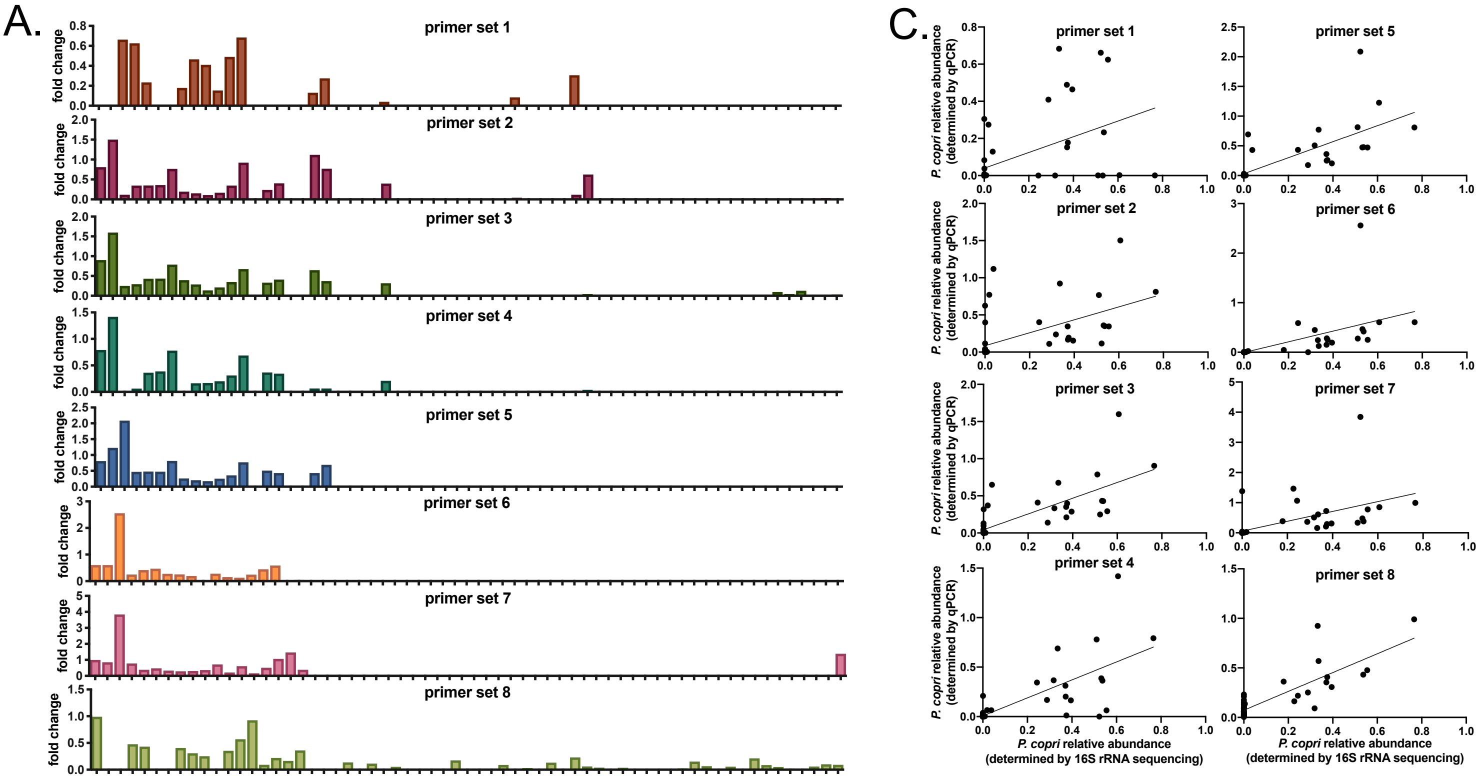
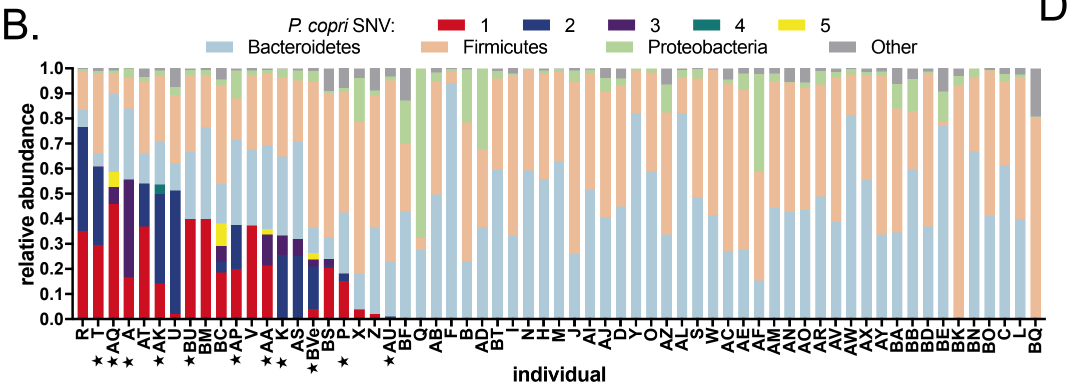
