PNAS

Supporting Information for

Human gut bacteria tailor extracellular vesicle cargo for the breakdown of diet- and host-derived glycans

Mariana G. Sartorio[1], Evan J. Pardue[1], Nichollas E. Scott[2] and Mario F. Feldman[1*].

[1] Department of Molecular Microbiology, Washington University School of Medicine, Saint Louis, MO, 63110, USA.
[2] Department of Microbiology and Immunology, The Peter Doherty Institute for Infection and Immunity, University of Melbourne, Parkville, Victoria, Australia.

Mario F. Feldman
Email: mariofeldman@wustl.edu

This PDF file includes:

Figures S1 to S11
Tables S1
Materials and Methods
Legends for Movie S1
Legends for Datasets S1 to S9
References

Other supporting materials for this manuscript include the following:

Movie S1
Datasets S1 to S9

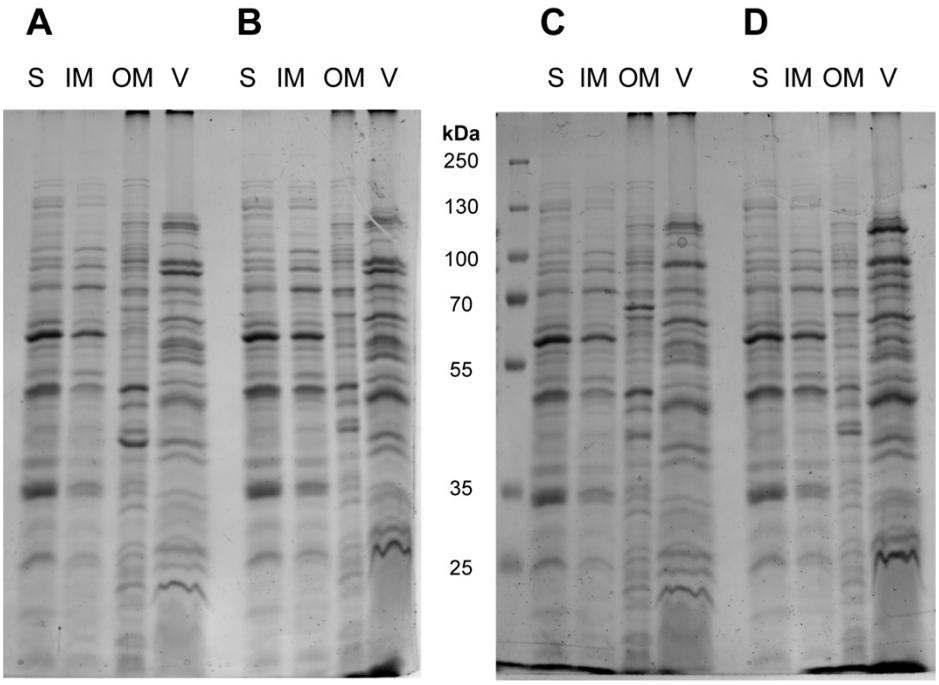

Fig. S1. Coomassie blue staining after SDS-PAGE of subcellular fractionations (10 μg) of *Bt* expressing A. Inulinase-GFP, B. BF_1581-GFP, C. OmpF-mCherry, and D. BT_2844-mCherry. References: S, soluble fraction; IM, inner membrane fraction; OM, outer membrane fraction; V, OMV fraction.

