Supplementary Information

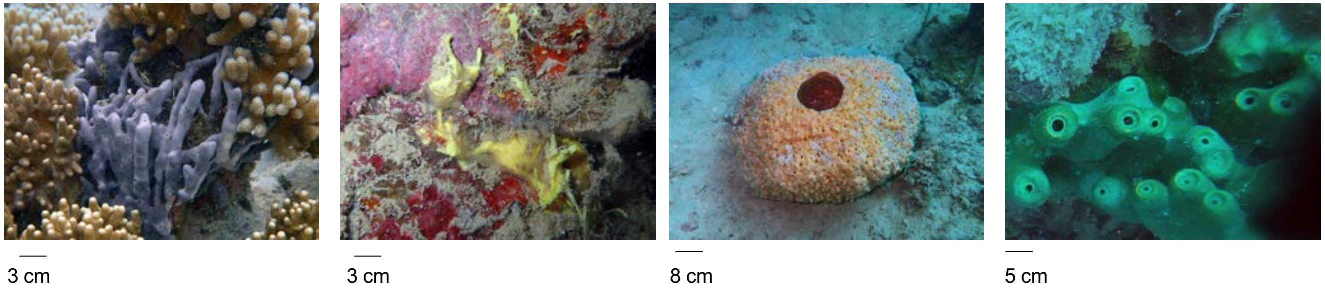
Pangenomic Comparison of Globally Distributed Poribacteria Associated with Sponge Hosts and Marine Particles

Sheila Podell, Jessica M. Blanton, Alexander Neu, Vinayak Agarwal, Jason S. Biggs, Bradley S. Moore, Eric E. Allen

Contents

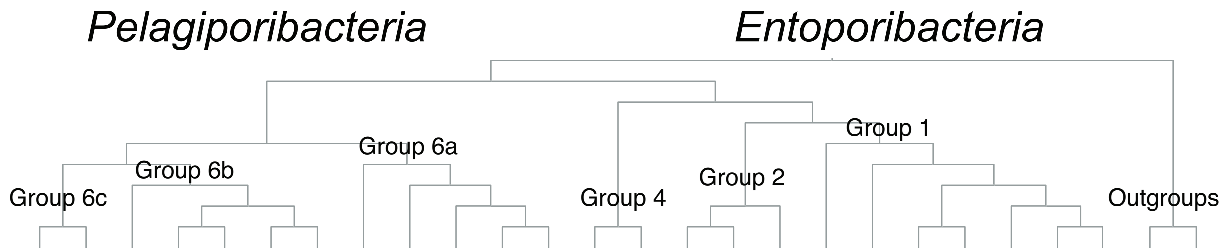
Supplementary Figure S1. **Sponge host sample characteristics.** *Dysidea granulosa, Pseudoceratina sp.,* and *Melophlus* sarasinorum photos by Dr. Jason Biggs, University of Guam. *Agelas tubulata* photo by Tamara Schwent of the Sirenas Collection team.

	Dysidea granulosa	***Pseudoceratina sp.***	***Melophlus sarasinorum***	***Agelas tubulata***
Sample name	GUM098	GUM146	GUM22	SBM112
NCBI Biosample	SAMN04939370	SAMN08482400	SAMN08501426	SAMN08501975
collection date	7/1/2015	7/7/2015	6/29/2015	1/13/2014
sample location	Piti Bomb Holes Guam	San Louis Beach Santa Rita, Guam	Western Shoals Apra Harbor, Guam	St. Thomas Virgin Islands
lattitude	13.345738 N	13.4428 N	13.451324 N	18.309517 N
longitude	144.638186 E	144.6487 E	144.655623 E	64.96135 W
depth	5 ft	28 ft	29 ft	40 ft



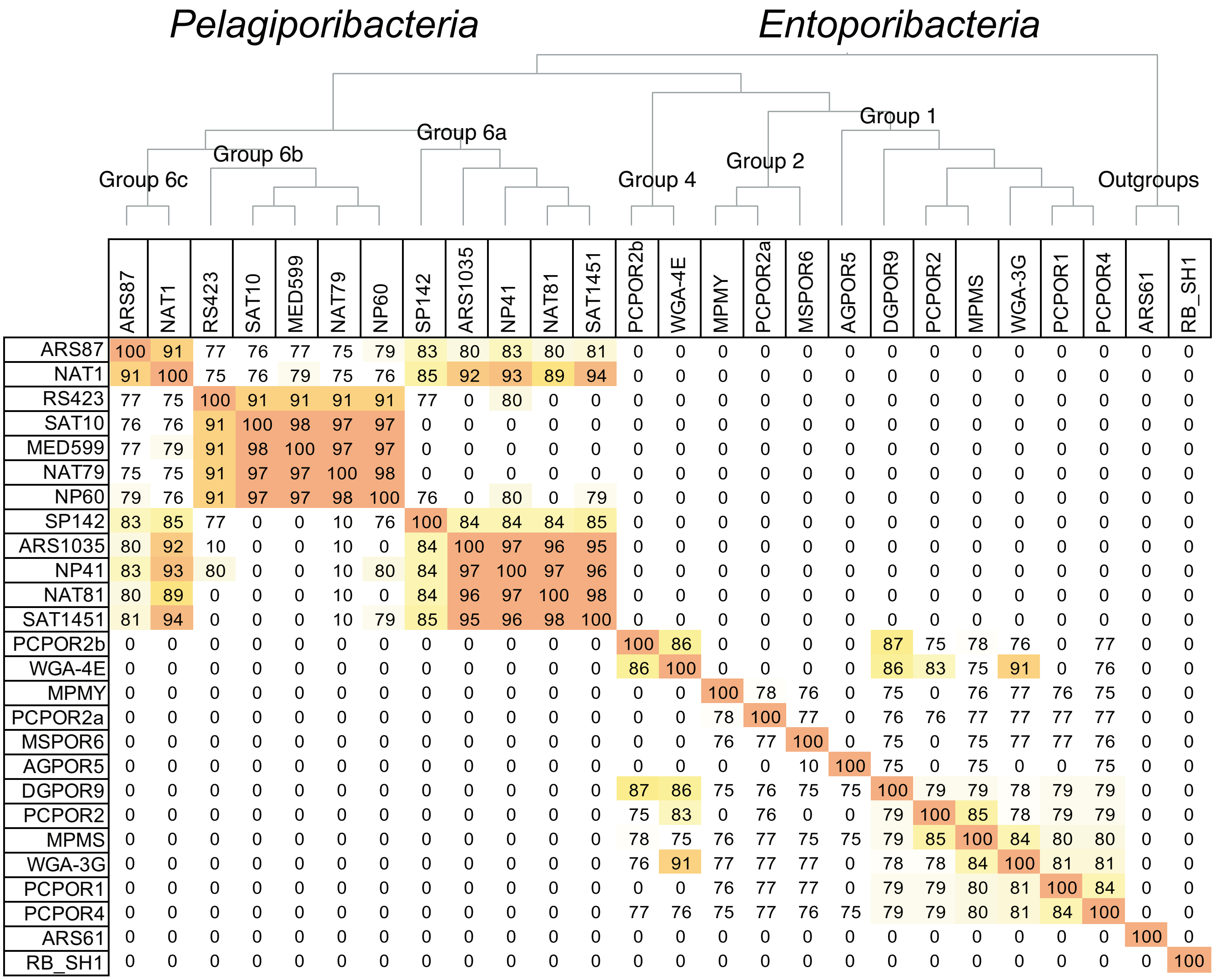
Supplementary Figure S2. **16S rRNA tree**. Blue text indicates Tara Oceans metagenome-assembled genomes. Red text indicates sequences obtained from sponge tissue-derived MAGs and SAGs. Bolded text marks sequences generated by the current study, with host abbreviations *Agelas tubulata,* AG; *Dysidea granulosa,* DG; *Melophlus sarasinorum,* MS; *Pseudocerataina sp*, PC; *Aplysina aerophoba,* WGA MPMY, and MPMS. Tara Oceans genomes are italicized, with geographic abbreviations ARS, Arabian Sea; RS, Red Sea; MED Mediterranean Sea; NAT, North Atlantic; SAT, South Atlantic; NP, North Pacific; SP, South Pacific. Codes G1-G5 (appended with underscores) designate 16S rRNA gene fragments previously classified by other investigators [1, 2]. Outgroup abbreviation Rhodopirellula indicates *Rhodopirellula baltica* strain SH1. Minimum 16S gene length was 700 nt.