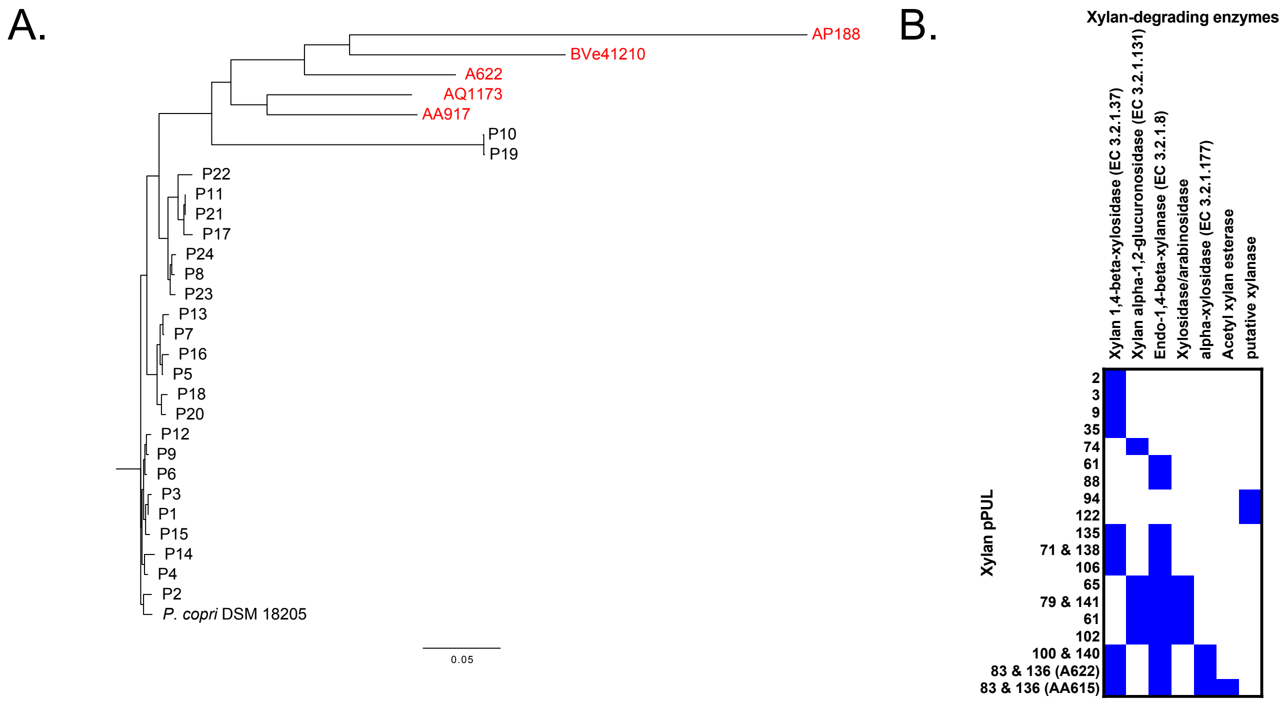




