

Supplemental Table 1: Nucleic acid sequences obtained from MCRA. The table gives the designation, GenBank accession number, e-value, and similarity (in %) to the closest hit in the core and environmental database of GenBank based on tblastx comparisons of nucleic acid sequences obtained from MCRA. Data on location and sample depth for the hits in env_nt were retrieved from the CAMERA database.

Designation (GenBank)	Database					
	nr			env_nt		
	e-value	Similarity	Hit	e-value	Similarity	Hit
VIII-1_3 (ET671172)	3E-60	81.1	AP007255 (Magnetospirillum magneticum AMB-1; leucyl-tRNA synthetase)	2E-126	97.4	AACY023198799 (Mar. metagen. ctg_1101668006150; Gulf of Maine, North American East Coast, coastal, 1 m depth)
VIII-1_4 (ET671173)	3E-33	56-75	CP000377 (Silicibacter sp. TM1040; RNA-binding S4)	5E-89	77.5-92.6	AACY020035295 (Mar. metagen. 1096626807910; Browns Bank, Gulf of Maine, North American East Coast, coastal, 1 m depth)
VIII-1_5 (ET671174)	1E-17	47.2-56.8	BX294155 (Rhodopirellula baltica SH 1; probably glutamine amidotransferase)	7E-68	98.1-100	AACY020398708 (Mar. metagen. 1096626733890; station 11, Sargasso Sea, open ocean, 5 m depth)
VIII-2_1 (ET671175)	1E-60	25.6-82.4	AB260935 (Uncult. bacterium DNA; salt-stress tolerance related region)	5E-70	47.9-95.8	AACY023636459 (Mar. metagen. ctg_1101668443810; Galapagos Islands, warm seep, Roca Redonda, 19 m depth)

VIII-3_2 (ET671176)	1E-12	33.3-69	CP000679 (Caldicellulosiruptor saccharolyticus DSM 8903; aminotransferase)	3E-20	49.3-71.4	ABEF01050569 (Mar. metagen. HOTS_Contig50569; North Pacific Subtropical Gyre, Hawaii Ocean Time Series station ALOHA, 4000 m depth)
VIII-4_3 (ET671177)	7E-30	33.9-68.4	CP000774 (Parvibaculum lavamentivorans DS-1; anion transporter)	3E-30	39.4-60.5	AACY023246928 (Mar. metagen. ctg_1101668054279; Newport Harbor RI, North American East Coast, coastal, 1 m depth)
VIII-4_4 (ET671178)	-	-	-	6E-22	59.5-89.1	ABEF01044912 (Mar. metagen. HOTS_Contig44912; North Pacific Subtropical Gyre, Hawaii Ocean Time Series station ALOHA, 4000 m depth)
VIII-4_5 (ET671179)	3E-16	29-63.3	CP000804 (Roseiflexus castenholzii DSM 13941; aldehyde dehydrogenase)	2E-16	55-63.9	AACY022768255 (Mar. metagen. ctg_1101667175606; Panama Canal, Lake Gatun, fresh water, 2 m depth)
					43.6-61	AACY023450775 (Mar. metagen. ctg_1101668258126; Gulf of Panama, Eastern Tropical Pacific, coastal, 1.6 m depth)
VIII-6_2 (ET671180)	1E-40	88.9-100	AY822609 (Prorocentrum micans; large subunit rRNA)	7E-32	71.3-91.1	AACY022407764 (Mar. metagen. 1933981; hydrostation S, Sargasso Sea, open ocean, 5 m depth)
IX-1_3 (ET671171)	-	-	-	-	-	-
X-1_2 (ET671181)	2E-87	62-78.7	COP000435 (Synechococcus sp. CC9311; hypothetical protein)	7E-79	42.5-75.8	AACY023218879 (Mar. metagen. ctg_1101668026230; outside Halifax, Nova Scotia, coastal, 2 m depth)

X-2_1 (ET671182)	4E-11	34.5	BA000040 (Bradyrhizobium japonicum USDA 110; probable mandelate racemase)	2E-25	44.8	AACY023248598 (Mar. metagen. ctg_1101668055949; Newport Harbor RI, North American East Coast, coastal, 1 m depth)
X-3_1 (ET671183)	1E-61	31.1-65.5	CP000356 (Sphingopyxis alaskensis RB2256; leucyl- tRNA synthetase)	2E-120	85.3-96.8	AACY023256279 (Mar. metagen. ctg_1101668063630; Newport Harbor RI, North American East Coast, coastal, 1 m depth)
X-3_2 (ET671184)	1E-17	48.5-57.9	CR378663 (Photobacterium profundum SS9; putative glycyl-tRNA synthetase, beta chain)	5E-65	77.3-100	AACY022263224 (Mar. metagen. 1091141390923; Newport Harbor RI, North American East Coast, coastal, 1 m depth)
X-3_3 (ET671185)	3E-25	47.8-97.3	AY455824 (Alexandrium lusitanicum; 181T large subunit rRNA)	3E-25	51.7-94.6	AACY022511047 (Mar. metagen. 1966319; hydrostation S, Sargasso Sea, open ocean, 5 m depth)
X-3_4 (ET671186)	-	-	-	3E-27	41.5-81.8	AACY023206424 (Mar. metagen. ctg_1101668013775; Gulf of Maine, North American East Coast, coastal, 1 m depth)

Supplemental Table 2: Nucleic acid sequences obtained from MOPA. The table gives the designation, GenBank accession number, e-value, and similarity (in %) to the closest hit in the core and environmental database of GenBank based on tblastx comparisons of nucleic acid sequences obtained from MOPA. Data on location and sample depth for the hits in env_nt were retrieved from the CAMERA database.

