradans 2-40	LG	6.62	0.13	5.27	0.10	1.46	0.02	0.19	0.00	0.02	0.00
Marinobacter algicola DG893	LG	8.47	0.12	7.24	0.10	3.42	0.04	1.45	0.02	0.39	0.00
marine actinobacterium PHSC20C1	LG	6.82	0.10	5.68	0.08	1.98	0.03	0.24	0.00	0.07	0.00
Oceanicaulis alexandrii HTCC2633	LG	6.80	0.10	5.64	0.08	2.90	0.04	1.94	0.02	1.71	0.02
Campylobacterales bacterium GD 1	LG	3.04	0.10	2.32	0.08	0.75	0.01	0.19	0.00	0.10	0.00
Verrucomicrobiae bacterium DG1235	LG	5.28	0.10	4.35	0.08	1.16	0.02	0.08	0.00	0.00	0.00
Nitrosococcus oceani ATCC 19707	LG	5.45	0.10	3.83	0.07	0.63	0.01	0.14	0.00	0.00	0.00
Nitrosococcus oceani C-27	LG	5.37	0.10	3.81	0.07	0.60	0.01	0.12	0.00	0.00	0.00
Psychromonas ingrahamii 37	LG	4.71	0.10	3.34	0.07	0.51	0.01	0.07	0.00	0.00	0.00
Idiomarina baltica OS145	LG	7.36	0.10	6.22	0.08	3.96	0.05	3.30	0.04	3.08	0.04
Aurantimonas sp. SI85-9A1	LG	5.48	0.10	4.73	0.08	2.05	0.04	0.49	0.01	0.22	0.01
Microscilla marina ATCC 23134	LG	3.47	0.09	2.66	0.07	0.56	0.02	0.12	0.01	0.00	0.00
Marinobacter sp. ELB17	LG	5.57	0.09	4.37	0.07	1.46	0.02	0.23	0.00	0.00	0.00
Hyphomonas neptunium ATCC 15444	LG	5.53	0.09	4.56	0.07	1.58	0.02	0.26	0.01	0.00	0.00
Petrogoga mobilis SJ95	LG	2.65	0.08	1.33	0.06	0.08	0.00	0.00	0.00	0.00	0.00
Lentisphaera araneosa HTCC2155	LG	4.38	0.08	3.41	0.06	0.70	0.01	0.19	0.00	0.08	0.00
Reinekea sp. MED297	LG	4.20	0.08	3.43	0.06	0.56	0.01	0.03	0.00	0.00	0.00
Pedobacter sp. BAL39	LG	4.33	0.08	3.31	0.06	0.66	0.01	0.28	0.00	0.06	0.00
Psychromonas sp. CNPT3	LG	3.82	0.08	2.60	0.05	0.31	0.00	0.05	0.00	0.02	0.00
Moritella sp. PE36	LG	3.50	0.08	2.39	0.05	0.30	0.01	0.05	0.00	0.02	0.00
gamma proteobacterium HTCC5015	LG	4.85	0.07	3.56	0.05	1.23	0.02	0.51	0.01	0.33	0.00
Shewanella frigidimarina NCIMB 400	LG	3.93	0.07	2.73	0.04	0.76	0.01	0.33	0.00	0.21	0.00
Blastopirellula marina DSM 3645	LG	3.82	0.07	3.21	0.06	0.65	0.01	0.07	0.00	0.01	0.00
Aciduliprofundum boonei T469	LG	2.06	0.07	1.34	0.05	0.23	0.03	0.03	0.00	0.00	0.00
Shewanella benthica KT99	LG	1.36	0.07	0.76	0.06	0.04	0.05	0.02	0.05	0.02	0.05
Cyanobium sp. PCC 7001	LG	4.20	0.06	4.00	0.06	2.58	0.04	1.06	0.01	0.14	0.00
Pirellula sp	LG	3.66	0.06	3.23	0.06	0.68	0.01	0.03	0.00	0.00	0.00
Photobacterium profundum 3TCK	LG	2.43	0.06	1.53	0.05	0.18	0.02	0.02	0.01	0.02	0.01
Photobacterium sp. SKA34	LG	3.21	0.06	1.93	0.04	0.40	0.00	0.11	0.00	0.04	0.00
Vibrio fischeri MJ11	LG	3.19	0.06	2.24	0.04	0.47	0.01	0.04	0.00	0.02	0.00
Vibrio angustum S14	LG	2.87	0.06	2.06	0.04	0.59	0.01	0.35	0.00	0.27	0.00

<i>Prochlorococcus marinus</i> str. MIT 9303	LG	3.49	0.05	2.95	0.05	1.30	0.02	0.53	0.01	0.14	0.00
<i>Planctomyces maris</i> DSM 8797	LG	2.88	0.05	2.19	0.04	0.25	0.00	0.04	0.00	0.00	0.00
<i>Mariprofundus ferrooxydans</i> PV-1	LG	2.15	0.05	1.55	0.04	0.36	0.01	0.04	0.00	0.00	0.00
<i>Trichodesmium erythraeum</i> IMS101	LG	1.68	0.05	1.28	0.04	0.59	0.02	0.12	0.00	0.04	0.00
<i>Prosthecochloris aestuarii</i> DSM 271	LG	2.08	0.05	1.51	0.04	0.27	0.01	0.07	0.00	0.03	0.00
<i>Cyanothece</i> sp. CCY0110	LG	1.48	0.05	1.12	0.04	0.45	0.02	0.14	0.01	0.02	0.01
<i>Shewanella woodyi</i> ATCC 51908	LG	2.68	0.04	1.78	0.03	0.34	0.00	0.05	0.00	0.00	0.00
<i>Vibrio shilonii</i> AK1	LG	2.34	0.04	1.63	0.03	0.38	0.00	0.15	0.00	0.04	0.00
<i>Parvularcula bermudensis</i> HTCC2503	LG	2.97	0.04	2.02	0.03	0.54	0.01	0.25	0.00	0.15	0.00
<i>Stenotrophomonas</i> sp. SKA14	LG	3.80	0.04	3.64	0.04	3.11	0.04	2.69	0.03	1.90	0.02
<i>Caldanaerobacter</i> DSM 12653	LG	2.58	0.04	1.46	0.02	0.28	0.01	0.20	0.00	0.15	0.00
<i>Chlorobium phaeobacteroides</i> BS1	LG	2.17	0.04	1.60	0.03	0.38	0.01	0.09	0.00	0.00	0.00
<i>Prochlorococcus marinus</i> MIT9313	LG	2.96	0.04	2.61	0.04	1.09	0.01	0.34	0.00	0.19	0.00
<i>Shewanella</i> MR-4	LG	3.84	0.04	3.49	0.04	3.17	0.03	2.98	0.03	2.43	0.03
<i>Photobacterium profundum</i> SS9	LG	2.13	0.04	1.43	0.03	0.26	0.01	0.01	0.00	0.00	0.00
<i>Shewanella</i> MR-7	LG	3.54	0.04	3.40	0.04	2.95	0.03	2.73	0.03	2.17	0.02
<i>Nitrococcus mobilis</i> Nb-231	LG	2.25	0.04	1.58	0.03	0.31	0.01	0.05	0.00	0.02	0.00
<i>Janibacter</i> sp. HTCC2649	LG	2.21	0.04	1.98	0.03	0.93	0.02	0.14	0.01	0.00	0.00
<i>Magnetococcus</i> MC-1	LG	1.71	0.04	1.31	0.03	0.32	0.01	0.07	0.00	0.01	0.00
<i>Rhodopseudomonas palustris</i> TIE 1	LG	2.64	0.04	2.21	0.03	0.90	0.01	0.31	0.01	0.08	0.00
<i>Vibrio parahaemolyticus</i> 16	LG	2.90	0.04	2.45	0.03	1.35	0.01	0.37	0.00	0.05	0.00
<i>Synechococcus</i> sp. WH 5701	LG	2.60	0.04	2.35	0.03	1.33	0.02	0.48	0.01	0.08	0.00
<i>Carnobacterium</i> sp. AT7	LG	2.36	0.03	1.40	0.02	0.21	0.00	0.10	0.00	0.00	0.00
<i>Vibrionales bacterium</i> SWAT-3	LG	3.01	0.03	2.66	0.03	1.97	0.02	1.60	0.02	1.27	0.01
<i>Hydrogenivirga</i> sp. 128-5-R1-1	LG	1.27	0.03	0.82	0.02	0.05	0.00	0.00	0.00	0.00	0.00
<i>Nitrobacter</i> sp. Nb-311A	LG	2.04	0.03	1.66	0.02	0.49	0.01	0.12	0.00	0.03	0.00
<i>Fulvimarina pelagi</i> HTCC2506	LG	2.08	0.03	1.24	0.02	0.30	0.00	0.04	0.00	0.00	0.00
<i>Cyanothece</i> ATCC 51142	LG	1.28	0.03	1.01	0.02	0.30	0.00	0.06	0.00	0.00	0.00
<i>Plesiocystis pacifica</i> SIR-1	LG	1.34	0.02	1.18	0.02	0.35	0.00	0.05	0.00	0.00	0.00
<i>Shewanella denitrificans</i> OS217	LG	1.41	0.02	0.77	0.01	0.12	0.00	0.00	0.00	0.00	0.00
<i>Vibrio alginolyticus</i> 12G01	LG	1.78	0.02	1.19	0.02	0.49	0.01	0.14	0.00	0.11	0.00
<i>Vibrio splendidus</i> 12B01	LG	1.64	0.02	1.10	0.02	0.36	0.00	0.13	0.00	0.03	0.00
<i>Bacillus</i> sp. B14905	LG	1.33	0.02	0.74	0.01	0.08	0.00	0.00	0.00	0.00	0.00
<i>Vibrio campbellii</i> AND4	LG	1.52	0.02	1.03	0.01	0.62	0.01	0.32	0.00	0.02	0.00
<i>Shewanella loihica</i> PV-4	LG	1.77	0.02	1.35	0.02	0.34	0.00	0.13	0.00	0.00	0.00
<i>Shewanella baltica</i> OS185	LG	1.79	0.02	1.49	0.02	1.18	0.01	1.02	0.01	0.93	0.01
<i>Chlorobaculum parvum</i> NCIB 8327	LG	1.23	0.02	0.90	0.02	0.20	0.00	0.00	0.00	0.00	0.00
<i>Shewanella baltica</i> OS195	LG	1.74	0.02	1.49	0.02	1.16	0.01	1.09	0.01	0.95	0.01
<i>Synechococcus</i> sp. PCC 7335	LG	0.82	0.02	0.59	0.01	0.19	0.01	0.04	0.00	0.00	0.00
<i>Shewanella baltica</i> OS155	LG	1.51	0.02	1.20	0.01	0.98	0.01	0.84	0.01	0.69	0.01
<i>Nodularia spumigena</i> CCY9414	LG	0.85	0.02	0.73	0.01	0.32	0.01	0.08	0.00	0.01	0.00
<i>Acaryochloris marina</i> MBIC11017	LG	0.66	0.02	0.48	0.01	0.19	0.01	0.08	0.00	0.00	0.00
<i>Bacillus</i> sp. SG-1	LG	1.02	0.02	0.65	0.01	0.24	0.00	0.15	0.00	0.02	0.00
<i>Vibrio</i> sp. MED222	LG	1.25	0.02	0.84	0.01	0.25	0.00	0.04	0.00	0.01	0.00
<i>Shewanella baltica</i> OS223	LG	1.27	0.02	1.17	0.01	0.82	0.01	0.74	0.01	0.64	0.01
<i>Thermococcus barophilus</i> MP	LG	0.75	0.01	0.43	0.01	0.00	0.00	0.00	0.00	0.00	0.00
<i>Bacillus</i> sp. NRRL B-14911	LG	0.63	0.01	0.40	0.01	0.16	0.00	0.00	0.00	0.00	0.00
<i>Lyngbya</i> sp. PCC 8106	LG	0.76	0.01	0.41	0.01	0.07	0.00	0.00	0.00	0.00	0.00
<i>Synechococcus elongatus</i> PCC 7942	LG	0.79	0.01	0.63	0.01	0.13	0.00	0.08	0.00	0.00	0.00
<i>Microcoleus chthonoplastes</i> PCC 7420	LG	0.55	0.01	0.38	0.01	0.10	0.00	0.02	0.00	0.00	0.00
<i>Geobacillus kaustophilus</i> HTA426	LG	0.56	0.01	0.34	0.01	0.12	0.00	0.01	0.00	0.01	0.00
<i>Thermococcus</i> sp. AM4	LG	0.43	0.01	0.31	0.01	0.01	0.00	0.00	0.00	0.00	0.00
<i>Microcystis aeruginosa</i> NIES 843	LG	0.59	0.01	0.38	0.01	0.08	0.00	0.01	0.00	0.00	0.00
<i>Haloarcula marismortui</i> ATCC 43049	LG	0.51	0.01	0.28	0.00	0.02	0.00	0.00	0.00	0.00	0.00
<i>Desulfovibrio vulgaris</i> Miyazaki F	LG	0.60	0.01	0.50	0.01	0.19	0.00	0.00	0.00	0.00	0.00
<i>Marinitoga piezophila</i> KA3	LG	0.46	0.01	0.12	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<i>Thermotoga</i> RQ2	LG	0.54	0.01	0.30	0.00	0.04	0.00	0.00	0.00	0.00	0.00
<i>Vibrio harveyi</i> ATCC BAA-1116	LG	0.37	0.00	0.30	0.00	0.15	0.00	0.11	0.00	0.07	0.00

(ST3)

All protein categories that are identified as being differentially distributed across HG and LG sets in both original data (i.e. using raw counts) and normalized data (i.e. using genome size). The Wilcoxon Rank Sum test statistic was used to assess differential distribution of a protein category and a rejection threshold was calculated after correction for multiple testing with the false discovery rate set to 0.005. There are 568 protein categories identified as differentially distributed and these are organized into higher order functional groups - transcriptional regulation, transport, metabolism, biosynthesis, motility, chemotaxis, secretion, degradation, photosynthesis, and repair and replication. The table lists the p-values of the Wilcoxon Rank Sum test for both original and normalized data, along with the slopes of the trend lines. The graphical representations of these protein categories are given in the supplementary figures SF3.