D.

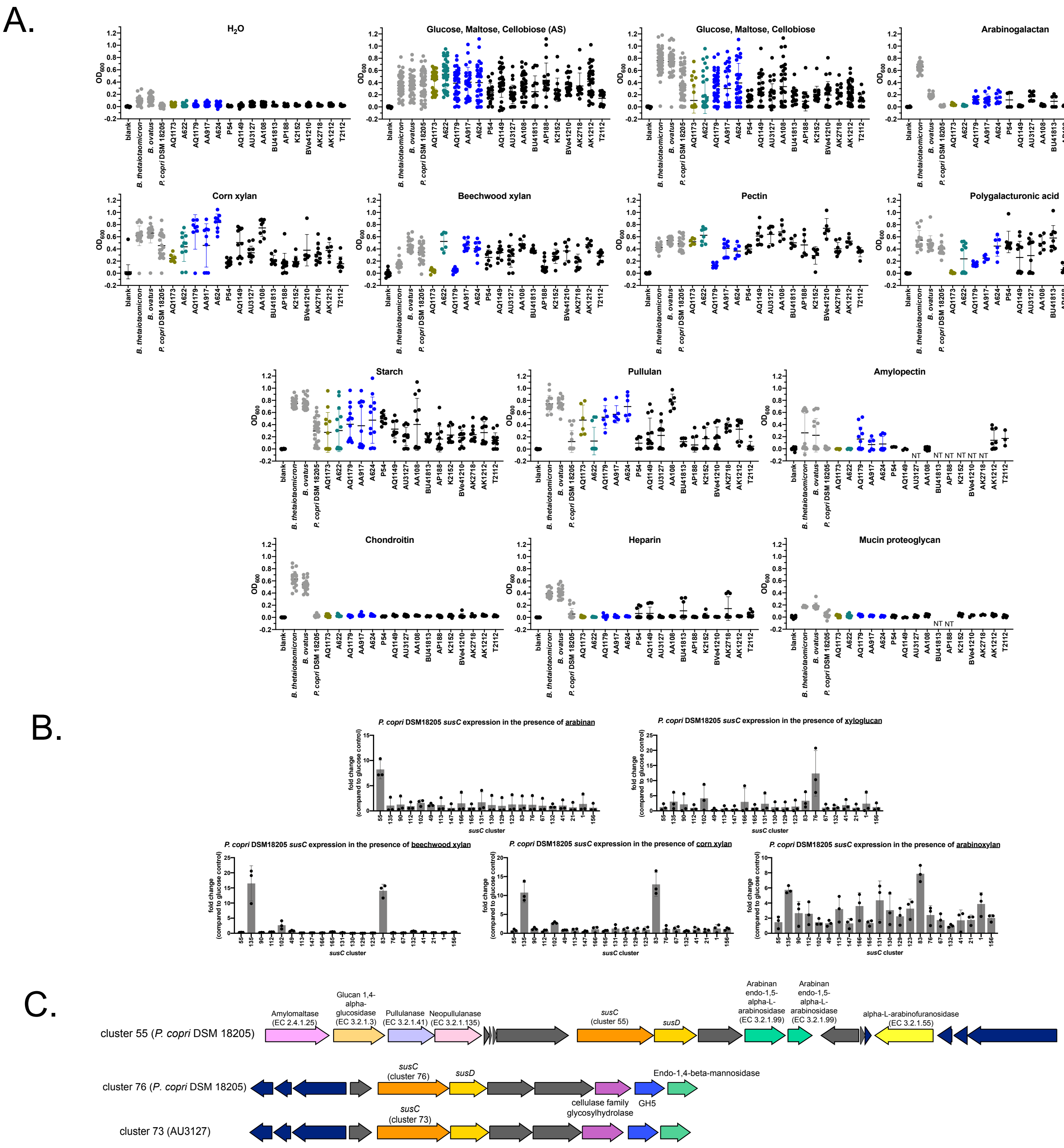
Donor	Number of *P. copri* isolates	Health status
K	6	Healthy control
AK	28	Healthy control
AU	1	Healthy control
A	2	NORA
P	2	NORA
AA	3	NORA
T	17	NORA
AP	10	NORA
AQ	8	NORA
BU	3	NORA
BVe	3	NORA
Total	83	

Figure S1. Screen for and isolation of *P. copri* from human stool; related to Figure 1
A. Eight sets of primers were designed to amplify conserved regions of the *P. copri* genome. Fecal DNA from each donor was screened by qPCR using each primer set (**Methods**). **B.** Relative abundance of *P. copri* 16S SNVs in human stool quantified by 16S sequencing. Percentage of total reads from 16S rRNA gene sequencing is shown. See also File S1. Stars in (**A**) and (**B**) indicate individuals from whom *P. copri* was isolated. **C.** Correlation of *P. copri* abundance determined by 16S sequencing and qPCR using each qPCR primer set. **D.** Number of *P. copri* isolates collected from individual donors. Health status is shown. NORA: new onset rheumatoid arthritis.

