Supporting Information

Jones *et al.* 10.1073/pnas.0804437105

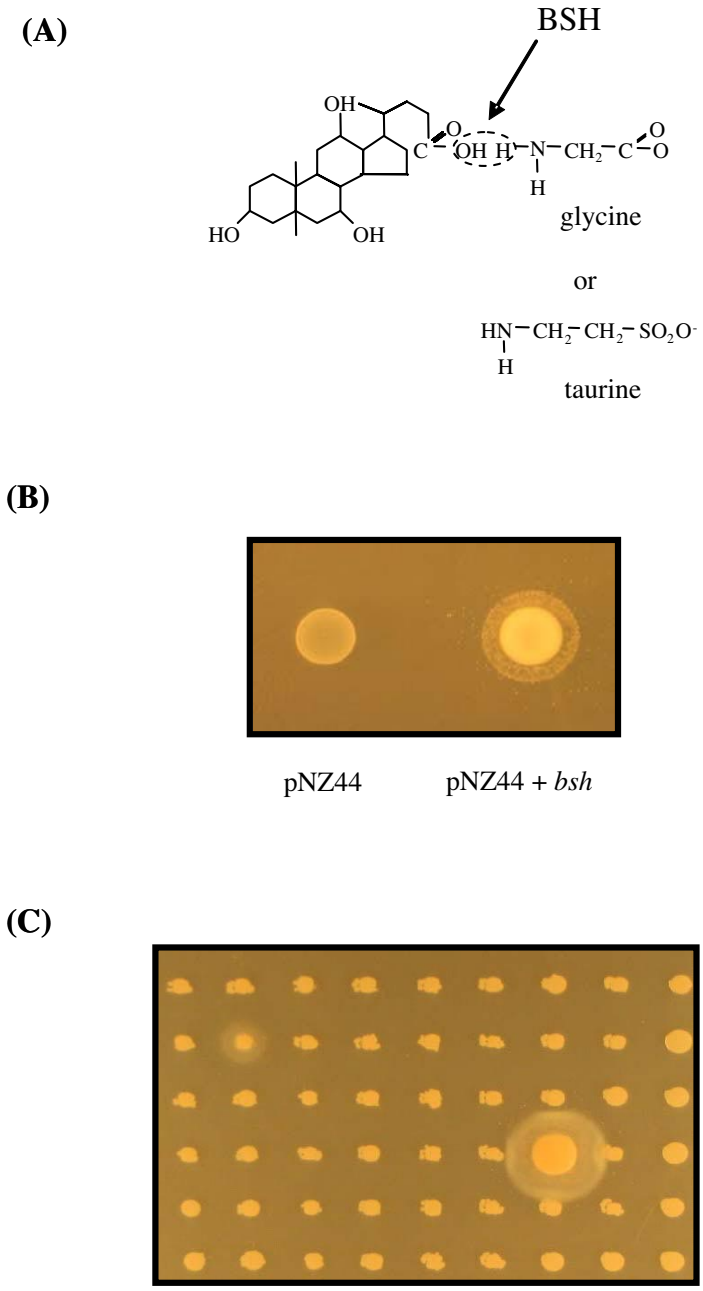

Fig. S1. Reaction catalysed by BSH and example of metagenomic library screening. (*A*) Reaction catalyzed by BSH enzymes. BSHs cleave the peptide linkage of bile acids to liberate the amino acid side chain from the steroid core. The resulting deconjugated free bile acids precipitate at low pH. (*B*) Creation of a BSH-positive *Escherichia coli* strain. *Bsh* from *Lactobacillus plantarum* was cloned into pNZ44 and used as a positive control to confirm that BSH could be detected in our functional metagenomic screens. A bile agar plate assay was used to examine the bile salt hydrolase activity of the resulting strain, and subsequently screen our metagenomic library. *E. coli* transformed with empty vector was used as a negative control. Cultures that were grown in LB broth were spotted (10 μl) onto LB-bile agar supplemented with 0.5% (wt/vol) taurodeoxycholic acid (TDCA; Sigma T0875). Strains with bile salt hydrolase activity were surrounded with a halo of precipitated deconjugated bile acids. (*C*) An example of a plate from our functional metagenomic screen for BSH-positive clones. Our metagenomic library was screened by using a Genetix Q-PIX2 XT colony-picking robot to replicate metagenomic clones onto LB-bile agar. Clones expressing BSH activity exhibited a halo of precipitated free bile acids around the colony.