<i>Description</i>	<i>Category</i>	<i>P-value for HG vs. LG comparison with original data</i>	<i>Slope of trend line for original data</i>	<i>P-value for HG vs. LG comparison with normalized data</i>	<i>Slope of trend line for normalized data</i>
Transcriptional regulation					
[T] COG0642 Signal transduction histidine kinase	COG	7.27E-11	Negative	1.95E-09	Negative
[K] COG0583 Transcriptional regulator	COG	5.71E-07	Negative	3.86E-05	Negative
[T] COG2199 FOG: GGDEF domain	COG	5.79E-09	Negative	1.22E-08	Negative
[TK] COG0745 Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain	COG	1.44E-10	Negative	1.07E-05	Negative
[T] COG2204 Response regulator containing CheY-like receiver, AAA-type ATPase, and DNA-binding domains	COG	3.09E-08	Negative	1.17E-05	Negative
[T] COG5001 Predicted signal transduction protein containing a membrane domain, an EAL and a GGDEF domain	COG	7.98E-07	Negative	1.88E-06	Negative
[T] COG3706 Response regulator containing a CheY-like receiver domain and a GGDEF domain	COG	1.95E-10	Negative	4.21E-10	Negative
[T] COG2200 FOG: EAL domain	COG	5.21E-09	Negative	1.72E-08	Negative
[K] COG1309 Transcriptional regulator	COG	4.29E-07	Negative	9.95E-06	Negative
[T] COG4191 Signal transduction histidine kinase regulating C4-dicarboxylate transport system	COG	1.10E-07	Negative	7.55E-07	Negative
[T] COG2206 HD-GYP domain	COG	5.29E-06	Negative	1.03E-05	Negative
[K] COG1846 Transcriptional regulators	COG	1.68E-06	Negative	1.19E-05	Negative
[KT] COG3437 Response regulator containing a CheY-like receiver domain and an HD-GYP domain	COG	2.24E-04	Negative	4.04E-04	Negative
[K] COG0789 Predicted transcriptional regulators	COG	1.67E-07	Negative	1.85E-06	Negative
[T] COG1639 Predicted signal transduction protein	COG	1.64E-04	Negative	1.56E-04	Negative
[T] COG0631 Serine/threonine protein phosphatase	COG	5.44E-05	Negative	1.36E-04	Negative
[KE] COG1167 Transcriptional regulators containing a DNA-binding HTH domain and an aminotransferase domain (MocR family) and their eukaryotic orthologs	COG	5.63E-05	Negative	1.70E-04	Negative
[KG] COG1349 Transcriptional regulators of sugar metabolism	COG	2.86E-06	Negative	8.70E-06	Negative
[T] COG2905 Predicted signal-transduction protein containing cAMP-binding and CBS domains	COG	2.36E-07	Negative	6.55E-05	Negative
[K] COG2186 Transcriptional regulators	COG	5.24E-05	Negative	1.55E-04	Negative
[K] COG1737 Transcriptional regulators	COG	6.75E-05	Negative	3.29E-04	Negative

[K] COG4977 Transcriptional regulator containing an amidase domain and an AraC-type DNA-binding HTH domain	COG	2.04E-05	Negative	1.39E-04	Negative
[O] COG1764 Predicted redox protein, regulator of disulfide bond formation	COG	1.99E-05	Negative	1.19E-04	Negative
[T] COG4566 Response regulator	COG	1.89E-04	Negative	2.52E-04	Negative
[K] COG2183 Transcriptional accessory protein	COG	5.75E-06	Negative	6.81E-04	Negative
[K] COG1802 Transcriptional regulators	COG	1.16E-05	Positive	2.24E-05	Positive
PF00072 Response_reg: response regulator receiver domain	HMM	2.36E-10	Negative	1.20E-08	Negative
PF02518 HATPase_c: ATPase, histidine kinase-, DNA gyrase B-, and HSP90-like domain protein	HMM	3.89E-10	Negative	1.31E-07	Negative
PF00990 GGDEF: GGDEF domain	HMM	3.64E-10	Negative	8.99E-10	Negative
TIGR00254 GGDEF: diguanylate cyclase (GGDEF) domain	HMM	4.04E-10	Negative	9.18E-10	Negative
PF00512 HisKA: His Kinase A (phosphoacceptor) domain	HMM	5.43E-10	Negative	1.15E-07	Negative
PF00672 HAMP: HAMP domain	HMM	1.23E-11	Negative	7.98E-09	Negative
TIGR00229 sensory_box: PAS domain S-box	HMM	1.49E-13	Negative	1.32E-12	Negative
PF00583 Acetyltransf_1: acetyltransferase, GNAT family	HMM	2.00E-07	Negative	3.19E-04	Negative
PF00126 HTH_1: transcriptional regulator, LysR family	HMM	1.03E-07	Negative	9.66E-06	Negative
PF08448 PAS_4: PAS fold	HMM	8.12E-15	Negative	3.60E-14	Negative
PF00989 PAS: PAS fold	HMM	4.11E-12	Negative	2.28E-10	Negative
PF03466 LysR_substrate: LysR substrate binding domain	HMM	6.85E-07	Negative	5.35E-05	Negative
PF00563 EAL: cyclic diguanylate phosphodiesterase (EAL) domain	HMM	6.35E-09	Negative	1.25E-08	Negative
PF08447 PAS_3: PAS fold	HMM	1.00E-11	Negative	1.01E-10	Negative
PF01590 GAF: GAF domain	HMM	4.50E-09	Negative	5.22E-06	Negative
PF00486 Trans_reg_C: transcriptional regulatory protein, C-terminal	HMM	4.00E-08	Negative	7.66E-04	Negative
PF08281 Sigma70_r4_2: Sigma-70, region 4	HMM	3.06E-05	Negative	7.58E-04	Negative
PF00440 TetR_N: transcriptional regulator, TetR family	HMM	4.64E-08	Negative	1.15E-06	Negative
PF02954 HTH_8: transcriptional regulator, Fis family	HMM	1.58E-06	Negative	3.47E-05	Negative
PF00165 HTH_AraC: transcriptional regulator, AraC family	HMM	2.52E-05	Negative	5.54E-04	Negative
PF01381 HTH_3: Helix-turn-helix	HMM	8.15E-08	Negative	3.36E-05	Negative
PF01627 Hpt: Hpt domain	HMM	6.93E-11	Negative	5.03E-10	Negative
PF08279 HTH_11: HTH domain	HMM	1.12E-10	Negative	3.01E-09	Negative
PF01047 MarR: transcriptional regulator, MarR family	HMM	4.54E-07	Negative	1.03E-05	Negative
PF00392 GntR: transcriptional regulator, GntR family	HMM	2.79E-06	Negative	3.98E-04	Negative

PF08659 KR: KR domain	HMM	1.33E-06	Negative	1.56E-04	Positive
TIGR01509 HAD-SF-IA-v3: HAD-superfamily hydrolase, subfamily IA, variant 3	HMM	1.77E-08	Negative	1.50E-04	Negative
PF00376 MerR: transcriptional regulator, MerR family	HMM	3.34E-08	Negative	6.65E-06	Negative
PF08220 HTH_DeoR: DeoR-like helix-turn-helix domain	HMM	1.88E-07	Negative	6.09E-07	Negative
PF04967 HTH_10: HTH DNA binding domain	HMM	5.25E-05	Negative	1.35E-04	Negative
PF00176 SNF2_N: SNF2 family N-terminal domain	HMM	2.40E-05	Negative	4.23E-04	Negative
PF01965 DJ-1_Pfpl: DJ-1/Pfpl family	HMM	2.14E-07	Negative	1.60E-05	Negative
PF08269 Cache_2: Cache domain	HMM	2.52E-05	Negative	3.08E-05	Negative
PF01464 SLT: Transglycosylase SLT domain	HMM	3.88E-07	Negative	5.68E-05	Negative
PF02895 H-kinase_dim: Signal transducing histidine kinase, homodimeric domain	HMM	3.80E-09	Negative	7.95E-09	Negative
PF02082 Rrf2: Transcriptional regulator	HMM	1.28E-08	Negative	4.88E-04	Negative
PF03459 TOBE: TOBE domain	HMM	3.03E-07	Negative	1.83E-06	Negative
PF09278 MerR-DNA-bind: MerR, DNA binding	HMM	9.89E-08	Negative	5.22E-07	Negative
PF03929 PepSY_TM: PepSY-associated TM helix	HMM	1.78E-05	Negative	1.72E-04	Negative
PF00455 DeoR: transcriptional regulator, DeoR family	HMM	1.22E-06	Negative	3.83E-06	Negative
PF02311 AraC_binding: AraC-like ligand binding domain	HMM	2.22E-04	Negative	3.20E-04	Negative
PF01418 HTH_6: transcriptional regulator, RpiR family	HMM	2.85E-05	Negative	7.29E-05	Negative
PF00771 FHIPEP: FHIPEP family	HMM	1.35E-07	Negative	2.06E-06	Negative
PF09383 NIL: NIL domain	HMM	3.10E-04	Negative	3.43E-04	Negative
PF03924 CHASE: CHASE domain	HMM	5.84E-06	Negative	6.62E-06	Negative
PF04316 FlgM: anti-sigma-28 factor, FlgM	HMM	4.26E-04	Negative	6.66E-04	Negative
PF01706 FliG_C: FliG C-terminal domain	HMM	2.42E-08	Negative	8.71E-07	Negative
TIGR01460 HAD-SF-IIA: HAD-superfamily hydrolase, subfamily IIA	HMM	7.42E-05	Negative	1.38E-04	Negative
PF05170 AsmA: AsmA family	HMM	1.73E-04	Negative	7.12E-04	Negative
PF08298 AAA_PrkA: PrkA AAA domain	HMM	1.67E-04	Negative	2.39E-04	Negative
PF04293 SpoVR: SpoVR like protein	HMM	2.04E-04	Negative	2.86E-04	Negative
TIGR01419 nitro_reg_IIA: PTS IIA-like nitrogen-regulatory protein PtsN	HMM	3.48E-04	Negative	4.11E-04	Negative
PF04066 MrpF_PhaF: Multiple resistance and pH regulation protein F (MrpF / PhaF)	HMM	4.78E-05	Negative	1.87E-04	Negative
PF05992 SbmA_BacA: SbmA/BacA-like family	HMM	7.24E-05	Positive	1.56E-06	Positive
PF09084 NMT1: NMT1/THI5 like	HMM	2.29E-05	Negative	4.10E-05	Negative
Signal Transduction Two-component system [PATH:ko02020]	KEGG pathway	1.75E-10	Negative	2.13E-05	Negative

Transport in

[E] COG0531 Amino acid transporters	COG	3.90E-06	Negative	2.22E-04	Negative
[G] COG1879 ABC-type sugar transport system, periplasmic component	COG	2.02E-05	Negative	3.80E-05	Negative
[E] COG4608 ABC-type oligopeptide transport system, ATPase component	COG	6.99E-05	Negative	9.21E-05	Negative
[GT] COG1762 Phosphotransferase system mannitol/fructose-specific IIA domain (Ntr-type)	COG	9.56E-07	Negative	6.06E-05	Negative
[P] COG0598 Mg ²⁺ and Co ²⁺ transporters	COG	2.99E-06	Negative	3.26E-04	Negative
[G] COG1925 Phosphotransferase system, HPr-related proteins	COG	1.10E-06	Negative	9.11E-05	Negative
[M] COG1292 Choline-glycine betaine transporter	COG	8.87E-05	Negative	6.85E-04	Negative
[G] COG1129 ABC-type sugar transport system, ATPase component	COG	4.48E-04	Negative	6.86E-04	Negative
PF00532 Peripla_BP_1: Periplasmic binding proteins and sugar binding domain of the LacI family	HMM	6.66E-05	Negative	2.70E-04	Negative
TIGR01494 ATPase_P-type: ATPase, P-type (transporting), HAD superfamily, subfamily IC	HMM	3.51E-11	Negative	7.69E-08	Negative
PF08352 oligo_HPY: Oligopeptide/dipeptide transporter, C-terminal region	HMM	7.41E-07	Negative	1.70E-05	Negative
TIGR01727 oligo_HPY: oligopeptide/dipeptide ABC transporter, ATP-binding protein, C-terminal domain	HMM	1.24E-06	Negative	4.39E-06	Negative
PF00497 SBP_bac_3: bacterial extracellular solute-binding protein, family 3	HMM	7.84E-07	Negative	5.58E-05	Negative
PF00391 PEP-utilizers: PEP-utilizing enzyme, mobile domain	HMM	9.90E-11	Negative	9.61E-05	Negative
PF00359 PTS_EIIA_2: phosphoenolpyruvate-dependent sugar phosphotransferase system, EIIA 2	HMM	1.50E-07	Negative	1.57E-05	Negative
TIGR01297 CDF: cation diffusion facilitator family transporter	HMM	1.08E-10	Negative	9.09E-08	Negative
PF03480 SBP_bac_7: Bacterial extracellular solute-binding protein, family 7	HMM	4.16E-05	Positive	3.80E-04	Positive
PF00324 AA_permease: Amino acid permease	HMM	4.66E-06	Negative	4.85E-04	Negative
PF02896 PEP-utilizers_C: PEP-utilizing enzyme, TIM barrel domain	HMM	2.46E-09	Negative	7.76E-04	Negative
TIGR00933 2a38: potassium uptake protein, TrkH family	HMM	1.88E-04	Negative	3.74E-04	Negative
PF00381 PTS-HPr: PTS HPr component phosphorylation site	HMM	2.42E-06	Negative	1.53E-04	Negative
TIGR01003 PTS_HPr_family: phosphocarrier, HPr family	HMM	7.91E-07	Negative	4.08E-05	Negative

