Supporting Information

O'Connor et al. 10.1073/pnas.1413110111

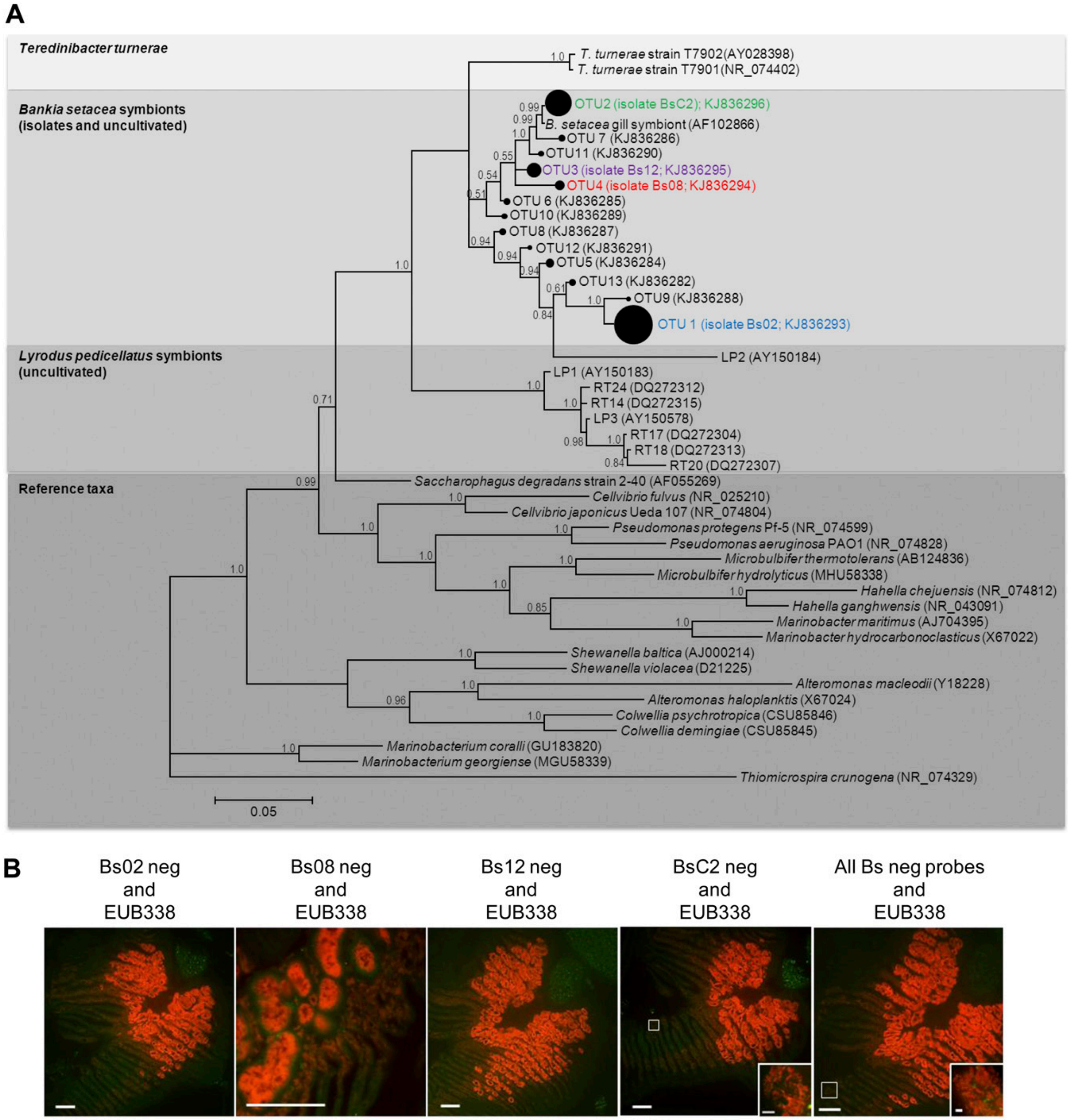
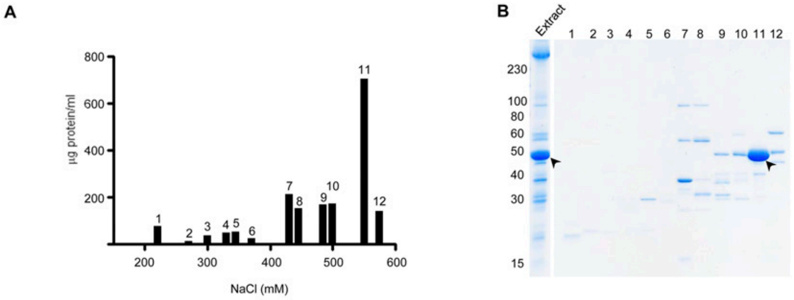