Designation (GenBank)	Database					
	nr			env_nt		
	e-value	Similarity	Hit	e-value	Similarity	Hit
V-1_1 (ET671187)	5E-104	47.5-100	CP000435 (<i>Synechococcus</i> sp. CC9311; possible methylpurine-DNA glycosylase)	5E-75	31-90.3	AACY023868928 (Mar. metagen. ctg_1101668676279; North Seamore Island, Galapagos Islands, coastal, 2.1 m depth)
V-1_5 (ET671188)	-	-	-	-	-	-
V-2_1 (ET671189)	7E-20	23.1-44.7	AM412317 (<i>Clostridium botulinum</i> A str. ATCC 3502; sodium/solute symporter)	9E-46	43.2-54.6	ABEF01031803 (Mar. metagen. HOTS_Contig31803; North Pacific Subtropical Gyre, Hawaii Ocean Time Series station ALOHA, 4000 m depth)
V-2_2 (ET671190)	-	-	-	-	-	-
V-3_1-1 (ET671191)	3E-11	34.2-57.7	AP009552 (<i>Microcystis aeruginosa</i> NIES-843; HEAT domain protein repeat-containing protein)	4E-07	36.8-45.2	AACY023785540 (Mar. metagen. ctg_1101668592891; mangrove on Isabella Island, Galapagos Islands, 0.1 m depth)

V-3_1-2 (ET671192)	3E-23	35.7-63.3	AM286690 (Alcanivorax borkumensis SK2; TRAP-type transport system, large permease component)	6E-72	36.8-85.3	AACY020453397 (Mar. metagen. 1096626604794; hydrostation S, Sargasso Sea, open ocean, 5 m depth)
V-4_3 (ET671193)	4E-54	23.1-58.3	CP000453 (Alkalilimnicola ehrlichei MLHE-1; lipopolysaccharide heptosyltransferase II)	1E-48	38.9-54.3	AAFZ01019829 (Mar. metagen. AHAA7073.y1; microbial mat from gray whale carcass in the Pacific Ocean, 1674 m depth, Santa Cruz Basin)
					32.7-55	AACY022537614 (Mar. metagen. 475709; Sargasso Station 11, Sargasso Sea, open ocean, 5 m depth)
V-5_1 (ET671194)	3E-109	48.6-93.6	CP000435 (Synechococcus sp. CC9311; possible methylpurine-DNA glycosylase)	2E-80	35.7-84	AACY023868928 (Mar. metagen. ctg_1101668676279; North Seamore Island, Galapagos Islands, coastal, 2.1 m depth)
V-6_2 (ET671195)	-	-	-	1E-15	48.8-54.9	AACY020754335 (Mar. metagen. 1021535; Sargasso Stations 3, Sargasso Sea, open ocean, 5 m depth)
V-7_1 (ET671196)	3E-19	23.1-46.2	AM412317 (Clostridium botulinum A str. ATCC 3502; sodium/solute symporter)	1E-40	40.5-59.5	ABEF01031803 (Mar. metagen. HOTS_Contig31803; North Pacific Subtropical Gyre, Hawaii Ocean Time Series station ALOHA, 4000 m depth)
V-7_2 (ET671197)	-	-	-	-	-	-

V-8_1 (ET671198)	1E-80	23.2-74.4	AY458634 (Uncultured marine bacterium 314 clone EBAC750-09G06)	1E-93	31.1-78.6	AACY020083014 (Mar. metagen. 1096626095075; off Key West FL, Caribbean Sea, coastal, 1.7 m depth)
VI-1_3 (ET671199)	3E-07	40-44.8	CP000471 (Magnetococcus sp. MC-1; putative outer membrane adhesin like protein)	7E-25	40-80	AACY022498683 (Mar. metagen. 1095515462673; upwelling, Fernandina Island, Galapagos Islands, coastal upwelling, 12 m depth)
VI-3_1 (ET671200)	4E-57	29.9-62.5	BA000012 (Mesorhizobium loti MAFF303099; carbon-monoxide dehydrogenase medium chain)	2E-69	35.9-70.6	AACY023726705 (Mar. metagen. ctg_1101668534056; upwelling, Fernandina Island, Galapagos Islands, coastal upwelling, 12 m depth)
VI-3_2 (ET671201)	3E-23	35.7-63.3	AM286690 (Alcanivorax borkumensis SK2; TRAP-type transport system, large permease component)	7E-71	36.8-84.6	AACY020453397 (Mar. metagen. 1096626604794; hydrostation S, Sargasso Sea, open ocean, 5 m depth)
VI-4_2 (ET671202)	1E-05	34.2	BA000045 (Gloeobacter violaceus PCC 7421; glr1803)	9E-08	51-70	ABEF01024173 (Mar. metagen. HOTS_Contig24173; North Pacific Subtropical Gyre, Hawaii Ocean Time Series station ALOHA, 4000 m depth)
VII-1_4 (ET671203)	7E-110	45.9-92.7	CP000435 (Synechococcus sp. CC9311; possible methylpurine-DNA glycosylase)	2E-80	33.3-84	AACY023868928 (Mar. metagen. ctg_1101668676279; North Seamore Island, Galapagos Islands, coastal, 2.1 m depth)
VII-1_5 (ET671204)	3E-06	61.5-86.7	AC136729 (Mus musculus chromosome 3 clone RP23-175I24)	-	-	-
		46.7-100	AC007381 (Homo sapiens BAC clone RP11-158I21 from 2)			
VII-1_6 (ET671205)	-	-	-	-	-	-

VII-2_1 (ET671206)	9E-112	48.6-93.6	CP000435 (Synechococcus sp. CC9311; possible methylpurine-DNA glycosylase)	9E-81	35.7-84.9	AACY023868928 (Mar. metagen. ctg_1101668676279; North Seamore Island, Galapagos Islands, coastal, 2.1 m depth)
VII-2_3 (ET671207)	1E-30	48.8-66.3	CP000362 (Roseobacter denitrificans OCh 114; glutathione synthase)	2E-44	93.7-100	AACY020456071 (Mar. metagen. 1096626608810; Northern Gulf of Maine, North American East Coast, coastal, 1 m depth)
VII-3_1 (ET671208)	7E-22	25.4-43	AM412317 (Clostridium botulinum A str. ATCC 3502; sodium/solute symporter)	5E-47	37.8-56.3	ABEF01031803 (Mar. metagen. HOTS_Contig31803; North Pacific Subtropical Gyre, Hawaii Ocean Time Series station ALOHA, 4000 m depth)
VII-3_2 (ET671209)	6E-19	34.6-48.4	CP000820 (Frankia sp. EAN1pec; amidohydrolase 3)	4E-56	73.2-96.9	ABEF01017894 (Mar. metagen. HOTS_Contig17894; North Pacific Subtropical Gyre, Hawaii Ocean Time Series station ALOHA, 4000 m depth)

Supplemental Table 3: Nucleic acid sequences obtained from VCRA. The table gives the designation, GenBank accession number, e-value, and similarity (in %) to the closest hit in the core and environmental database of GenBank based on tblastx comparisons of nucleic acid sequences obtained from VCRA. Data on location and sample depth for the hits in env_nt were retrieved from the CAMERA database.