TIGR01511 ATPase-IB1_Cu: copper-translocating P-type ATPase	HMM	4.39E-09	Negative	5.67E-04	Negative
TIGR00842 bcct: transporter, betaine/carnitine/choline transporter (BCCT) family	HMM	4.31E-06	Negative	1.63E-05	Negative
TIGR01417 PTS_I_fam: phosphoenolpyruvate-protein phosphotransferase	HMM	1.51E-06	Negative	4.22E-04	Negative
PF01544 CorA: CorA-like Mg ²⁺ transporter protein	HMM	1.12E-05	Negative	7.13E-04	Negative
PF06629 MipA: MltA-interacting protein MipA	HMM	2.01E-04	Negative	2.24E-04	Negative
PF02308 MgtC: putative Mg ²⁺ transporter-C (MgtC) family	HMM	1.85E-04	Negative	4.44E-04	Negative
PF08402 TOBE_2: TOBE domain	HMM	3.40E-06	Negative	9.58E-05	Positive
PF06472 ABC_membrane_2: ABC transporter transmembrane region 2	HMM	2.02E-04	Positive	5.87E-06	Positive
Membrane Transport Other ion-coupled transporters	KEGG pathway	1.70E-10	Negative	2.17E-06	Negative
Membrane Transport Phosphotransferase system (PTS) [PATH:ko02060]	KEGG pathway	1.56E-06	Negative	7.32E-05	Negative
Membrane Transport Other transporters	KEGG pathway	2.43E-08	Negative	4.00E-04	Negative
Transport out					
[G] COG2814 Arabinose efflux permease	COG	4.16E-08	Negative	6.08E-05	Negative
[V] COG1566 Multidrug resistance efflux pump	COG	2.25E-08	Negative	1.12E-06	Negative
[V] COG0841 Cation/multidrug efflux pump	COG	1.82E-08	Negative	1.41E-05	Negative
[P] COG2217 Cation transport ATPase	COG	1.62E-09	Negative	1.04E-05	Negative
[P] COG0474 Cation transport ATPase	COG	1.39E-04	Negative	1.87E-04	Negative
[P] COG0053 Predicted Co/Zn/Cd cation transporters	COG	3.18E-08	Negative	1.22E-04	Negative
[V] COG0842 ABC-type multidrug transport system, permease component	COG	8.42E-06	Negative	1.24E-05	Negative
[P] COG1230 Co/Zn/Cd efflux system component	COG	1.79E-07	Negative	2.97E-06	Negative
[P] COG1275 Tellurite resistance protein and related permeases	COG	3.20E-04	Negative	3.43E-04	Negative
PF00873 ACR_tran: RND transporter, Hydrophobe/Amphiphile Efflux-1 (HAE1) and Heavy Metal Efflux (HME) families, permease protein	HMM	1.17E-08	Negative	2.58E-05	Negative
TIGR01730 RND_mfp: efflux transporter, RND family, MFP subunit	HMM	5.68E-07	Negative	1.02E-04	Negative
PF01545 Cation_efflux: cation efflux family protein	HMM	2.21E-10	Negative	1.59E-07	Negative
TIGR01525 ATPase-IB_hvy: heavy metal translocating P-type ATPase	HMM	3.92E-10	Negative	7.93E-06	Negative
PF00690 Cation_ATPase_N: Cation transporter/ATPase, N-terminus	HMM	2.08E-04	Negative	2.77E-04	Negative

TIGR00914 2A0601: heavy metal efflux pump, CzcA family	HMM	2.47E-05 Negative	1.21E-04 Negative
TIGR00915 2A0602: RND transporter, hydrophobe/amphiphile efflux-1 (HAE1) family	HMM	1.13E-06 Negative	1.91E-05 Negative
TIGR00711 efflux_EmrB: drug resistance MFS transporter, drug:H ⁺ antiporter-1 (14 Spanner) (DHA2) family	HMM	2.63E-05 Negative	4.53E-05 Negative
TIGR01512 ATPase-IB2_Cd: cadmium-translocating P-type ATPase	HMM	1.22E-04 Negative	3.76E-04 Negative
TIGR01845 outer_NodT: efflux transporter, outer membrane factor (OMF) lipoprotein, NodT family	HMM	5.22E-05 Negative	1.43E-04 Negative
TIGR00932 2a37: transporter, monovalent cation:proton antiporter-2 (CPA2) family	HMM	2.40E-04 Negative	8.06E-04 Negative
TIGR01189 ccmA: heme ABC exporter, ATP-binding protein CcmA	HMM	7.97E-06 Negative	6.93E-04 Negative
Sodium/H⁺ antiporter			
[P] COG1006 Multisubunit Na ⁺ /H ⁺ antiporter, MnhC subunit	COG	1.59E-05 Negative	3.46E-04 Negative
[P] COG1863 Multisubunit Na ⁺ /H ⁺ antiporter, MnhE subunit	COG	4.72E-05 Negative	1.82E-04 Negative
[P] COG1320 Multisubunit Na ⁺ /H ⁺ antiporter, MnhG subunit	COG	6.00E-05 Negative	2.23E-04 Negative
PF04039 MnhB: Domain related to MnhB subunit of Na ⁺ /H ⁺ antiporter	HMM	1.59E-05 Negative	3.37E-04 Negative
TIGR01300 CPA3_mnhG_phaG: monovalent cation/proton antiporter, MnhG/PhaG subunit	HMM	3.36E-05 Negative	1.24E-04 Negative
PF03334 PhaG_MnhG_YufB: Na ⁺ /H ⁺ antiporter subunit	HMM	5.97E-05 Negative	2.20E-04 Negative
PF01899 MNHE: Na ⁺ /H ⁺ ion antiporter subunit	HMM	4.72E-05 Negative	1.82E-04 Negative
Electron transport			
PF01654 Bac_Ubq_Cox: Bacterial Cytochrome Ubiquinol Oxidase	HMM	6.59E-08 Negative	4.27E-06 Negative
PF02322 Cyto_ox_2: cytochrome oxidase subunit II	HMM	3.34E-07 Negative	1.36E-05 Negative
PF00284 Cytochrom_B559a: Lumenal portion of Cytochrome b559, alpha (gene psbE) subunit	HMM	4.45E-04 Positive	6.46E-05 Positive
Energy Metabolism Oxidative phosphorylation [PATH:ko00190]	KEGG pathway	1.68E-04 Negative	1.33E-04 Positive
Iron transport			
[P] COG0609 ABC-type Fe ³⁺ -siderophore transport system, permease component	COG	3.37E-10 Negative	1.00E-08 Negative
[P] COG0614 ABC-type Fe ³⁺ -hydroxamate transport system, periplasmic component	COG	2.04E-05 Negative	7.14E-04 Negative

[P] COG4773 Outer membrane receptor for ferric coprogen and ferric-rhodotorulic acid	COG	3.33E-04	Negative	5.38E-04	Negative
[S] COG3182 Uncharacterized iron-regulated membrane protein	COG	8.65E-05	Negative	2.22E-04	Negative
[S] COG3016 Uncharacterized iron-regulated protein	COG	6.92E-05	Negative	1.68E-04	Negative
[P] COG3487 Uncharacterized iron-regulated protein	COG	3.12E-04	Negative	4.85E-04	Negative
PF01032 FecCD: ABC transporter, iron chelate uptake transporter (FeCT) family, permease protein	HMM	8.28E-08	Negative	6.38E-06	Negative
TIGR01783 TonB-siderophore: TonB-dependent siderophore receptor	HMM	1.44E-04	Negative	6.57E-04	Negative
TIGR01785 TonB-hemin: TonB-dependent heme/hemoglobin receptor family protein	HMM	3.89E-04	Negative	4.11E-04	Negative
PF08021 FAD_binding_9: Siderophore-interacting FAD-binding domain	HMM	3.22E-04	Negative	6.76E-04	Negative
TIGR00754 bfr: bacterioferritin	HMM	3.26E-04	Negative	7.34E-04	Negative
M00370 Iron complex transport system	KEGG Complex Module	2.83E-07	Negative	6.29E-06	Negative
Uptake of four carbon dicarboxylic acids					
[G] COG1593 TRAP-type C4-dicarboxylate transport system, large permease component	COG	1.95E-04	Positive	5.57E-04	Positive
[G] COG1638 TRAP-type C4-dicarboxylate transport system, periplasmic component	COG	2.10E-04	Positive	5.57E-04	Positive
[G] COG3090 TRAP-type C4-dicarboxylate transport system, small permease component	COG	1.45E-04	Positive	4.51E-04	Positive
PF06808 DctM: TRAP transporter, DctM-like membrane protein	HMM	2.20E-06	Positive	5.45E-05	Positive
PF04290 DctQ: TRAP transporter, DctQ-like membrane protein	HMM	2.86E-05	Positive	3.88E-04	Positive
TIGR00786 dctM: TRAP transporter, DctM subunit	HMM	6.85E-06	Positive	4.60E-05	Positive
PF03595 C4dic_mal_tran: C4-dicarboxylate transporter/malic acid transport protein	HMM	4.61E-04	Negative	4.91E-04	Negative
TIGR02122 TRAP_TAXI: TRAP transporter solute receptor, TAXI family	HMM	8.28E-05	Negative	6.18E-04	Positive
Anaerobic metabolism					
[O] COG1180 Pyruvate-formate lyase-activating enzyme	COG	7.07E-05	Negative	1.22E-04	Negative
[CP] COG0651 Formate hydrogenlyase subunit 3/Multisubunit Na ⁺ /H ⁺ antiporter, MnhD subunit	COG	3.36E-06	Negative	2.97E-05	Negative

[C] COG1902 NADH:flavin oxidoreductases, Old Yellow Enzyme family	COG	6.63E-06 Negative	2.60E-04 Negative
[P] COG0725 ABC-type molybdate transport system, periplasmic component	COG	8.77E-10 Negative	1.72E-07 Negative
[C] COG1271 Cytochrome bd-type quinol oxidase, subunit 1	COG	2.47E-07 Negative	1.06E-05 Negative
[C] COG1294 Cytochrome bd-type quinol oxidase, subunit 2	COG	3.39E-07 Negative	1.26E-05 Negative
[P] COG2223 Nitrate/nitrite transporter	COG	1.66E-04 Negative	7.65E-04 Negative
[P] COG4149 ABC-type molybdate transport system, permease component	COG	1.49E-06 Negative	9.41E-05 Negative
[C] COG1251 NAD(P)H-nitrite reductase	COG	5.51E-05 Negative	5.47E-04 Negative
[C] COG2010 Cytochrome c, mono- and diheme variants	COG	9.23E-06 Negative	1.97E-04 Negative
PF00158 Sigma54_activat: Sigma-54 interaction domain	HMM	6.73E-07 Negative	1.69E-05 Negative
PF01022 HTH_5: transcriptional regulator, ArsR family	HMM	3.36E-07 Negative	1.94E-04 Negative
PF01558 POR: 2- oxoacid:ferredoxin/flavodoxin oxidoreductases, gamma subunit	HMM	2.11E-04 Negative	6.25E-04 Negative
TIGR01256 modA: molybdate ABC transporter, periplasmic molybdate- binding protein	HMM	5.17E-09 Negative	1.01E-06 Negative
TIGR00203 cydB: cytochrome d ubiquinol oxidase, subunit II	HMM	8.93E-06 Negative	9.06E-05 Negative
TIGR02141 modB_ABC: molybdate ABC transporter, permease protein	HMM	5.43E-08 Negative	1.99E-05 Negative
PF00309 Sigma54_AID: sigma-54 factor, Activator interacting domain (AID)	HMM	2.98E-06 Negative	7.34E-05 Negative
TIGR02374 nitri_red_nirB: nitrite reductase [NAD(P)H], large subunit	HMM	1.20E-04 Negative	2.81E-04 Negative
TIGR02378 nirD_assim_sml: nitrite reductase [NAD(P)H], small subunit	HMM	1.20E-04 Negative	2.81E-04 Negative
M00282 Cytochrome bd complex	KEGG Complex Module	7.74E-07 Negative	2.71E-05 Negative
M00319 Molybdate transport system	KEGG Complex Module	2.37E-07 Negative	4.59E-06 Negative
Basic metabolism			
M00001 Glycolysis (Embden-Meyerhof pathway), glucose => pyruvate	KEGG Pathway Module	2.18E-05 Negative	7.86E-07 Positive
M00003 Gluconeogenesis, oxaloacetate => fructose-6P	KEGG Pathway Module	8.52E-07 Negative	4.99E-06 Positive
M00002 Glycolysis, core module involving three-carbon compounds	KEGG Pathway Module	7.17E-05 Negative	1.04E-07 Positive