Fig. S1. Gill endosymbiont community of *B. setacea*. (*A*) Phylogeny of cultivated and uncultivated endosymbionts of *B. setacea* inferred by Baysian analysis of 16S rRNA gene sequences (parent tree for Fig. 2*A*). OTUs were identified from among four isolate and 352 cloned 16S rRNA sequences by clustering at 99% identity. For OTUs with more than two members, one representative sequence (isolates and Genbank accession numbers are in parentheses) was included in the analysis: Circles at branch tips are proportional to the fraction of isolate and clone sequences contained in that OTU (Tables S3 and S4). Posterior probabilities are shown at nodes. (Scale bar: 0.05 substitutions per nucleotide position.) (*B*) Negative controls for FISH on gill tissue sections shown in Fig. 2*B*. Five gill tissue sections from *B. setacea* were dual-labeled with the indicated isolate-specific negative control probe (green) plus a probe broadly targeting the domain bacteria (EUB338, red). In this figure, images from the red and green channels are merged. Although strong hybridization was observed with probe EUB338 (positive control), no hybridization was detected for the four negative (neg) control probes. (Scale bar: 100 μm.) (*Insets*) Details of boxed region in the corresponding section. (Scale bar: 5 μm.)



C

Gene ID	Source	Detection method*	Modular structure	N-terminal sequence	Predicted activity
KJ861991	Bs02	MS/N	CBM2-CBM10-GH5_53	^{279}GSGTFRVDSQG^{289}	endo-1, 4-beta-glucanase (EC 3.2.1.4) or cellodextrinase (EC 3.2.1.74)
KJ861972	Bs12	MS/N	CBM2-CBM10-GH6	^{293}DATGLDNPFVGA^{304}	endo-1,4-beta-glucanase (EC 3.2.1.4) or cellobiohydrolase (EC 3.2.1.91)
KJ861960	Bs12	MS	CBM2-CBM10-GH5_53		endo-1, 4-beta-glucanase (EC 3.2.1.4) or cellodextrinase (EC 3.2.1.74)

Fig. S2. Isolation and identification of cecum proteins. Cecum contents were pooled from 12 specimens of *B. setacea* and separated over a Qiagen Q column. Protein peaks (1–12) were collected. (*A*) Protein elution as a function of NaCl concentration. (*B*) Denaturing gel electrophoresis (SDS/PAGE) showing total cecum extract and proteins contained in peaks 1–12 (5 µL per lane). Molecular weight markers (in kilodaltons) are shown on the left. Arrowheads indicate bands corresponding to the GH5_53 and GH6 catalytic modules from proteins KJ861991 (Bs02) and KJ861972 (Bs12). (*C*) Table describing endosymbiont-encoded proteins in fraction 11. MS, LC/MS/MS; N, N-terminal sequencing.

Table S1. Phylogenetic composition of the *B. setacea* gill endosymbiont community: Pyrotag analysis of phylum/class-level diversity

Phylum/class	Percentage of total	No. of reads
Acidobacteria	0.01	216
Bacteroidetes	0.00	45
Proteobacteria	0.99	25,637
Alphaproteobacteria	0.07	1,763
Deltaproteobacteria	0.00	2
Epsilonproteobacteria	0.01	165
Gammaproteobacteria	0.91	23,707
Spirochetes	0.00	22
Miscellaneous	0.00	14
Total		25,934

Table S2. Phylogenetic composition of the *B. setacea* gill endosymbiont community: Cluster analyses of 16S clone library using Qiime de novo

Cluster	Sequences in cluster, %	Isolates included in cluster
Five largest OTUs clustered at 99% identity		
1	46	Bs02
2	23.6	BsC2
3	6.8	Bs12
4	3.1	Bs08
5	2.3	—
Total number of OTUs		37
Number of OTUs represented by a single sequence		19
Five largest OTUs clustered at 98% identity		
1	49.1	Bs02
2	36.9	BsC2, Bs12
3	3.4	Bs08
4	4.8	—
5	2.0	—
Total number of OTUs		11
Number of OTUs represented by a single sequence		3
Five largest OTUs clustered at 97% identity		
1	52.3	Bs02
2	42.9	Bs08, Bs12, BsC2
3	4.3	—
4	0.3	—
5	0.3	—
Total number of OTUs		5
Number of OTUs represented by a single sequence		2

Table S3. *B. setacea* gill endosymbiont metagenome statistics

Genome size	Gene count	Scaffold count	Scaffolds with at least one gene, %	Largest scaffold	No. of scaffolds over 10,000 bp	No. of bases in 100 largest scaffolds	Coding density, %	%GC content
Metagenome*								
26,539,887	43,560	38,227	69	172,446	297	4,787,625	73.66	47.1

*Sequenced on the Illumina and Roche Diagnostics 454 GS Titanium technology platforms.