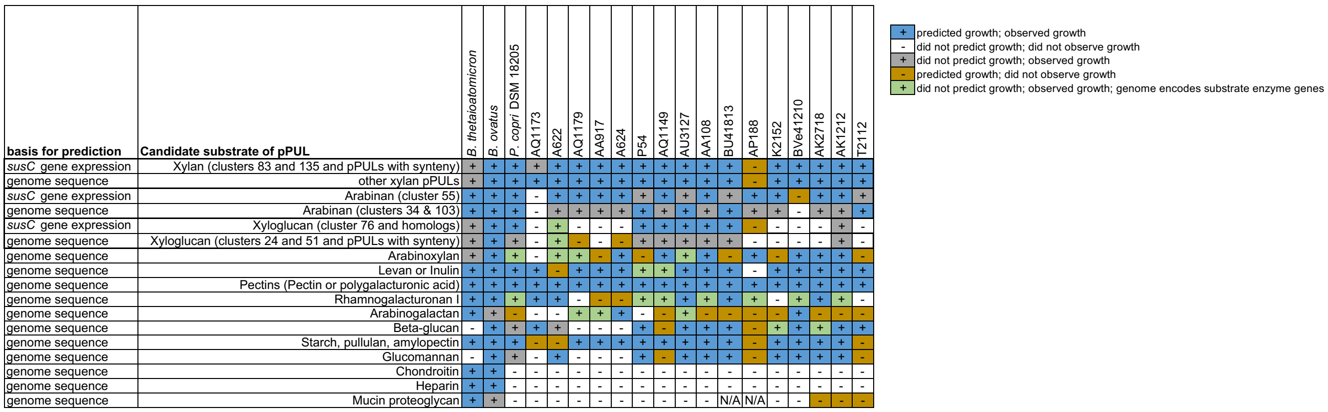


Xylan pPUL	Xylan 1,4-beta-xylosidase (EC 3.2.1.37)	Xylan alpha-1,2-glucuronosidase (EC 3.2.1.131)	Endo-1,4-beta-xylanase (EC 3.2.1.8)	Xylosidase/arabinosidase	alpha-xylosidase (EC 3.2.1.177)	Acetyl xylan esterase	putative xylanase
2	■						
3	■						
9	■						
35	■						
74		■					
61			■				
88			■				
94							■
122							■
135	■		■				
71 & 138	■		■				
106	■		■				
65		■	■	■			
79 & 141		■	■	■			
61		■	■	■			
102		■	■	■			
100 & 140	■		■		■		
83 & 136 (A622)	■		■		■		
83 & 136 (AA615)	■		■		■	■	

Figure S2. Alignment of *Prevotella* 16S V1-V3 sequences; related to Figure 2
A. Representative isolates were selected and the full length 16S sequence was determined by Sanger sequencing. 16S hypervariable V1-V3 sequences were aligned with *Prevotella* 16S V1-V3 sequences described in De Filippis, et al. (2016), labeled as P1-P23, using Clustal Omega. A simple phylogeny is shown. Isolates from this study are highlighted in red. **B.** Xylan-degrading enzyme annotations in each pPUL are shown. Presence of each xylan-degrading enzyme in each pPUL is indicated by a blue box.



D.