M00246 Heme biosynthesis, glutamate => protoheme/siroheme	KEGG Pathway Module	4.21E-04 Negative	9.15E-09 Positive
M00695 Urea transport system	KEGG Complex Module	3.70E-04 Positive	6.64E-05 Positive
Translation Other translation proteins	KEGG pathway	9.78E-05 Negative	3.69E-07 Positive
Carbohydrate Metabolism Glycolysis / Gluconeogenesis [PATH:ko00010]	KEGG pathway	2.93E-07 Negative	2.81E-04 Positive
Transcription RNA polymerase [PATH:ko03020]	KEGG pathway	3.88E-05 Negative	4.39E-05 Positive
Carbohydrate Metabolism Pentose phosphate pathway [PATH:ko00030]	KEGG pathway	2.31E-05 Negative	5.51E-05 Positive
Cell Growth and Death Cell division	KEGG pathway	7.39E-05 Negative	6.99E-05 Positive
DNA metabolism			
M00597 DNA polymerase III complex	KEGG Complex Module	3.89E-05 Negative	1.70E-05 Positive
RNA metabolism			
PF04073 YbaK: YbaK / prolyl-tRNA synthetases associated domain	HMM	3.10E-07 Negative	1.39E-04 Negative
TIGR00342 TIGR00342: thiamine biosynthesis/tRNA modification protein Thil	HMM	7.09E-05 Negative	9.33E-05 Negative
TIGR02152 D_ribokin_bact: ribokinase	HMM	1.59E-04 Negative	1.98E-04 Negative
TIGR02651 RNase_Z: ribonuclease Z	HMM	9.45E-05 Positive	5.36E-06 Positive
Organic phosphorous metabolism			
PF00481 PP2C: protein phosphatase 2C	HMM	1.21E-04 Negative	3.44E-04 Negative
PF01895 PhoU: PhoU family	HMM	1.07E-08 Negative	6.39E-04 Negative
TIGR02135 phoU_full: phosphate transport system regulatory protein PhoU	HMM	1.48E-09 Negative	4.53E-04 Negative
TIGR01449 PGP_bact: phosphoglycolate phosphatase, bacterial	HMM	1.25E-05 Negative	1.91E-05 Negative
PF07831 PYNP_C: Pyrimidine nucleoside phosphorylase C-terminal domain	HMM	5.87E-06 Negative	6.11E-05 Negative
Ribosome metabolism/biosynthesis			
[J] COG2813 16S RNA G1207 methylase RsmC	COG	2.98E-06 Negative	2.08E-05 Negative
Exopolysaccharide biosynthesis			
[M] COG3206 Uncharacterized protein involved in exopolysaccharide biosynthesis	COG	8.12E-06 Negative	4.36E-04 Negative
Biosynthesis			
PF02568 Thil: Thiamine biosynthesis protein (Thil)	HMM	1.12E-04 Negative	2.24E-04 Negative
TIGR01853 lipid_A_lpxD: UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase	HMM	1.71E-04 Positive	2.40E-12 Positive

M00034 Tryptophan biosynthesis, chorismate => tryptophan	KEGG Pathway Module	9.19E-05 Negative	3.68E-09 Positive
M00091 Adenine nucleotide biosynthesis, IMP => ADP/dADP,ATP/dATP	KEGG Pathway Module	2.63E-06 Negative	6.31E-06 Positive
M00092 Guanine nucleotide biosynthesis, IMP => GDP/dGDP,GTP/dGTP	KEGG Pathway Module	7.60E-07 Negative	1.73E-05 Positive
M00033 Chorismate biosynthesis, phosphoenolpyruvate + erythrose-4P =>	KEGG Pathway Module	5.15E-06 Negative	3.99E-07 Positive
M00197 C40 isoprenoid biosynthesis, GGAP => lycopene	KEGG Pathway Module	4.33E-04 Positive	7.67E-06 Positive
M00198 C40 isoprenoid biosynthesis, lycopene => beta-carotene	KEGG Pathway Module	2.68E-07 Positive	8.02E-11 Positive
M00014 Glyoxylate biosynthesis, glycine => glyoxylate	KEGG Pathway Module	7.53E-07 Positive	2.54E-08 Positive
M00159 Fatty acid biosynthesis, initiation	KEGG Pathway Module	3.83E-04 Negative	8.46E-09 Positive
M00203 Glyceroglycolipid biosynthesis	KEGG Pathway Module	1.28E-05 Positive	5.25E-06 Positive
Nucleotide Metabolism Purine metabolism [PATH:ko00230]	KEGG pathway	8.88E-07 Negative	1.02E-05 Positive
Nucleotide Metabolism Pyrimidine metabolism [PATH:ko00240]	KEGG pathway	2.01E-05 Negative	2.18E-08 Positive
Amino Acid Metabolism Alanine and aspartate metabolism [PATH:ko00252]	KEGG pathway	4.01E-05 Negative	2.56E-08 Positive
Amino Acid Metabolism Phenylalanine, tyrosine and tryptophan biosynthesis [PATH:ko00400]	KEGG pathway	1.18E-06 Negative	2.78E-09 Positive
Biosynthesis of Secondary Metabolites Carotenoid biosynthesis [PATH:ko00906]	KEGG pathway	7.82E-07 Positive	1.76E-10 Positive
Metabolism of Cofactors and Vitamins Nicotinate and nicotinamide metabolism [PATH:ko00760]	KEGG pathway	3.08E-06 Negative	3.80E-04 Positive
Amino Acid Metabolism Lysine biosynthesis [PATH:ko00300]	KEGG pathway	8.16E-07 Negative	7.55E-05 Positive
Translation Aminoacyl-tRNA biosynthesis [PATH:ko00970]	KEGG pathway	8.11E-07 Negative	1.98E-08 Positive
Chemotaxis			
[NT] COG0840 Methyl-accepting chemotaxis protein	COG	2.59E-10 Negative	2.50E-10 Negative
[NT] COG0835 Chemotaxis signal transduction protein	COG	3.44E-10 Negative	4.44E-10 Negative
[NT] COG0643 Chemotaxis protein histidine kinase and related kinases	COG	2.28E-10 Negative	4.54E-10 Negative

[NT] COG1352 Methylase of chemotaxis methyl-accepting proteins	COG	2.21E-10	Negative	1.30E-09	Negative
[NT] COG2201 Chemotaxis response regulator containing a CheY-like receiver domain and a methylesterase domain	COG	2.99E-09	Negative	1.88E-08	Negative
[NT] COG1871 Chemotaxis protein; stimulates methylation of MCP proteins	COG	3.12E-04	Negative	3.84E-04	Negative
PF00015 MCPsignal: Methyl-accepting chemotaxis protein (MCP) signaling domain	HMM	2.64E-10	Negative	2.56E-10	Negative
PF01584 CheW: CheW-like domain	HMM	2.06E-10	Negative	2.50E-10	Negative
PF01739 CheR: CheR methyltransferase, SAM binding domain	HMM	2.26E-09	Negative	5.70E-08	Negative
PF01339 CheB_methylest: protein-glutamate methylesterase CheB	HMM	8.49E-10	Negative	4.61E-09	Negative
PF03705 CheR_N: CheR methyltransferase, all-alpha domain	HMM	1.15E-08	Negative	6.48E-08	Negative
PF03975 CheD: CheD	HMM	4.61E-04	Negative	5.63E-04	Negative
Cell Motility Bacterial chemotaxis [PATH:ko02030]	KEGG pathway	1.19E-10	Negative	8.78E-10	Negative
Motility					
[N] COG1344 Flagellin and related hook-associated proteins	COG	4.37E-08	Negative	1.11E-07	Negative
[N] COG1580 Flagellar basal body-associated protein	COG	5.49E-05	Negative	5.83E-05	Negative
[N] COG4786 Flagellar basal body rod protein	COG	1.07E-07	Negative	4.63E-07	Negative
[NUO] COG1516 Flagellin-specific chaperone FliS	COG	2.59E-04	Negative	5.29E-04	Negative
[N] COG1291 Flagellar motor component	COG	5.92E-08	Negative	4.28E-07	Negative
[N] COG1345 Flagellar capping protein	COG	1.93E-04	Negative	4.39E-04	Negative
[N] COG1256 Flagellar hook-associated protein	COG	3.10E-08	Negative	7.95E-07	Negative
[NU] COG1677 Flagellar hook-basal body protein	COG	4.91E-05	Negative	1.56E-04	Negative
[NU] COG1298 Flagellar biosynthesis pathway, component FlhA	COG	1.38E-07	Negative	2.93E-06	Negative
[NU] COG1338 Flagellar biosynthesis pathway, component FliP	COG	2.90E-08	Negative	9.06E-07	Negative
[N] COG1536 Flagellar motor switch protein	COG	2.54E-08	Negative	6.70E-07	Negative
[N] COG1749 Flagellar hook protein FlgE	COG	9.03E-07	Negative	8.91E-06	Negative
[NU] COG1684 Flagellar biosynthesis pathway, component FliR	COG	1.33E-07	Negative	3.46E-06	Negative
[NU] COG1377 Flagellar biosynthesis pathway, component FlhB	COG	1.19E-07	Negative	2.86E-06	Negative
[NU] COG1987 Flagellar biosynthesis pathway, component FliQ	COG	1.24E-07	Negative	2.90E-06	Negative
[N] COG1843 Flagellar hook capping protein	COG	6.93E-08	Negative	2.14E-06	Negative

[N] COG1558 Flagellar basal body rod protein	COG	9.62E-08 Negative	2.76E-06 Negative
[N] COG1815 Flagellar basal body protein	COG	1.24E-07 Negative	2.95E-06 Negative
[N] COG1419 Flagellar GTP-binding protein	COG	1.87E-04 Negative	5.84E-04 Negative
[N] COG1706 Flagellar basal-body P-ring protein	COG	9.50E-07 Negative	1.14E-05 Negative
[S] COG2995 Uncharacterized paraquat-inducible protein A	COG	1.28E-04 Negative	1.59E-04 Negative
[N] COG1868 Flagellar motor switch protein	COG	3.35E-06 Negative	2.92E-05 Negative
[NO] COG1261 Flagellar basal body P-ring biosynthesis protein	COG	7.26E-06 Negative	5.83E-05 Negative
[N] COG2063 Flagellar basal body L-ring protein	COG	7.01E-07 Negative	1.12E-05 Negative
PF00460 Flg_bb_rod: Flagella basal body rod protein	HMM	2.24E-07 Negative	6.72E-07 Negative
PF00669 Flagellin_N: Bacterial flagellin N-terminus	HMM	9.20E-08 Negative	2.30E-07 Negative
TIGR03506 FlgEFG_subfam: flagellar hook-basal body proteins	HMM	8.06E-08 Negative	1.45E-06 Negative
PF00700 Flagellin_C: Bacterial flagellin C-terminus	HMM	4.96E-07 Negative	1.01E-06 Negative
PF07196 Flagellin_IN: Flagellin hook IN motif	HMM	2.84E-04 Negative	3.20E-04 Negative
PF03748 FliL: flagellar basal body-associated protein FliL	HMM	1.58E-06 Negative	4.28E-06 Negative
PF00813 FliP: FliP family	HMM	3.08E-08 Negative	7.74E-07 Negative
PF02465 FliD_N: flagellar hook-associated protein 2	HMM	1.76E-04 Negative	4.18E-04 Negative
TIGR00328 flhB: flagellar biosynthetic protein FlhB	HMM	2.19E-05 Negative	7.36E-05 Negative
TIGR00205 fliE: flagellar hook-basal body complex protein FliE	HMM	2.81E-05 Negative	8.99E-05 Negative
PF07195 FliD_C: Flagellar hook-associated protein 2 C-terminus	HMM	3.03E-04 Negative	6.69E-04 Negative
PF02049 FliE: flagellar hook-basal body complex protein FliE	HMM	1.33E-05 Negative	6.06E-05 Negative
TIGR01026 flil_yscN: ATPase FliI/YscN family	HMM	1.87E-07 Negative	3.55E-06 Negative
TIGR00206 fliF: flagellar M-ring protein FliF	HMM	5.77E-08 Negative	5.90E-07 Negative
TIGR01400 fliR: flagellar biosynthetic protein FliR	HMM	7.93E-05 Negative	2.05E-04 Negative
TIGR01396 FlgB: flagellar basal-body rod protein FlgB	HMM	2.44E-04 Negative	5.17E-04 Negative
TIGR02492 flgK_ends: flagellar hook-associated protein FlgK	HMM	4.14E-08 Negative	1.32E-06 Negative
TIGR00765 yihY_not_rbn: YihY family protein	HMM	1.35E-07 Negative	4.04E-05 Negative
TIGR03496 Flil_clade1: flagellar protein export ATPase FliI	HMM	3.39E-04 Negative	8.05E-04 Negative