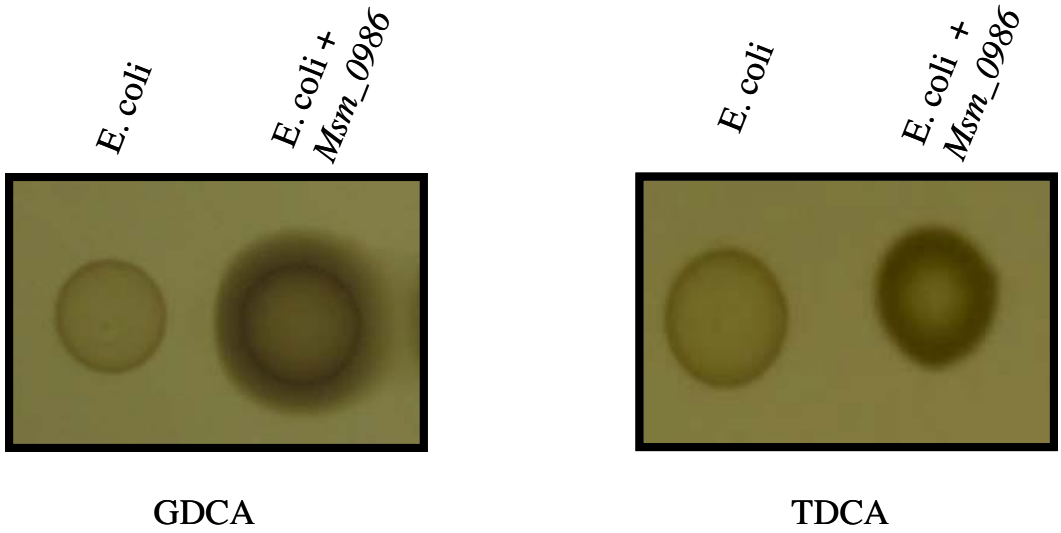

Fig. S2. Cloning of BSH from *Methanobrevibacter smithii* 3142. *Msm_0986* (NC_009515) from *M. smithii* was cloned in *E. coli*. Cells transformed with empty vector were used as a negative control. Cultures that were grown in LB broth were spotted (10 μl) onto LB-bile agar supplemented with either 0.5% (wt/vol) taurodeoxycholic acid (TDCA; Sigma T0875) or glycodeoxycholic acid (GDCA; Sigma G9910). BSH activity is indicated by white halos of precipitate-free bile acids.