Supplementary Figure S3a: **16S rRNA gene similarity matrix**. Areas shaded in darker colors indicate closer evolutionary relationships. Abbreviation RB_SH1 indicates *Rhodopirellula baltica* strain SH1. Minimum 16S gene length was 700 nt. Values > 97% typically define members of the same species, while values between 95%-97% suggest membership in the same genus. Percent identity values were calculated based on the longest alignment > 400 nt obtained by BLASTN search. Zero values indicate relationships that could not be determined due to fragmented or missing 16S rRNA genes.



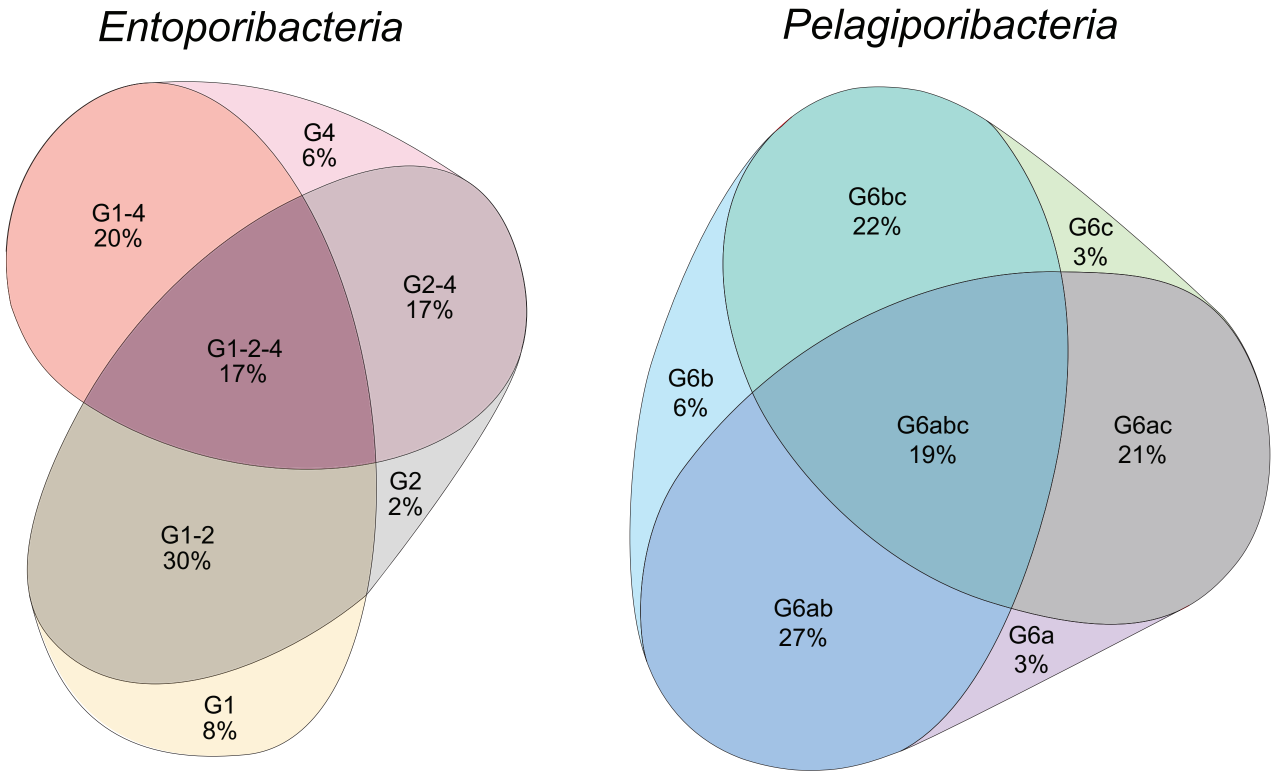
	ARS87	NAT1	RS423	SAT10	MED599	NAT79	NP60	SP142	ARS1035	NP41	NAT81	SAT1451	PCPOR2b	WGA-4E	MPMY	PCPOR2a	MSPOR6	AGPOR5	DGPOR9	PCPOR2	MPMS	WGA-3G	PCPOR1	PCPOR4	ARS61	RB_SH1
ARS87	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NAT1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RS423	0	0	100	100	99	99	100	0	92	89	92	91	0	89	0	0	89	88	84	91	0	0	88	88	87	83
SAT10	0	0	100	100	100	99	100	0	92	89	92	91	0	89	0	0	89	88	88	91	0	0	87	88	87	84
MED599	0	0	99	100	100	99	100	0	92	89	92	91	0	89	0	0	89	88	88	90	0	0	87	88	87	84
NAT79	0	0	99	99	99	100	100	0	92	89	92	91	0	89	0	0	89	88	88	91	0	0	87	88	87	84
NP60	0	0	100	100	100	100	0	0	92	89	92	91	0	89	0	0	89	88	88	91	0	0	87	88	87	84
SP142	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ARS1035	0	0	92	92	92	92	92	0	100	97	100	99	87	91	0	0	90	89	0	93	0	0	89	89	84	83
NP41	0	0	89	89	89	89	89	0	97	100	97	97	87	88	0	0	87	87	0	88	0	0	86	86	81	82
NAT81	0	0	92	92	92	92	92	0	100	97	100	99	87	91	0	0	90	89	0	93	0	0	88	88	84	82
SAT1451	0	0	91	91	91	91	91	0	99	97	99	100	87	90	0	0	90	89	0	92	0	0	88	88	83	83
PCPOR2b	0	0	0	0	0	0	0	0	87	87	87	87	100	98	0	0	85	84	85	86	0	0	86	86	83	0
WGA-4E	0	0	89	89	89	89	89	0	91	88	91	90	98	100	0	0	85	88	86	91	0	0	89	89	85	84
MPMY	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PCPOR2a	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MSPOR6	0	0	89	89	89	89	89	0	90	87	90	90	85	94	0	0	100	96	95	97	0	0	98	98	86	87
AGPOR5	0	0	88	88	88	88	88	0	89	87	89	89	84	88	0	0	96	100	95	94	0	0	95	95	87	85
DGPOR9	0	0	84	88	88	88	88	0	0	0	0	0	85	86	0	0	95	95	100	95	0	0	96	96	0	0
PCPOR2	0	0	91	91	90	91	91	0	93	88	93	92	86	91	0	0	97	94	95	100	0	0	96	96	87	85
MPMS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
WGA-3G	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PCPOR1	0	0	88	87	87	87	87	0	89	86	88	88	86	89	0	0	98	95	96	96	0	0	100	99	86	86
PCPOR4	0	0	88	88	88	88	88	0	89	86	88	88	86	89	0	0	98	95	96	96	0	0	99	100	86	87
ARS61	0	0	87	87	87	87	87	0	84	81	84	83	83	85	0	0	86	83	0	87	0	0	86	86	100	81
RB_SH1	0	0	83	84	84	84	84	0	83	82	82	83	0	84	0	0	87	85	0	85	0	0	86	87	81	100