TIGR02473 flagell_FliJ: flagellar export protein FliJ	HMM	2.29E-04 Negative	4.10E-04 Negative
TIGR02480 fliN: flagellar motor switch protein FliN	HMM	3.01E-05 Negative	1.08E-04 Negative
TIGR01103 fliP: flagellar biosynthetic protein FliP	HMM	8.67E-08 Negative	1.98E-06 Negative
PF08345 YscJ_FliF_C: Flagellar M-ring protein C-terminal	HMM	1.78E-07 Negative	2.72E-06 Negative
TIGR01395 FlgC: flagellar basal-body rod protein FlgC	HMM	3.05E-07 Negative	3.03E-06 Negative
TIGR01402 fliQ: flagellar biosynthetic protein FliQ	HMM	2.47E-04 Negative	5.39E-04 Negative
PF07559 FlaE: Flagellar basal body protein FlgE	HMM	3.55E-05 Negative	7.88E-05 Negative
PF02120 Flg_hook: flagellar hook-length control protein	HMM	9.53E-05 Negative	4.17E-04 Negative
PF03963 FlgD: flagellar hook capping protein	HMM	1.93E-07 Negative	4.37E-06 Negative
TIGR01398 FlhA: flagellar biosynthesis protein FlhA	HMM	3.45E-07 Negative	6.53E-06 Negative
PF02154 FliM: flagellar motor switch protein FliM	HMM	9.95E-06 Negative	3.86E-05 Negative
TIGR03499 FlhF: flagellar biosynthetic protein FlhF	HMM	8.63E-05 Negative	1.27E-04 Negative
PF02119 FlgI: flagellar P-ring protein FlgI	HMM	9.61E-07 Negative	1.29E-05 Negative
TIGR02488 flgG_G_neg: flagellar basal-body rod protein FlgG	HMM	3.71E-07 Negative	6.79E-06 Negative
TIGR03170 flgA_cterm: flagella basal body P-ring formation protein FlgA	HMM	3.15E-06 Negative	3.01E-05 Negative
TIGR01397 fliM_switch: flagellar motor switch protein FliM	HMM	1.49E-04 Negative	4.07E-04 Negative
PF02107 FlgH: flagellar L-ring protein FlgH	HMM	1.31E-06 Negative	1.78E-05 Negative
PF07859 Abhydrolase_3: alpha/beta hydrolase fold	HMM	2.28E-04 Negative	7.91E-04 Negative
TIGR02490 flgF: flagellar basal-body rod protein FlgF	HMM	1.02E-05 Negative	2.40E-05 Negative
TIGR00207 fliG: flagellar motor switch protein FliG	HMM	1.08E-04 Negative	4.42E-04 Negative
Cell Motility Flagellar assembly [PATH:ko02040]	KEGG pathway	6.88E-07 Negative	2.53E-06 Negative
Pili formation			
[U] COG4962 Flp pilus assembly protein, ATPase CpaF	COG	4.31E-06 Negative	5.50E-06 Negative
[U] COG4965 Flp pilus assembly protein TadB	COG	3.35E-05 Negative	4.91E-05 Negative
[NU] COG2064 Flp pilus assembly protein TadC	COG	1.10E-04 Negative	1.77E-04 Negative
[U] COG4963 Flp pilus assembly protein, ATPase CpaE	COG	1.76E-04 Negative	2.88E-04 Negative
[U] COG3745 Flp pilus assembly protein CpaB	COG	2.26E-04 Negative	4.19E-04 Negative

PF07238 PilZ: Type IV pilus assembly protein PilZ	HMM	3.90E-07 Negative	6.32E-07 Negative
TIGR03177 pilus_cpaB: Flp pilus assembly protein CpaB	HMM	1.42E-04 Negative	2.53E-04 Negative
Secretion			
[R] COG2244 Membrane protein involved in the export of O-antigen and teichoic acid	COG	4.06E-05 Negative	7.94E-04 Negative
[NU] COG1157 Flagellar biosynthesis/type III secretory pathway ATPase	COG	1.87E-07 Negative	3.55E-06 Negative
[NU] COG1886 Flagellar motor switch/type III secretory pathway protein	COG	1.38E-05 Negative	6.01E-05 Negative
[NU] COG1766 Flagellar biosynthesis/type III secretory pathway lipoprotein	COG	1.93E-07 Negative	2.29E-06 Negative
[U] COG3505 Type IV secretory pathway, VirD4 components	COG	1.13E-04 Negative	1.32E-04 Negative
[U] COG4964 Flp pilus assembly protein, secretin CpaC	COG	1.33E-04 Negative	1.99E-04 Negative
[U] COG3451 Type IV secretory pathway, VirB4 components	COG	1.92E-04 Negative	1.98E-04 Negative
TIGR01409 TAT_signal_seq: Tat (twin-arginine translocation) pathway signal sequence	HMM	3.63E-06 Negative	2.51E-04 Negative
PF00437 GSPII_E: Type II/IV secretion system protein	HMM	1.13E-07 Negative	5.15E-04 Negative
PF00482 GSPII_F: Bacterial type II secretion system protein F domain	HMM	8.77E-09 Negative	4.35E-05 Negative
PF02810 SEC-C: SEC-C motif	HMM	2.57E-07 Negative	7.89E-05 Negative
PF07660 STN: Secretin and TonB N terminus short domain	HMM	1.22E-04 Negative	3.58E-04 Negative
PF01514 YscJ_FliF: Secretory protein of YscJ/FliF family	HMM	9.19E-08 Negative	1.41E-06 Negative
PF01312 Bac_export_2: FlhB/HrpN/YscU/SpaS family	HMM	6.39E-08 Negative	1.20E-06 Negative
PF01311 Bac_export_1: Bacterial export proteins, family 1	HMM	2.58E-07 Negative	4.15E-06 Negative
PF01313 Bac_export_3: Bacterial export proteins, family 3	HMM	1.85E-07 Negative	3.43E-06 Negative
TIGR01843 type_I_hlyD: type I secretion membrane fusion protein, HlyD family	HMM	3.45E-05 Negative	7.00E-05 Negative
Folding, Sorting and Degradation Type II secretion system [PATH:ko03090]	KEGG pathway	2.85E-07 Negative	1.29E-04 Negative
Folding, Sorting and Degradation Type III secretion system [PATH:ko03070]	KEGG pathway	5.46E-07 Negative	1.01E-06 Negative
Folding, Sorting and Degradation Protein export [PATH:ko03060]	KEGG pathway	1.48E-04 Negative	6.13E-08 Positive
Repair and replication			
[L] COG0419 ATPase involved in DNA repair	COG	4.49E-07 Negative	1.02E-05 Negative
[L] COG0415 Deoxyribodipyrimidine photolyase	COG	1.17E-04 Positive	3.43E-11 Positive

[L] COG0420 DNA repair exonuclease	COG	2.66E-10	Negative	1.43E-07	Negative
[L] COG1112 Superfamily I DNA and RNA helicases and helicase subunits	COG	3.93E-05	Negative	5.61E-05	Negative
[J] COG1514 2'-5' RNA ligase	COG	2.35E-06	Negative	1.24E-05	Negative
PF01609 Transposase_11: transposase, IS4 family	HMM	1.65E-07	Negative	2.07E-07	Negative
PF00665 rve: integrase core domain	HMM	5.85E-07	Negative	1.23E-06	Negative
PF02371 Transposase_20: transposase, IS116/IS110/IS902 family	HMM	4.16E-04	Negative	5.28E-04	Negative
PF02384 N6_Mtase: N-6 DNA Methylase	HMM	5.80E-05	Negative	2.30E-04	Negative
PF01527 Transposase_8: transposase	HMM	2.14E-05	Negative	3.54E-05	Negative
PF00009 GTP_EFTU: Elongation factor Tu GTP binding domain	HMM	1.61E-05	Negative	3.10E-07	Positive
PF03144 GTP_EFTU_D2: Elongation factor Tu domain 2	HMM	1.13E-05	Negative	4.00E-06	Positive
PF01751 Toprim: Toprim domain	HMM	2.67E-04	Negative	5.70E-06	Positive
PF01548 Transposase_9: transposase	HMM	4.25E-04	Negative	4.85E-04	Negative
PF02834 2_5_RNA_ligase: 2',5' RNA ligase family	HMM	1.70E-05	Negative	9.46E-05	Negative
TIGR02258 2_5_ligase: 2'-5' RNA ligase	HMM	1.85E-04	Negative	4.24E-04	Negative
Replication and Repair Mismatch repair [PATH:ko03430]	KEGG pathway	1.95E-09	Negative	2.50E-04	Positive
Replication and Repair DNA replication [PATH:ko03030]	KEGG pathway	6.24E-05	Negative	1.10E-07	Positive
Replication and Repair Base excision repair [PATH:ko03410]	KEGG pathway	9.27E-07	Negative	2.64E-07	Positive
Proteases					
[RTKL] COG0515 Serine/threonine protein kinase	COG	5.80E-06	Negative	3.09E-05	Negative
[O] COG0501 Zn-dependent protease with chaperone function	COG	5.58E-06	Negative	3.33E-04	Negative
[O] COG1067 Predicted ATP-dependent protease	COG	7.50E-05	Negative	7.79E-05	Negative
[R] COG0693 Putative intracellular protease/amidase	COG	3.22E-05	Negative	5.87E-04	Negative
[OU] COG1585 Membrane protein implicated in regulation of membrane protease activity	COG	2.75E-05	Negative	2.22E-04	Negative
[T] COG2766 Putative Ser protein kinase	COG	1.68E-04	Negative	2.35E-04	Negative
[E] COG3340 Peptidase E	COG	1.26E-04	Positive	1.59E-06	Positive
PF00069 Pkinase: protein kinase domain	HMM	4.00E-06	Negative	5.91E-06	Negative
PF06798 PrkA: PrkA serine protein kinase	HMM	2.03E-04	Negative	2.91E-04	Negative
Clan MO	MEROPS	8.90E-13	Negative	9.01E-08	Negative
Clan PC	MEROPS	1.88E-05	Negative	2.75E-04	Positive
Clan MK	MEROPS	6.66E-05	Negative	7.60E-05	Positive
Clan CL	MEROPS	3.27E-06	Negative	2.86E-04	Negative
Family M23	MEROPS	8.90E-13	Negative	9.01E-08	Negative
Family M22	MEROPS	6.66E-05	Negative	7.60E-05	Positive
Family C56	MEROPS	5.67E-07	Negative	6.16E-04	Positive

Family M75	MEROPS	1.68E-04 Negative	3.71E-04 Negative
TCA cycle			
PF00291 PALP: Pyridoxal-phosphate dependent enzyme	HMM	2.60E-04 Negative	9.81E-06 Positive
PF06039 Mqo: Malate:quinone oxidoreductase (Mqo)	HMM	2.33E-04 Positive	3.59E-06 Positive
Carboxysome			
PF08936 CsoSCA: Carboxysome Shell Carbonic Anhydrase	HMM	5.17E-06 Positive	1.49E-06 Positive
TIGR02701 shell_carb_anhy: carboxysome shell carbonic anhydrase	HMM	5.17E-06 Positive	1.49E-06 Positive
TIGR02703 carboxysome_A: carboxysome peptide A	HMM	5.17E-06 Positive	1.49E-06 Positive
TIGR02704 carboxysome_B: carboxysome peptide B	HMM	5.17E-06 Positive	1.49E-06 Positive
Photosynthesis			
[H] COG1232 Protoporphyrinogen oxidase	COG	3.74E-04 Negative	4.11E-04 Negative
[H] COG1239 Mg-chelatase subunit ChII	COG	4.03E-04 Positive	9.34E-06 Positive
PF05834 Lycopene_cycl: Lycopene cyclase protein	HMM	6.76E-07 Positive	1.55E-09 Positive
PF00421 PSII: Photosystem II protein	HMM	4.65E-04 Positive	1.62E-04 Positive
TIGR01790 carotene-cycl: lycopene cyclase family protein	HMM	3.63E-04 Positive	2.30E-04 Positive
PF00283 Cytochrom_B559: Cytochrome b559, alpha (gene psbE) and beta (gene psbF) subunits	HMM	3.49E-04 Positive	9.77E-05 Positive
PF02419 PsbL: PsbL protein	HMM	1.39E-04 Positive	2.62E-05 Positive
PF05151 PsbM: Photosystem II reaction centre M protein (PsbM)	HMM	4.28E-05 Positive	6.66E-06 Positive
PF07444 Ycf66_N: Ycf66 protein N-terminus	HMM	2.37E-04 Positive	4.17E-05 Positive
PF02532 PsbI: Photosystem II reaction center I protein (PSII 4.8 kDa protein)	HMM	1.39E-04 Positive	2.39E-05 Positive
TIGR03038 PS_II_psbM: photosystem II reaction center protein PsbM	HMM	7.88E-05 Positive	1.14E-05 Positive
PF01653 DNA_ligase_aden: NAD-dependent DNA ligase adenylation domain	HMM	3.26E-04 Positive	2.83E-09 Positive
TIGR03039 PS_II_CP47: photosystem II chlorophyll-binding protein CP47	HMM	3.90E-04 Positive	6.46E-05 Positive
PF01788 PsbJ: PsbJ	HMM	3.90E-04 Positive	6.46E-05 Positive
TIGR02656 cyanin_plasto: plastocyanin	HMM	1.39E-04 Positive	2.45E-05 Positive
PF04839 PSRP-3_Ycf65: Plastid and cyanobacterial ribosomal protein (PSRP-3 / Ycf65)	HMM	3.90E-04 Positive	6.46E-05 Positive
TIGR02730 carot_isom: carotene isomerase	HMM	3.90E-04 Positive	6.46E-05 Positive
TIGR02731 phytoene_desat: phytoene desaturase	HMM	3.90E-04 Positive	6.46E-05 Positive
TIGR02732 zeta_caro_desat: carotene 7,8-desaturase	HMM	3.90E-04 Positive	6.46E-05 Positive

