nature microbiology

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41564-020-0683-3

In the format provided by the authors and unedited.

Spatially distinct physiology of *Bacteroides fragilis* within the proximal colon of gnotobiotic mice

Gregory P. Donaldson[1,5]✉, Wen-Chi Chou[2,5], Abigail L. Manson[2], Peter Rogov[2], Thomas Abeel[2,3], James Bochicchio[2], Dawn Ciulla[2], Alexandre Melnikov[2], Peter B. Ernst[4], Hiutung Chu[4], Georgia Giannoukos[2], Ashlee M. Earl[2]✉ and Sarkis K. Mazmanian[1]✉

[1]Divison of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA. [2]Infectious Disease and Microbiome Program, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [3]Delft Bioinformatics Lab, Delft University of Technology, Delft, the Netherlands. [4]Department of Pathology, University of California, San Diego, CA, USA. [5]These authors contributed equally to this work: Gregory P. Donaldson, Wen-Chi Chou. ✉e-mail: gdonaldson@rockefeller.edu; aearl@broadinstitute.org; sarkis@caltech.edu

Supplementary Table 1 | Hybrid selection results in substantial enrichment for bacterial reads. RNA-Seq read counts and mapping statistics shown with and without hybrid selection. Average values of three replicates and standard deviations are indicated.

	Without hybrid selection			**With hybrid selection**		
Sample	**Lumen**	**Mucus**	**Tissue**	**Lumen**	**Mucus**	**Tissue**
Total RNAseq reads	30,037,711 ± 4,838,289	30,851,659 ± 380,774	30,284,943 ± 3,142,595	33,719,927 ± 5,268,780	30,509,921 ± 2,638,387	31,786,088 ± 5,102,116
% Reads aligned	70 ± 2.4	67.5 ± 2.1	67.2 ± 0.6	95.8 ± 0.7	89.5 ± 0.9	86.7 ± 0.7
% Reads mapped to *B. fragilis* genome	50.4 ± 7.3	0.6 ± 0.2	0.1 ± 0.0	84 ± 7.0	28.6 ± 6.8	15.4 ± 3.2
Average % of aligned reads mapping to the *B. fragilis* genome	72.2 ± 11.1	0.83 ± 0.3	0.23 ± 0.06	87.6 ± 6.7	32.0 ± 7.3	17.8 ± 3.6
% Reads mapped to mouse genome	19.5 ± 8.2	66.9 ± 2.1	67.1 ± 0.6	11.9 ± 6.3	60.9 ± 6.0	71.2 ± 2.5
Average % of aligned reads mapping to the mouse genome	27.8 ± 11.1	99.2 ± 0.3	99.8 ± 0.06	12.4 ± 6.7	68.0 ± 7.3	82.2 ± 3.6
Median of % read coverage over *B. fragilis* genes	100 ± 0.0	31.9 ± 8.5	8.3 ± 4.0	100 ± 0.0	95 ± 4.3	79.1 ± 4.3

Supplementary Table 2. Correlation (Pearson's r) values between biological replicates, as well between HS and non-HS samples for individual mice. For the lumen samples with the highest bacterial abundance, the HS samples correlate with the non-HS sample as well as biological replicates correlate with one another (0.99-1.00 correlation value). However, in the samples with lower bacterial abundance (tissue and mucus), the correlation between biological replicates stays high (0.98-1.00), whereas the correlation between HS and non-HS samples drops slightly (0.95-0.97), primarily due to detectable transcripts in HS samples which were undetected in non-HS samples.

Site	**Sample**	**Correlations between biological replicates**						**Correlations bt. HS & non-HS**
Lumen								
		non-HS			HS			
Lumen		mouse 1	mouse 2	mouse 3	mouse 1	mouse 2	mouse 3	
	mouse 1	1.00	1.00	0.99	1.00	0.99	0.99	1.00
	mouse 2	1.00	1.00	1.00	0.99	1.00	1.00	0.99
	mouse 3	0.99	1.00	1.00	0.99	1.00	1.00	0.99
Mucus								
		non-HS			HS			
Mucus		mouse 1	mouse 2	mouse 3	mouse 1	mouse 2	mouse 3	
	mouse 1	1.00	0.99	0.99	1.00	0.99	0.98	0.97
	mouse 2	0.99	1.00	1.00	0.99	1.00	1.00	0.96
	mouse 3	0.99	1.00	1.00	0.98	0.99	1.00	0.95
Tissue								
		non-HS			HS			
Tissue		mouse 1	mouse 2	mouse 3	mouse 1	mouse 2	mouse 3	
	mouse 1	1.00	0.99	0.99	1.00	0.99	0.99	0.97
	mouse 2	0.99	1.00	0.99	0.99	1.00	1.00	0.96
	mouse 3	0.99	0.99	1.00	0.99	1.00	1.00	0.95