basis for prediction	Candidate substrate of pPUL	*B. thetaiotaomicron*	*B. ovatus*	*P. copri* DSM 18205	AQ1173	A622	AQ1179	AA917	A624	P54	AQ1149	AU3127	AA108	BU41813	AP188	K2152	BVe41210	AK2718	AK1212	T2112
susC gene expression	Xylan (clusters 83 and 135 and pPULs with synteny)	⬛+	🟦+	🟦+	⬛+	🟦+	🟦+	🟦+	🟦+	🟦+	🟦+	🟦+	🟦+	🟦+	🟧-	🟦+	🟦+	🟦+	🟦+	🟦+
genome sequence	other xylan pPULs	⬛+	🟦+	🟦+	🟦+	🟦+	🟦+	🟦+	🟦+	🟦+	🟦+	🟦+	🟦+	🟦+	🟧-	🟦+	🟦+	🟦+	🟦+	🟦+
susC gene expression	Arabinan (cluster 55)	🟦+	🟦+	🟦+	⬜-	🟦+	🟦+	🟦+	🟦+	⬛+	🟦+	⬛+	🟦+	⬛+	🟦+	🟦+	🟧-	🟦+	🟦+	⬛+
genome sequence	Arabinan (clusters 34 & 103)	🟦+	🟦+	🟦+	⬜-	⬛+	⬛+	⬛+	⬛+	🟦+	⬛+	🟦+	⬛+	🟦+	⬛+	⬛+	⬜-	⬛+	⬛+	🟦+
susC gene expression	Xyloglucan (cluster 76 and homologs)	⬛+	🟦+	🟦+	⬜-	🟩+	⬜-	⬜-	⬜-	🟦+	🟦+	🟦+	🟦+	🟦+	🟧-	⬜-	⬜-	⬜-	⬛+	⬜-
genome sequence	Xyloglucan (clusters 24 and 51 and pPULs with synteny)	⬛+	🟦+	⬛+	⬜-	🟩+	🟧-	⬜-	🟧-	⬛+	⬛+	⬛+	⬛+	⬛+	⬜-	⬜-	⬜-	⬜-	⬛+	⬜-
genome sequence	Arabinoxylan	⬛+	🟦+	🟩+	⬜-	🟩+	🟩+	🟧-	🟦+	🟧-	🟦+	🟩+	🟦+	🟧-	🟦+	🟧-	🟦+	🟦+	🟦+	🟧-
genome sequence	Levan or Inulin	🟦+	🟦+	🟦+	🟦+	🟧-	🟦+	🟦+	🟦+	🟩+	🟩+	🟦+	🟦+	🟦+	⬜-	🟦+	🟦+	🟦+	🟦+	🟦+
genome sequence	Pectins (Pectin or polygalacturonic acid)	🟦+	🟦+	🟦+	🟦+	🟦+	🟦+	🟦+	🟦+	🟦+	🟦+	🟦+	🟦+	🟦+	🟦+	🟦+	🟦+	🟦+	🟦+	🟦+
genome sequence	Rhamnogalacturonan I	🟦+	🟦+	🟩+	🟦+	🟦+	⬜-	🟧-	🟧-	🟩+	🟩+	🟦+	🟩+	🟦+	🟩+	⬜-	🟩+	🟦+	🟩+	⬜-
genome sequence	Arabinogalactan	🟦+	⬛+	🟧-	⬜-	⬜-	🟩+	🟩+	🟦+	⬜-	🟧-	🟩+	🟧-	🟧-	🟧-	🟧-	🟦+	🟧-	🟧-	🟧-
genome sequence	Beta-glucan	⬜-	🟦+	⬛+	🟦+	⬛+	⬜-	⬜-	⬜-	🟦+	🟧-	🟦+	🟦+	🟦+	🟧-	🟩+	🟦+	🟩+	🟦+	🟦+
genome sequence	Starch, pullulan, amylopectin	🟦+	🟦+	🟦+	🟧-	🟧-	🟦+	🟦+	🟦+	🟦+	🟦+	🟦+	🟦+	🟦+	🟧-	🟦+	🟦+	🟦+	🟦+	🟧-
genome sequence	Glucomannan	⬜-	🟦+	⬛+	⬜-	🟦+	⬜-	⬜-	⬜-	🟦+	🟧-	🟦+	🟦+	🟦+	🟧-	🟦+	🟦+	🟦+	🟦+	🟧-
genome sequence	Chondroitin	🟦+	🟦+	⬜-	⬜-	⬜-	⬜-	⬜-	⬜-	⬜-	⬜-	⬜-	⬜-	⬜-	⬜-	⬜-	⬜-	⬜-	⬜-	⬜-
genome sequence	Heparin	🟦+	🟦+	⬜-	⬜-	⬜-	⬜-	⬜-	⬜-	⬜-	⬜-	⬜-	⬜-	⬜-	⬜-	⬜-	⬜-	⬜-	⬜-	⬜-
genome sequence	Mucin proteoglycan	🟦+	⬛+	⬜-	⬜-	⬜-	⬜-	⬜-	⬜-	⬜-	⬜-	⬜-	⬜-	N/A	N/A	⬜-	⬜-	🟧-	🟧-	🟧-