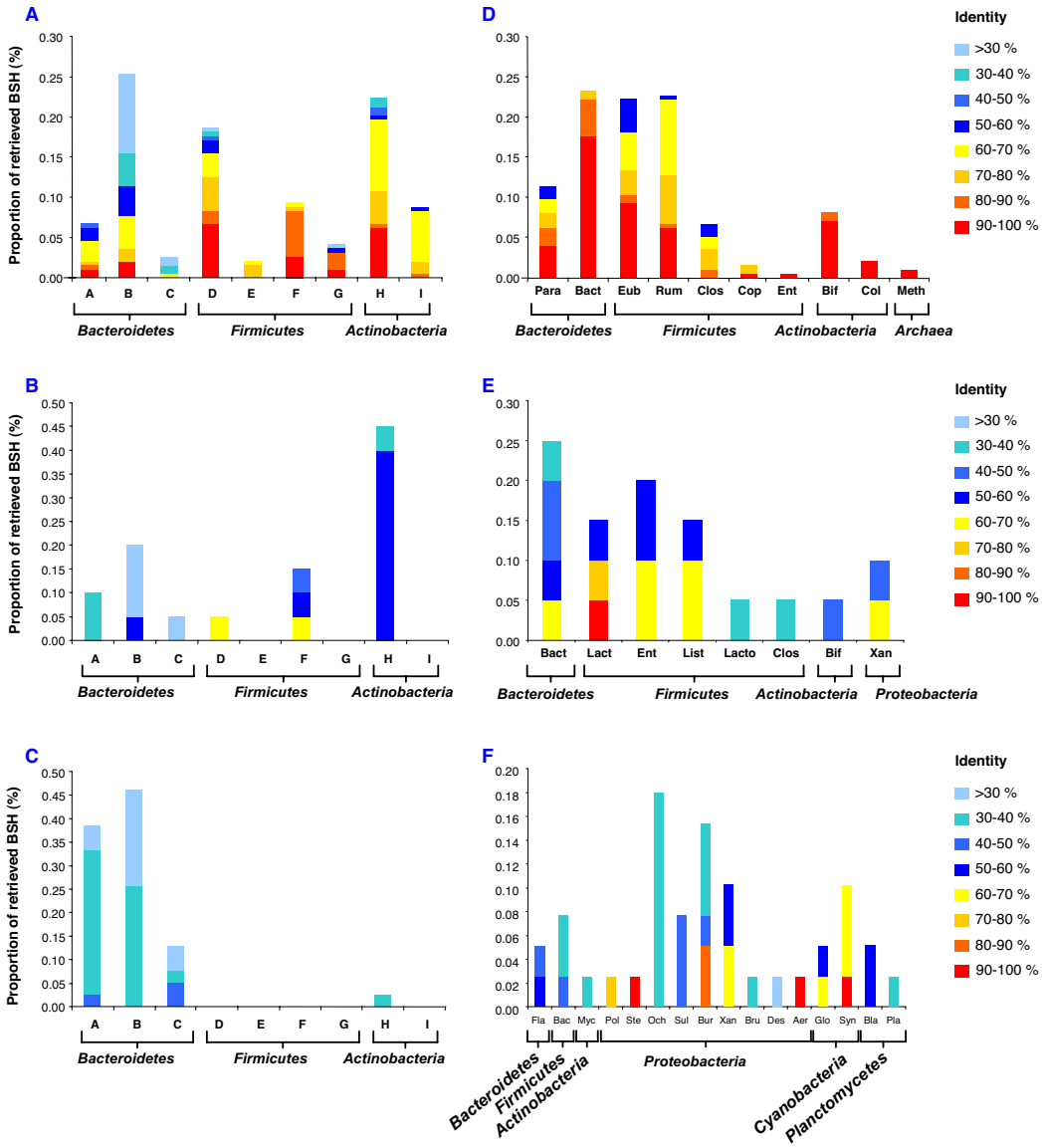

Fig. S3. Predicted origin of putative BSH and related proteins retrieved from human, murine, and Sargasso sea metagenomes. Amino acid sequences representative of each of our functional BSH types A-I (Table S3) were compared with 15 human gut metagenomes, Sargasso sea, soil, and the combined gut metagenomes of lean and obese mice by using tBlastn. Hits with significant e values ($1e^{-8}$ or lower) and a length of 30 aa or more were retrieved, and homology compared with BSH and related proteins in our extensive in-house database. Of the sequences retrieved 96.4% were >100 aa in length and 62.5% were >200 aa in length. Because this analysis was concerned with BSH ORFs only, rather than the putative phylogenetic origin of the entire fosmid insert, for the purposes of this analysis BSH types were affiliated to phylogenetic groups based on their homologies to proteins from bacterial species. Therefore, because BSH types E, F, G (unclassified by end-sequence-based phylogenetic affiliation) all exhibited high identity to proteins from species belonging to the Firmicute division (Table S3), for our comparative metagenomic analysis, these BSH types were affiliated with the Firmicutes division. For all other BSH types, phylogentic affiliation matched end-sequence-based grouping (Table S3). Bars represent percentage of recovered sequences with highest homology to functional BSH types or sequences from species with sequences represented in our extensive in house database. Colors within bars indicate per cent identity of recovered sequences to homologous sequences. (*A–C*) Proportion of retrieved amino acid sequences from human, murine, and Sargasso sea metagenomes, respectively, homologous to our functional BSH types A–I. The majority of sequences retrieved from human metagenomes showed high identity to our functional BSH and were distributed among BSH types derived from all affiliated bacterial divisions. (*D–F*) Homology of sequences recovered from human, murine, and Sargasso sea metagenomes respectively, to BSH and related proteins in various Bacterial and Archaeal genera represented in our extensive in house database. All retrieved sequences from human gut metagenomes exhibited homology to proteins from bacterial species identified in the human gut microbiota representing all major bacterial divisions and also archaeal species. Bacteria: Para, *Parabacteroides*; Bact, *Bacteroides*; Eub, *Eubacterium*; Rum, *Ruminococcus*; Clos, *Clostridium*; Cop, *Coprococcus*; Ent, *Enterococcus*; Bif, *Bifidobacteria*; Col, *Collinsella*; Lact, *Lactobacillus*; List, *Listeria*; Lacto, *Lactococcus;* Xan, *Xanthobacter*; Bac, *Bacillus*; Myc, *Mycobacterium*; Pol, *Polynucleobacter*; Ste, *Stenotrophomonas*; Och, *Ochrobacterium*; Sul, *Sulfurovum*; Bur, *Burkholderia*; Bru, *Brucella*; Fla, *Flavobacterium*; Des, *Desulfotalea*; Aer, *Aeromonas*; Glo, *Gloeobacter*; Syn, *Synechococcus;* Bla, *Blastopirellula*; Pla, *Planctomycetes*. Archaea: Meth, *Methanobrevibacter*.