Table S4. *B. setacea* gill endosymbiont isolate genome statistics

Genome size	Genome size	Gene count	Scaffold count	Scaffolds with at least one gene, %	Smallest scaffold	Largest scaffold	No. of scaffolds over 10,000 bp	No. of bases in scaffolds >10,000 bp	Coding density, %	%GC
Isolate genomes										
Bs02*	3,886,134	3,320	141	95.7	500	655,607	15	3,688,175	87.7	47.8
Bs08*	4,836,222	4,119	123	95.9	514	415,715	33	4,686,552	86.6	47.1
Bs12*	4,582,833	4,202	240	97.5	502	421,677	43	4,292,109	87.5	45.9
BsC2[†]	5,414,953	4,699	10	100	7,974	4,223,372	8	5,398,333	85.4	47.3

*Sequenced on the Illumina platform.

[†] Sequenced on the PacBio platform.

Table S5. Comparison of metagenome and isolate genome sequences: Metagenome reads mapped to assembled isolate genomes

	Bs02	Bs08	Bs12	BsC2	Unmapped reads	Total (mapped and unmapped)
No. of combined read counts mapped to isolate assembly	21,136,321	235,394	6,140,612	33,463,465	11,235,788	72,211,580
Percentage of total	27.65	0.29	8.39	46.06	17.61	100

Table S6. Comparison of metagenome and isolate genome sequences: Isolate reads mapped to each assembled isolate genome and the metagenome

	Isolate genomes				
Isolate reads	Bs02, %	Bs08, %	Bs12, %	BsC2, %	Metagenome,* %
Bs02[†]	95.85	0.51	1.16	0.42	88.43
Bs08[†]	0.41	95.93	0.88	0.61	7.99
Bs12[†]	0.63	0.88	90.12	0.40	78.92
BsC2[‡]	0.69	0.99	0.85	85.92	63.16

*Sequenced on Illumina and Roche Diagnostics 454 GS Titanium technology platforms.

[†] Sequenced on the Illumina platform.

[‡] Sequenced on the PacBio platform.

Table S7. Summary of carbohydrate-active modules types found in isolate genomes and gill endosymbiont metagenome

	Bs02	Bs08	Bs12	BsC2	Metagenome
CBMs					
Total number of CBMs	89	97	92	137	401
Number of CBM families	10	13	14	14	20
CBMs specific for wood components, %	92.1	89.7	90.0	92.7	90.2
GH modules					
Total number of GH modules	84	94	83	128	734
Number of GH families	28	33	26	35	46
GH modules specific for wood components, %	85.7	79.8	78.3	73.4	78.1
CE modules					
Total number of CE modules	7	16	13	21	116
Number of CE families	5	8	7	7	11
CE modules specific for wood components, %	100	93.8	92.3	76.2	88.8
Auxillary activity (AA) families					
AA10 (CBM33)	3	3	2	1	14
Glycosyl transferase (GT) modules					
Total number of GT modules	40	43	42	35	307
Number of GT families	14	14	16	15	17
PL modules					
Total number of PL modules	0	2	1	40	104
Number of PL families	0	2	1	5	5

Table S8. List of individual carbohydrate-active modules (by type) found in isolate genomes and gill endosymbiont metagenome