TIGR02654 circ_KaiB: circadian clock protein KaiB	HMM	3.90E-04 Positive	6.46E-05 Positive
TIGR03060 PS_II_psb29: photosystem II biogenesis protein Psp29	HMM	3.90E-04 Positive	6.33E-05 Positive
ATPases			
[C] COG0636 F0F1-type ATP synthase, subunit c/Archaeal/vacuolar-type H ⁺ -ATPase, subunit K	COG	1.26E-04 Positive	9.02E-05 Positive
PF00122 E1-E2_ATPase: E1-E2 ATPase	HMM	3.74E-11 Negative	8.74E-08 Negative
PF02874 ATP-synt_ab_N: ATP synthase alpha/beta family, beta-barrel domain	HMM	3.18E-05 Negative	2.95E-04 Positive
PF07724 AAA_2: ATPase family associated with various cellular activities (AAA)	HMM	1.01E-06 Negative	3.62E-04 Positive
DNAase			
[L] COG3593 Predicted ATP-dependent endonuclease of the OLD family	COG	4.57E-04 Negative	7.55E-04 Negative
[L] COG1722 Exonuclease VII small subunit	COG	3.11E-06 Negative	5.20E-05 Negative
[R] COG2961 Protein involved in catabolism of external DNA	COG	9.78E-06 Negative	6.82E-05 Negative
Multicopper oxidase			
[Q] COG2132 Putative multicopper oxidases	COG	1.89E-05 Negative	9.41E-05 Negative
PF07731 Cu-oxidase_2: Multicopper oxidase	HMM	1.38E-05 Negative	5.75E-05 Negative
PF07732 Cu-oxidase_3: Multicopper oxidase	HMM	2.34E-05 Negative	9.85E-05 Negative
PF00394 Cu-oxidase: Multicopper oxidase	HMM	9.45E-05 Negative	6.30E-04 Negative
Outer membrane protein			
[M] COG3713 Outer membrane protein V	COG	4.34E-04 Negative	4.66E-04 Negative
[M] COG2982 Uncharacterized protein involved in outer membrane biogenesis	COG	5.72E-05 Negative	3.43E-04 Negative
[M] COG0729 Outer membrane protein	COG	3.57E-07 Negative	1.14E-05 Negative
Xenobiotic degradation			
Xenobiotics Biodegradation and Metabolism Trinitrotoluene degradation [PATH:ko00633]	KEGG pathway	8.58E-07 Negative	6.81E-06 Negative
Miscellaneous			
[J] COG1670 Acetyltransferases, including N-acetylases of ribosomal proteins	COG	5.82E-05 Negative	5.58E-04 Negative
[L] COG2801 Transposase and inactivated derivatives	COG	1.63E-05 Negative	2.12E-05 Negative
[D] COG1192 ATPases involved in chromosome partitioning	COG	3.86E-09 Negative	2.88E-06 Negative
[T] COG0589 Universal stress protein UspA and related nucleotide-binding proteins	COG	3.41E-05 Negative	1.43E-04 Negative

[F] COG1051 ADP-ribose pyrophosphatase	COG	1.06E-08 Negative	5.38E-06 Negative
[L] COG1943 Transposase and inactivated derivatives	COG	2.06E-04 Negative	2.14E-04 Negative
[R] COG0110 Acetyltransferase (isoleucine patch superfamily)	COG	5.44E-06 Negative	5.51E-04 Negative
[V] COG0286 Type I restriction-modification system methyltransferase subunit	COG	6.46E-05 Negative	3.18E-04 Negative
[I] COG1502 Phosphatidylserine/phosphatidylglycerophosphate/cardiolipin synthases and related enzymes	COG	7.42E-09 Negative	4.12E-07 Negative
[C] COG0584 Glycerophosphoryl diester phosphodiesterase	COG	1.78E-09 Negative	4.43E-07 Negative
[V] COG0732 Restriction endonuclease S subunits	COG	8.35E-05 Negative	3.41E-04 Negative
[R] COG1752 Predicted esterase of the alpha-beta hydrolase superfamily	COG	2.56E-05 Negative	3.14E-04 Negative
[R] COG0701 Predicted permeases	COG	7.11E-08 Negative	1.71E-05 Negative
[IR] COG1597 Sphingosine kinase and enzymes related to eukaryotic diacylglycerol kinase	COG	6.21E-06 Negative	3.71E-05 Negative
[I] COG2267 Lysophospholipase	COG	2.76E-06 Negative	3.16E-05 Negative
[R] COG1092 Predicted SAM-dependent methyltransferases	COG	1.93E-06 Negative	3.64E-05 Negative
[O] COG1047 FKBP-type peptidyl-prolyl cis-trans isomerases 2	COG	5.93E-06 Negative	5.36E-04 Negative
[P] COG0753 Catalase	COG	1.51E-04 Negative	2.54E-04 Negative
[ER] COG1387 Histidinol phosphatase and related hydrolases of the PHP family	COG	3.79E-04 Negative	4.11E-04 Negative
[F] COG1816 Adenosine deaminase	COG	8.19E-05 Negative	1.96E-04 Negative
[R] COG0390 ABC-type uncharacterized transport system, permease component	COG	3.05E-05 Negative	4.27E-05 Negative
[R] COG1451 Predicted metal-dependent hydrolase	COG	5.31E-07 Negative	1.97E-04 Negative
[H] COG0301 Thiamine biosynthesis ATP pyrophosphatase	COG	2.52E-04 Negative	4.83E-04 Negative
[R] COG0121 Predicted glutamine amidotransferase	COG	7.76E-05 Negative	2.72E-04 Negative
[H] COG1893 Ketopantoate reductase	COG	2.81E-06 Negative	3.94E-05 Negative
[H] COG1488 Nicotinic acid phosphoribosyltransferase	COG	7.89E-06 Negative	3.86E-05 Negative
[M] COG3137 Putative salt-induced outer membrane protein	COG	2.78E-04 Negative	3.13E-04 Negative
[R] COG3568 Metal-dependent hydrolase	COG	6.73E-05 Negative	2.18E-04 Negative
[E] COG2902 NAD-specific glutamate dehydrogenase	COG	3.08E-05 Negative	7.63E-05 Negative
[R] COG3769 Predicted hydrolase (HAD superfamily)	COG	4.08E-04 Positive	8.90E-07 Positive
[O] COG4235 Cytochrome c biogenesis factor	COG	1.47E-05 Negative	3.02E-04 Negative

[G] COG1015 Phosphopentomutase	COG	7.41E-05	Negative	2.78E-04	Negative
[F] COG0213 Thymidine phosphorylase	COG	5.87E-06	Negative	6.11E-05	Negative
[C] COG4230 Delta 1-pyrroline-5-carboxylate dehydrogenase	COG	5.45E-05	Negative	8.14E-04	Negative
[S] COG1285 Uncharacterized membrane protein	COG	2.28E-04	Negative	5.35E-04	Negative
[F] COG0127 Xanthosine triphosphate pyrophosphatase	COG	4.97E-05	Positive	7.16E-10	Positive
[P] COG4558 ABC-type hemin transport system, periplasmic component	COG	1.68E-04	Negative	2.59E-04	Negative
[F] COG3613 Nucleoside 2-deoxyribosyltransferase	COG	2.99E-04	Positive	9.01E-05	Positive
[C] COG3488 Predicted thiol oxidoreductase	COG	9.17E-06	Negative	2.62E-05	Negative
[E] COG0687 Spermidine/putrescine-binding periplasmic protein	COG	1.50E-05	Negative	2.76E-04	Positive
PF00702 Hydrolase: haloacid dehalogenase-like hydrolase	HMM	2.11E-12	Negative	5.07E-06	Negative
PF00529 HlyD: auxiliary transport protein, membrane fusion protein (MFP) family	HMM	1.59E-08	Negative	1.76E-06	Negative
PF00582 Usp: universal stress protein family	HMM	2.26E-06	Negative	4.76E-04	Negative
TIGR01549 HAD-SF-IA-v1: HAD-superfamily hydrolase, subfamily IA, variant 1	HMM	3.85E-10	Negative	1.95E-06	Negative
PF01497 Peripla_BP_2: Periplasmic binding protein	HMM	5.11E-06	Negative	1.44E-04	Negative
PF01336 tRNA_anti: OB-fold nucleic acid binding domain	HMM	2.40E-04	Negative	1.08E-05	Positive
PF08238 Sel1: Sel1 repeat	HMM	1.44E-05	Negative	1.67E-04	Negative
PF01734 Patatin: phospholipase, patatin family	HMM	8.15E-06	Negative	7.68E-05	Negative
PF00575 S1: S1 RNA binding domain	HMM	9.63E-07	Negative	4.73E-04	Positive
PF01471 PG_binding_1: Putative peptidoglycan binding domain	HMM	2.18E-06	Negative	1.31E-05	Negative
PF00156 Pribosyltran: Phosphoribosyl transferase domain	HMM	1.48E-07	Negative	6.09E-04	Positive
PF02743 Cache_1: Cache domain	HMM	1.19E-05	Negative	1.25E-05	Negative
PF03009 GPD: glycerophosphodiester phosphodiesterase family	HMM	7.19E-10	Negative	2.80E-07	Negative
PF01052 SpoA: Surface presentation of antigens (SPOA) protein	HMM	1.10E-07	Negative	8.75E-07	Negative
PF00708 Acylphosphatase: acylphosphatase	HMM	1.25E-05	Negative	3.82E-04	Negative
PF00614 PLDc: phospholipase D active site domain protein	HMM	9.08E-08	Negative	5.11E-06	Negative
TIGR01552 phd_fam: prevent-host-death family protein	HMM	1.79E-04	Negative	4.21E-04	Negative
PF01814 Hemerythrin: Hemerythrin HHE cation binding domain	HMM	4.26E-05	Negative	1.65E-04	Negative
PF00724 Oxidored_FMN: oxidoreductase, FAD/FMN-binding	HMM	7.55E-06	Negative	6.29E-04	Negative

PF01957 NfeD: Nodulation efficiency protein D (NfeD)	HMM	1.15E-08 Negative	4.72E-07 Negative
PF04324 Fer2_BFD: BFD-like [2Fe-2S] binding domain	HMM	4.24E-08 Negative	4.82E-07 Negative
PF00781 DAGK_cat: Diacylglycerol kinase catalytic domain (presumed)	HMM	5.72E-06 Negative	3.18E-05 Negative
PF02782 FGGY_C: FGGY family of carbohydrate kinases, C-terminal domain	HMM	1.27E-07 Negative	7.86E-05 Negative
TIGR01764 excise: DNA binding domain, excisionase family	HMM	6.25E-06 Negative	9.31E-06 Negative
PF07969 Amidohydro_3: Amidohydrolase family	HMM	7.33E-06 Negative	4.17E-04 Negative
PF02558 ApbA: Ketopantoate reductase PanE/ApbA	HMM	2.66E-07 Negative	1.03E-05 Negative
TIGR00221 nagA: N-acetylglucosamine-6-phosphate deacetylase	HMM	9.42E-06 Negative	2.41E-05 Negative
PF00199 Catalase: catalase	HMM	2.76E-04 Negative	4.44E-04 Negative
PF00962 A_deaminase: Adenosine/AMP deaminase	HMM	8.30E-05 Negative	2.05E-04 Negative
PF08267 Meth_synt_1: Cobalamin-independent synthase, N-terminal domain	HMM	2.78E-05 Negative	3.50E-05 Negative
PF05996 Fe_bilin_red: Ferredoxin-dependent bilin reductase	HMM	1.46E-04 Positive	6.60E-05 Positive
PF05433 Rick_17kDa_Anti: Rickettsia 17 kDa surface antigen	HMM	1.06E-05 Negative	1.44E-04 Negative
TIGR01091 upp: uracil phosphoribosyltransferase	HMM	9.07E-08 Negative	8.51E-06 Negative
TIGR00619 sbcd: exonuclease SbcCD, D subunit	HMM	3.86E-06 Negative	8.59E-06 Negative
PF03445 DUF294: Putative nucleotidyltransferase DUF294	HMM	4.52E-05 Negative	1.81E-04 Negative
PF00448 SRP54: SRP54-type protein, GTPase domain	HMM	3.74E-05 Negative	2.04E-05 Positive
PF00042 Globin: Globin	HMM	1.96E-05 Negative	2.33E-05 Negative
PF08546 ApbA_C: Ketopantoate reductase PanE/ApbA C terminal	HMM	6.47E-06 Negative	7.23E-05 Negative
TIGR00745 apbA_panE: 2-dehydropantoate 2-reductase	HMM	1.54E-05 Negative	1.59E-04 Negative
PF04403 PqiA: Paraquat-inducible protein A	HMM	1.69E-04 Negative	1.75E-04 Negative
PF02036 SCP2: SCP-2 sterol transfer family	HMM	9.33E-05 Negative	3.24E-04 Negative
TIGR01311 glycerol_kin: glycerol kinase	HMM	7.72E-09 Negative	2.76E-04 Negative
TIGR01371 met_syn_B12ind: 5-methyltetrahydropteroyltriglutamate--homocysteine S-methyltransferase	HMM	1.59E-04 Negative	1.98E-04 Negative
TIGR02199 rfaE_dom_II: bifunctional protein RfaE, domain II	HMM	2.29E-04 Negative	5.93E-04 Negative
PF09371 Tex_N: Tex-like protein N-terminal domain	HMM	5.74E-06 Negative	7.35E-04 Negative
PF06094 AIG2: AIG2-like family	HMM	9.90E-07 Negative	1.82E-05 Negative