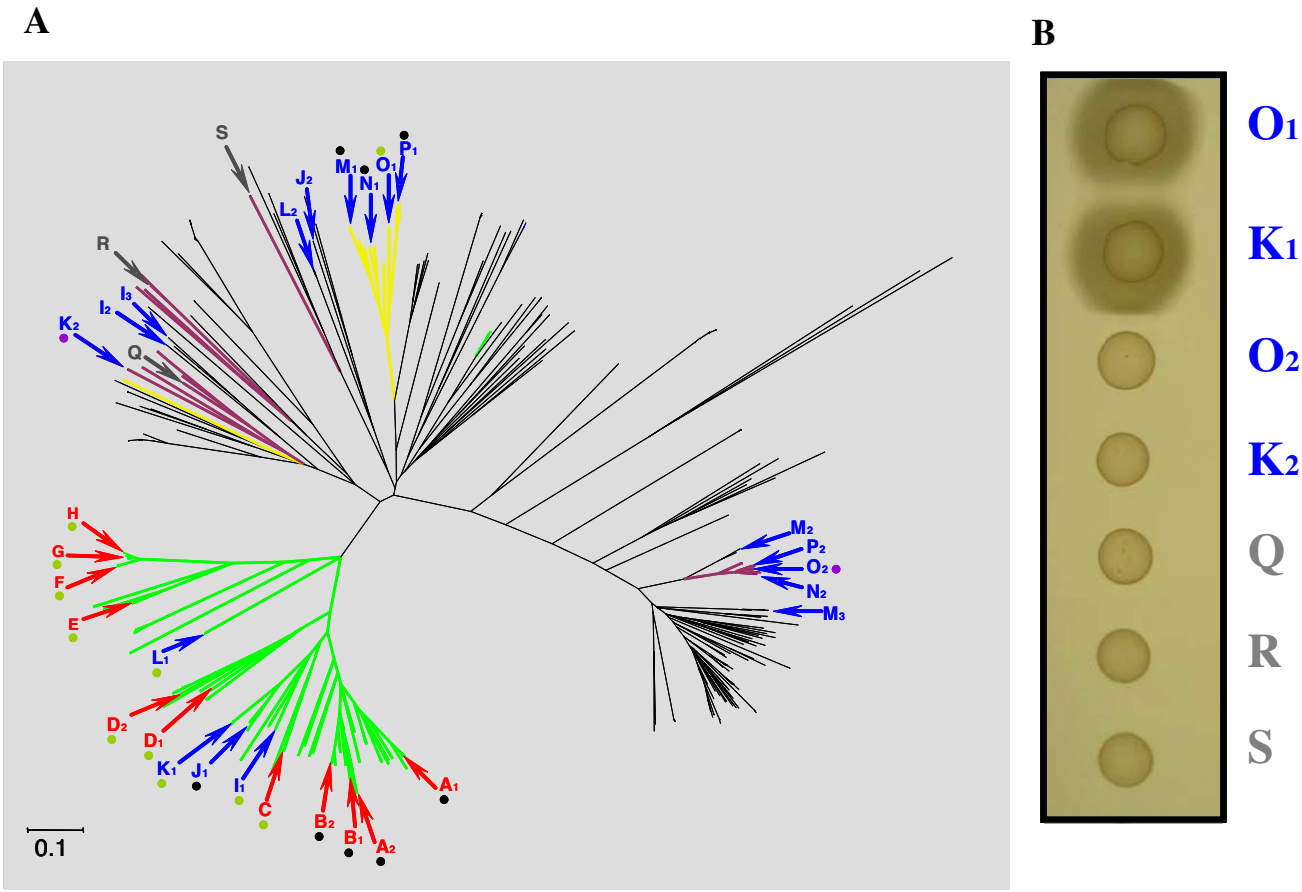

Fig. S4. Prediction of BSH or PVA activity based on phylogenetic analysis of amino acid sequences. (*A*) Examples of species respresented in our phylogenetic analysis that are predicted or confirmed to encode both BSH and PVA activity, or BSH activity only. Green branches represent clades containing sequences with proven activity against glyco-CAB and tauro-CBA, whereas yellow branches indicate clades containing sequences with proven activity against tauro-CBA only. Purple branches represent clades with sequences which have been proven to exhibit no BSH activty. Yellow circle indicates sequences with proven BSH activity (Table S2), black circle indicates sequences with predicted BSH activity based on >60% identity to our functional BSH, and magenta circle indicates sequences proven to exhibit no BSH activity. Underlined text indicates GenBank accession numbers for sequences. Red arrows indicate sequences from species predicted or confirmed to encode BSH activity only: A1, A2, *Eubacterium ventriosum* ATCC27560: *ZP_02027297.1, ZP_02026836.1,* respectively; B1, B2, *Ruminococcus obeum* ATCC29174: *EDM88905.1*, *EDN78955.1,* respectively; C, *Methanobrevibacter smithii* ATCC35061, *ABQ87191.1*; D1, D2, *Lactobacillus acidophilus* NCFM/ ATCC700396: *AAV42751.1, AAV42923.1*, respectively; E, *Collinsella aerofaciens* ATCC25986, *ZP 01771587.1*; F, *Bifidobacterium adolescentis* L2-32, *EDN82839.1*; G, *Bifidobacterium bifidum* ATCC15696, *AAT11512.1*; H, *Bifidobacterium longum* SBT2928, *AAF67801.1*. Blue arrows indicate sequences from species predicted or confirmed to encode both BSH and PVA activity: I1, I2, I3, *Lactobacillus plantarum* WCFS1: *CAD65617.1*, *NP_783921.1*, *NP_786598.1*, respectively; J1, J2, *Enterococcus faecalis* V583: *AAO80370.1*, *NP_816620.1*, respectively; K1, K2, *Listeria monocytogenes* EGD-e: *CAD00145.1*, *NP_463975.1*, respectively; L1, L2, *Clostridium perfringens* Str.13: *BAB80415.1*, *BAB80798.1*, respectively; M1, M2, M3, *Parabacteroides distasonis* ATCC8503: *ABR43305.1*, *ABR42134.1*, *ABR42537.1*, respectively; N1, N2, *Bacteroides uniformis* ATCC8429: *EDO53654.1*, *EDO55730.1*, respectively; O1, O2, *Bacteroides vulgatus* ATCC8482: *ABR40351.1*, *ABR41596.1,* respectively; P1, P2; *Bacteroides ovatus* ATCC8483: *EDO11155.1*, *EDO11390.1*, respectively. Gray arrow indicates sequences from species predicted and confirmed not to encode BSH activity: Q, *Bacillus licheniformis* DSM 13, *AAU43030*; R, *Lactococcus lactis* subsp. *cremoris* MG1363, *CAL98005*; S, *Salmonella typhimurium* LT2, *AAL20459*. (*B*) Confirmation of activity of selected sequences. BSH activity for selected sequences was predicted based on location in our phylogenetic tree. Predicted activity was confirmed by subcloning of relevant ORFs into the expression vector pNZ44 and assessing clone activity by using LB-bile agar with 0.5% (wt/vol) TDCA (Sigma T0875). BSH activity was indicated by halos of precipitated deconjugated bile acids around activity clones. The designation of clones shown corresponds to sequences identified in *A*. O1, O2, *Bacteroides vulgatus* ATCC8482: *ABR40351.1*, *ABR41596.1*; respectively, K1, K2, *Listeria monocytogenes* EGD-e: *CAD00145.1*, *NP_463975.1*, respectively. Q, *Bacillus licheniformis* DSM 13, *AAU43030*; R, *Lactococcus lactis* subsp. *cremoris* MG1363, *CAL98005*; S, *Salmonella typhimurium* LT2, *AAL20459*.