Designation (GenBank)	Database					
	nr			env_nt		
	e-value	Similarity	Hit	e-value	Similarity	Hit
I-1_2 (ET671210)	3E-95	31.1-50	DQ395179 (<i>Bartonella tami</i> strain; Th239 heat shock protein (groEL) gene)	3E-126	31.8-79.4	AACY023386073 (Mar. metagen. ctg_1101668193424; off Key West FL, Caribbean Sea, coastal, 1.7 m depth)
I-1_3 (ET671211)	-	-	-	-	-	-
I-1_5 (ET671212)	-	-	-	2E-43	35.9-56.3	AACY023952884 (Mar. metagen. ctg_1101668760235; hydrostation S, Sargasso Sea, open ocean, 5 m depth)
I-1_6 (ET671213)	1E-09	51.2-65	CP000394 (<i>Granulibacter bethesdensis</i> CGDNIH1; peptide methionine sulfoxide reductase msrB)	1E-39	45.2-95	AACY020332633 (Mar. metagen. 1096626412155; 201 miles from F. Polynesia, Tropical South Pacific, open ocean, 30 m depth)
I-2_2 (ET671214)	-	-	-	1E-65	44-80.5	AACY020622849 (Mar. metagen. 3005708; hydrostation S, Sargasso Sea, open ocean, 5 m depth)

I-3_2 (ET671215)	7E-07	38.2-46.3	AY940168 (Cyanophage P-SSM4)	2E-17	34.7-56.5	ABEF01038247 (Mar. metagen. HOTS_Contig38247; North Pacific Subtropical Gyre, Hawaii Ocean Time Series station ALOHA, 4000 m depth)
I-3_3 (ET671216)	-	-	-	7E-29	41.9-62.5	AACY023145713 (Mar. metagen. ctg_1101667553064; Sargasso Station 13, Sargasso Sea, open ocean, 5 m depth)
I-4_1 (ET671217)	-	-	-	2E-65	37.9-79.7	AACY020622849 (Mar. metagen. 3005708; hydrostation S, Sargasso Sea, open ocean, 5 m depth)
I-4_2-1 (ET671218)	-	-	-	8E-29	57.7-100	AACY023043482 (Mar. metagen. ctg_1101667450833; information not available)
I-4_2-2 (ET671219)	5E-22	39-69.1	CP000084 (Candidatus Pelagibacter ubiquus HTCC1062; glycine cleavage system protein P2 gcvP)	3E-43	80.8-98.6	ABEF01053919 (Mar. metagen. HOTS_Contig53919; North Pacific Subtropical Gyre, Hawaii Ocean Time Series station ALOHA, 4000 m depth)
I-5_4-2 (ET671220)	-	-	-	3E-09	32.5-56.8	AACY020468158 (Mar. metagen. 1096626625736; Block Island NY, North American East Coast, coastal, 1 m depth)
I-6_1 (ET671221)	4E-05	27.6-75	XM_814922 (Trypanosoma cruzi strain CL Brener; mucin associated surface protein)	2E-30	47.4-63.3	AACY020067345 (Mar. metagen. 1096626076666; off Nags Head NC, North American East Coast, coastal, 2.1 m depth)

I-6_2 (ET671222)	2E-19	34-62.1	CP000884 (<i>Delftia acidovorans</i> SPH-1; constituent protein, prophage)	5E-49	39.3-75.9	AACY022660460 (Mar. metagen. ctg_1101667067811; off Nags Head NC, North American East Coast, coastal, 2.1 m depth)
I-6_3 (ET671223)	-	-	-	9E-15	57.6-57.9	AACY022079556 (Mar. metagen. 1093015312178; off Nags Head NC, North American East Coast, coastal, 2.1 m depth)
I-7_2 (ET671224)	-	-	-	-	-	-
I-9_1 (ET671225)	-	-	-	2E-08	44.5-56.5	AACY021905291 (Mar. metagen. 1092344255434; Dirty Rock, Cocos Island, Eastern Tropical Pacific, fringing reef, 1.1 m depth)
II-1_2 (ET671226)	-	-	-	4E-20	51.2-60.4	AACY023258480 (Mar. metagen. ctg_1101668065831; Newport Harbor RI, North American East Coast, coastal, 1 m depth)
II-1_3 (ET671227)	-	-	-	2E-26	28.8-81	AACY020330726 (Mar. metagen. 1096626410185; 201 miles from F. Polynesia, Tropical South Pacific, open ocean, 30 m depth)
II-1_4 (ET671228)	-	-	-	2E-37	58.6-90.5	AACY022045099 (Mar. metagen. 1092969800169; Rangirora Atoll, Polynesia Archipelagos, coral reef atoll, 1 m depth)
II-3_3 (ET671229)	-	-	-	-	-	-