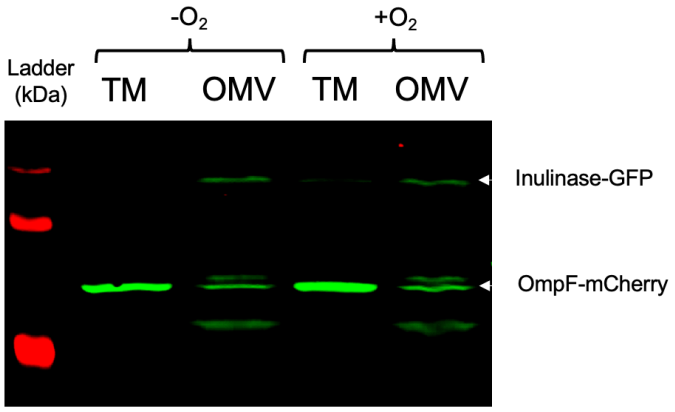

Fig. S2. Western blots after SDS-PAGE of total membranes (TM) and OMVs from *Bt* co-expressing Inulinase-GFP and OmpF-mCherry. Bacteria were swabbed from BHI agar plates into MM supplemented with 0.5% (w/v) of glucose and cultured for 20 hs in anaerobic chamber at 37 °C. Cultures were then maintained in anaerobiosis ($-O_2$) or exposed for 4 hours to atmospheric conditions ($+O_2$) prior to fractionation. Anti-His and anti-mCherry antibodies were employed to identify GFP and mCherry chimeric markers, respectively.

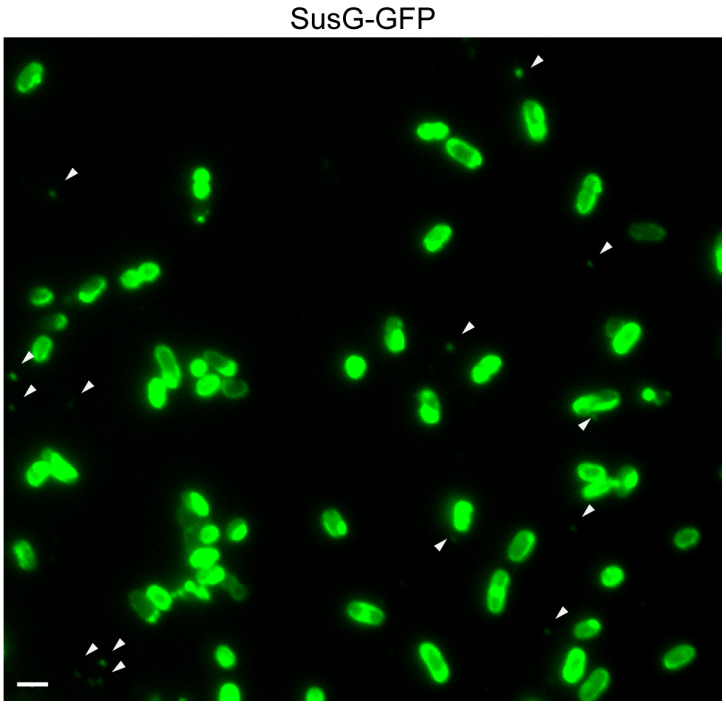

Fig. S3. Representative widefield fluorescence microscopy images of OMV chimeric marker SusG-GFP showing OMVs and localization at defined foci on the bacterial surface. Scale bar: 2 µm.