Table S1. Tentative phylogenetic affiliation of clones retrieved in functional metagenomic analysis

Affiliated bacterial division*	% of clones affiliated to each division	Average G + C content of assigned clones†	Average G + C of sequenced representatives‡	Range of G + C content of sequenced representatives§
Actinobacteria	8.9	59.34	63.88	46.3–74.2
Bacteroidetes	14.4	49.22	43.23	32.5–66.1
Firmicute	30	44.1	34.65	23.8–55.8
Spirochete	2.2	42.30	36.13	27.8–52.8
Proteobacteria	1.1	50.75	48.86	16.6–74.9
Unclassified¶				
Total	43.3	-	-	-
No significant hits	8.9			
Different divisions	7.8			
Criteria not met	26.7			

*Fosmid clones recovered from function-driven metagenomics were subjected to a tentative phylogenetic affiliation based on end sequence data. Results from tBlastX searches were filtered by using the following criteria: minimum 35% identity over 50 amino acids or more and an e-value of 1^{-15} or lower. Sequences with hits matching these criteria were used to assign clones to a bacterial division based on species hit. Where search results fit criteria but hit species from different bacterial divisions, clones were not assigned.

†End sequences from all clones assigned to a bacterial division were used to estimate an average G + C content, and strengthen phylogenetic assignment.

‡G + C contents of sequenced genomes in each bacterial division to which clones were affiliated were calculated.

§Range of G + C contents observed for complete bacterial genomes from each bacterial division available through National Center for Biotechnology Information (NCBI).

In addition to G + C contents, the phylogenetic affiliations are further strengthened by the following additional observations: (*i*) Complete BSH ORFs along with flanking regions obtained from 19 clones associated with various bacterial divisions exhibited homologies to sequences from bacterial species that matched the end sequence-based phylogenetic affiliation in all cases (Table S3). (*ii*) The complete nucleotide sequences for three fosmid inserts, and a contiguous 14-kb region of a fourth clone exhibited consistent phylogenetic affiliation across the length of the sequence, and matched the end sequence-based groupings for these clones (data not shown). (*iii*) One clone, designated FM5 (Table S2) was found to encode a 23S rRNA gene with high identity to a 23S rRNA gene from *Rubrobacter xylanophilus* DSM 9941 (76% identity, e = 0.0), which is a member of the Actinobacteria, showed similar identity to 23S rRNA genes from other Actinobacteria, and matched the end sequence-based phylogenetic affiliation from this clone.

¶Clones that could not be classified were composed of three distinct subgroups: (*i*) Clones in which end sequences did not generate any significant hits during tBlastX searches and potentially represent novel uncultured members of the gut microbiota, (*ii*) clones in which both end sequences generated significant hits that satisfied the criteria but were homologous to species belonging to different bacterial divisions, and (*iii*) clones in which end sequences did not satisfy the criteria.