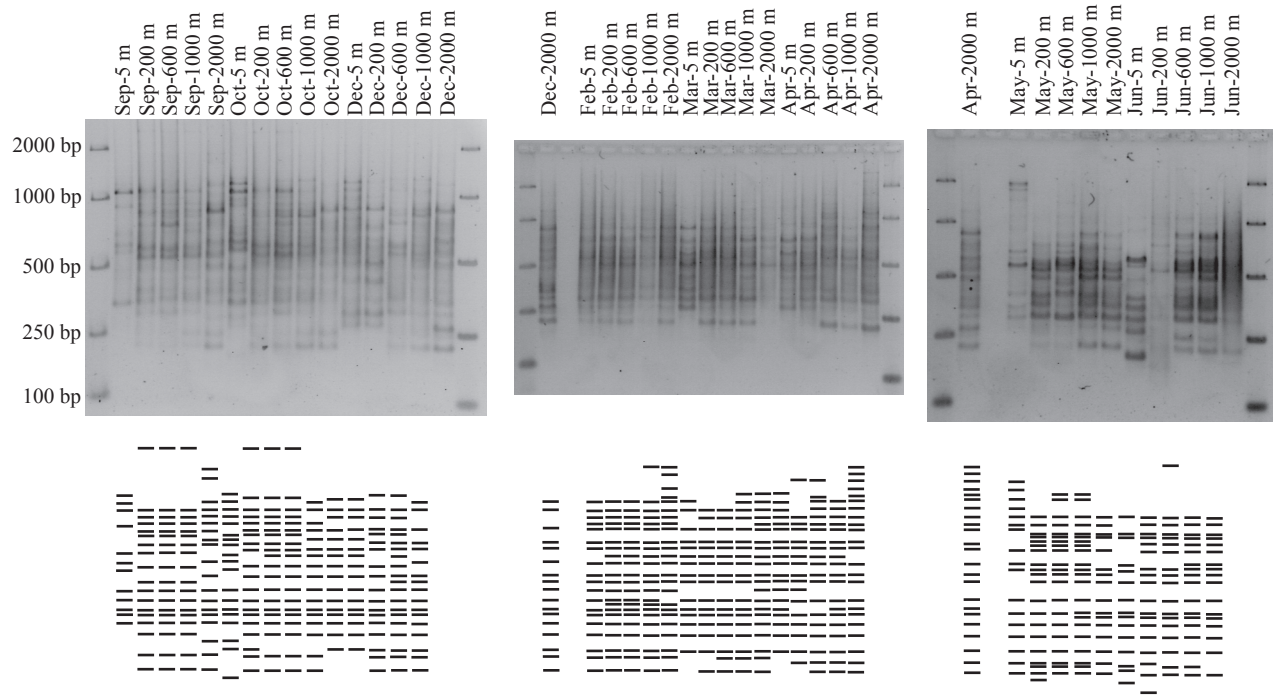
II-3_4-1 (ET671230)	-	-	-	4E-50	41-72.6	AACY021634592 (Mar. metagen. 1182675; Sargasso Stations 3, Sargasso Sea, open ocean, 5 m depth)
II-3_4-2 (ET671231)	1E-52	38-94.6	CP000435 (Synechococcus sp. CC9311; mannose-1- phosphate guanylyltransferase/mannose-6- phosphate isomerase)	1E-51	58.2-100	AACY022611423 (Mar. metagen. ctg_1101667018774; outside Halifax, Nova Scotia, North American East Coast, coastal, 2 m depth)
II-5_2 (ET671232)	-	-	-	3E-66	44-83.5	AACY020622849 (Mar. metagen. 3005708; hydrostation S, Sargasso Sea, open ocean, 5 m depth)

Supplemental Table 4: Nucleic acid sequences obtained from VOPA. The table gives the designation, GenBank accession number, e-value, and similarity (in %) to the closest hit in the core and environmental database of GenBank based on tblastx comparisons of nucleic acid sequences obtained from VOPA. Data on location and sample depth for the hits in env_nt were retrieved from the CAMERA database.

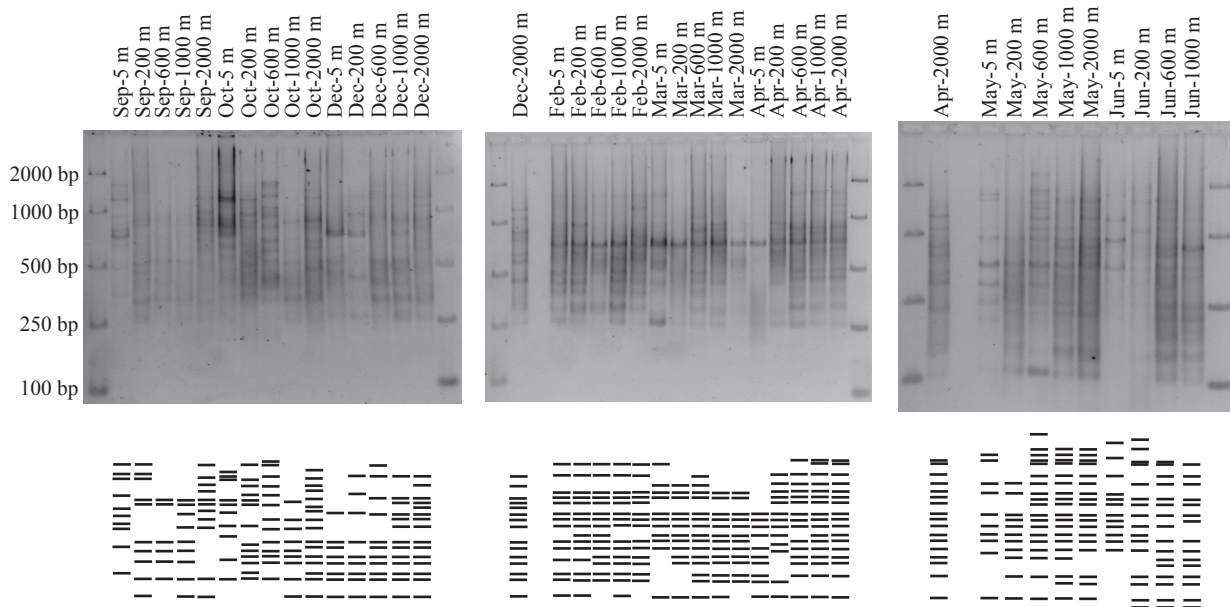
Designation (GenBank)	Database					
	nr			env_nt		
	e-value	Similarity	Hit	e-value	Similarity	Hit
III-1_3 (ET671236)	-	-	-	9E-54	52.1-75.7	AACY020456584 (Mar. metagen. 1096626609421; outside Halifax, Nova Scotia, North American East Coast, coastal, 2 m depth)
III-2_3 (ET671237)	-	-	-	1E-27	65.6-95.2	AACY023731909 (Mar. metagen. ctg_1101668539260; upwelling, Fernandina Island, Galapagos Islands, coastal upwelling, 12 m depth)
III-2_4 (ET671238)	-	-	-	2E-18	35.1-56.6	AACY023354615 (Mar. metagen. ctg_1101668161966; off Nags Head NC, North American East Coast, coastal, 2.1 m depth)
III-3_2 (ET671239)	-	-	-	-	-	-
III-4_1 (ET671240)	-	-	-	3E-23	40.5-50	AACY021537048 (Mar. metagen. 1092343162191; Dirty Rock, Cocos Island, Eastern Tropical Pacific, fringing reef, 1.1 m depth)