PF05088 Bac_GDH: Bacterial NAD-glutamate dehydrogenase	HMM	6.04E-05 Negative	1.56E-04 Negative
PF05940 NnrS: NnrS protein	HMM	4.92E-05 Negative	6.59E-05 Negative
PF08342 PPM_N: Phosphopentomutase N-terminal	HMM	7.41E-05 Negative	2.78E-04 Negative
TIGR02644 Y_phosphoryl: pyrimidine-nucleoside phosphorylase	HMM	5.79E-05 Negative	2.40E-04 Negative
PF00814 Peptidase_M22: glycoprotease family	HMM	4.87E-08 Negative	3.98E-05 Positive
TIGR01696 deoB: phosphopentomutase	HMM	9.37E-05 Negative	3.77E-04 Negative
TIGR01228 hutU: urocanate hydratase	HMM	4.90E-06 Negative	1.18E-05 Negative
PF04965 GPW_gp25: Gene 25-like lysozyme	HMM	4.26E-04 Negative	6.35E-04 Negative
PF01725 Ham1p_like: Ham1 family	HMM	2.71E-05 Positive	7.39E-10 Positive
PF04095 NAPRTase: Nicotinate phosphoribosyltransferase (NAPRTase) family	HMM	8.43E-05 Negative	3.27E-04 Negative
TIGR00553 pabB: aminodeoxychorismate synthase, component I	HMM	4.13E-06 Negative	4.94E-05 Negative
TIGR03160 cobT_DBIPRT: nicotinate-nucleotide--dimethylbenzimidazole phosphoribosyltransferase	HMM	3.97E-05 Negative	3.51E-04 Negative
TIGR02472 sucP_syn_N: sucrose-phosphate synthase, putative, glycosyltransferase domain	HMM	1.39E-04 Positive	2.68E-05 Positive
TIGR02505 RTPR: ribonucleoside-triphosphate reductase, adenosylcobalamin-dependent	HMM	2.23E-05 Positive	5.36E-06 Positive
PF09506 Salt_tol_Pase: Glucosylglycerol-phosphate phosphatase (Salt_tol_Pase)	HMM	4.28E-05 Positive	9.19E-06 Positive
TIGR02399 salt_tol_Pase: glucosylglycerol 3-phosphatase	HMM	4.28E-05 Positive	9.19E-06 Positive
TIGR00385 dsbE: periplasmic protein thiol:disulfide oxidoreductases, DsbE subfamily	HMM	3.37E-05 Negative	7.34E-04 Negative
TIGR02463 MPGP_rel: mannosyl-3-phosphoglycerate phosphatase homolog	HMM	5.94E-07 Positive	5.29E-07 Positive
PF04853 Invertase_neut: Plant neutral invertase	HMM	2.62E-05 Positive	5.93E-06 Positive
TIGR01262 maiA: maleylacetoacetate isomerase	HMM	3.74E-05 Negative	9.81E-05 Negative
PF02392 Ycf4: Ycf4	HMM	3.90E-04 Positive	6.46E-05 Positive
TIGR03279 cyano_FeS_chp: putative FeS-containing Cyanobacterial-specific oxidoreductase	HMM	3.90E-04 Positive	6.46E-05 Positive
PF09375 Peptidase_M75: Imelysin	HMM	2.76E-04 Negative	4.84E-04 Negative
PF07698 7TM-7TMR_HD: 7TM receptor with intracellular HD hydrolase	HMM	4.15E-04 Positive	2.83E-05 Positive
PF07729 FCD: FCD domain	HMM	3.66E-05 Negative	7.10E-05 Positive

M00342 Ribose transport system	KEGG Complex Module	1.34E-04 Negative	1.36E-04 Negative
M00238 Jasmonic acid biosynthesis	KEGG Pathway Module	1.81E-04 Negative	6.87E-04 Negative
M00361 Octopine/nopaline transport system	KEGG Complex Module	1.05E-05 Positive	1.05E-05 Positive
M00211 Diphosphatidylglycerol biosynthesis, CDP-glycerol => cardiolipin	KEGG Pathway Module	3.55E-05 Negative	1.90E-04 Negative
Unclassified			
[S] COG2606 Uncharacterized conserved protein	COG	5.58E-05 Negative	7.26E-04 Negative
[S] COG1432 Uncharacterized conserved protein	COG	1.36E-06 Negative	5.12E-05 Negative
[S] COG1720 Uncharacterized conserved protein	COG	1.25E-05 Negative	1.94E-05 Negative
[S] COG1872 Uncharacterized conserved protein	COG	5.88E-06 Negative	1.01E-05 Negative
[S] COG2719 Uncharacterized conserved protein	COG	1.35E-04 Negative	1.96E-04 Negative
[S] COG4487 Uncharacterized protein conserved in bacteria	COG	9.41E-08 Positive	1.33E-08 Positive
[S] COG2718 Uncharacterized conserved protein	COG	1.67E-04 Negative	2.35E-04 Negative
[P] COG3213 Uncharacterized protein involved in response to NO	COG	4.92E-05 Negative	6.59E-05 Negative
[S] COG3326 Predicted membrane protein	COG	1.96E-04 Negative	2.38E-04 Negative
[S] COG2832 Uncharacterized protein conserved in bacteria	COG	1.42E-05 Negative	5.20E-05 Negative
[R] COG4240 Predicted kinase	COG	3.87E-04 Positive	1.62E-06 Positive
[S] COG3556 Predicted membrane protein	COG	7.88E-05 Positive	1.80E-05 Positive
[S] COG2911 Uncharacterized protein conserved in bacteria	COG	2.39E-07 Negative	1.09E-04 Negative
[S] COG2989 Uncharacterized protein conserved in bacteria	COG	5.83E-05 Negative	2.62E-04 Negative
[S] COG2954 Uncharacterized protein conserved in bacteria	COG	3.26E-04 Positive	3.83E-06 Positive
[S] COG2354 Uncharacterized protein conserved in bacteria	COG	7.17E-06 Negative	7.68E-05 Negative
PF06429 DUF1078: Domain of unknown function (DUF1078)	HMM	7.13E-08 Negative	8.52E-07 Negative
PF07864 DUF1651: Protein of unknown function (DUF1651)	HMM	8.67E-08 Positive	7.16E-08 Positive
PF03773 DUF318: Predicted permease	HMM	1.50E-07 Negative	2.83E-05 Negative
PF01936 DUF88: Protein of unknown function	HMM	1.17E-07 Negative	3.07E-06 Negative
PF01980 UPF0066: Uncharacterised protein family	HMM	1.24E-05 Negative	1.94E-05 Negative

PF03649 UPF0014: Uncharacterised protein family (UPF0014)	HMM	5.83E-05 Negative	7.69E-05 Negative
PF04463 DUF523: Protein of unknown function (DUF523)	HMM	1.72E-04 Negative	2.07E-04 Negative
TIGR00252 TIGR00252: conserved hypothetical protein TIGR00252	HMM	1.24E-05 Negative	1.37E-04 Negative
PF07878 DUF1662: Protein of unknown function (DUF1662)	HMM	9.77E-08 Positive	9.13E-08 Positive
TIGR00104 TIGR00104: conserved hypothetical protein TIGR00104	HMM	3.76E-05 Negative	5.71E-05 Negative
TIGR03643 TIGR03643: conserved hypothetical protein TIGR03643	HMM	2.65E-09 Positive	8.88E-15 Positive
PF02594 DUF167: Uncharacterized ACR, YggU family COG1872	HMM	2.78E-06 Negative	5.33E-06 Negative
PF04285 DUF444: Protein of unknown function (DUF444)	HMM	2.03E-04 Negative	2.86E-04 Negative
PF06961 DUF1294: Protein of unknown function (DUF1294)	HMM	1.32E-04 Negative	1.65E-04 Negative
PF08843 DUF1814: Domain of unknown function (DUF1814)	HMM	4.69E-04 Negative	4.91E-04 Negative
PF04304 DUF454: Protein of unknown function (DUF454)	HMM	1.49E-04 Negative	3.78E-04 Negative
PF07566 DUF1543: Domain of Unknown Function (DUF1543)	HMM	2.99E-04 Positive	1.19E-04 Positive
PF04378 DUF519: Protein of unknown function (DUF519)	HMM	1.28E-05 Negative	8.36E-05 Negative
TIGR01620 hyp_HI0043: conserved hypothetical protein TIGR01620	HMM	5.53E-05 Negative	1.26E-04 Negative
PF08854 DUF1824: Domain of unknown function (DUF1824)	HMM	1.39E-04 Positive	2.62E-05 Positive
PF08848 DUF1818: Domain of unknown function (DUF1818)	HMM	2.37E-04 Positive	4.08E-05 Positive
PF06537 DUF1111: Protein of unknown function (DUF1111)	HMM	2.42E-05 Negative	6.11E-05 Negative
PF05421 DUF751: Protein of unknown function (DUF751)	HMM	3.90E-04 Positive	6.46E-05 Positive
PF03382 DUF285: Mycoplasma protein of unknown function, DUF285	HMM	4.16E-04 Positive	2.31E-04 Positive
TIGR00159 TIGR00159: conserved hypothetical protein TIGR00159	HMM	2.97E-04 Positive	8.94E-06 Positive
PF05661 DUF808: Protein of unknown function (DUF808)	HMM	9.46E-06 Negative	9.26E-05 Negative

(ST4)

Table of cell volume versus genome size together with the literature references.

Cell volume versus genome size:

Table showing reported cell volume ranges and reported genome size ranges for various marine groups. For the last five groups, cell volumes are indicated for both low nutrient and high nutrient conditions.

Organism/Group	Reported cell volume range (μm^3)	References	Reported genome size range (Mbp)
<i>Candidatus pelagibacter ubique</i> (SAR11)	0.01	¹	1.3-1.5
Marine prochlorococcus	0.06-0.25	²⁻⁴	1.6-2.7
Marine Synechococcus	0.4-2.5	^{5,6}	2.2-6
Marine Vibrio	1, 4-7	⁷	4.2-6
Marine Shewanella	1, 4-7	⁷	4.3-5.5
Photobacterium	1, 4-7	⁷	5-6.4
Roseobacter	1, 4-7	⁷	4.1-4.6
Silicibacter	1, 4-7	⁷	3.2-4.7

References

- 1 Rappe, M. S., Connon, S. A., Vergin, K. L. & Giovannoni, S. J. Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature* **418**, 630-633, doi:10.1038/nature00917 nature00917 [pii] (2002).
- 2 Dufresne, A. *et al.* Genome sequence of the cyanobacterium Prochlorococcus marinus SS120, a nearly minimal oxyphototrophic genome. *Proc Natl Acad Sci U S A* **100**, 10020-10025, doi:10.1073/pnas.1733211100 1733211100 [pii] (2003).
- 3 Campbell, L. & Nolla, H. a. The importance of *Prochlorococcus* to community structure in the central North Pacific. *Limnol. Oceanog.* **39**, 954-961 (1994).
- 4 Partensky, F. & Garczarek, L. *Prochlorococcus*: Advantages and limits of minimalism. *Ann. Rev. Mar. Sci.* **2**, 305-331 (2010).
- 5 Binder, B. J. & Liu, Y. C. Growth rate regulation of rRNA content of a marine synechococcus (Cyanobacterium) strain. *Appl Environ Microbiol* **64**, 3346-3351 (1998).
- 6 Haverkamp, T. H. *et al.* Colorful microdiversity of Synechococcus strains (picocyanobacteria) isolated from the Baltic Sea. *ISME J* **3**, 397-408, doi:ismej2008118 [pii] 10.1038/ismej.2008.118 (2009).
- 7 Ingraham, J. L., Maaloe, O. & Neidhardt, F. C. *Growth of the Bacterial Cell.* 435 (Sinauer, 1983).

(ST5)

Correlation of CRISPR presence with the lifestyle of the organisms.