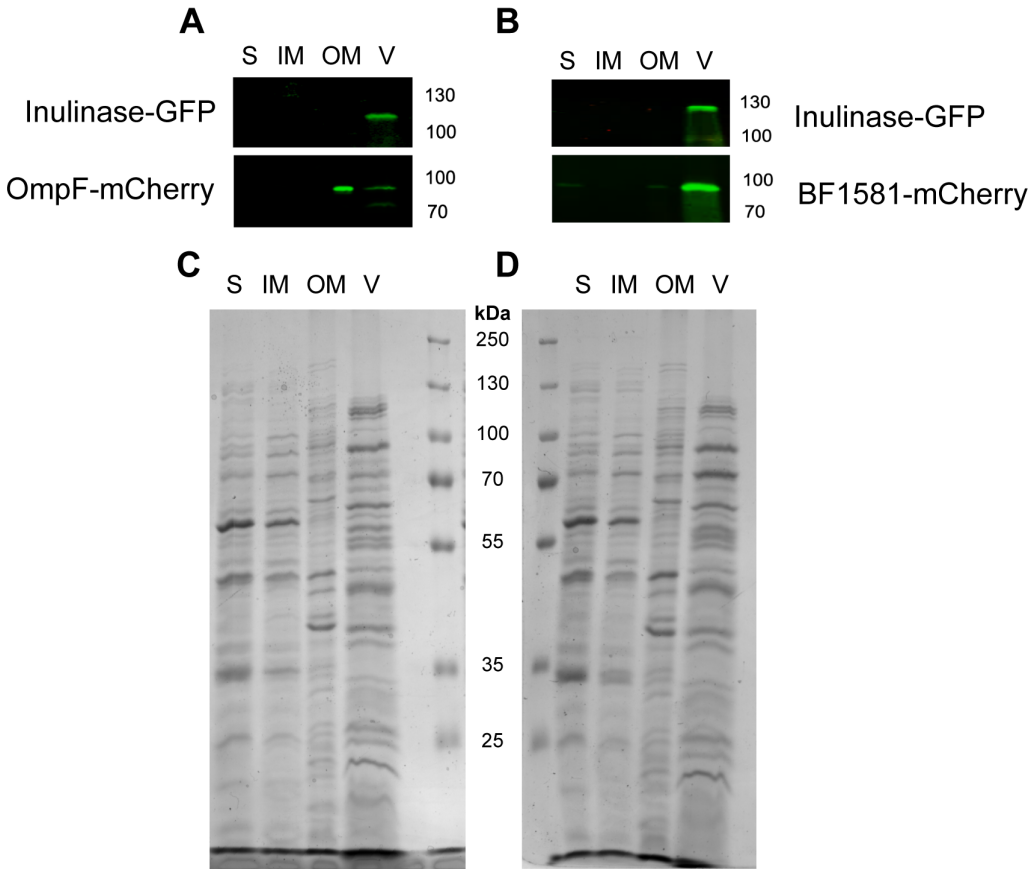

Fig. S4. Western blots and Coomasie blue staining after SDS-PAGE of subcellular fractions of *Bt* co-expressing Inulinase-GFP and OmpF-mCherry (A and C, respectively), or Inulinase-GFP and BF_1581-GFP (B and D, respectively). References: soluble fraction (S), inner membrane (IM), outer membrane (OM) and OMVs (V). Anti-His and anti-mCherry antibodies were employed to identify GFP and mCherry chimeric markers, respectively.

