

Table S2. Environmental-PCR product identifiers and metadata from corresponding sampling sites. The *Bathycoccus* PRP8 full intein sequence from the targeted metagenome is on contig C1684, Genbank accession number: AFUW01000096.

Clone identifier	Genbank identifier	Insertion type	Insert length (nt)	Cruise name	Site	Latitude	Longitude	Sampling date	Depth (m)	Temp (°C)	NO ₃ +NO ₂ (μM)	PO ₄ (μM)	Salinity (psu)
TAi1PRP8	JX625117	intron	135	SJ0609	Tropical Atlantic	12°22'N	27°14'W	12-Jul-2006	64	21.7	0.2	0.00	35.90
TAi2PRP8	JX625118	intron	135	SJ0609	Tropical Atlantic	12°22'N	27°14'W	12-Jul-2006	64	21.7	0.2	0.00	35.90
TAi3cPRP8*	JX625119	intron	135	SJ0609	Tropical Atlantic	12°22'N	27°14'W	12-Jul-2006	64	21.7	0.2	0.00	35.90
TAi3dPRP8*	JX625119	intron	135	SJ0609	Tropical Atlantic	12°22'N	27°14'W	12-Jul-2006	64	21.7	0.2	0.00	35.90
TAfiPRP8	JX625120	intein	1686	SJ0609	Tropical Atlantic	12°22'N	27°14'W	12-Jul-2006	64	21.7	0.2	0.00	35.90
CNPi1PRP8	JX625121	intron	135	KM0715	Central N. Pacific (ALOHA)	22°45'N	158°00'W	17-Aug-2007	150	21.7	0.38	0.06	35.13
ENP1PRP8	JX625122	no insert	0	CN207	Eastern N. Pacific	37°7'N	123°29'W	10-Oct-2007	10	15.6	0.51	0.61	33.15
NAS1PRP8	JX625123	no insert	0	OC374	N. Atlantic Slope	39°44'N	69°39'W	5-Mar-2001	2	11.0	n.d.	n.d.	34.40
SS1PRP8	JX625124	no insert	0	OC374	Sargasso Sea (BATS)	32°10'N	64°30'W	13-Mar-2001	60	20.0	n.d.	n.d.	36.68
SS2PRP8	JX625125	no insert	0	OC374	Sargasso Sea (BATS)	32°10'N	64°30'W	13-Mar-2001	60	20.0	n.d.	n.d.	36.68
SS3PRP8	JX625126	no insert	0	OC374	Sargasso Sea (BATS)	32°10'N	64°30'W	13-Mar-2001	60	20.0	n.d.	n.d.	36.68
SSfi1PRP8	JX625127	intein	1665	OC374	Sargasso Sea (BATS)	32°10'N	64°30'W	13-Mar-2001	60	20.0	n.d.	n.d.	36.68
SSfi2PRP8	JX625128	intein	1686	OC374	Sargasso Sea (BATS)	32°10'N	64°30'W	13-Mar-2001	60	20.0	n.d.	n.d.	36.68
SSfi3PRP8	JX625129	intein	1686	OC374	Sargasso Sea (BATS)	32°10'N	64°30'W	13-Mar-2001	60	20.0	n.d.	n.d.	36.68

Abbreviations: *Sequences are 100% identical and deposited as a single Genbank entry, the corresponding PCR was run at 62°C annealing temperature; n.d., not determined; Sequence names: TA, tropical Atlantic Ocean (nearest land mass was Cape Verde); CNP, Central North Pacific (Station ALOHA), ENP, Eastern N. Pacific (CalCOFI line 67); NAS, North Atlantic Slope off USA; SS, Sargasso Sea, N. Atlantic Subtropical Gyre (at BATS site); codes in sequence names following site specification: i, intron containing; fi, full intein containing.