Supplementary Figure S3b: **Average nucleotide identity (ANI) similarity matrix**. Areas shaded in darker colors indicate closer evolutionary relationships. Abbreviation RB_SH1 indicates *Rhodopirellula baltica* strain SH1. AAI values > 95% typically define members of the same species, while values between 88%-95% suggest membership in the same genus. Zeros indicate values < 75%, where the ANI metric has been shown to lose its reliability, and average amino acid identity (AAI) is recommended to be used instead [3].

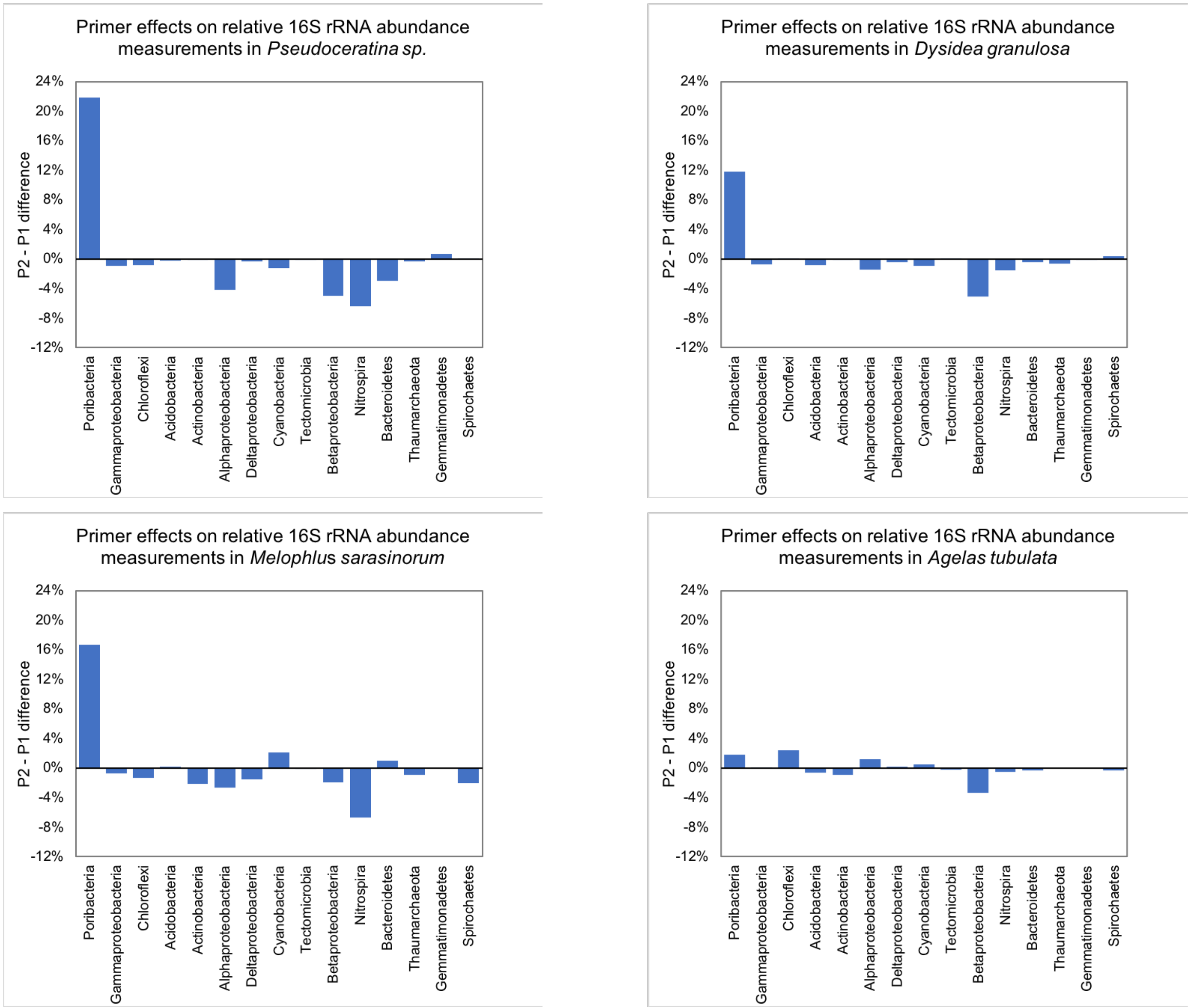


	ARS87	NAT1	RS423	SAT10	MED599	NAT79	NP60	SP142	ARS1035	NP41	NAT81	SAT1451	PCPOR2b	WGA-4E	MPMY	PCPOR2a	MSPOR6	AGPOR5	DGPOR9	PCPOR2	MPMS	WGA-3G	PCPOR1	PCPOR4	ARS61	RB_SH1
ARS87	100	91	77	76	77	75	79	83	80	83	80	81	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NAT1	91	100	75	76	79	75	76	85	92	93	89	94	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RS423	77	75	100	91	91	91	91	77	0	80	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SAT10	76	76	91	100	98	97	97	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MED599	77	79	91	98	100	97	97	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NAT79	75	75	91	97	97	100	98	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NP60	79	76	91	97	97	98	100	76	0	80	0	79	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SP142	83	85	77	0	0	10	76	100	84	84	84	85	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ARS1035	80	92	10	0	0	10	0	84	100	97	96	95	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NP41	83	93	80	0	0	10	80	84	97	100	97	96	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NAT81	80	89	0	0	0	10	0	84	96	97	100	98	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SAT1451	81	94	0	0	0	10	79	85	95	96	98	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PCPOR2b	0	0	0	0	0	0	0	0	0	0	0	0	100	86	0	0	0	0	87	75	78	76	0	77	0	0
WGA-4E	0	0	0	0	0	0	0	0	0	0	0	0	86	100	0	0	0	0	86	83	75	91	0	76	0	0
MPMY	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	78	76	0	75	0	76	77	76	75	0	0
PCPOR2a	0	0	0	0	0	0	0	0	0	0	0	0	0	0	78	100	77	0	76	76	77	77	77	77	0	0
MSPOR6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	76	77	100	0	75	0	75	77	77	76	0	0
AGPOR5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	100	75	0	75	0	0	75	0	0
DGPOR9	0	0	0	0	0	0	0	0	0	0	0	0	87	86	75	76	75	75	100	79	79	78	79	79	0	0
PCPOR2	0	0	0	0	0	0	0	0	0	0	0	0	75	83	0	76	0	0	79	100	85	78	79	79	0	0
MPMS	0	0	0	0	0	0	0	0	0	0	0	0	78	75	76	77	75	75	79	85	100	84	80	80	0	0
WGA-3G	0	0	0	0	0	0	0	0	0	0	0	0	76	91	77	77	77	0	78	78	84	100	81	81	0	0
PCPOR1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	76	77	77	0	79	79	80	81	100	84	0	0
PCPOR4	0	0	0	0	0	0	0	0	0	0	0	0	77	76	75	77	76	75	79	79	80	81	84	100	0	0
ARS61	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0
RB_SH1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100

Supplementary Figure S4: ***Entoporibacteria* and *Pelagiporibacteria* subgroup protein family clusters**. Shared families are defined as ProteinOrtho clusters found in one or more genomes from each different lineage. Subgroup overlap labels refer to groups of genomes defined by taxonomic placement in both the concatenated multilocus tree (Figure 2) and average amino acid identity clustering cladogram (Figure 3).