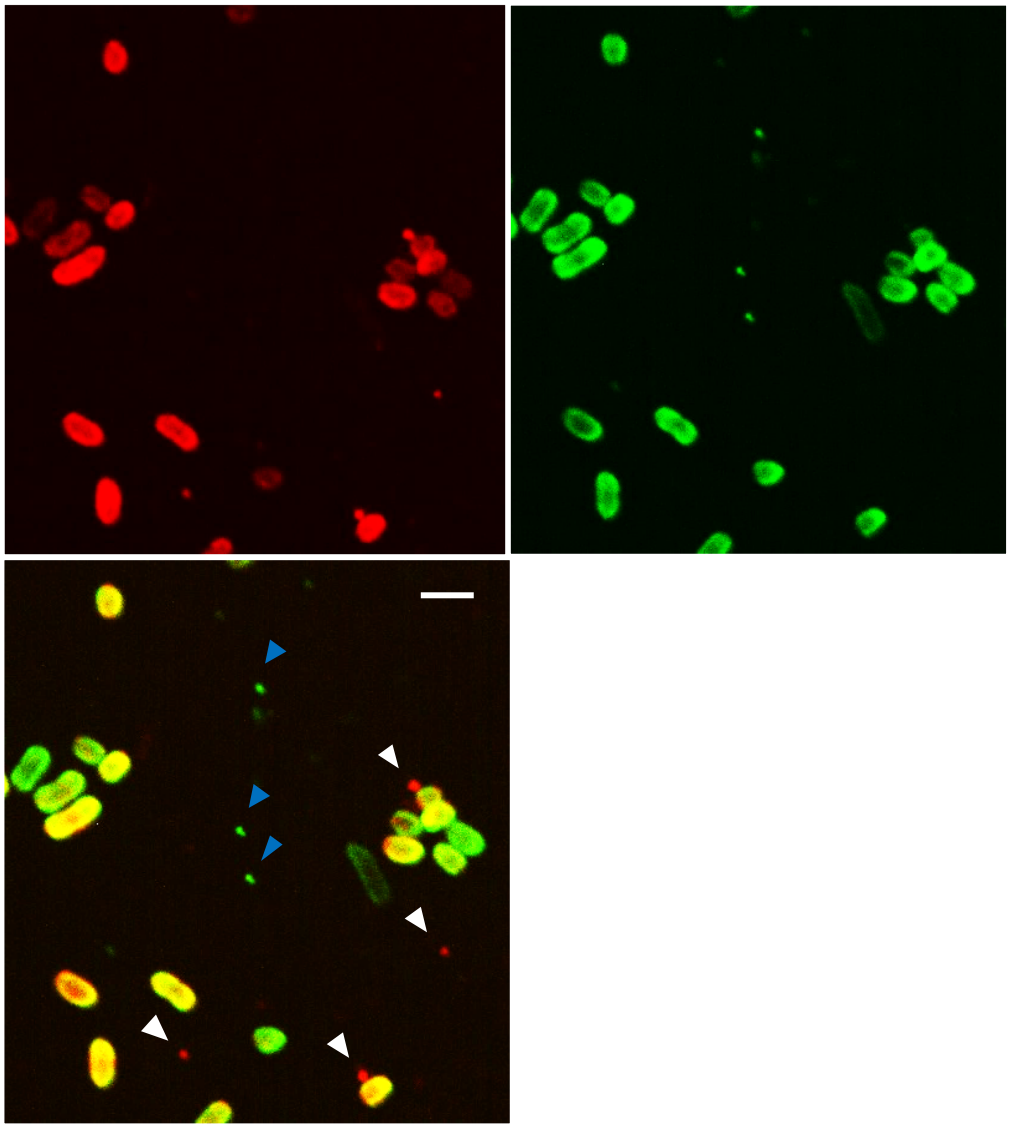

Fig. S5 Representative widefield fluorescence microscopy images of *Bt* co-expressing Inulinase-mCherry and OmpF-GFP. Late stationary cultures of 24 h growth in minimal media in the presence of glucose were analyzed. Cell debris (green, blue arrows) can be distinguished from bonafide OMVs (red, white arrows). Scale bar: 2μm.

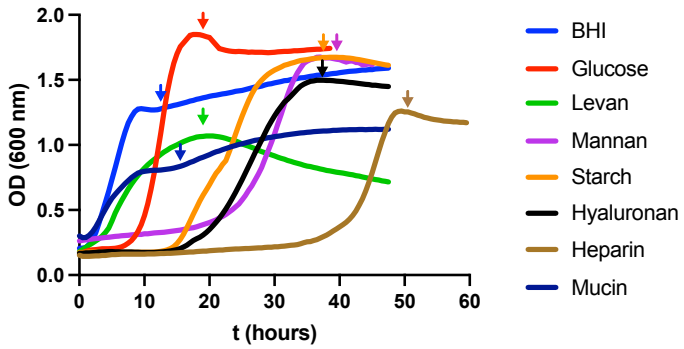

Fig. S6. Representative growth curves of *Bt* in BHI or minimal media with the indicated carbon sources. OMVs were harvested after reaching stationary phase for each condition (indicating with arrows).