Module	Bs02	Bs08	Bs12	BsC2	Metagenome	Module	Bs02	Bs08	Bs12	BsC2	Metagenome	Module	Bs02	Bs08	Bs12	BsC2	Metagenome
GHs																	
GH1	0	0	0	0	2	GH26	4	4	4	3	25	GH72	0	0	0	0	1
GH2	4	0	5	4	27	GH27	1	2	1	2	12	GH73	2	2	2	2	7
GH3	6	7	7	8	44	GH28	0	0	0	6	17	GH74	1	1	0	3	16
GH5	17	19	11	14	135	GH30	0	1	0	1	3	GH77	1	2	1	1	11
GH6	7	3	2	1	20	GH31	1	1	0	2	8	GH79	1	1	1	0	4
GH8	1	2	1	1	3	GH33	0	0	0	0	2	GH88	0	0	1	0	2
GH9	5	6	5	6	33	GH35	2	1	1	1	10	GH97	0	0	0	2	8
GH10	6	6	9	11	59	GH42	0	0	0	2	3	GH94	3	3	3	3	22
GH11	4	4	3	1	10	GH43	2	4	3	18	56	GH95	0	0	0	1	2
GH12	0	1	0	1	3	GH44	0	0	0	1	3	GH130	0	0	0	3	0
GH13	3	4	3	3	37	GH45	1	1	1	1	6	GH103	1	1	1	1	6
GH15	0	1	0	0	3	GH51	1	2	1	1	3	GH105	0	1	0	8	25
GH16	2	2	1	2	15	GH52	1	0	0	0	3	GH108	0	1	0	0	0
GH18	0	0	0	0	2	GH53	1	1	5	2	20	GH109	0	0	0	0	2
GH23	4	6	9	7	29	GH62	0	1	0	0	0	GH115	1	1	1	4	22
GH24	0	1	0	0	2	GH67	1	1	1	1	11						
CEs																	
CE1	2	3	2	0	23	CE6	0	1	1	2	5	CE14	0	0	0	0	1
CE2	1	1	2	3	10	CE8	0	0	0	3	6	CE15	1	4	3	0	10
CE3	2	3	2	3	12	CE11	0	1	1	2	6						
CE4	1	2	2	2	25	CE12	0	1	0	6	18						
PLs																	
PL1	0	0	0	22	53	PL9	0	1	1	4	12	PL11	0	0	0	2	14
PL3	0	1	0	6	14	PL10	0	0	0	6	11						
GTs																	
GT1	1	1	2	1	10	GT19	1	1	1	1	9	GT35	1	2	1	1	17
GT2	12	10	15	9	83	GT22	0	0	1	0	0	GT39	0	1	1	1	2
GT4	12	17	10	11	104	GT25	1	0	1	1	3	GT51	1	2	2	2	7
GT5	1	1	1	1	6	GT26	1	1	1	1	11	GT81	1	1	1	1	8
GT9	3	1	1	1	9	GT28	1	1	1	1	9	GT83	2	3	2	2	16
GT10	0	0	0	0	2	GT30	2	1	1	1	11						
CBMs																	
CBM2	11	17	19	34	85	CBM13	0	2	0	7	18	CBM50	5	5	5	6	16
CBM3	0	2	1	2	4	CBM15	0	0	0	0	2	CBM56	0	0	0	0	2
CBM4	1	0	4	4	13	CBM22	0	2	2	5	6	CBM57	0	2	1	1	7
CBM5	4	2	1	23	29	CBM32	0	0	0	2	4	CBM60	4	5	11	3	22
CBM6	1	1	1	0	10	CBM35	0	2	0	34	37	CBM61	0	0	1	0	1
CBM9	1	1	2	1	8	CBM36	2	0	0	0	1						
CBM10	58	53	39	12	107	CBM48	2	3	3	3	15						
AA																	
AA10	3	3	2	1	14												

Table S9. Endosymbiont-encoded proteins found in *B. setacea* gill tissue by COG category

