

Supplemental Material

Marine cyanophages demonstrate biogeographic patterns throughout the global ocean

Sijun Huang[#], Si Zhang, Nianzhi Jiao, Feng Chen[#]

[#] Address correspondence to SH, E-mail: huangsijun@scsio.ac.cn or FC, E-mail: chenf@umces.edu

Table S1. Summary of marine cyanopodovirus and cyanomyovirus genomes tested in this study.

Group	Phage	Original host	Genome size (bp)	%G+C	Accession	Publication
Cyanopodovirus						
	P-RSP2	<i>Prochlorococcus</i> MIT9302	42257	34.0%	HQ332139	(1)
MPP-A	P60	<i>Synechococcus</i> WH7805	46675	53.3%	AF338467	(2)
	Syn5	<i>Synechococcus</i> WH8019	46214	55.0%	EF372997	(3)
	P-SSP9	<i>Prochlorococcus</i> SS120	46997	40.5%	HQ316584	(1)
	S-CBP2	<i>Synechococcus</i> CB0208	46237	55.0%	KC310806	unpublished
	S-CBP42	<i>Synechococcus</i> WH7803	45218	54.6%	KC310805	unpublished
MPP-B	P-SSP7	<i>Prochlorococcus</i> MED4	44970	38.8%	AY939843	(4)
	P-SSP5	<i>Prochlorococcus</i> MIT9515	47055	39.2%	GU071100	unpublished
	P-RSP5	<i>Prochlorococcus</i> NATL1A	47741	38.7%	GU071102	(1)
	P-HP1	<i>Prochlorococcus</i> NATL2A	47536	39.9%	GU071104	(1)
	P-SSP2	<i>Prochlorococcus</i> MIT9312	45890	37.9%	GU071107	(1)
	P-GSP1	<i>Prochlorococcus</i> MED4	44945	39.6%	HQ332140	(1)
	P-SSP3	<i>Prochlorococcus</i> MIT9312	46198	37.9%	HQ332137	(1)
	P-SSP10	<i>Prochlorococcus</i> NATL2A	47325	39.2%	HQ337022	(1)
	P-SSP11	<i>Prochlorococcus</i> MIT9515	47039	39.2%	HQ634152	(1)
	S-CBP1	<i>Synechococcus</i> CB0101	46547	47.6%	KC310802	unpublished
	S-CBP3	<i>Synechococcus</i> CB0101	45871	47.0%	KC310803	unpublished
	S-CBP4	<i>Synechococcus</i> CB0101	44147	44.4%	KC310804	unpublished
	S-RIP1	<i>Synechococcus</i> WH8101	44892	42.9%	HQ317388	unpublished
	S-RIP2	<i>Synechococcus</i> WH7803	45728	47.3%	HQ317389	unpublished
Cyanomyovirus						
I	S-PM2	<i>Synechococcus</i> WH7803	196280	37.8%	AJ630128	(5)
	S-RSM4	<i>Synechococcus</i> WH7803	194454	41.0%	NC_013085	(6)
	Syn1	<i>Synechococcus</i> WH8101	191195	41.0%	NC_015288	(7)
II	P-SSM2	<i>Prochlorococcus</i> NATL1A	252401	35.5%	NC_006883	(4)
	P-SSM5	<i>Prochlorococcus</i> NATL2A	252013	35.5%	HQ632825	(8)