Supplementary Figure S5 : 16S rRNA primer effects on relative abundance of different taxa in four different sponge hosts. Abbreviations: P1, 16S primer set 515FB/806RB; P2, 16S primer set 515Fsp/806Rsp.



Primer effects on relative 16S rRNA abundance measurements in *Pseudoceratina sp.*

Primer effects on relative 16S rRNA abundance measurements in *Dysidea granulosa*

Primer effects on relative 16S rRNA abundance measurements in *Melophlus sarasinorum*

Primer effects on relative 16S rRNA abundance measurements in *Agelas tubulata*

Supplementary Figure S6a: **16S rRNA primer mismatches**. Calculated based on 16S rRNA sequence alignments to full-length, metagenomically assembled Poribacteria genome sequences.

genome name	27F [4]	341F [5]	515Fb [6]	515Fsp *	784F [7]	785R [5]	806Rb [6]	806Rsp *	1061R [8]	1492R [4]	sub-group
PCPOR1	3	0	1	0	2	2	1	0	0	0	*Entoporibacteria* G1
PCPOR2	3	0	1	0	2	2	1	0	0	0	*Entoporibacteria* G1
PCPOR4	3	0	1	0	2	2	1	0	0	0	*Entoporibacteria* G1
AGPOR5	3	0	1	0	2	2	1	0	0	0	*Entoporibacteria* G2
WGA-4E	3	0	1	0	2	1	1	0	0	0	*Entoporibacteria* G4
SAT1451	3	0	1	0	2	1	0	0	0	0	*Pelagiporibacteria* G6a
NAT81	3	0	1	0	2	1	0	0	0	0	*Pelagiporibacteria* G6a
ARS1035	3	0	1	0	2	1	0	0	0	0	*Pelagiporibacteria* G6a
NP41	3	0	1	0	2	1	0	0	0	0	*Pelagiporibacteria* G6b
SAT10	3	0	1	0	2	0	0	0	0	0	*Pelagiporibacteria* G6b
NAT79	3	0	1	0	2	0	0	0	0	0	*Pelagiporibacteria* G6b
NP60	3	0	1	0	2	0	0	0	0	0	*Pelagiporibacteria* G6b
RS423	3	0	1	0	2	0	0	0	0	0	*Pelagiporibacteria* G6b
MED599	3	0	1	0	2	0	0	0	0	0	*Pelagiporibacteria* G6b
ARS61	3	0	1	0	0	0	0	0	0	0	unclassified

* specifically designed for this study.

Supplementary Figure S6b: **16S rRNA sequence mismatch localizations for primer 515FB.** This table contains additional partial sequences that could not be included in Supplementary Figure S6a because they did not cover all primers. Sequence identifiers beginning with "Ga02" and "POR" refer to JGI Integrated Microbial Genome Pipeline locus ids.

sponge host	**sequence_id**	**sequence**	**reference**
	515FB == 515F-Y	GTGYCAGCMGCCGCGGTAA	
Agelas tubulata	AGPOR5 Ga0227391_12124	GTGCCAGCAGCTGCGGTAA	This study
Dysidea granulosa	DGPOR9 Ga0236299_105113	GTGCCAGCAGCTGCGGTAA	This study
Melophlus sarasinorum	MSPOR6 Ga0207211_14795	GTGCCAGCAGCTGCGGTAA	This study
Melophlus sarasinorum	MSPOR6 Ga0207211_134131	GTGCCAGCAGCTGCGGTAA	This study
Pseudoceratina sp.	PCPOR1 Ga0206387_116928	GTGCCAGCAGCTGCGGTAA	This study
Pseudoceratina sp.	PCPOR2 Ga0207196_1125	GTGCCAGCAGCTGCGGTAA	This study
Pseudoceratina sp.	PCPOR2 Ga0207196_15322	GTGCCAGCAGCTGCGGTAA	This study
Pseudoceratina sp.	PCPOR2b Ga0206366_1461	GTGCCAGCAGCTGCGGTAA	This study
Pseudoceratina sp.	PCPOR4 Ga0207198_1187265	GTGCCAGCAGCTGCGGTAA	This study
Aplysina aerophoba	WGA-A3 POR_0001.00000030	GTGCCAGCAGCTGCGGTAA	Kamke 2013 [9]
Aplysina aerophoba	WGA-4C POR4C_0180.00000040	GTGCCAGCAGCTGCGGTAA	Kamke 2013 [9]
Aplysina aerophoba	WGA-4E POR4E_0050.00000080	GTGCCAGCAGCTGCGGTAA	Kamke 2013 [9]
Aplysina aerophoba	WGA-4G POR4G_0091.00000100	GTGCCAGCAGCTGCGGTAA	Kamke 2013 [9]
Aplysina aerophoba	AY713479.1	GTGCCAGCAGCTGCGGTAA	Fieseler 2006 [10]
Aplysina aerophoba	AY485280.1	GTGCCAGCAGCTGCGGTAA	Fieseler 2006 [10]
Astrosclera willeyana	HE985121.1	GTGCCAGCAGCTGCGGTAA	Karlinska-Batres 2013 [11]
Vaceletia crypta	HE817802.1	GTGCCAGCAGCTGCGGTAA	Karlinska-Batres 2013 [11]
Vaceletia crypta	HE817855.1	GTGCCAGCAGCTGCGGTAA	Karlinska-Batres 2013 [11]
Vaceletia crypta	HE817858.1	GTGCCAGCAGCTGCGGTAA	Karlinska-Batres 2013 [11]
Xestospongia testudinaria	JN596759.1	GTGCCAGCAGCTGCGGTAA	Montalvo 2011 [12]
Ircinia fasciculata	JX206505.1	GTGCCAGCAGCTGCGGTAA	Erwin 2012 [13]
Aplysina cauliformis	KF286093.1	GTGCCAGCAGCTGCGGTAA	Olsen 2014 [14]

Supplementary Figure 6c: 16S rRNA sequence mismatch localizations for primer 806RB. This table contains additional partial sequences that could not be included in Supplementary Figure S6a because they did not cover all primers. Sequence identifiers beginning with "Ga02" and "POR" refer to JGI Integrated Microbial Genome Pipeline locus ids.