Accession no.	Source	Gene annotation	Accession no.	Source	Gene annotation
Energy production and conversion (C)					
KJ943312	Bs08	Malate dehydrogenase	KJ943309	Bs08	Nitrogenase molybdenum-iron protein β-chain
KJ943310	Bs08	Flavodoxin, long chain	KJ943335	BsC2	2-Oxoglutarate dehydrogenase
KJ943307	Bs08	Nitrogenase iron protein	KJ943336	BsC2	Isopropylmalate dehydrogenase
KJ943308	Bs08	Nitrogenase molybdenum-iron protein α-chain	KJ943359	BsC2	Malate dehydrogenase
Amino acid metabolism and transport (E)					
KJ943298	Bs02	Nitrogen regulatory protein PII	KJ943355	BsC2	Ketol acid reductoisomerase
KJ943324	Bs12	Nitrogen regulatory protein PII	KJ943338	BsC2	Nitrogen regulatory protein PII
KJ943333	BsC2	D-3-phosphoglycerate dehydrogenase	KJ943339	BsC2	Nitrogen regulatory protein PII
KJ943344	BsC2	L-glutamine synthase	KJ943271	M	ATP/guanidophosphotransferase, C-terminal
Carbohydrate metabolism and transport (G)					
KJ861987	Bs02	CBM33-CBM10-CBM10	KJ943287	Bs02	Glyceraldehyde-3-phosphate dehydrogenase
KJ861991	Bs02	CBM2-CBM10-GH5_53	KJ943314	Bs08	Glyceraldehyde-3-phosphate dehydrogenase
KJ861985	Bs02	CBM10-CBM10-CBM60-GH10	KJ943311	Bs08	Transaldolase
KJ861994	Bs02	CBM2-CBM10-GH6	KJ943300	Bs08	Xylose isomerase
KJ861990	Bs02	GH5_8-CBM10-CBM10-CBM10-GH6	KJ943330	Bs12	Glyceraldehyde phosphate dehydrogenase
KJ861982	Bs08	CBM2-CBM10-GH5_53	KJ943331	Bs12	Transaldolase
KJ861978	Bs08	CE6-CBM10-CBM10-CBM60-GH10	KJ943334	BsC2	Transaldolase
KJ861963	Bs12	CBM10-CBM10-GH10	KJ943332	BsC2	Pyruvate kinase
KJ861962	Bs12	CBM10-CBM10-X241	KJ943341	BsC2	Xylose isomerase
KJ861972	Bs12	CBM2-CBM10-GH6	KJ943354	BsC2	Fructose-bisphosphate aldolase (EC 4.1.2.1)
KJ861973	Bs12	CBM33-CBM10	KJ943275	M	D-glucuronate isomerase
			KJ943270	M	Transaldolase
Translation, ribosome structure and biogenesis (J)					
KJ943293	Bs02	Ribosomal protein S7	KJ943321	Bs12	Ribosomal protein S16
KJ943294	Bs02	Translation elongation factor 1A	KJ943317	Bs12	Ribosomal protein L7/L12
KJ943288	Bs02	Ribosomal protein L19	KJ943348	BsC2	Ribosomal protein S7
KJ943292	Bs02	Ribosomal protein L7/L12	KJ943315	Bs12	Ribosomal protein L14
KJ943304	Bs08	Ribosomal protein L7/L12	KJ943351	BsC2	Ribosomal protein L14P
KJ943305	Bs08	Translation elongation factor TU	KJ943352	BsC2	Ribosomal protein L6P
KJ943316	Bs12	Ribosomal protein L3	KJ943347	BsC2	Ribosomal protein L12P
KJ943323	Bs12	Ribosomal protein S1P	KJ943357	BsC2	Ribosomal protein L19
KJ943322	Bs12	Ribosomal protein L19	KJ943349	BsC2	Translation elongation factor 1A
Transcription (K)					
KJ943280	Bs02	Transcription termination factor Rho	KJ943301	Bs08	Transcription termination factor Rho
KJ943295	Bs02	Cold shock DNA-binding module	KJ943328	BS12	Transcription termination factor Rho
Replication, recombination, and repair (L)					
KJ943284	Bs02	Bacterial nucleoid DNA binding protein HU	KJ943318	Bs12	Bacterial nucleoid DNA binding protein
KJ943313	Bs08	Bacterial DNA-binding protein	KJ943358	BsC2	Bacterial nucleoid DNA binding protein
Cell wall structure, biogenesis, and outer membrane (M)					
KJ943326	Bs12	Putative salt-induced outer membrane protein	KJ943274	M	Outer membrane protein (porin)
KJ943353	BsC2	UDP glucose pyrophosphorylase			
Molecular chaperones and related functions (O)					
KJ943279	Bs02	FKBP-type peptidyl-prolyl *cis/trans*-isomerase	KJ943320	Bs12	FKBP-type peptidyl-prolyl *cis/trans*-isomerases
KJ943291	Bs02	Chaperonin GroL	KJ943337	BsC2	Chaperonin GroEL
KJ943299	Bs08	Chaperon protein DnaK	KJ943346	BsC2	FKBP-type peptidyl prolyl *cis/trans*-isomerase
KJ943306	Bs08	Chaperonin GroL	KJ943269	M	ATP-dependent protease Clp, ATPase subunit
Inorganic ion transport and metabolism (P)					
KJ943282	Bs02	TonB-dependent receptor	KJ943327	Bs12	Outer membrane receptor proteins, Fe transport
KJ943289	Bs02	TonB-dependent receptor	KJ943356	BsC2	TonB-dependent receptor
KJ943286	Bs02	TonB-dependent receptor	KJ943343	BsC2	Cyanate hydratase

Table S9. Cont.

Accession no.	Source	Gene annotation	Accession no.	Source	Gene annotation
KJ943283	Bs02	Bacterioferritin	KJ943345	BsC2	TonB-dependent receptor
KJ943281	Bs02	TonB-dependent receptor	KJ943342	BsC2	Phosphate ABC transporter
KJ943302	Bs08	TonB-dependent receptor	KJ943276	M	TonB-dependent receptor
KJ943303	Bs08	TonB-dependent receptor	KJ943272	M	TonB-dependent receptor
KJ943319	Bs12	TonB-dependent receptor	KJ943277	M	Outer membrane receptor for ferrienterochelin
KJ943329	Bs12	Outer membrane receptor for ferrienterochelin			
Secondary metabolites biosynthesis, transport, and catabolism (Q)/lipid transport and metabolism (I)					
KJ943350	BsC2	Dehydrogenases with different specificities			
No functional prediction(s)					
KJ943285	Bs02	DUF2780	KJ943325	Bs12	Hypothetical protein
KJ943297	Bs02	DUF3450	KJ943340	BsC2	Hypothetical protein
KJ943273	M	Hypothetical protein			
Intracellular trafficking, secretion, and vesicular transport (U)					
KJ943290	Bs02	Type IV pilus secretin	2503998172	Bs02	MotA/TolQ/ExbB proton channel
KJ943278	Bs02	General secretion pathway/protein C			

M, found in metagenome only.