🟦+ predicted growth; observed growth
⬜- did not predict growth; did not observe growth
⬛+ did not predict growth; observed growth
🟧- predicted growth; did not observe growth
🟩+ did not predict growth; observed growth; genome encodes substrate enzyme genes

Figure S3. Growth of *P. copri* isolates on polysaccharides; related to Figure 3
A. OD600 was measured for indicated isolates and polysaccharides after ~24h, ~48h, ~72h anaerobic incubation at 37 degrees C. Data indicate mean + SD and represent 2-3 experiments. Each condition was tested in duplicate or triplicate. Data show the highest OD from 24, 48, or 72h for each condition for each experiment. NT: not tested. **B.** *P. copri* DSM 18205 was grown to mid-log phase in the presence of the indicated polysaccharide substrates. Levels of *susC* transcripts were quantified by qRT-PCR. Graphs show fold change compared to a glucose control and indicate mean + SD. **C.** Schematic of PULs with *susC* genes that were upregulated in the presence arabinan and xyloglucan. PATRIC annotations are indicated above the PUL. Genes with annotations determined by homology are annotated beneath the PUL. Genes with the same annotation are the same color. Genes annotated as hypothetical proteins are gray. Genes with annotations unlikely to be involved in carbohydrate metabolism are colored dark blue. **D.** Growth data were matched to growth predictions based on the presence or absence of pPULs in the genomes. Growth was considered positive if the OD was statistically significant, as determined by one-way ANOVA and post hoc Dunnett's test using H2O (no carbon source) as the control (**Methods**).

Table S3. Summary of *P. copri* isolate growth by *P. copri* complex clade; related to Figure 3

	clade			
Substrate	B	D	C	A
Corn xylan	+	+	+	+/-
Beechwood xylan	-	+	+/-	+/-
Arabinoxylan	-	+	+	+/-
Levan	+	-	-	+/-
Inulin	+	-	+	+/-
Pectin	+	+	+/-	+/-
Polygalacturonic acid	-	-	+/-	+/-
Rhamnogalacturonan I	+	+	-	+/-
Arabinan	-	+	+	+/-
Arabinogalactan	-	-	-	+/-
Xyloglucan	-	+	-	+/-
Beta-glucan	+	+	-	+/-
Starch	+	+	+	+/-
Pullulan	+	-	+	+/-
Amylopectin	-	-	-	+/-
Glucomannan	-	+	-	+/-
Chondroitin	-	-	-	-
Heparin	-	-	-	-
Mucin proteoglycan	-	-	-	-

Table S4. Oligonucleotides used in screens; related to Methods, Figures S1, 3D, and S3B

screen	primer ID	oligonucleotide sequence (5' --> 3')	source
P. copri screen (**Figures S1A** and **S1C**) and *susC* screen (**Figure 3D** and **Figure S3B**)	Universal 16S Primer 340F	ACTCCTACGGGAGGCAGCAGT	Scher et al., 2013
	Universal 16S Primer 515R	ATTACCGCGGCTGCTGGC	Scher et al., 2013
P. copri screen (**Figures S1A** and **S1C**)	primer_set_1_F	CCGGACTCCTGCCCCTGCAA	Scher et al., 2013
	primer_set_1_R	GTTGCGCCAGGCACTGCGAT	Scher et al., 2013
	primer_set_2_F	AATAGCTTTGGTGGCGCTTT	This study
	primer_set_2_R	AGGTTCCCTCACCGCATACA	This study
	primer_set_3_F	TTGGGTCGCCTGATGTTACC	This study
	primer_set_3_R	TCTTCAGCAGAAGCGTTGGT	This study
	primer_set_4_F	AGGACGTTGACCAGCAGAAG	This study
	primer_set_4_R	GCATAGCGGTTTTGGCTGAG	This study
	primer_set_5_F	GAGACGGGCTGGAGTGATTT	This study
	primer_set_5_R	TCCGTGTGGCTTTCATGGAG	This study
	primer_set_6_F	ACCGTAAGGAGCGTGTTCTG	This study
	primer_set_6_R	GTCGCGATTCTCTACTCCCG	This study
	primer_set_7_16S_F	TTTGGAGACAATGACGCCCT	This study
	primer_set_7_16S_R	ACTTAAGCCGACACCTCACG	This study
	primer_set_8_16S_F	CACRGTAAACGATGGATGCC	Scher et al., 2013
	primer_set_8_16S_R	GGTCGGGTTGCAGACC	Scher et al., 2013
susC screen (**Figure 3D** and **Figure S3B**)	susC_123_PUL1116 F_1	CAAGTAAAGCCCGCAAGGAC	This study
	susC_123_PUL1116 R_1	TTACCCTGAAGTGCTACCGC	This study
	susC_129_PUL1802 F_1	TGGCTCTGTCAAGCGTAAGG	This study
	susC_129_PUL1802 R_1	TGCCTTTCCCTGCATACCTG	This study
	susC_130_PUL762 F_1	TGCTCGATGAGGTGGTTGTC	This study
	susC_130_PUL762 R_1	TAACACCCGCAACCTTACCC	This study
	susC_131_PUL1281 F_1	CGTGGCATTGTTCGTCATCG	This study
	susC_131_PUL1281 R_1	TCAGAGCCTCACCATTCTGC	This study
	susC_132_PUL1287 F_1	TGCTGAAGACAACGAACCGA	This study
	susC_132_PUL1287 R_1	TTTCACAACGTGCTTCTGCG	This study
	susC_133_PUL484 F_1	TCGAGGGTGGCGACAATAAC	This study
	susC_133_PUL484 R_1	CCATTAGAGGCACGGCTACC	This study
	susC_135_PUL1745 F_1	TACGTGGGTGGTTGGAATGG	This study
	susC_135_PUL1745 R_1	TCTCGCTAGGCATCACGTTG	This study
	susC_102_PUL1218 F_2	GGTGCTACGCTCGTGATTTC	This study
	susC_102_PUL1218 R_2	TCACCACCCACATTGGCTAC	This study
	susC_112_PUL792 F_1	TGACATGCAGAGTAGCGGTC	This study