sponge host	**sequence_id**	**sequence**	**reference**
	806RB	GGACTACNVGGGTWTCTAAT	
Agelas tubulata	AGPOR5 Ga0227391_12124a	GGACTAGCGGGGTATCTAAT	This study
Dysidea granulosa	DGPOR9 Ga0236299_105113	GGACTACCGGGGTATCTAAT	This study
Melophlus sarasinorum	MSPOR6 Ga0207211_1479	GGACTAGCGGGGTATCTAAT	This study
Melophlus sarasinorum	MSPOR6 Ga0207211_134131	GGACTAGCGGGGTATCTAAT	This study
Pseudoceratina sp.	PCPOR1 Ga0206387_116928	GGACTAGCGGGGTATCTAAT	This study
Pseudoceratina sp.	PCPOR2 Ga0207196_1125	GGACTAGCGGGGTATCTAAT	This study
Pseudoceratina sp.	PCPOR2 Ga0207196_15322	GGACTAGCGGGGTATCTAAT	This study
Pseudoceratina sp.	PCPOR2b Ga0207198_1187265	GGACTACCGGGGTATCTAAT	This study
Pseudoceratina sp.	PCPOR4 Ga0207198_1187265	GGACTAGCGGGGTATCTAAT	This study
Aplysina aerophoba	WGA-A3 POR_0001.00000030	GGACTAGCGGGGTATCTAAT	Kamke 2013 [9]
Aplysina aerophoba	WGA-4C POR4C_0180.00000040	GGACTAGCGGGGTATCTAAT	Kamke 2013 [9]
Aplysina aerophoba	WGA-4E POR4E_0050.00000080	GGACTACCGGGGTATCTAAT	Kamke 2013 [9]
Aplysina aerophoba	WGA-4G POR4G_0091.00000100	GGACTAGCGGGGTATCTAAT	Kamke 2013 [9]
Aplysina aerophoba	AY713479.12035.13581	GGACTAGCGGGGTATCTAAT	Fieseler 2006 [10]
Aplysina aerophoba	AY485280.1	GGACTAGCGGGGTATCTAAT	Fieseler 2006
Astrosclera willeyana	HE985121.1	GGACTAGCGGGGTATCTAAT	Karlinska-Batres 2013
Vaceletia crypta	HE817802.1	GGACTACCGGGGTATCTAAT	Karlińska-Batres 2013
Vaceletia crypta	HE817855.1	GGACTAGCGGGGTATCTAAT	Karlińska-Batres 2013
Vaceletia crypta	HE817858.1	GGACTACCGGGGTATCTAAT	Karlińska-Batres 2013
Xestospongia testudinaria	JN596759.1	GGACTAGCGGGGTATCTAAT	Montalvo 2011 [12]
Ircinia fasciculata	JX206505.1	GGACTAGCGGGGTATCTAAT	Erwin 2012 [13]
Aplysina cauliformis	KF286093.1	GGACTAGCGGGGTATCTAAT	Olsen 2014 [14]

Supplementary Table S1. **Additional MIMAG properties of assembled genomes.** Abbreviations: SAG, single-cell genome; MAG metagenomic assembly genome. IMG, JGI Integrated Microbial Genome Pipeline . Taxonomic identification method for all genomes was "multi-marker".

Genome	**NCBI accession number**	**IMG accession number**	**project type**	**assembly software**	**annotation pipeline**	**MIMAG Genome Quality Rating**	**MAG binning software**	**MAG binning parameters**	**SAG cell isolation**	**SAG lysis**	**SAG ampli-fication**
PCPOR4	PYIY00000000	2744054414	MAG	Celera Asssembler v. 8.3	PROKKA + IMG	High	Custom	Pct GC, coverage, DarkHorse taxonomy	na	na	na
PCPOR2	PYJB00000000	2744054412	MAG	Celera Asssembler v. 8.3	PROKKA + IMG	Med	Custom	Pct GC, coverage, DarkHorse taxonomy	na	na	na
DGPOR9	PYJF00000000	2765235806	MAG	Celera Asssembler v. 8.3	PROKKA + IMG	Low	Custom	Pct GC, coverage, DarkHorse taxonomy	na	na	na
PCPOR2b	PYIZ00000000	2740892555	MAG	Celera Asssembler v. 8.3	PROKKA + IMG	High	Custom	Pct GC, coverage, DarkHorse taxonomy	na	na	na
MSPOR6	PYJD00000000	2744055035	MAG	Celera Asssembler v. 8.3	PROKKA + IMG	Med	Custom	Pct GC, coverage, DarkHorse taxonomy	na	na	na
PCPOR1	PYJC00000000	2740892559	MAG	Celera Asssembler v. 8.3	PROKKA + IMG	Med	Custom	Pct GC, coverage, DarkHorse taxonomy	na	na	na
AGPOR5	PYJE00000000	2758568595	MAG	Celera Asssembler v. 8.3	PROKKA + IMG	Med	Custom	Pct GC, coverage, DarkHorse taxonomy	na	na	na

Genome	NCBI accession number	IMG accession number	project type	assembly software	annotation pipeline	MIMAG Genome Quality Rating	MAG binning software	MAG binning parameters	SAG cell isolation	SAG lysis	SAG ampli-fication
PCPOR2a	PYJA00000000	2744054413	MAG	Celera Asssembler v. 8.3	PROKKA + IMG	Med	Custom	Pct GC, coverage, DarkHorse taxonomy	na	na	na
MPMS	MPMS00000000	na	MAG	SPAdes v. 3.5.0	PROKKA + IMG	Low	CONCOCT v. 0.4.0	coverage, composiiton	na	na	na
MPMY	MPMY00000000	na	MAG	SPAdes v. 3.5.0	PROKKA + IMG	Med	CONCOCT v. 0.4.0	coverage, composiiton	na	na	na
WGA-3G	ASZN01000001	2265129009	SAG	Newbler vers. 1.1.03.24	PROKKA + IMG	Med	na	na	flow cytometry	freeze/thaw + alkaline	MDA
WGA-4E	AQTV00000000	2265129006	SAG	Newbler vers. 1.1.03.24	PROKKA + IMG	Med	na	na	flow cytometry	freeze/thaw + alkaline	MDA
WGA-4C	APGO00000000	2265129007	SAG	Newbler vers. 1.1.03.24	PROKKA + IMG	Low	na	na	flow cytometry	freeze/thaw + alkaline	MDA
WGA-4G	AQPC00000000	2265129008	SAG	Newbler vers. 1.1.03.24	PROKKA + IMG	Low	na	na	flow cytometry	freeze/thaw + alkaline	MDA
WGA-A3	ADFK00000000	2265129010	SAG	Newbler vers. 1.1.03.24	PROKKA + IMG	Low	na	na	flow cytometry	freeze/thaw + alkaline	MDA
NP60	PBTM00000000	2775506747	MAG	Minimus2	PROKKA + IMG	High	BinSanity	coverage, composiiton	na	na	na
SP142	PAQF00000000	2775506749	MAG	Minimus2	PROKKA + IMG	High	BinSanity	coverage, composiiton	na	na	na