Table S10.   Endosymbiont-encoded proteins identified in cecum content of *B. setacea*

					Top BLAST hit to catalytic module			
Accession no.	Source	Linker type	Modular structure	Catalytic module(s)	Annotation	Identity, %	Coverage, %	Accession no.
GHs								
KJ861991	Bs02	S	CBM2-CBM10-GH5_53	GH5_53	Endoglucanase (*S. degradans* 2-40)	75	100	WP_011469891
KJ861979	Bs08	S	CBM2-CBM10-GH5_53	GH5_53	GH family 5 (*T. turnerae*)	74	99	WP_018013554.1
KJ861982	Bs08	S	CBM2-CBM10-GH5_53	GH5_53	Hypothetical protein (*Gilvimarinus chinensis*)	71	99	WP_020210767.1
KJ861960	Bs12	S	CBM2-CBM10-GH5_53	GH5_53	GH family 5 module-containing protein (*T. turnerae* T7901)	73	100	YP_003072082.1
KJ861976	Bs12	Pt	CBM2-CBM10-GH5_53	GH5_53	GH family 5 (*T. turnerae*)	75	100	WP_018013554.1
KJ861955	BsC2	S	CBM2-CBM10-GH5_53	GH5_53	GH family 5 (*T. turnerae*)	72	99	WP_019601273.1
KJ861968	Bs12	S	GH5_8-CBM10-CBM10-CBM10	GH5_8	β-Mannosidase (*T. turnerae*)	73	100	WP_019605211.1
KJ861980	Bs08	S	GH5_8-CBM10	GH5_8	β-Mannosidase (*T. turnerae*)	74	100	WP_019605211.1
KJ861981	Bs08	S	GH5_2-CBM10-CBM2	GH5_2	Endoglucanase (*Paenibacillus* sp. KSM-N145)	57	97	BAF62085.1
KJ861970	Bs12	S	GH5_2-CBM10-CBM2	GH5_2	Hypothetical protein, partial (*T. turnerae*)	73	100	WP_018277780.1
KJ861994	Bs02	S	CBM2-CBM10-GH6	GH6	Hypothetical protein (*T. turnerae*)	76	100	WP_018013966.1
KJ861971	Bs12	pt	CBM2-CBM10-GH6	GH6	Cellobiohydrolase (*T. turnerae* T7901)	76	98	YP_003075283.1
KJ861972	Bs12	s/pt	CBM2-CBM10-GH6	GH6	Cellobiohydrolase A (1,4-β-cellobiosidase A)-like protein (*S. degradans* 2-40)	76	99	YP_527744.1
KJ861995	M		GH6	GH6	Cellobiohydrolase A (1,4-β-cellobiosidase A)-like protein (*S. degradans* 2-40)	77	98	YP_527744.1
KJ861989	Bs02	s	GH9-CBM10-CBM10	GH9	Endoglucanase (*T. turnerae*)	68	99	WP_018414401.1
KJ861969	Bs12	pt	GH9-CBM10	GH9	Endoglucanase (*T. turnerae*)	80	100	WP_018277081.1
KJ861985	Bs02	s	CBM10-CBM10-CBM60-GH10	GH10	Hypothetical protein (*T. turnerae*)	89	99	WP_019604285.1
KJ861963	Bs12	s	CBM10-CBM10-GH10	GH10	β-1,4-xylanase–like protein (*S. degradans* 2-40)	74	100	YP_525657.1
KJ861967	Bs12	pt	GH10-CBM60-CBM60-CBM10	GH10	β-1,4-xylanase (*Cohnella laeviribosi*)	59	97	WP_019004722.1
KJ861961	Bs12	s	CBM2-CBM10-GH45	GH45	Cellulase (*T. turnerae* T7901)	75	100	YP_003074740.1
KJ861958	Bs12	pt	CBM61-CBM60-GH53	GH53	GH, family 53-likely arabinogalactan 1,4-β-galactosidase (*Deinococcus gobiensis* I-0)	44	89	YP_006271586.1
KJ861966	Bs12	s	GH53-CBM60-CBM60-CBM10	GH53	Arabinogalactan endo-1,4-β-galactosidase (*S. degradans* 2-40)	63	100	YP_529177.1
CEs								
KJ861983	Bs08	s	CE1-CBM10-CBM10-CBM10	CE1	Hypothetical protein (*Catelliglobosispora koreensis*)	69	100	WP_020523700.1
KJ861956	Bs12	s	CE3-CBM10-CBM10-CBM10	CE3	Hypothetical protein (*Nonomuraea* sp. ATCC 39727)	67	100	CAD91236.1
KJ861957	Bs12	s/pt	CE3-CBM10-CBM10	CE3	Hypothetical protein (*T. turnerae*)	79	100	WP_019602878.1
KJ861975	Bs12	pt	CBM60-CBM22-CE15	CE15	Cellulose or protein-binding module–containing protein (*C. japonicus* Ueda107)	61	98	YP_001981339.1
AAs								
KJ861987	Bs02	s	AA10(CBM33)-CBM10-CBM10	AA10	Chitin-binding protein (*T. turnerae*)	68	97	WP_019605580.1
KJ861973	Bs12	s	AA10(CBM33)-CBM10-CBM10	AA10	Chitin-binding protein (*T. turnerae*)	70	96	WP_019605580.1
KJ861974	Bs12	pt	AA10(CBM33)-CBM10	AA10	CBM family 33 and 10 module-containing protein (*T. turnerae* T7901)	68	98	YP_003071742.1
Modules of unknown function								
KJ861959	Bs12	s	CBM2-CBM10-X122	X122	Putative cellulose-binding protein (*S. degradans* 2-40)	82	99	YP_526124.1
KJ861992	Bs02	s	CBM10-CBM10-CBM10-X241	X241	Hypothetical protein W97_07402 (*Coniosporium apollinis* CBS 100218)	50	97	EON67905.1