III-5_2 (ET671241)	-	-	-	7E-09	56.3-90	AACY022469673 (Mar. metagen. 1095403453722; coastal Floreana, Galapagos Islands, coastal, 2 m depth)
III-6_2 (ET671242)	-	-	-	7E-06	62.5	AACY021162966 (Mar. metagen. 1095901088495; 201 miles from F. Polynesia, Tropical South Pacific, open ocean, 30 m depth)
III-7_2 (ET671243)	-	-	-	-	-	-
III-8_1 (ET671244)	-	-	-	-	-	-
III-8_2 (ET671245)	8E-08	32.8-43.9	AM774415 (Halobacterium salinarum R1; chitinase)	5E-11	43	AACY021391124 (Mar. metagen. 1633614; Sargasso Station 13, Sargasso Sea, open ocean, 5 m depth)
III-9_1 (ET671246)	-	-	-	-	-	-
III-10_1 (ET671233)	1E-06	41.5	CP000158 (Hyphomonas neptunium ATCC 15444; deoxyuridine 5'-triphosphate nucleotidohydrolase)	5E-23	35.2-72.1	AACY023689306 (Mar. metagen. ctg_1101668496657; upwelling, Fernandina Island, Galapagos Islands, coastal upwelling, 12 m depth)
III-12_1 (ET671234)	-	-	-	-	-	-
III-12_2 (ET671235)	-	-	-	2E-12	30.2-70.7	AACY020095727 (Mar. metagen. 1096626109423; Gulf of Mexico, Caribbean Sea, coastal sea, 2 m depth)

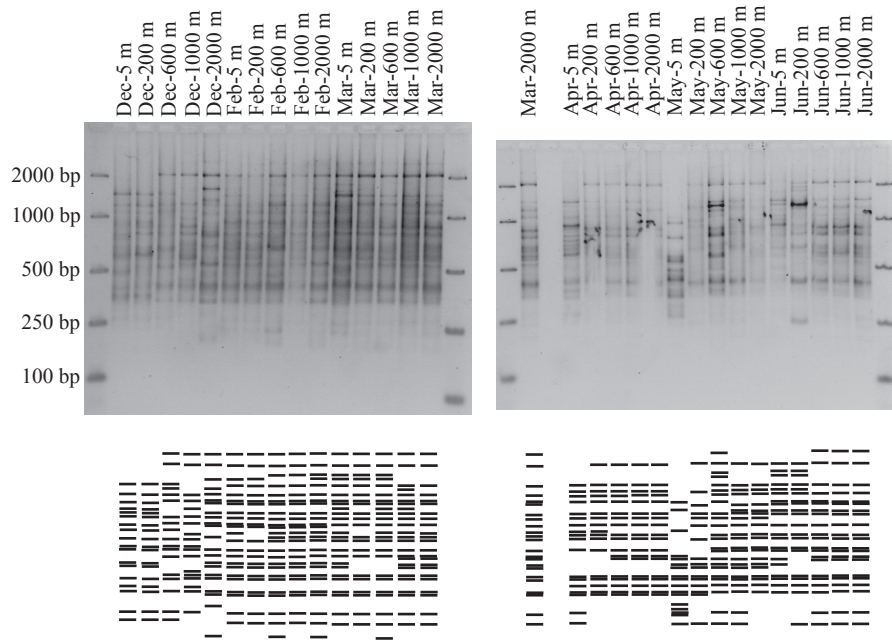
IV-1_1 (ET671247)	2E-13	47.4-57.1	AE015451 (Pseudomonas putida KT2440; tail tubular protein B, prophage)	2E-60	37.2-69.2	AACY023199791 (Mar. metagen. ctg_1101668007142; Gulf of Maine, North American East Coast, coastal, 1 m depth)
IV-1_2 (ET671248)	-	-	-	-	-	-
IV-1_3 (ET671249)	-	-	-	-	-	-
IV-1_4 (ET671250)	-	-	-	-	-	-
IV-2_1 (ET671251)	-	-	-	-	-	-
IV-3_2 (ET671252)	-	-	-	3E-42	55.4-82.1	ABEF01000394 (Mar. metagen. HOTS_Contig394; North Pacific Subtropical Gyre, Hawaii Ocean Time Series station ALOHA, 4000 m depth)
IV-5_1 (ET671253)	-	-	-	2E-46	51.9-64.9	ABEF01018562 (Mar. metagen. HOTS_Contig18562; North Pacific Subtropical Gyre, Hawaii Ocean Time Series station ALOHA, 4000 m depth)
IV-6_1 (ET671254)	3E-12	36.1	CP000884 (Delftia acidovorans SPH-1; hypothetical protein-constituent protein, prophage)	1E-37	34.9-81.6	AACY021762271 (Mar. metagen. 1092963713904; off Nags Head NC, North American East Coast, coastal, 2.1 m depth)