Genome	NCBI accession number	IMG accession number	project type	assembly software	annotation pipeline	MIMAG Genome Quality Rating	MAG binning software	MAG binning parameters	SAG cell isolation	SAG lysis	SAG ampli-fication
RS423	PBVO00000000	2775506743	MAG	Minimus2	PROKKA + IMG	Med	BinSanity	coverage, composiiton	na	na	na
MED599	NZIS00000000	2775506746	MAG	Minimus2	PROKKA + IMG	High	BinSanity	coverage, composiiton	na	na	na
SAT10	PAYP00000000	2775506750	MAG	Minimus2	PROKKA + IMG	High	BinSanity	coverage, composiiton	na	na	na
NAT79	NZUT00000000	2775506744	MAG	Minimus2	PROKKA + IMG	High	BinSanity	coverage, composiiton	na	na	na
SAT1451	PAVB00000000	2775506788	MAG	Minimus2	PROKKA + IMG	Med	BinSanity	coverage, composiiton	na	na	na
NP41	PBXZ00000000	2775506745	MAG	Minimus2	PROKKA + IMG	High	BinSanity	coverage, composiiton	na	na	na
NAT81	NZUQ00000000	2775506748	MAG	Minimus2	PROKKA + IMG	High	BinSanity	coverage, composiiton	na	na	na
ARS61	NYZG00000000	2775506754	MAG	Minimus2	PROKKA + IMG	Med	BinSanity	coverage, composiiton	na	na	na
NAT1	PACG00000000	2775506753	MAG	Minimus2	PROKKA + IMG	Med	BinSanity	coverage, composiiton	na	na	na
ARS1035	NZDP00000000	2775506752	MAG	Minimus2	PROKKA + IMG	Med	BinSanity	coverage, composiiton	na	na	na
ARS87	NYYD00000000	2775506751	MAG	Minimus2	PROKKA + IMG	Low	BinSanity	coverage, composiiton	na	na	na

Supplementary Table S2. **Tara oceans genome metadata.** Adapted from supplementary data tables linked to Tully et al [19] merged with supplementary information from Sunegawa et al [20] and NCBI Short Read Achive (SRA) metadata [21]. Biosample read source names separated by underscores signify the following abbreviatons: Tara Oceans station identifier (e.g. tara038), filter size fraction (prot=0.8-5.0 µm, bact =0.22-1.6 µm), and environmental feature (meso=mesopelagic zone, dcm=deep chlorophyll maximum layer, epi =epipelagic mixing layer).

genome name	Genbank accession	NCBI Biosample id	Location	num bio-samples co-assembled	Most abundant biosample read source	pct reads mapped to most abund. source	Date/Time	Latitude [degrees North]	Longitude [degrees East]	Sampling depth [m]
ARS87	NYYD00000000	SAMN07619756	Arabian Sea	14	tara038_prot_meso	81.4%	2010-03-16T06:13:30	19.0351	64.5638	340
ARS61	NYZG00000000	SAMN07619755	Arabian Sea	36	tara038_prot_meso	76.5%	2010-03-16T06:13:30	19.0351	64.5638	340
ARS1035	NZDP00000000	SAMN07619754	Arabian Sea	36	tara037_prot_meso	67.3%	2010-03-12T17:08:47	20.8457	63.5851	600
MED599	NZIS00000000	SAMN07619757	Mediterranean Sea	20	tara030_prot_dcm	67.3%	2009-12-15T15:23:37	33.9235	32.8118	70
NAT81	NZUQ00000000	SAMN07619760	North Atlantic Ocean	74	tara149_prot_meso	36.9%	2012-03-01T18:50:26	34.0771	-49.8233	740
NAT79	NZUT00000000	SAMN07619759	North Atlantic Ocean	33	tara151_prot_dcm	36.9%	2012-03-09T17:42:29	36.1811	-28.9373	80
NAT1	PACG00000000	SAMN07619758	North Atlantic Ocean	177	tara146_prot_meso	36.8%	2012-02-16T12:01:25	34.6663	-71.2907	640
SP142	PAQF00000000	SAMN07619766	South Pacific Ocean	18	tara125_prot_epi	96.6%	2011-08-09T17:05:33	-8.8999	-142.5461	140
SAT1451	PAVB00000000	SAMN07619765	South Atlantic Ocean	38	tara078_bact_meso	21.9%	2010-11-05T12:33:27	-30.1471	-43.2915	800
SAT10	PAYP00000000	SAMN07619764	South Atlantic Ocean	30	tara076_bact_dcm	38.6%	2010-10-16T16:58:37	-21.0292	-35.3498	150
NP60	PBTM00000000	SAMN07619762	North Pacific Ocean	41	tara137_prot_dcm	88.9%	2011-12-03T14:31:20	14.2025	-116.6433	375
RS423	PBVO00000000	SAMN07619763	Red Sea	74	tara032_bact_dcm	30.6%	2010-01-11T14:17	23.4183	37.245	80
NP41	PBXZ00000000	SAMN07619761	North Pacific Ocean	30	tara138_prot_meso	43.0%	2011-12-11T20:37:22	6.3559	-103.0598	450

Supplementary Table S3. **List of marker genes used to construct multi-locus taxonomic tree**.

function	**aln len (aa)**
DNA-directed_RNA_polymerase_subunit_alpha	349
protein_translocase_subunit_SecY	457
ribosomal_protein_L14	127
ribosomal_protein_L15	158
ribosomal_protein_L16	138
ribosomal_protein_L17	135
ribosomal_protein_L18	122
ribosomal_protein_L2	274
ribosomal_protein_L21	120
ribosomal_protein_L22	121
ribosomal_protein_L23	99
ribosomal_protein_L24	117
ribosomal_protein_L29	234
ribosomal_protein_L3	218
ribosomal_protein_L4	217
ribosomal_protein_L5	188
ribosomal_protein_L6	227
ribosomal_protein_S10	103
ribosomal_protein_S11	124
ribosomal_protein_S12	215
ribosomal_protein_S13	183
ribosomal_protein_S15	189
ribosomal_protein_S17	103
ribosomal_protein_S19	125
ribosomal_protein_S20	127
ribosomal_protein_S3	288
ribosomal_protein_S5	176
ribosomal_protein_S8	161

Supplementary Table S4: Shared versus unique pathways/functions in *Entoporibacteria* and *Pelagiporibacteria*.

	Function	**Entoporibacteria**	**Pelagiporibacteria**	**Key gene(s)**
Central metabolism	glycolysis	+	+	Glyceraldehyde-3-phosphate dehydrogenase
	tricarboxylic acid cycle	+	+	Aconitate hydratase
	oxidative pentose phosphate pathway	+	+	glucose 6-phosphate dehydrogenase, 6-phosphogluconolactonase, 6-phosphogluconate dehydrogenase
	non-oxidative pentose phosphate pathway	+	+	Ribose-5-phosphate isomerase, Ribulose 5-Phosphate 3-Epimerase, transaldolase, transketolase
	oxidative phosphorylation	+	+	Cytochrome c oxidase
	Entner-Doudoroff	+	+	glucose 1-dehydrogenase, gluconolactonase
	Wood–Ljungdahl pathway (reductive acetyl CoA)	+	+	Carbon monoxide dehydrogenase/acetyl-CoA synthase. methyltransferases
	assimilatory sulfate reduction	+	+	sulfite reductase, APS reductase (Adenylyl-sulfate reductase), ATP-sulfurylase (sulfate adenylyltransferase, adenyltransferase/sulfurtransferase MoeZ), DMSO reductase
	assimilatory denitrification	+	+	assimilatory (ferredoxin)nitrate reductase (NarB); assimilatory nitrite reductase (NirA) ;
	ammonia scavenging	+	+	glutamine synthase
Vitamins/cofactors	cobalamin (B12)	+	+	cobalamin synthesis protein cobW