	S-SM2	<i>Synechococcus</i> WH8017	190789	40.0%	NC_015279	(7)
	S-SSM7	<i>Synechococcus</i> WH8109	232878	39.0%	NC_015287	(7)
III	P-RSM1	<i>Prochlorococcus</i> MIT 9303	177211	40.2%	HQ634175	(8)
	P-RSM3	<i>Prochlorococcus</i> NATL2A	178750	36.7%	HQ634176	(8)
	P-RSM4	<i>Prochlorococcus</i> MIT 9303	176428	38.0%	NC_015283	(7)
	P-RSM6	<i>Prochlorococcus</i> NATL2A	192497	39.3%	HQ634193	(8)
	P-SSM3	<i>Prochlorococcus</i> NATL2A	179063	36.7%	HQ337021	(8)
	P-SSM4	<i>Prochlorococcus</i> NATL2A	178249	36.7%	NC_006884	(4)
	P-SSM7	<i>Prochlorococcus</i> NATL1A	182180	37.0%	NC_015290	(7)
	S-ShM2	<i>Synechococcus</i> WH8102	179563	41.0%	NC_015281	(7)
	S-SM1	<i>Synechococcus</i> WH6501	174079	41.0%	NC_015282	(7)
	S-SSM2	<i>Synechococcus</i> WH8102	179980	41.1%	JF974292	(8)
	S-SSM4	<i>Synechococcus</i> WH8018	182801	39.4%	HQ316583	(8)
	S-SSM5	<i>Synechococcus</i> WH8102	176184	40.0%	NC_015289	(7)
	Syn10	<i>Synechococcus</i> WH8017	177103	40.6%	HQ634191	(8)
	Syn19	<i>Synechococcus</i> WH8109	175230	41.0%	NC_015286	(7)
	Syn2	<i>Synechococcus</i> WH8012	175596	41.3%	HQ634190	(8)
	Syn30	<i>Synechococcus</i> WH7803	178807	39.9%	HQ634189	(8)
	Syn33	<i>Synechococcus</i> WH7803	174285	40.0%	NC_015285	(7)
Syn9	<i>Synechococcus</i> WH8012	177300	40.5%	NC_008296	(9)	
IV	MED4-213	<i>Prochlorococcus</i> MED4	180977	37.8%	HQ634174	(8)
	P-HM1	<i>Prochlorococcus</i> MED4	181044	38.0%	NC_015280	(7)
	P-HM2	<i>Prochlorococcus</i> MED4	183806	38.0%	NC_015284	(7)

Table S2. Information of metagenomes tested in this study.

Metagenome project	ID of sample tested in this study	Latitude & Longitude	Number of reads	Size fraction (µm)	Sampling region	Habitat	Collection date	Reference	Accession number at CAMERA
Global Ocean Sampling (GOS)	GS000a, GS000b, GS000c, GS000d, GS001a, GS001b, GS001c, GS002, GS003, GS004, GS006, GS007, GS008, GS009, GS010, GS011, GS012, GS013, GS014, GS015, GS016, GS017, GS018, GS019, GS021, GS022, GS023, GS025, GS026, GS027, GS028, GS029, GS031, GS034, GS035, GS036, GS037, GS040, GS042, GS043, GS047, GS048a, GS048b, GS049, GS050, GS051, GS108, GS108a, GS108b, GS109, GS110a, GS110b, GS111, GS112, GS112a, GS112b, GS113, GS114, GS115, GS116, GS117a, GS117b, GS119, GS120, GS121, GS122a, GS122b, GS123, GS148.	Global	10,999,635	0.1-0.8 (mostly)	Global oceans	Many typical marine environments, including estuary, coastal water, open ocean and reef	2004-2005	(10, 11)	CAM_PROJ_GOS
HOT	HF_SMPL_HOT179_125M_SG HF_SMPL_HOT179_25M_SG HF_SMPL_HOT179_75M_SG JGI_SMPL_HF10_10-07-02 JGI_SMPL_HF130_10-06-02 JGI_SMPL_HF200_10-06-02 JGI_SMPL_HF70_10-07-02	22.75N, 158.00W	477,879	0.22 - 1.6	Pacific gyre, HOT station	Open ocean	2002-2006	(12-14)	CAM_PROJ_HOT
MarineVirome	SCUMS_SMPL_BBC	49.705N, 124.35167W	416,456	0-0.2	Bay of British Columbia	Coastal water	1996-2004	(15)	CAM_PROJ_MarineVirome
	SCUMS_SMPL_SAR	32.166668N, 64.5W	399,343	0-0.2	Sargasso Sea	Open ocean	Jun-2005		
	SCUMS_SMPL_GOM	28.258333N, 87.67333W 26.205N, 93.941666W	263,908	0-0.2	Gulf of Mexico	Oceanic water	1996-2001		