Supplementary Table 3. (separate file). List of outlier genes numbered in Extended Data 4. Most of these genes are very short (median length 110 nucleotides), and most of them are tRNA and 5S rRNA genes. An "x" means that, for this gene, the difference between the HS and non-HS values was larger than three standard deviations. There are 8 instances where the same gene is observed as an outlier in more than one condition (lumen, mucus, and tissue). Column B indicates the numbering in Extended Data 4. Gene annotations, predicted Pfam protein domains, and gene lengths are also shown here.

Supplementary Table 4. (separate file) For the differential expression analysis of lumen and mucus in Fig. 1e, normalized read counts in all samples (with and without hybrid selection) for those genes identified as differentially expressed in one or both analyses (HS or non-HS).

Supplementary Table 5. (separate file) For the differential expression analysis of lumen and tissue in Fig. 1e, normalized read counts in all samples (with and without hybrid selection) for those genes identified as differentially expressed in one or both analyses (HS or non-HS).

Supplementary Table 6. (separate file). Full differential expression analysis for all genes, for each of 3 comparisons. A) Lumen vs. Mucus; B) Lumen vs. Tissue; C) Mucus vs. Tissue (n = 3 animals, adjusted p values calculated using edgeR).

Supplementary Table 7: Significantly enriched functional domains in the 68 genes differentially expressed in mucus compared to lumen (white background), 99 genes differentially expressed in tissue compared to lumen (light grey background) or the 130 genes representing the union of these two sets (dark grey background). We calculated Pearson correlation coefficients to determine if the expression of genes (in TPM) were comparable between two samples. We used the Wilcoxon signed-rank test to compare the expression of a gene family in two different samples (n = 3 animals, FDR < 0.05).

Comparison	Pfam functional domain	adjusted p-value (FDR)	# differentially expressed genes	# total genes in genome
Mucus v Lumen	Glycosyl transferases group 1	0.004	3	28
Mucus v Lumen	Bacterial DNA-binding protein	0.0059	2	16
Mucus v Lumen	GTB	0.0242	3	51
Tissue v Lumen	Elongation factor Tu domain 2	0.0028	2	7
Tissue v Lumen	50S ribosome-binding GTPase	0.0036	3	16
Tissue v Lumen	Elongation factor Tu GTP binding domain	0.0069	2	10
Tissue v Lumen	Glycosyl transferases group 1	0.0185	3	28
Tissue v Lumen	Type I phosphodiesterase / nucleotide pyrophosphatase	0.0324	2	19
Mucus/Tissue v Lumen	Elongation factor Tu domain 2	0.0045	2	7
Mucus/Tissue v Lumen	50S ribosome-binding GTPase	0.0066	3	16
Mucus/Tissue v Lumen	Sigma-70 region 2	0.0106	5	43
Mucus/Tissue v Lumen	Elongation factor Tu GTP binding domain	0.0106	2	10
Mucus/Tissue v Lumen	Sigma-70, region 4	0.0115	5	44
Mucus/Tissue v Lumen	ECF sigma factor	0.0128	2	11

Mucus/Tissue v Lumen	Glycosyl transferases group 1	0.0287	3	28
Mucus/Tissue v Lumen	Bacterial DNA-binding protein	0.0294	2	16
Mucus/Tissue v Lumen	Bacterial regulatory proteins, luxR family	0.0396	2	18
Mucus/Tissue v Lumen	Type I phosphodiesterase / nucleotide pyrophosphatase	0.0453	2	19

Supplementary Table 8. (separate file). Annotations for full *B. fragilis* NCTC9343 genome with original and modern locus IDs and combined functional annotations from several databases, including CAZy[1], KEGG[2], EC numbers, Pfam[3], TIGRFAMs[4], and SWISS-PROT[5].

Supplementary Table 9. Primers used for mutagenesis, cloning, and qPCR.