	Function	Ento-poribacteria	Pelagi-poribacteria	Key gene(s)
	biotin	+	+	biotin synthase
	thiamine	+	+	thiamine biosynthesis protein ThiI
Nutrient sources	mixed acid fermentation	+	+	lactate dehydrogenase, pyruvate dehydrogenase, alcohol dehydrogenase, pyruvate-formate lyase, PEP carboxylase, malate dehydrogenase, Fumarate hydratase, fumarate reductase, menaquinone
	citrate fermentation	+	+	citrate lyase/ 2-dehydro-3-deoxyglucarate aldolase
	urea degradation	+	+	urease, ABC urease transport gene cluster (UreA, UreB, UreC)
	algal wall degradation	+	+	pectate lyase
	algal wall degradation II	-	+	endo-1,4-beta-xylanase
	propane/butane-diol utilization	+	+	1,3-propanediol dehydrogenase, (R,R)-butanediol dehydrogenase, dihydrodiol dehydrogenase. Propanediol utilization protein PduA, propionaldehyde dehydrogenase
	methyl-citrate cycle (feeds propionyl-CoA into pyruvate)	+	-	[group 4 only] 2-methylcitrate synthase, 2-methycitrate dehydratase
	trimethylamine (TMA) production	-	+	[group 6c only] Choline trimethylamine-lyase
	Glutamate-ammonia-ligase adenylyltransferase	+	+	uses ammonia as a nitrogen source for glutamate synthesis. ammonia might be generated by denitrification, rather than importing, except that there is an ammonia transporter

	Function	Ento-poribacteria	Pelagi-poribacteria	Key gene(s)
Gram-negative envelope	periplasm	+	+	TonB-dependent outer membrane receptor, preprotein translocase subunit SecA
	translocases	+	+	YajC, SecA, SecE, SecY, Sec-independent
	Sterol biosynthesis	+	+	Squalene--hopene cyclase, oxidosqualene cyclase
	membrane coat-like proteins	+	+	gas vesicle protein, Biopolymer transport protein ExbB
	peptidoglycan biosynthesis	+	+	Undecaprenyldiphospho-muramoylpentapeptide beta-N-acetylglucosaminyltransferase, N-acetylneuraminate lyase
	lipopolysaccharide biosynthesis	+	+	3-deoxy-D-manno-octulosonic acid transferase
	phospho-inositol-linked lipids	+	+	myo-inositol-1-phosphate synthase, phosphatidylinositol mannoside acyltransferase, CDP-diacylglycerol--inositol 3-phosphatidyltransferase, inosose isomerase
	capsule assembly	+	+	Capsule biosynthesis protein CapA, capsule assembly protein Wzi
	O-antigen synthesis	-	+	perosamine synthase
Type II secretion	type II protein secretion	+	+	type II secretion system proteins C, D, E, F, G, I, PulK

	Function	Ento-poribacteria	Pelagi-poribacteria	Key gene(s)
Motility	flagellar assembly	-	+	Flagellar basal body rod protein FlgB,C,D,G,Flagellar biosynthesis protein FlhA,B,F,Flagellar biosynthetic protein FliP,Q,R,Flagellar brake protein YcgR,Flagellar filament core protein,Flagellar hook protein,Flagellar hook-basal body complex protein FliE,Flagellar L-ring protein,Flagellar M-ring protein,Flagellar P-ring protein,Flagellar protein FliS,Flagellin,Flagellum site-determining protein YlxH,
	chemotaxis	-	+	Chemotaxis protein CheW, Chemotaxis protein methyltransferase, Chemotaxis protein PomA
Sporulation	dispersal, dessication protection	+	+	stage II sporulation proteins E and M, inner spore coat protein H, SpoIIAA, SpoIVB , SpoVE, Spore protein SP21
Osmoprotection	ectoine metabolism	+	+	Ectoine dioxygenase (multiple copies, ectoine_ectB: diaminobutyrate--2-oxoglutarate aminotransferase - 1st step in ectoine biosynthesis, ectoine_eutB: ectoine utilization protein EutB, ectoine_eutD: ectoine utilization protein EutD,
	trehalose import	+	+	Trehalose transport system permease protein SugA, trehalose utilisation protein
	glycine betaine metabolism	+	+	Choline-sulfatase

	Function	Ento-poribacteria	Pelagi-poribacteria	Key gene(s)
	osmotic stress sensing	+	+	Small Conductance Mechanosensitive Ion Channel (MscS) Family
Bacterial proteasome	protein turnover	+	+	Proteasome subunit alpha, Proteasome subunit beta, Pup deamidase/depupylase, Pup--protein ligase
Adhesion	adhesins characteristic of eukaryotic cell surfaces	+	+	concanavalin A-like lectin, Ig-like domain, internalin-A, leucine-rich repeats, fibronectin type III domain, PH (Pleckstrin Homology) domain
	sialic acid decoration/cleavage	+	+	N-acetylneuraminate lyase
Extracellular matrix degradation	cellulosome (cell wall-anchored extracellular enzyme complexes)	+	+	dockerin, cohesin domain-containing protein
	GAG (glucose amino glycan) degradation			N-acetylglucosamine-6-sulfatase, N-acetyl-galactosamine-4-sulfatase, N-sulfoglucosamine-sulfohydrolase
	arylsulfatase	+	+	
	cerebroside-sulfatase	+	+	
	collagenase	+	+	
	ceramidase	+	+	
	chitinase	+	+	
	heparinase	+	-	
DNA exchange	DNA uptake competence	+	+	Competence protein ComM, ComE operon protein 1
	type IV pilus assembly protein PilM, PilB, PilC, PilF	+	+	
	plasmid support	+	+	F plasmid transfer operon protein TraF

	Function	Ento-poribacteria	Pelagi-poribacteria	Key gene(s)
Phage defense	Restriction enzymes/modification methylases	+	+	See Supplementary Table 5
	Modification methylases (restriction associated)	+	+	See Supplementary Table 5
	CRISPR-associated	+	+	Cas, Cmr families
Bacterial toxins	bacteriocins	+	+	subtilosin biosynthesis protein AlbA, ABC-type bacteriocin/lantibiotic exporter
	toxin de-activation	+	+	Macrolide export, Daunorubicin/doxorubicin resistance, penicillin amidase, microcystin degradation

Supplementary Table S5: List of HMM pattern identifiers used to identify endonucleases and transposases

Transposase HMMs

Accession	**Name**	**Description**
PF04693.11	DDE_Tnp_2	Archaeal putative transposase ISC1217
PF03400.12	DDE_Tnp_IS1	IS1 transposase
PF12762.6	DDE_Tnp_IS1595	ISXO2-like transposase domain
PF07592.10	DDE_Tnp_ISAZ013	Rhodopirellula transposase DDE domain
PF13340.5	DUF4096	Putative transposase of IS4/5 family (DUF4096)
PF13022.5	HTH_Tnp_1_2	Helix-turn-helix of insertion element transposase
PF13542.5	HTH_Tnp_ISL3	Helix-turn-helix domain of transposase family ISL3
PF10551.8	MULE	MULE transposase domain
PF01385.18	OrfB_IS605	Probable transposase
PF07282.10	OrfB_Zn_ribbon	Putative transposase DNA-binding domain
PF12784.6	PDDEXK_2	PD-(D/E)XK nuclease family transposase
PF04754.11	Transposase_31	Putative transposase, YhgA-like
PF13005.6	zf-IS66	zinc-finger binding domain of transposase IS66