Correlation of CRISPR presence with the lifestyle of the organisms:

Organism	Lifestyle features	Notes from source	Source
I. All CRISPR-rich genomes (20 genomes). <i>All genomes with more than two arrays and more than three CAS proteins</i>			
Caminibacter mediatlanticus TB-2	AT	walls of an active deep-sea hydrothermal vent chimney	1
Chlorobium phaeobacteroides BS1		hemocline of the Black Sea; extremely dilute population; adapted to extreme low-light photosynthesis	2,3
Cyanothece sp. ATCC 51142	A	isolated from the intertidal sands of the Texas Gulf coast in enrichment conditions designed to favor rapid growth	4,5
Dinoroseobacter shibae DFL 12	A	phototrophic bacterium isolated from a marine dinoflagellate (eukaryotic algae)	6
Flavobacterium sp. MED217		flavobacteria represent bacterial populations that are likely to mediate a substantial proportion of the carbon flow and nutrient turnover in the sea during and following algal blooms; isolated from a surface seawater sample from the Bay of Blanes; consumes protein	7
Haloarcula marismortui ATCC 43049	S	Hypersaline, Dead Sea	8
Hydrogenivirga sp. 128-5-R1-1	T	deep-sea vent	9
Lyngbya sp. PCC 8106	A	benthic marine (sediment), forms mats	10
Marinitoga piezophila KA3	T	deep-sea hydrothermal vent chimney	11
Marinomonas sp. MED121		1 m depth Mediterranean sea	12
Microcoleus chthonoplastes PCC 7420	A	salt marsh, very frequent in microbial mats	13
Microcystis aeruginosa NIES-843	A	Bloom-forming, colony/aggregate-forming	14
Microscilla marina ATCC 23134	A	M. marina closely related to Cytophaga sequences in environmental clone libraries of particle-associated bacteria; very closely related sequences in	15

		libraries from Trichodesmium colonies;	
Nitrobacter sp. Nb-311A		NB311A was isolated by Stan Watson in 1968 from surface waters in the tropical Eastern Atlantic Ocean off the coast of West Africa	16
Parvularcula bermudensis HTCC2503		sea-water sample was collected from a depth of 10 m Sargasso Sea	17
Prosthecochloris aestuarii DSM 271	AS	same as SK413; originally isolated from the hydrogen sulfide-rich mud of a shallow lagoon with elevated salt concentration	18,19
Thermococcus barophilus MP	T	barophilic and hyperthermophilic archaeon isolated under high hydrostatic pressure from a deep-sea hydrothermal vent	20
Thermosiphon africanus TCF52B	T	hydrothermal vent; a lot of LGT with archaea and firmicutes	21
Thermococcus sp. AM4	T	hydrothermal vent	22,23
Thermotoga sp. RQ2	T	Geothermally heated seafloor in the Azores	24
II. Sample of CRISPR-free genomes (13 genomes) <i>Randomly selected 10% of all genomes that have neither CRISPR arrays nor CAS genes and that assembled into 20 or fewer contigs</i>			
Synechococcus sp. BL107		picoplankton	25
Roseobacter GAI101		collected in Coastal seawater; isolation method - Nonselectively on low-nutrient seawater medium	26
Prochlorococcus marinus str. MIT 9312		picoplankton	27
Erythrobacter sp. NAP1		Representative strain of open ocean aerobic anoxygenic photosynthetic bacteria	28,29
Candidatus Pelagibacter ubique HTCC1062		picoplankton	30
Sphingopyxis alaskensis RB2256		picoplankton	31
Polaribacter irgensii 23-P		relatives of this strain are cosmopolitan members of the global marine bacterioplankton community	32,33
Fulvimarina pelagi HTCC2506		A sea-water sample was collected from a depth of 10 m in Sargasso Sea	34

Cyanobium sp. PCC 7001		isolated from intertidal waters off Long Island	35
Erythrobacter litoralis HTCC2594		isolated from a depth of 10 m in the Sargasso Sea	36
Shewanella sp. MR-4		5 m depth of the Black Sea	37,38
Trichodesmium erythraeum IMS101	A	Large blooms ("red tides"); forms colonies (in symbiosis with a larger related species); releases toxin that kills natural predators; phage is postulated to be a major agent of mortality; responsible for significant part of nitrogen release into the food chain	39-41
Flavobacteria bacterium BBFL7		This is strain BF-7 with extremely high ectoprotease activity from the reference paper; a free-floating isolate obtained after an artificially induced bloom at the lab; original samples were collected from the surface 1 km offshore of San Diego	42,43

Supplementary Table 5. Table showing reported lifestyle data for groups of genomes with and without CRISPR system. The column Lifestyle Features encodes any reported data that is different from a picoplanktonic lifestyle:

A – associated with particles, surfaces or host organisms, or forms dense colonies/mats/blooms

S – lives in hypersaline environments

T – lives in high-temperature environments

1. Voordeckers, J.W., Starovoytov, V. & Vetriani, C. Caminibacter mediatlanticus sp. nov., a thermophilic, chemolithoautotrophic, nitrate-ammonifying bacterium isolated from a deep-sea hydrothermal vent on the Mid-Atlantic Ridge. *Int J Syst Evol Microbiol* **55**, 773-779 (2005).
2. Overmann, J. & Manske, A.K. ANOXYGENIC PHOTOTROPHIC BACTERIA IN THE BLACK SEA CHEMOCLINE. *Past and Present Water Column Anoxia* 523-541 (2006). at <http://dx.doi.org/10.1007/1-4020-4297-3_20>
3. Chlorobium phaeobacteroides BS1 - Home. at <<http://genome.jgi-psf.org/chlpb/chlpb.home.html>>
4. Reddy, K.J., Haskell, J.B., Sherman, D.M. & Sherman, L.A. Unicellular, aerobic nitrogen-fixing cyanobacteria of the genus Cyanothece. *J Bacteriol* **175**, 1284-1292 (1993).
5. Welsh, E.A. et al. The genome of Cyanothece 51142, a unicellular diazotrophic cyanobacterium important in the marine nitrogen cycle. *Proceedings of the National Academy of Sciences* **105**, 15094-15099 (2008).
6. Dinoroseobacter shibae Genome Project Result. at <<http://www.ncbi.nlm.nih.gov/sites/entrez?Db=genomeprj&cmd=ShowDetailView&TermToSearch=17417>>
7. Pinhassi, J. et al. Leeuwenhoekiella blandensis sp. nov., a genome-sequenced marine member of the family Flavobacteriaceae. *Int J Syst Evol*

- Microbiol* **56**, 1489-1493 (2006).
8. Baliga, N.S. et al. Genome sequence of *Haloarcula marismortui*: a halophilic archaeon from the Dead Sea. *Genome Res* **14**, 2221-2234 (2004).
 9. *Hydrogenivirga* sp. 128-5-R1-1 - Microbial Genome Sequencing Project - Gordon and Betty Moore Foundation. at <http://www.moore.org/microgenome/detail.aspx?id=166>
 10. *Lyngbya* sp. CCY9616 (= PCC8106) - Microbial Genome Sequencing Project - Gordon and Betty Moore Foundation. at <http://www.moore.org/microgenome/detail.aspx?id=70>
 11. Alain, K. et al. *Marinitoga piezophila* sp. nov., a rod-shaped, thermo-piezophilic bacterium isolated under high hydrostatic pressure from a deep-sea hydrothermal vent. *Int. J. Syst. Evol. Microbiol* **52**, 1331-1339 (2002).
 12. *Marinomonas* sp. MED121 - Microbial Genome Sequencing Project - Gordon and Betty Moore Foundation. at <http://www.moore.org/microgenome/detail.aspx?id=56>
 13. *Microcoleus chthonoplastes* PCC7420 - Microbial Genome Sequencing Project - Gordon and Betty Moore Foundation. at <http://www.moore.org/microgenome/detail.aspx?id=146>
 14. Kaneko, T. et al. Complete Genomic Structure of the Bloom-forming Toxic Cyanobacterium *Microcystis aeruginosa* NIES-843. *DNA Res* dsm026 (2008).doi:10.1093/dnares/dsm026
 15. *Microscilla marina* ATCC23134 - Microbial Genome Sequencing Project - Gordon and Betty Moore Foundation. at <http://www.moore.org/microgenome/detail.aspx?id=39>
 16. *Nitrobacter* sp. Nb-311A - Microbial Genome Sequencing Project - Gordon and Betty Moore Foundation. at <http://www.moore.org/microgenome/detail.aspx?id=77>
 17. Cho, J. & Giovannoni, S.J. *Parvularcula bermudensis* gen. nov., sp. nov., a marine bacterium that forms a deep branch in the {alpha}-Proteobacteria. *Int J Syst Evol Microbiol* **53**, 1031-1036 (2003).
 18. Gorlenko, V.M. A new phototrophic green sulfur bacterium-- *Prosthecochloris aestuarii* nov. gen. nov. sp. *Z Allg Mikrobiol* **10**, 147-149 (1970).
 19. *Prosthecochloris aestuarii* DSM 271 - Home. at <http://genome.jgi-psf.org/proae/proae.home.html>
 20. Marteinsson, V.T. et al. *Thermococcus barophilus* sp. nov., a new barophilic and hyperthermophilic archaeon isolated under high hydrostatic pressure from a deep-sea hydrothermal vent. *Int J Syst Bacteriol* **49**, 351-359 (1999).
 21. Nesbø, C.L. et al. The genome of *Thermosipho africanus* TCF52B: lateral genetic connections to the Firmicutes and Archaea. *J. Bacteriol* **191**, 1974-1978 (2009).
 22. Sokolova, T.G. et al. The first evidence of anaerobic CO oxidation coupled with H₂ production by a hyperthermophilic archaeon isolated from a deep-sea hydrothermal vent. *Extremophiles* **8**, 317-323 (2004).
 23. *Thermococcus* sp. AM4 - Microbial Genome Sequencing Project - Gordon and Betty Moore Foundation. at <http://www.moore.org/microgenome/detail.aspx?id=159>

24. *Thermotoga* sp. RQ2 - Home. at <http://genome.jgi-psf.org/the_r/the_r.home.html>
25. *Synechococcus* sp. BL107 - Microbial Genome Sequencing Project - Gordon and Betty Moore Foundation. at <<http://www.moore.org/microgenome/detail.aspx?id=62>>
26. *Roseobacter* sp. GAI-101 - Microbial Genome Sequencing Project - Gordon and Betty Moore Foundation. at <<http://www.moore.org/microgenome/detail.aspx?id=112>>
27. *Prochlorococcus marinus* MIT 9312 - Home. at <http://genome.jgi-psf.org/pro_2/pro_2.home.html>
28. Koblížek, M. et al. Isolation and characterization of *Erythrobacter* sp. strains from the upper ocean. *Arch. Microbiol* **180**, 327-338 (2003).
29. *Erythrobacter* sp. NAP1 - Microbial Genome Sequencing Project - Gordon and Betty Moore Foundation. at <<http://www.moore.org/microgenome/detail.aspx?id=15>>
30. Giovannoni, S.J. et al. Proteorhodopsin in the ubiquitous marine bacterium SAR11. *Nature* **438**, 82-85 (2005).
31. Lauro, F.M. et al. The genomic basis of trophic strategy in marine bacteria. *Proc Natl Acad Sci U S A* **106**, 15527-15533 (2009).
32. Gosink, J.J., Woese, C.R. & Staley, J.T. *Polaribacter* gen. nov., with three new species, *P. irgensii* sp. nov., *P. franzmannii* sp. nov. and *P. filamentus* sp. nov., gas vacuolate polar marine bacteria of the Cytophaga-Flavobacterium-Bacteroides group and reclassification of 'Flectobacillus glomeratus' as *Polaribacter glomeratus* comb. nov. *Int J Syst Bacteriol* **48**, 223-235 (1998).
33. *Polaribacter irgensii* 23-P - Microbial Genome Sequencing Project - Gordon and Betty Moore Foundation. at <<http://www.moore.org/microgenome/detail.aspx?id=46>>
34. Cho, J. & Giovannoni, S.J. *Fulvimarina pelagi* gen. nov., sp. nov., a marine bacterium that forms a deep evolutionary lineage of descent in the order 'Rhizobiales'. *Int J Syst Evol Microbiol* **53**, 1853-1859 (2003).
35. *Cyanobium* sp. PCC7001 - Microbial Genome Sequencing Project - Gordon and Betty Moore Foundation. at <<http://www.moore.org/microgenome/detail.aspx?id=134>>
36. Oh, H., Giovannoni, S.J., Ferriera, S., Johnson, J. & Cho, J. Complete genome sequence of *Erythrobacter litoralis* HTCC2594. *J. Bacteriol* **191**, 2419-2420 (2009).
37. Venkateswaran, K. et al. Polyphasic taxonomy of the genus *Shewanella* and description of *Shewanella oneidensis* sp. nov. *Int J Syst Bacteriol* **49**, 705-724 (1999).
38. *Shewanella* sp. MR-4 - Home. at <http://genome.jgi-psf.org/she_4/she_4.home.html>
39. Capone, D.G., Zehr, J.P., Paerl, H., Bergman, B. & Carpenter, E.J. *Trichodesmium*, a Globally Significant Marine Cyanobacterium. *Science* **276**, 1221-1229 (1997).
40. *Trichodesmium erythraeum* IMS101 - Home. at <<http://genome.jgi->

psf.org/trier/trier.home.html>

41. Hewson, I., Govil, S.R., Capone, D.G., Carpenter, E.J. & Fuhrman, J.A. Evidence of Trichodesmium viral lysis and potential significance for biogeochemical cycling in the oligotrophic ocean. *Aquat Microb Ecol* **36**, 1-8 (2004).
42. J, M., DC, S., GF, S. & F, A. Variability in ectohydrolytic enzyme activities of pelagic marine bacteria and its significance for substrate processing in the sea. *Aquat Microb Ecol* **10**, 223-230 (1996).
43. Unnamed Flavobacterium BBFL7 - Microbial Genome Sequencing Project - Gordon and Betty Moore Foundation. at <<http://www.moore.org/microgenome/detail.aspx?id=3>>