		27.615N, 96.81167W 25.895N, 85.18W							
	SCUMS_SMPL_Arctic	70.69334N, 136.42334W 73.16167N, 159.44667W	688,590	0-0.2	Arctic Ocean	Coastal water	Sep-Oct-2002		
BroadPhage	CAM_SMPL_000966 CAM_SMPL_000994 CAM_SMPL_000995 CAM_SMPL_001000	38.967N, 76.378W	752,938	0-0.2	Chesapeake Bay	Estuary	30-Jul-2007		CAM_PROJ_BroadPhage
	CAM_SMPL_000722 CAM_SMPL_000723 CAM_SMPL_000724 CAM_SMPL_000725 CAM_SMPL_000726 CAM_SMPL_000727	32.867N, 117.257W	947,196	0-0.2	Scripps pier	Coastal water	7-Apr-2009	(16)	
	CAM_SMPL_000801 CAM_SMPL_000816 CAM_SMPL_000823 CAM_SMPL_000837	22.75N, 158.00W	985,754	0-0.2	Pacific gyre HOT station	Open ocean	11-Mar-2010		
	CAM_SMPL_000961 CAM_SMPL_000990	33.55N, 118.4W	519,987	0-0.2	Southern California Bight	Coastal water	9-Aug-2009		
	CAM_SMPL_000972 CAM_SMPL_001015	54.1873N, 7.8964E	690,747	0-0.2	Helgoland, North Sea	Coastal water	14-Apr-2009		
	CAM_SMPL_000999 CAM_SMPL_000987 CAM_SMPL_000959 CAM_SMPL_000980 CAM_SMPL_000958 CAM_SMPL_000988 CAM_SMPL_001013	45.00018S, 61.29991W 49.45S, 60.2825W 52.3638S, 60.16867W	1,361,811	0-0.2	Argentine Sea	Coastal water	30-Dec-2008		
	CAM_SMPL_001011	50.00N, 145.00W	325,857	0-0.2	Subarctic open ocean	Open ocean	9-Jun-2009	(17)	
	CAM_SMPL_000806	27.7737N, 82.6215W	202,505	0-0.2	Tampa Bay	Estuary	4-Mar-2009		

Note: For the BroadPhage and HOT projects, only samples collected from euphotic zone were tested in this study.

Table S3. Information of the six single-copy housekeeping genes of *Prochlorococcus* and marine *Synechococcus* genomes (N = 27) tested as reference sequences in metagenomic read recruitment.