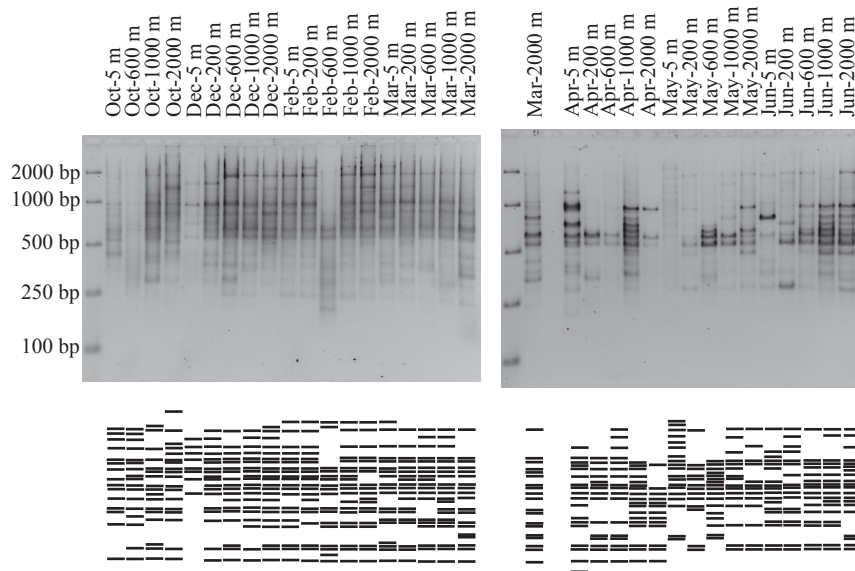
Supplemental Figure 1: Gel images and corresponding drawings of the banding patterns of MCRA. The figure shows original gel images and the drawings of the banding patterns that were used to obtain the schematic representation of the RAPD-PCR banding patterns shown in Suppl. Fig. 5. Negative controls (PCR amplifications without target DNA) did not indicate contamination and are not shown.



Supplemental Figure 2: Gel images and corresponding drawings of the banding patterns of MOPA. The figure shows original gel images and the drawings of the banding patterns that were used to obtain the schematic representation of the RAPD-PCR banding patterns shown in Suppl. Fig. 6. Negative controls (PCR amplifications without target DNA) did not indicate contamination and are not shown.

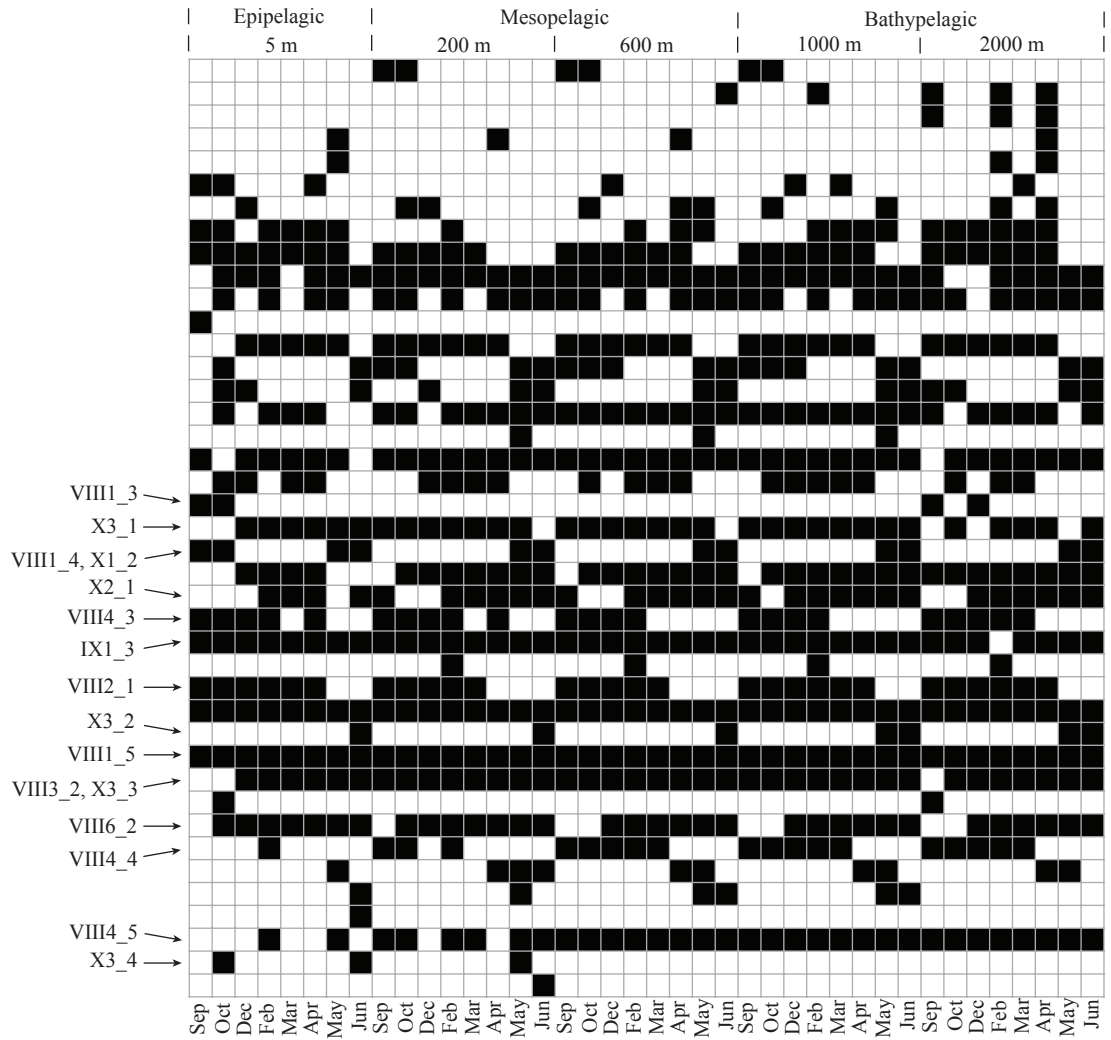


Supplemental Figure 3: Gel images and corresponding drawings of the banding patterns of VCRA. The figure shows original gel images and the drawings of the banding patterns that were used to obtain the schematic representation of the RAPD-PCR banding patterns shown in Suppl. Fig. 7. Negative controls (PCR amplifications without target DNA) did not indicate contamination and are not shown.