Table S10. Cont.

Accession no.	Source	Linker type	Modular structure	Catalytic module(s)	Top BLAST hit to catalytic module			
					Annotation	Identity, %	Coverage, %	Accession no.
KJ861962	Bs12	s	CBM10-CBM10-X241	X241	Hypothetical protein W97_07402 (*Coniosporium apollinis* CBS 100218)	49	91	EON67905.1
Multimodular								
KJ861990	Bs02	s	GH5_8-CBM10-CBM10-CBM10-GH6	GH5_8	Putative secreted β-mannosidase (*S. degradans* 2-40)	77	99	YP_525540.1|
				GH6	Cellobiohydrolase A (1,4-β-cellobiosidase A)-like protein (*S. degradans* 2-40)	76	100	YP_527744.1
KJ861986	Bs02	s/pt	CE1-CBM60-CBM10-CBM10-CE15	CE1	Esterase (*Micromonospora lupini* str. Lupac 08)	69	100	WP_007461339.1
				CE15	Cellulose or protein-binding module–containing protein (*C. japonicus* Ueda107)	58	98	YP_001981339.1
KJ861978	Bs08	s	CE6-CBM10-CBM10-CBM60-GH10	CE6	Hypothetical protein (*T. turnerae*)	83	95	WP_018275443.1
				GH10	Hypothetical protein (*T. turnerae*)	86	100	WP_019604285.1
KJ861965	Bs12	s	CE6-CBM60-CBM10-CBM60-GH10	CE6	Hypothetical protein (*T. turnerae*)	81	100	WP_018275443.1
				GH10	β-1,4-xylanase–like protein (*S. degradans* 2-40)	76	97	YP_525657.1
KJ861977	Bs08	s	AA10(CBM33)-CBM10-X122-CBM10	AA10	Chitin-binding protein (*C. japonicus* Ueda107)	71	99	YP_001983593.1
				X122	Hypothetical protein (*T. turnerae*)	62	96	WP_019603562.1
KJ861988	Bs02	s	CBM2-X122-pfam04264	X122	Hypothetical protein (*T. turnerae*)	72	98	WP_019603562.1
				Pfam 04264	Hypothetical protein Sde_1182 (*S. degradans* 2-40)	58	99	YP_526656.1
KJ861964	Bs12	s	CBM2-X122-pfam08450-CBM57-pfam07995-CBM2-X132	X122	Hypothetical protein (*T. turnerae*)	74	97	WP_018013849.1
				Pfam 08450	Hypothetical protein (*T. turnerae*)	69	100	WP_019605594.1
				Pfam 07995	Hypothetical protein (*T. turnerae*)	76	100	WP_018276651.1
				X132	Hypothetical protein (*T. turnerae*)	76	99	WP_018417022.1
KJ861993	Bs02	s	GH11-CBM60-CE4-CBM10	GH11	Endo-β-1,4-xylanase (*Cellvibrio mixtus*)	88	100	CAA88761.1
				CE4	Xylanase/chitin deacetylase-like protein (*S. degradans* 2-40)	85	95	YP_528530.1
KJ861984	Bs08	s	GH11-CBM60-CE4-CBM10	GH11	Endo-1,4-β-xylanase (*Cellvibrio* sp. BR)	89	100	WP_007639046.1
				CE4	Xylanase/chitin deacetylase-like protein (*S. degradans* 2-40)	87	97	YP_528530.1
Non–PCWP-active								
2503998993	Bs02		TonB-dependent receptor		TonB-dependent receptor (*Simiduia agarivorans* SA1 = DSM 21679)	43	99	YP_006916014.1

Gene identification numbers are italicized to indicate that the gene sequence is incomplete. M, found in metagenome only; pt, proline- and threonin-rich linker; s, serine-rich linker.