Clade	Strain	Max. identity to other clade (%) ^a						Accession	Publication
		<i>recA</i>	<i>atpD</i>	<i>gyrB</i>	<i>rpoB</i>	<i>tuf</i>	<i>nrdJ</i>		
HLI	<i>Prochlorococcus marinus</i> MED4	88	84	83	84	87	83	BX548174	(18)
HLI	<i>Prochlorococcus marinus</i> MIT 9515	85	85	82	84	87	83	CP000552	(19)
HLII	<i>Prochlorococcus marinus</i> AS9601	87	84	83	84	87	83	CP000551	(19)
HLII	<i>Prochlorococcus marinus</i> MIT 9215	86	84	82	83	86	83	CP000825	(19)
HLII	<i>Prochlorococcus marinus</i> MIT 9301	88	85	83	84	87	83	CP000576	(19)
HLII	<i>Prochlorococcus marinus</i> MIT 9312	87	83	84	84	87	83	CP000111	(20)
LLI	<i>Prochlorococcus marinus</i> NATL1A	80	79	81	80	82	82	CP000553	(19)
LLI	<i>Prochlorococcus marinus</i> NATL2A	79	79	81	80	82	82	CP000095	(19)
LLII	<i>Prochlorococcus marinus</i> SS120	82	81	81	81	82	85	AE017126	(21)
LLIII	<i>Prochlorococcus marinus</i> MIT 9211	82	81	81	81	82	85	CP000878	(19)
LLIV	<i>Prochlorococcus marinus</i> MIT 9303	82	83	79	81	83	82	CP000554	(19)
LLIV	<i>Prochlorococcus marinus</i> MIT 9313	81	82	79	81	84	82	BX548175	(18)
Subcluster 5.1-I	<i>Synechococcus</i> sp. CC9311	82	83	82	82	88	77	CP000435	(22)
Subcluster 5.1-I	<i>Synechococcus</i> sp. WH 8016	81	85	82	82	88	80	AGIK00000000	unpublished
Subcluster 5.1-II	<i>Synechococcus</i> sp. CC9605	84	90	85	89	94	83	CP000110	(23)
Subcluster 5.1-II	<i>Synechococcus</i> sp. WH 8109	86	90	85	88	95	83	CP006882	unpublished
Subcluster 5.1-III	<i>Synechococcus</i> sp. WH 8102	86	90	85	89	88	82	BX548020	(24)
Subcluster 5.1-IV	<i>Synechococcus</i> sp. BL107	81	84	81	83	87	80	AATZ00000000	(23)
Subcluster 5.1-IV	<i>Synechococcus</i> sp. CC9902	80	85	80	83	88	79	CP000097	(23)
Subcluster 5.1-IX	<i>Synechococcus</i> sp. RS9916	85	89	85	87	95	80	AAUA00000000	(23)
Subcluster 5.1-V	<i>Synechococcus</i> sp. WH 7803	89	89	88	88	94	86	CT971583	(23)
Subcluster 5.1-VI	<i>Synechococcus</i> sp. WH 7805	89	89	88	88	94	86	AAOK00000000	(23)
Subcluster 5.1-VIII	<i>Synechococcus</i> sp. RS9917	87	86	86	87	90	83	AANP00000000	(23)
Subcluster 5.2	<i>Synechococcus</i> sp. WH 5701	87	86	84	84	87	82	AANO00000000	(23)
Subcluster 5.2	<i>Synechococcus</i> sp. CB0101	87	89	84	87	90	81	ADXL01000000	(25)
Subcluster 5.2	<i>Synechococcus</i> sp. CB0205	84	89	85	86	91	82	ADXM00000000	unpublished
Subcluster 5.3	<i>Synechococcus</i> sp. RCC307	86	89	82	86	91	80	CT978603	(23)

Average of the identities (%) ^a	84.5	85.3	83.0	84.3	88.0	82.2	
Standard deviation of the identities (%) ^a	3.0	3.4	2.4	2.8	4.1	2.0	

^a A sequence of one of the six single-copy housekeeping genes from a picocyanobacterial strain was compared to the sequences of this gene from strains in the other clades and the maximum identity was then recorded. The data show the maximum identities between sequences from different picocyanobacterial clades.