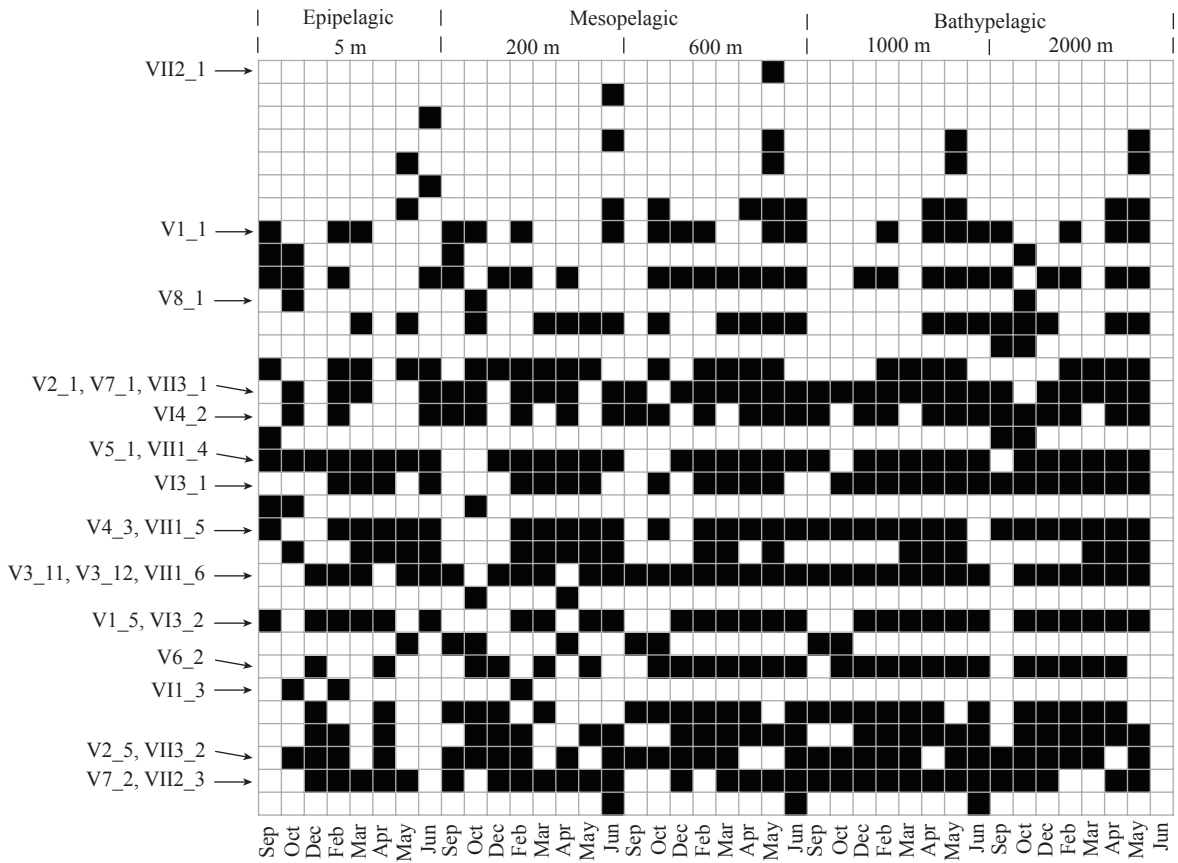
Supplemental Figure 4: Gel images and corresponding drawings of the banding patterns of VOPA. The figure shows original gel images and the drawings of the banding patterns that were used to obtain the schematic representation of the RAPD-PCR banding patterns shown in Suppl. Fig. 8. Negative controls (PCR amplifications without target DNA) did not indicate contamination and are not shown.

MCRA



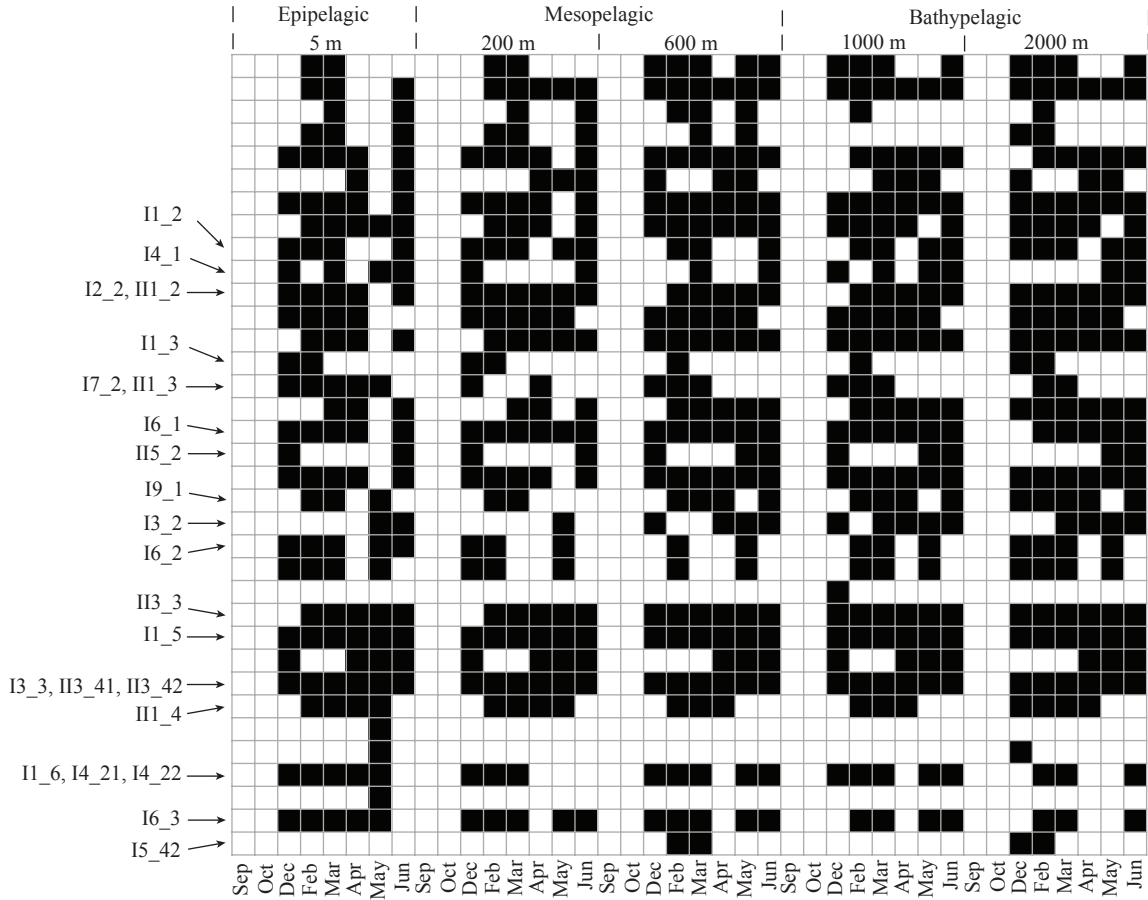
Supplemental Figure 5: Schematic representation of the RAPD-PCR banding patterns obtained from MCRA. Presence and absence of bands is indicated by black and white squares, respectively. Additionally, the figure indicates the position and designation of excised and subsequently sequenced bands.