Primer	Purpose	Sequence
BF gyrB QF3	qPCR (endogenous)	GTGAATGAGGACGGCAGTTT
BF gyrB QR3	qPCR (endogenous)	CTCGATGGGGATGTTTTGTT
BF1252 QF1	qPCR	TAGACCCTGCATGTGGTTCG
BF1252 QR1	qPCR	TCACATCCACTCCTGATGCT
BF3379 QF1	qPCR	AGCTGAGTTCAACAAAGATGCA
BF3379 QR1	qPCR	ATGAAGCTTTACGGGCACGA
BF3086 QF1	qPCR	TATCCGGACGCAGCATCTTC
BF3086 QR1	qPCR	CCTTCCCGCTACCCGATAAC
BF3134 QF1	qPCR	CTTCTACCGCAACCTGCTGA
BF3134 QR1	qPCR	CGTACCGTTCAACAGGACGA
ahpC QF1	qPCR	CACTGTTCGTCTCCGTTCCA
ahpC QF2	qPCR	GGGTTCCTTCTGTCTTCGCA
ccfC QF	qPCR	GATGAACTGATAGCCCATTA
ccfC QR	qPCR	TAGCGATGACTAAAGGTGTT
PSA flip QF2	qPCR	TTGTATCCGCAAGGGAGAGA
PSA flip QR2	qPCR	CGCTCCATACTGCCCATATT
PSB flip QF1	qPCR	GCTTTTGGCTTAATGCTTGTTGG
PSB flip QR1	qPCR	GCCTAGAAGTACAATTAGCCCGA
PSC flip QF2	qPCR	TGTTTGGTGGCTGCTACTTG
PSC flip QR2	qPCR	AGGTGAAGTTTGAAGCCAAGG
PSG flip QF2	qPCR	CAAGTACACCTGTCAGTAGTTTGC
PSG flip QR2	qPCR	GCAACTTCCAATTCCTAACAAAAGA
3134 flank 1	KO BF3134	**CGCTCTAGAACTAGTGGATC**CGCGATTGGCTACTCAAAGC
3134 flank 1	KO BF3134	**GAGTGCCGTTACTTTTCCGC**ACCAGAAGGGCGGATCGATTT
3134 flank 2	KO BF3134	**AAATCGATCCGCCCTTCTGG**TGCGGAAAAGTAACGGCACTC
3134 flank 2	KO BF3134	**TTCCTGCAGCCCGGGGGATC**CTTGCTGTGAAAGGTGTGCC
3086 flank 1	KO BF3086	**CGCTCTAGAACTAGTGGATC**GCGACGAATTGGTATGTGCC
3086 flank 1	KO BF3086	**ATGACGGACGATCATACGGC**CCCGACGGAAATGCACAATC
3086 flank 2	KO BF3086	**GATTGTGCATTTCCGTCGGG**GCCGTATGATCGTCCGTCAT
3086 flank 2	KO BF3086	**TTCCTGCAGCCCGGGGGATC**ACTCGTCGATGCCCTGTTTT
pst amy 5	complement BF3134	AA**CTGCAG**CGACAGTATCATGGAGCGCT
pst amy 3	complement BF3134	AA**CTGCAG**CGGGTGAAATCCTTTGTGCC
pst sulf 5	complement BF3086	AA**CTGCAG**TGCGTTCTGGAAATCGTGGA
pst sulf 3	complement BF3086	AA**CTGCAG**TGGATTTCTGTTTGGAGTGGGA

Supplementary Table 10. (separate file) Summary of statistical analysis methods and results for comparisons in Fig. 1-4 and Extended Data 8-10. For each comparison the table shows the two groups, the type of statistical test, type of post-hoc correction, effect size and confidence interval thereof, n numbers for each group, degrees of freedom, and the exact, adjusted p-value. For violin and box plots, descriptive statistics are provided as well (minimum, maximum, and quartiles).

Supplementary References

1. Kaoutari, A. E., Armougom, F., Gordon, J. I., Raoult, D. & Henrissat, B. The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nat. Rev. Microbiol.* **11,** 497–504 (2013).
2. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* **45,** D353–D361 (2017).
3. Finn, R. D. *et al.* The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* **44,** D279–85 (2016).
4. Haft, D. H. *et al.* TIGRFAMs: a protein family resource for the functional identification of proteins. *Nucleic Acids Res* **29,** 41–43 (2001).
5. The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res* **45,** D158–D169 (2017).