Table S2. BSH activity of our metagenomic clones and previously characterized BSH enzymes

Domain/division	Clones/origin*	Tauro-CBA†	Glyco-CBA†	Human bile†	Source/reference
Bacteria					
Actinobacteria					
	FM 5	+	+	+	This study
	FM 6	+	+	+	This study
	FM 7	+	+	+	This study
	FM 17	+	+	+	This study
	FM 75	+	+	+	This study
	FM 78	+	+	+	This study
	FM 89	+	+	+	This study
	FM 98	+	+	+	This study
	Bifidobacterium longum SBT2928 (AAF67801)	+	+	ND	2
	Bifidobacterium adolescentis ATCC15705 (AAX86039)	+	+	ND	3
	Bifidobacterium bifidum ATCC11863	+	+	ND	4
Bacteroidetes					
	FM1	+	-	-	This study
	FM16	+	-	-	This study
	FM20	+	+	-	This study
	FM26	+	-	-	This study
	FM57	+	-	-	This study
	FM63	+	+	-	This study
	FM65	+	-	-	This study
	FM71	+	-	-	This study
	FM74	+	+	-	This study
	FM92	+	-	-	This study
	FM97	+	-	-	This study
	B. vulgatus VI-31	+	-	ND	5
	B. vulgatus ATCC8482	+	-	ND	This study
Firmicutes					
	FM 4	+	+	+	This study
	FM 29	+	+	+	This study
	FM 79	+	+	+	This study
	FM 85	+	+	+	This study
	FM 95	+	+	+	This study
	FM 2	+	+	+	This study
	FM 3	+	+	+	This study
	FM 8	+	+	+	This study
	FM 13	+	+	+	This study
	FM 14	+	+	+	This study
	FM 15	+	+	+	This study
	FM 18	+	+	+	This study
	FM 19	+	+	+	This study
	FM 22	+	+	+	This study
	FM 24	+	+	+	This study
	FM 25	+	+	+	This study
	FM 31	+	+	+	This study
	FM 39	+	+	+	This study
	FM 43	+	+	+	This study
	FM 44	+	+	+	This study
	FM 52	+	+	-	This study
	FM 56	+	+	+	This study
	FM 61	+	+	+	This study
	FM 62	+	+	+	This study
	FM 70	+	+	+	This study
	FM 76	+	+	+	This study
	FM 77	+	+	+	This study
	FM 81	+	+	+	This study
	FM 83	+	+	+	This study
	FM 87	+	+	+	This study
	FM 88	+	+	+	This study
	FM 96	+	+	+	This study
	FM 100	+	+	+	This study
	FM 101	+	+	+	This study
	FM 72	+	+	+	This study
	FM 86	+	+	+	This study

Domain/division	Clones/origin*	Tauro-CBA†	Glyco-CBA†	Human bile†	Source/reference
	FM 91	+	+	+	This study
	Listeria monocytogenes EGDe (CAD00145)	+	+	+	6, 7
	*Clostridium perfringens*13 (P54965)	+	+	ND	8
	Enterococcus faecium FAIR-E345 (AAP20760)	+	ND	ND	9
	Lactobacillus acidophilus NCFM (AAV42751 BshA)	-	+	ND	10
	Lactobacillus acidophilus NCFM (AAV42923 BshB)	+	+	ND	10
	Lactobacillus acidophilus KS-13 (AAD03709)	+	ND	ND	11
	Lactobacillus plantarum 80 (AAB24746)	+	+	ND	1
	Lactobacillus plantarum WCFS1 (CAD65617 Bsh1)	+	+	+	This study
	Lactobacillus johnsonni 100–100 alpha (AAG22541)	+	ND	ND	12
	Lactobacillus johnsonni 100–100 beta (AAC34381)	+	ND	ND	13
Spirochete					
	FM 42	+	+	-	This study
Proteobacteria					
	Brucella abortus 2308	ND	+	ND	14
Archaea					
Euryarchaeota	*Methanobrevibacter smithii* 3142	+	+	ND	This study

*FM, Functional Metagenomic. Indicates fosmid clones retrieved through our functional metagenomic analysis. National Centre for Biotechnology Information accession numbers of BSH enzymes are given in parentheses.

†Bile salt hydrolase activity was examined by using a plate assay as described (1). Media were supplemented with either 0.5% (wt/vol) tauro-CBA (TDCA, taurodeoxycholic acid; Sigma T0875), 0.5% (wt/vol) glyco-CBA (GDCA, glycodeoxycholic acid; Sigma G9910) or 3% (v/v) human bile (obtained from gallbladders at laparascopic cholecystectomy). Bile salt hydrolase activity was indicated by halos of precipitated deconjugated bile acids. ND = activity not determined.