MOPA



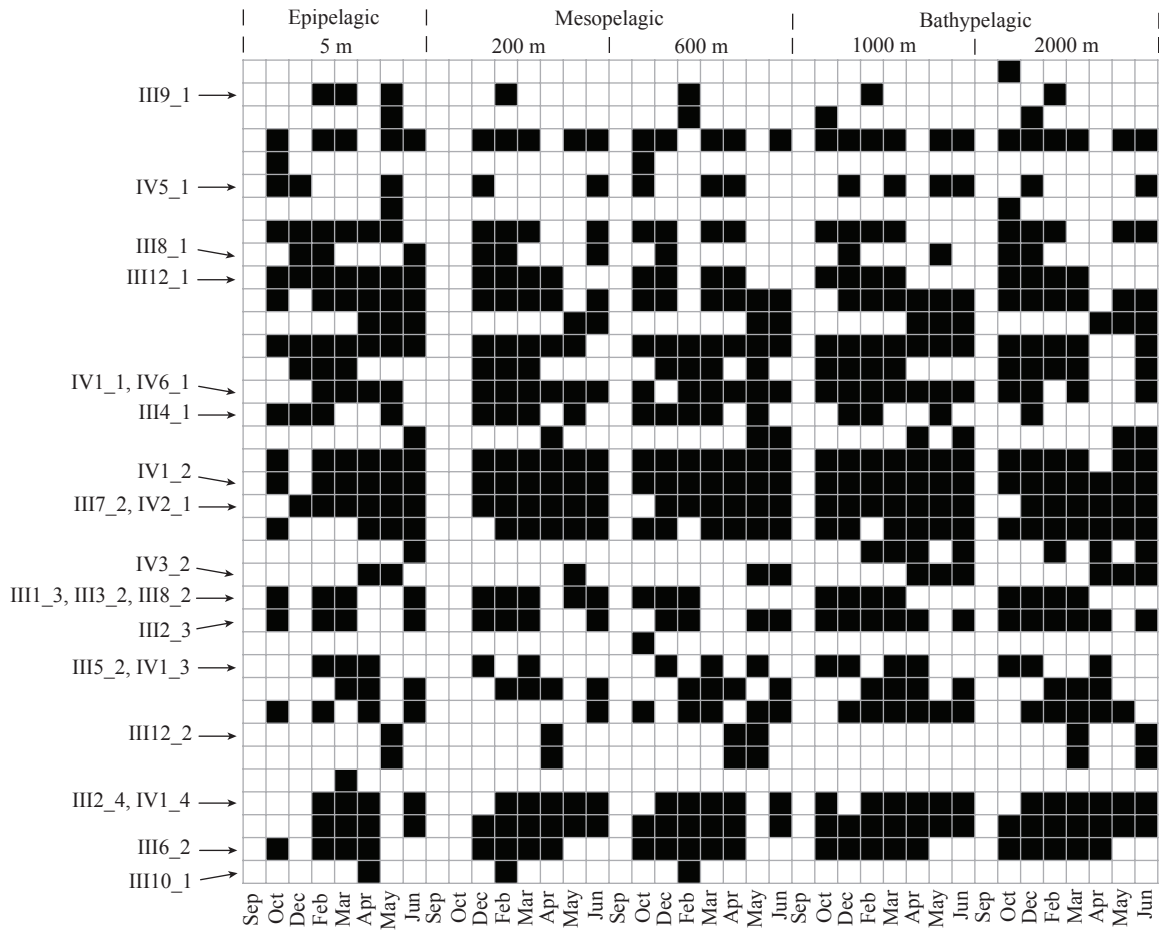
Supplemental Figure 6: Schematic representation of the RAPD-PCR banding patterns obtained from MOPA. Presence and absence of bands is indicated by black and white squares, respectively. Additionally, the figure indicates the position and designation of excised and subsequently sequenced bands.

VCRA



Supplemental Figure 7: Schematic representation of the RAPD-PCR banding patterns obtained from VCRA. Presence and absence of bands is indicated by black and white squares, respectively. Additionally, the figure indicates the position and designation of excised and subsequently sequenced bands.

VOPA



Supplemental Figure 8: Schematic representation of the RAPD-PCR banding patterns obtained from VOPA. Presence and absence of bands is indicated by black and white squares, respectively. Additionally, the figure indicates the position and designation of excised and subsequently sequenced bands.