Endonuclease HMMs

Accession	**Name**	**Description**
PF07582.11	AP_endonuc_2_N	AP endonuclease family 2 C terminus
PF06616.10	BsuBI_PstI_RE	BsuBI/PstI restriction endonuclease C-terminus
PF13358.5	DDE_3	DDE superfamily endonuclease
PF13546.5	DDE_5	DDE superfamily endonuclease
PF13359.5	DDE_Tnp_4	DDE superfamily endonuclease
PF04556.11	DpnII	DpnII restriction endonuclease
PF02963.15	EcoRI	Restriction endonuclease EcoRI
PF10576.8	EndIII_4Fe-2S	Iron-sulfur binding domain of endonuclease III
PF09195.10	Endonuc-BglII	Restriction endonuclease BglII
PF09194.9	Endonuc-BsobI	Restriction endonuclease BsobI
PF09124.9	Endonuc-dimeris	T4 recombination endonuclease VII, dimerisation
PF09233.10	Endonuc-EcoRV	Restriction endonuclease EcoRV
PF04493.13	Endonuclease_5	Endonuclease V
PF02945.14	Endonuclease_7	Recombination endonuclease VII
PF01223.22	Endonuclease_NS	DNA/RNA non-specific endonuclease

Accession	**Name**	**Description**
PF10107.8	Endonuc_Holl	Endonuclease related to archaeal Holliday junction resolvase
PF03372.22	Exo_endo_phos	Endonuclease/Exonuclease/phosphatase family
PF14529.5	Exo_endo_phos_2	Endonuclease-reverse transcriptase
PF01844.22	HNH	HNH endonuclease
PF13391.5	HNH_2	HNH endonuclease
PF13392.5	HNH_3	HNH endonuclease
PF13395.5	HNH_4	HNH endonuclease
PF14279.5	HNH_5	HNH endonuclease
PF13613.5	HTH_Tnp_4	Helix-turn-helix of DDE superfamily endonuclease
PF00961.18	LAGLIDADG_1	LAGLIDADG endonuclease
PF03161.12	LAGLIDADG_2	LAGLIDADG DNA endonuclease family
PF04471.11	Mrr_cat	Restriction endonuclease
PF13156.5	Mrr_cat_2	Restriction endonuclease
PF15515.5	MvaI_BcnI	MvaI/BcnI restriction endonuclease family
PF01939.15	NucS	Endonuclease NucS
PF05367.10	Phage_endo_I	Phage endonuclease I
PF11463.7	R-HINP1I	R.HinP1I restriction endonuclease
PF09549.9	RE_Bpu10I	Bpu10I restriction endonuclease
PF09516.9	RE_CfrBI	CfrBI restriction endonuclease
PF14511.5	RE_EcoO109I	Type II restriction endonuclease EcoO109I
PF09554.9	RE_HaeII	HaeII restriction endonuclease
PF09519.9	RE_HindVP	HindVP restriction endonuclease
PF09568.9	RE_MjaI	MjaI restriction endonuclease
PF09565.9	RE_NgoFVII	NgoFVII restriction endonuclease
PF09522.9	RE_R_Pab1	R.Pab1 restriction endonuclease
PF09569.9	RE_ScaI	ScaI restriction endonuclease
PF09520.9	RE_TdeIII	Type II restriction endonuclease, TdeIII
PF09572.9	RE_XamI	XamI restriction endonuclease
PF15650.5	Tox-REase-9	Restriction endonuclease fold toxin 9
PF05685.11	Uma2	Putative restriction endonuclease
PF03851.13	UvdE	UV-endonuclease UvdE
PF03852.14	Vsr	DNA mismatch endonuclease Vsr
PF04555.12	XhoI	Restriction endonuclease XhoI

References

1. Lafi, F.F., et al., *Widespread distribution of poribacteria in demospongiae.* Appl Environ Microbiol, 2009. **75**(17): p. 5695-9.
2. Steinert, G., et al., *Coexistence of poribacterial phylotypes among geographically widespread and phylogenetically divergent sponge hosts.* Environ Microbiol Rep, 2017.
3. Rodriguez-R, L.M., and Konstantinos T. Konstantinidis, *Bypassing Cultivation To Identify Bacterial Species.* Microbe, 2014. **9**(3): p. 111-8.
4. Lane, D.J., *16S/23S rRNA sequencing*, in *Nucleic acid techniques in bacterial systematics.* , E. Stackebrandt and M. Goodfellow, Editors. 1991, John Wiley and Sons,: New York, NY. p. 115-175.
5. Klindworth, A., et al., *Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies.* Nucleic Acids Res, 2013. **41**(1): p. e1.
6. Earth_Microbiome_Project. *Earth Microbiome Project Protocols and Standards :16S Illumina Amplicon Protocol.* 2016 November 2016 December 2017]; Available from: http://press.igsb.anl.gov/earthmicrobiome/protocols-and-standards/16s/.
7. Nossa CW, et al., *Design of 16S rRNA gene primers for 454 pyrosequencing of the human foregut microbiome.* World journal of gastroenterology: WJG, 2010. **16**: p. 4135-44.
8. Degnan, P.H. and H. Ochman, *Illumina-based analysis of microbial community diversity.* ISME J, 2012. **6**(1): p. 183-94.
9. Kamke, J., et al., *Single-cell genomics reveals complex carbohydrate degradation patterns in poribacterial symbionts of marine sponges.* ISME J, 2013. **7**(12): p. 2287-300.
10. Fieseler, L., et al., *Analysis of the first genome fragment from the marine sponge-associated, novel candidate phylum Poribacteria by environmental genomics.* Environ Microbiol, 2006. **8**(4): p. 612-24.
11. Karlinska-Batres, K. and G. Worheide, *Phylogenetic diversity and community structure of the symbionts associated with the coralline sponge Astrosclera willeyana of the Great Barrier Reef.* Microb Ecol, 2013. **65**(3): p. 740-52.
12. Montalvo, N.F. and R.T. Hill, *Sponge-associated bacteria are strictly maintained in two closely related but geographically distant sponge hosts.* Appl Environ Microbiol, 2011. **77**(20): p. 7207-16.
13. Erwin, P.M., et al., *Stability of sponge-associated bacteria over large seasonal shifts in temperature and irradiance.* Appl Environ Microbiol, 2012. **78**(20): p. 7358-68.
14. Olson, J.B., R.W. Thacker, and D.J. Gochfeld, *Molecular community profiling reveals impacts of time, space, and disease status on the bacterial community associated with the Caribbean sponge Aplysina cauliformis.* FEMS Microbiol Ecol, 2014. **87**(1): p. 268-79.
15. Parks, D.H., et al., *CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes.* Genome Res, 2015. **25**(7): p. 1043-55.
16. Slaby, B.M., et al., *Metagenomic binning of a marine sponge microbiome reveals unity in defense but metabolic specialization.* ISME J, 2017. **11**(11): p. 2465-2478.
17. Siegl, A., et al., *Single-cell genomics reveals the lifestyle of Poribacteria, a candidate phylum symbiotically associated with marine sponges.* ISME J, 2011. **5**(1): p. 61-70.
18. Tully, B.J., E.D. Graham, and J.F. Heidelberg, *The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans.* Sci Data, 2018. **5**: p. 170203.

19. Tully, B.J., E.D. Graham, and J.F. Heidelberg, *The Reconstruction of 2,631 Draft Metagenome-Assembled Genomes from the Global Oceans*, in *bioRxiv*. 2017.
20. Sunagawa, S., et al., *Ocean plankton. Structure and function of the global ocean microbiome.* Science, 2015. **348**(6237): p. 1261359.
21. Leinonen, R., et al., *The sequence read archive.* Nucleic Acids Res, 2011. **39**(Database issue): p. D19-21.