Figure S1

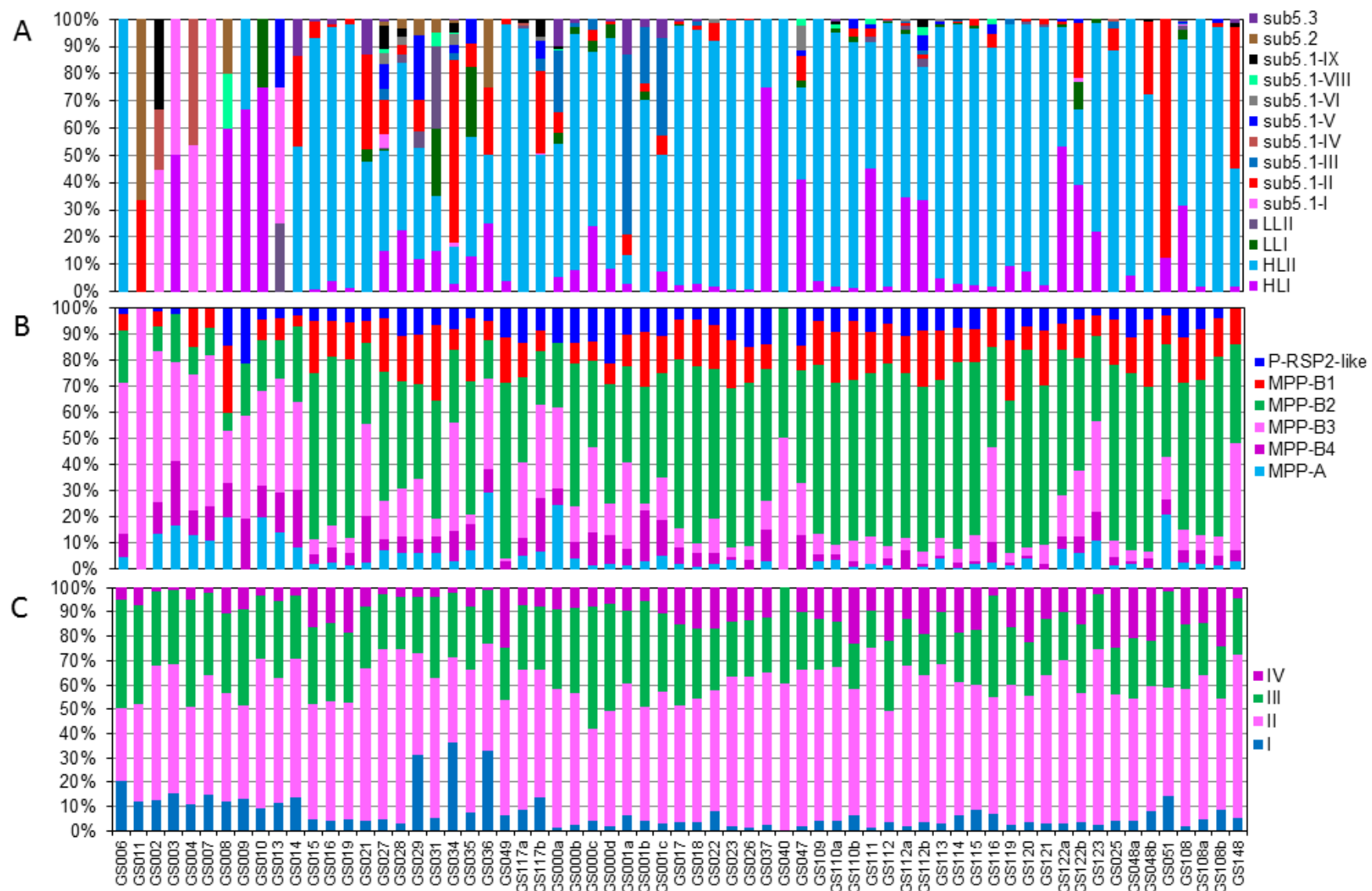


Figure S1. Relative abundances of picocyanobacterial (A), cyanopodovirus (B) and cyanomyovirus (C) genotypes at GOS sampling sites as revealed by metagenomic read recruitment against the GOS database. Four *Prochlorococcus* genotypes and 10 *Synechococcus* genotypes were found in the GOS database in this study. “sub5.1-I” stands for *Synechococcus* genotype Clade I in Marine Subcluster 5.1, and other terms follow the same naming rule.