Table S11. Sequences of primers and probes for amplification, hybridization, and cloning (5′–3′)

Primer/probe name	Sequence (5′–3′)
Primers for amplification and sequencing of 16S rRNA genes	
27-F	AGA GTT TGA TCM TGG CTC AG
530-F	GTG CCA GCM GCC GCG G
1055-F	ATG GCT GTC GTC AGC T
800-R	TAC CAG GGT ATC TAA TCC
1101-R	AGG GTT GCG CTC GTT
1492-R	TAC GGY TAC CTT GTT ACG ACT T
27-F′	AGA GTT TGA TCC TGG CTC AG
1391-R	GAC GGG CRG TGW GTR CA
4a-F	TCC GGT TGA TCC TGC CRG
926-F	AAA CTY AAA KGA ATT GAC GG
1392-R	ACG GGC GGT GTG TRC
Oligonucleotide probes for fluorescent in situ hybridization*	
Bs02-AF488	CAC AGA TAG CAG GTA TTA ACC GCT
Bs02neg-AF488	CAC AGA TAG **A**AG **A**TA TTA ACC GCT
Bs08-AF488	TCG AGT ATT AAT CAA CAC CCT TTC
Bs08neg-AF488	TCG AGT ATT AAT **A**AA **A**AC CCT TTC
Bs12-AF488	CTA GAG ATC GTT GAC TTG GTG GGC
Bs12neg-AF488	CTA GAG ATC GT**A** GA**A** TTG GTG GGC
BsC2-AF488	CTA GAG ATC GTA GTC TTG GTA GGC
BsC2neg-AF488	CTA GAG ATC G**A**A GT**A** TTG GTA GGC
EUB338-AF594	GCT GCC TCC CGT AGG AGT
Primers for cloning of GH5_53 catalytic domain	
GH5-F	CTA CCG TTG CGT TTG CTG GAT CGG GAA CTT TC
GH5-R	TTT CAC TTC ACA GGT CAC CCA AAG GAA TTA CAT ACC

Bold indicates mismatch in the negative control probe to the target sequence of the cognate probe. F, forward; R, reverse.

*Probe name (isolate label) is given in the list of oligonucleotide probes for fluorescent in situ hybridization.

Table S12. Sequences of amplification primers for gene-specific in vitro transcription/translation (5′–3′)

GenBank accession no.	Primer name (catalytic module)	Gene-specific sequence (5′–3′)*
Primers for amplification of templates for in vitro transcription-translation		
KJ861970	GH5_2-F	GAC GTA CCG CCA CTC ACC GTT AGT GGC
	GH5_2-R	GCC AAT ATC TAT CGT TGT TCC CCA G
KJ861968	GH5_8-F	GGT TTT TCG GTT TCA GGT ACT CAG CTG
	GH5_8-R	ATA CAC AGA TGC ACG CTG AGC AGT TTG C
KJ861985	GH10a-F	AGC CGT AAC TAC CCA CCT TTC TTC GTG G
	GH10a-R	TTA TAG TTA TCT ACC GAG GTA GTT CAT CAA CCA ATC
KJ861963	GH10b-F	CCA AGT TTT CCA GCA TTT TTT GTT GGT AAT
	GH10b-R	TTA TCG ACC TAGA TAA TCC ATT AAC CAA TC
KJ861993	GH11-F	CAA ACA CTT ACC TCC AAC CAA ACG GGC
	GH11-R	TTA GCC TTC CGA TAC GGT GAT ATC GGA G

*The sequence of the forward primers also includes (5′ to the gene-specific sequence): 5′-GCG AAT TAA TAC GAC TCA CTA TAG GGC TTA AGT ATA AGG AGG AAA AAA TATG-3′ encoding the T7 promoter, the Shine–Dalgarno sequence and the start codon. The sequence of the reverse primers also includes (5′ to the gene-specific sequence) a 3′ UTR for transcript stability: 5′-AAA CCC CTC CGT TTA GAG AGG GGT TAT GCT AG-3′ and a stop codon if needed.

Other Supporting Information Files

Dataset S1 (PDF)