1. Christiaens H, Leer RJ, Pouwels PH, Verstraete W (1992) Appl Environ Microbiol 58:3792–3798.
2. Tanaka, *et al.* (2000) *Appl Environ Microbiol* 66:2502–2512.
3. Kim, *et al.* (2005) *Biotechnol Lett* 27:817–822.
4. Kim, *et al.* (2004) *Appl Environ Microbiol* 70:5603–5612.
5. Kawamoto, *et al.* (1989) *J Biochem* 106:1049–1053.
6. Dussurget O, *et al.* (2002) *Mol Microbiol* 45:1095–1106.
7. Begley M, Sleator RD, Gahan CG, Hill C (2005) *Infect Immun* 73:894–904.
8. Coleman, Hudson (1995) *Appl Environ Microbiol* 61:2514–2520.
9. Wijaya, *et al.* (2004) *J Food Prot* 67:2772–2778.
10. McAuliffe, *et al.* (2005) *Appl Environ Microbiol* 71:4925–4929.
11. Moser, Savage (2001) *Appl Environ Microbiol* 67:3476–3480.
12. Elkins, *et al.* (2001) *Microbiology* 147:3403–3412.
13. Elkins, Savage (1998) *J Bacteriol* 180:4344–4349.
14. Delpino MV, *et al.* (2007) *Infect Immun* 75:299–305.

Table S3. Identification of sequences homologous to functional BSH obtained through function-driven metagenomic analysis

Division*	BSH type†	Homologous sequences‡	Identity§	Alignment to Ntn_PVA-conserved domains¶, %
Bacteroidetes	A (1 clone)	*Bacteriodes ovatus* ATCC8483, hypothetical protein, BACOVA_03057 ZP_2066063	80% (287/355)	100
	B (2 clones, 100%)	*Bacteroides uniformis* ATCC8429, hypothetical protein, BACUNI_02933 ZP_02071494.1	100% (361/361)	100
	C (1 clone)	*Bacteroides uniformis* ATCC8429, hypothetical protein, BACUNI_02933 ZP_02071494.1	67% (241/358)	100
Firmicutes	D (5 clones, 99–100%)	*Eubacterium ventriosum* ATCC27560, hypothetical protein, EUBVEN_02567 ZP_02027297.1	99% (326–327/329)	100
Unclassified	E (2 clones, 100%)	*Eubacterium ventriosum* ATCC27560, hypothetical protein, EUBVEN_02567 ZP_02027297.1	70% (234/333)	100
	F (3 clones, 99–100%)	*Ruminococcus obeum* ATCC29174, hypothetical protein, RUMOBE_03454 ZP_01965714.1	99–100% (302–322/322)	100
	G (1 clone)	*Ruminococcus obeum* ATCC29174, hypothetical protein,RUMOBE_00028 ZP_01962315.1	76% (250/325)	100
Actinobacteria	H (1 clone)	*Collinsella aerofaciens* ATCC25986, hypothetical protein, COLAER_00574 ZP_01771587	71% (219/307)	98.35
	I (3 clones, 99–100%)	*Bifidobacterium adolescentis* L2–32, bile salt hydrolase, BIFADO_01120 ZP_02028683	99% (344–345/347)	100%

*Tentative phylogenetic affiliation of clones with each BSH type based on end sequence data.
†Distinct BSH types identified from mutagenesis or subcloning of 19 clones obtained through functional metagenomic approach. Complete BSH ORF from each clone were compared and grouped according to amino acid identity. Figures in parenthes3s indicate the number of clones identified with each BSH type, then per cent identity between those clones.
‡BlastP search results showing top hits to representative sequences of each BSH type from our function driven metagenomic analysis.
§Percentage identity of homologous sequences from BlastP searches. All e values were $1e^{-117}$ or lower.
¶Percentage alignment of representative sequences from each BSH type to conserved domains of the Ntn_PVA family of proteins. All e-values were $4e^{-75}$ or lower. All sequences showed highest homology to conserved domains of the Ntn_PVA family, part of the wider Ntn_CGH-like family of proteins.