Reference

1. **Labrie S, Frois - Moniz K, Osburne M, Kelly L, Roggensack S, Sullivan M, Gearin G, Zeng Q, Fitzgerald M, Henn M.** 2013. Genomes of marine cyanopodoviruses reveal multiple origins of diversity. *Environ. Microbiol.* **15**:1356-1376.
2. **Chen F, Lu J.** 2002. Genomic sequence and evolution of marine cyanophage P60: a new insight on lytic and lysogenic phages. *Appl. Environ. Microbiol.* **68**:2589-2594.
3. **Pope WH, Weigele PR, Chang J, Pedulla ML, Ford ME, Houtz JM, Jiang W, Chiu W, Hatfull GF, Hendrix RW, King J.** 2007. Genome sequence, structural proteins, and capsid organization of the cyanophage Syn5: a "horned" bacteriophage of marine synechococcus. *J. Mol. Biol.* **368**:966-981.
4. **Sullivan MB, Coleman ML, Weigele P, Rohwer F, Chisholm SW.** 2005. Three *Prochlorococcus* cyanophage genomes: signature features and ecological interpretations. *PLoS Biol.* **3**:e144.
5. **Mann NH, Clokie MR, Millard A, Cook A, Wilson WH, Wheatley PJ, Letarov A, Krisch HM.** 2005. The genome of S-PM2, a "photosynthetic" T4-type bacteriophage that infects marine *Synechococcus* strains. *J. Bacteriol.* **187**:3188-3200.
6. **Millard AD, Zwirgmaier K, Downey MJ, Mann NH, Scanlan DJ.** 2009. Comparative genomics of marine cyanomyoviruses reveals the widespread occurrence of *Synechococcus* host genes localized to a hyperplastic region: implications for mechanisms of cyanophage evolution. *Environ. Microbiol.* **11**:2370-2387.
7. **Sullivan MB, Huang KH, Ignacio-Espinoza JC, Berlin AM, Kelly L, Weigele PR, DeFrancesco AS, Kern SE, Thompson LR, Young S, Yandava C, Fu R, Krastins B, Chase M, Sarracino D, Osburne MS, Henn MR, Chisholm SW.** 2010. Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. *Environ. Microbiol.* **12**:3035-3056.
8. **Kelly L, Ding H, Huang KH, Osburne MS, Chisholm SW.** 2013. Genetic diversity in cultured and wild marine cyanomyoviruses reveals phosphorus stress as a strong selective agent. *ISME J.* **7**:1827-1841.
9. **Weigele PR, Pope WH, Pedulla ML, Houtz JM, Smith AL, Conway JF, King J, Hatfull GF, Lawrence JG, Hendrix RW.** 2007. Genomic and structural analysis of Syn9, a cyanophage infecting marine *Prochlorococcus* and *Synechococcus*. *Environ. Microbiol.* **9**:1675-1695.
10. **Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S, Wu D, Eisen JA, Hoffman JM, Remington K, Beeson K, Tran B, Smith H, Baden-Tillson H, Stewart C, Thorpe J, Freeman J, Andrews-Pfannkoch C, Venter JE, Li K, Kravitz S, Heidelberg JF, Utterback T, Rogers YH, Falcon LI, Souza V, Bonilla-Rosso G, Eguiarte LE, Karl DM, Sathyendranath S, Platt T, Bermingham E, Gallardo V, Tamayo-Castillo G, Ferrari MR, Strausberg RL, Neilson K, Friedman R, Frazier M, Venter JC.** 2007. The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol.* **5**:e77.
11. **Yooseph S, Sutton G, Rusch DB, Halpern AL, Williamson SJ, Remington K, Eisen JA, Heidelberg KB, Manning G, Li W.** 2007. The Sorcerer II Global Ocean Sampling expedition: expanding the universe of protein families. *PLoS Biol.* **5**:e16.
12. **DeLong EF, Preston CM, Mincer T, Rich V, Hallam SJ, Frigaard NU, Martinez A, Sullivan MB, Edwards R, Brito BR, Chisholm SW, Karl DM.** 2006. Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **311**:496-503.
13. **Frias-Lopez J, Shi Y, Tyson GW, Coleman ML, Schuster SC, Chisholm SW, Delong EF.** 2008. Microbial community gene expression in ocean surface waters. *Proc Natl Acad Sci U S A* **105**:3805-3810.
14. **Martinez A, Tyson GW, Delong EF.** 2010. Widespread known and novel phosphonate utilization pathways in marine bacteria revealed by functional