susC screen (**Figure 3D** and **Figure S3B**)	susC_112_PUL792 R_1	GATGACACCATTGGCAGCAC	This study
	susC_147_PUL2366 F_1	GTGCAGCCCGCTATATGGAC	This study
	susC_147_PUL2366 R_1	CCACTGGGTTTTGTGTAGCG	This study
	susC_156_PUL610 F_1	CTCAGAGCGTTGACCCTAGC	This study
	susC_156_PUL610 R_1	TAATGCCCACTGCCTTACCG	This study
	susC_1_PUL1911 F_1	GATAATGGAGCTGCCATCCG	This study
	susC_1_PUL1911 R_1	CAGGCACTACAACAGAAGCG	This study
	susC_21_PUL1668 F_1	AGCTCTTGACGAGGTCGTTG	This study
	susC_21_PUL1668 R_1	TGGGTGATGTTGACACCAGG	This study
	susC_41_PUL1672 F_1	AAGGAGCTTGCTGACATCCC	This study
	susC_41_PUL1672R_1	TGATGTTTACACCGGCAAGC	This study
	susC_49_PUL1214 F_1	ATTTACCGGTTCTGCAGCCG	This study
	susC_49_PUL1214 R_1	CCAAAGCGTCTGTTACGTTGG	This study
	susC_55_PUL621 F_1	TCTTGGCTACAAGGAGCGTG	This study
	susC_55_PUL621 R_1	ACAGAACCGGTAAGGTCTGC	This study
	susC_67_PUL1810 F_1	CTGGCGAAACCTTGATTGGC	This study
	susC_67_PUL1810 R_1	GCCATTAGCAGCTTTCACGG	This study
	susC_76_PUL296 F_1	TGTCCTCAATGCTCAGACGG	This study
	susC_76_PUL296 R_1	CTGCACCATGACACTTGCAC	This study
	susC_83_PUL1747 F_1	TCTTGAAGACGACAACGCCA	This study
	susC_83_PUL1747 R_1	CCTGTGGTCTGCGCAATAGA	This study
	susC_90_PUL499 F_1	GCGCTCTGTACTGTTTGCAC	This study
	susC_90_PUL499 R_1	AATGACTGTCTCACCGGCAG	This study
	susC_165_PUL2751 F_1	TCGTGAAAGGTACCGTGCTG	This study
	susC_165_PUL2751 R_1	CCTCTTGTGCCACCATACCG	This study
	susC_166_PUL485 F_1	GAACCGGTTATTGGCGCATC	This study
	susC_166_PUL485 R_1	CTTGGCATTTCCACTAGGAGC	This study