- screening and metagenomic analyses. *Environ. Microbiol.* **12**:222-238.
15. **Angly FE, Felts B, Breitbart M, Salamon P, Edwards RA, Carlson C, Chan AM, Haynes M, Kelley S, Liu H.** 2006. The marine viromes of four oceanic regions. *PLoS Biol.* **4**:e368.
 16. **Hurwitz BL, Deng L, Poulos BT, Sullivan MB.** 2013. Evaluation of methods to concentrate and purify ocean virus communities through comparative, replicated metagenomics. *Environ. Microbiol.* **15**:1428-1440.
 17. **Hurwitz BL, Sullivan MB.** 2013. The Pacific Ocean Virome (POV): a marine viral metagenomic dataset and associated protein clusters for quantitative viral ecology. *PLoS One* **8**:e57355.
 18. **Rocap G, Larimer FW, Lamerdin J, Malfatti S, Chain P, Ahlgren NA, Arellano A, Coleman M, Hauser L, Hess WR, Johnson ZI, Land M, Lindell D, Post AF, Regala W, Shah M, Shaw SL, Steglich C, Sullivan MB, Ting CS, Tolonen A, Webb EA, Zinser ER, Chisholm SW.** 2003. Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* **424**:1042-1047.
 19. **Kettler GC, Martiny AC, Huang K, Zucker J, Coleman ML, Rodrigue S, Chen F, Lapidus A, Ferriera S, Johnson J, Steglich C, Church GM, Richardson P, Chisholm SW.** 2007. Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*. *PLoS Genet.* **3**:e231.
 20. **Coleman ML, Sullivan MB, Martiny AC, Steglich C, Barry K, Delong EE, Chisholm SW.** 2006. Genomic islands and the ecology and evolution of *Prochlorococcus*. *Science* **311**:1768-1770.
 21. **Dufresne A, Salanoubat M, Partensky F, Artiguenave F, Axmann IM, Barbe V, Duprat S, Galperin MY, Koonin EV, Le Gall F, Makarova KS, Ostrowski M, Oztas S, Robert C, Rogozin IB, Scanlan DJ, Tandeau de Marsac N, Weissenbach J, Wincker P, Wolf YI, Hess WR.** 2003. Genome sequence of the cyanobacterium *Prochlorococcus marinus* SS120, a nearly minimal oxyphototrophic genome. *Proc Natl Acad Sci U S A* **100**:10020-10025.
 22. **Palenik B, Ren Q, Dupont CL, Myers GS, Heidelberg JF, Badger JH, Madupu R, Nelson WC, Brinkac LM, Dodson RJ, Durkin AS, Daugherty SC, Sullivan SA, Khouri H, Mohamoud Y, Halpin R, Paulsen IT.** 2006. Genome sequence of *Synechococcus* CC9311: Insights into adaptation to a coastal environment. *Proc Natl Acad Sci U S A* **103**:13555-13559.
 23. **Dufresne A, Ostrowski M, Scanlan DJ, Garczarek L, Mazard S, Palenik BP, Paulsen IT, de Marsac NT, Wincker P, Dossat C, Ferriera S, Johnson J, Post AF, Hess WR, Partensky F.** 2008. Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. *Genome Biol* **9**:R90.
 24. **Palenik B, Brahamsha B, Larimer FW, Land M, Hauser L, Chain P, Lamerdin J, Regala W, Allen EE, McCarren J, Paulsen I, Dufresne A, Partensky F, Webb EA, Waterbury J.** 2003. The genome of a motile marine *Synechococcus*. *Nature* **424**:1037-1042.
 25. **Marsan D, Wommack KE, Ravel J, Chen F.** 2014. Draft genome sequence of *Synechococcus* sp. strain CB0101, isolated from the Chesapeake Bay estuary. *Genome announcements* **2**:e01111-01113.