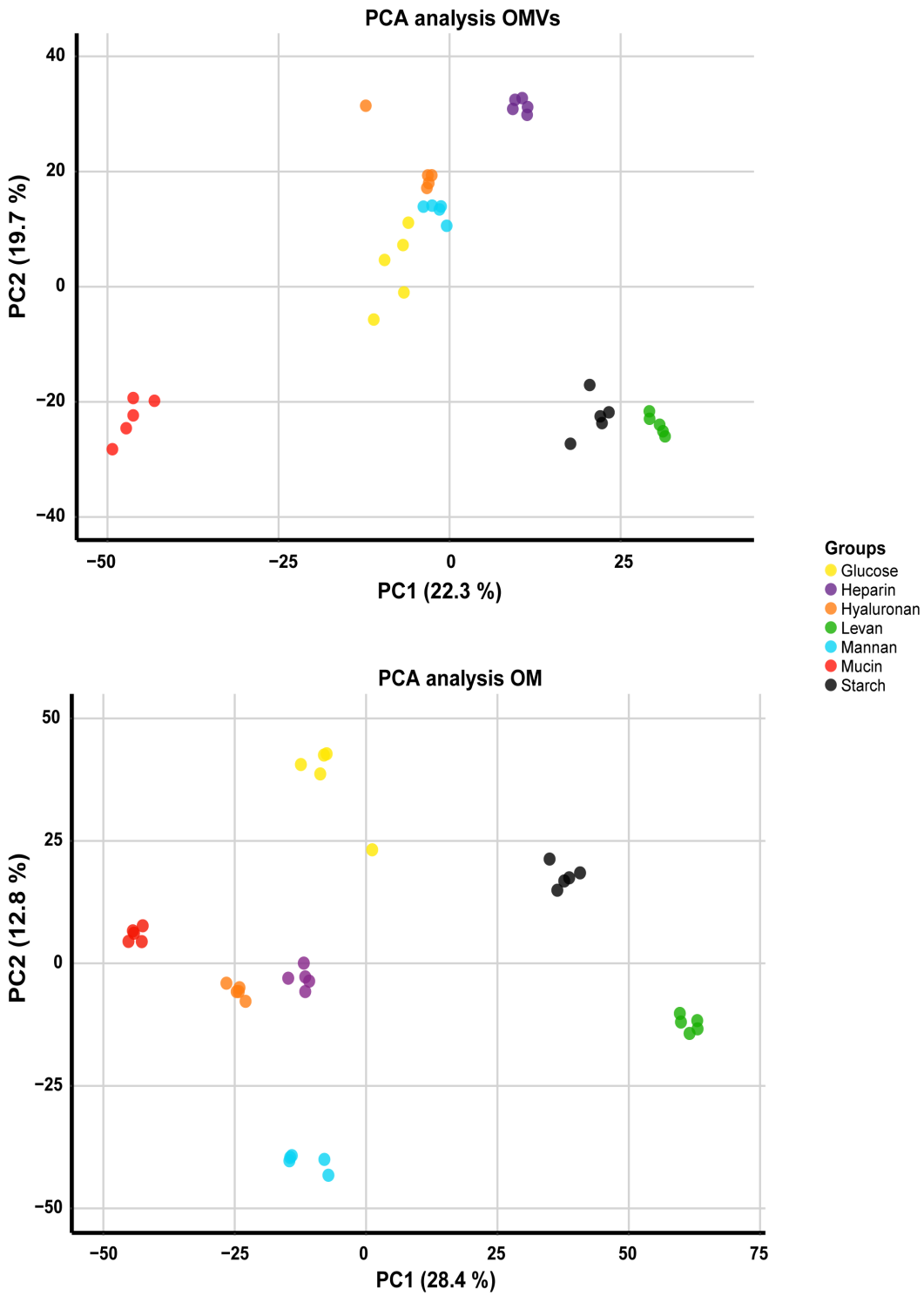

Fig. S7. Principal component analysis (PCA) of OMV (upper panel) and OM (lower panel) proteomes from *Bt* grown in minimal media supplemented with the indicated carbon sources. Five biological replicates were performed for each condition.

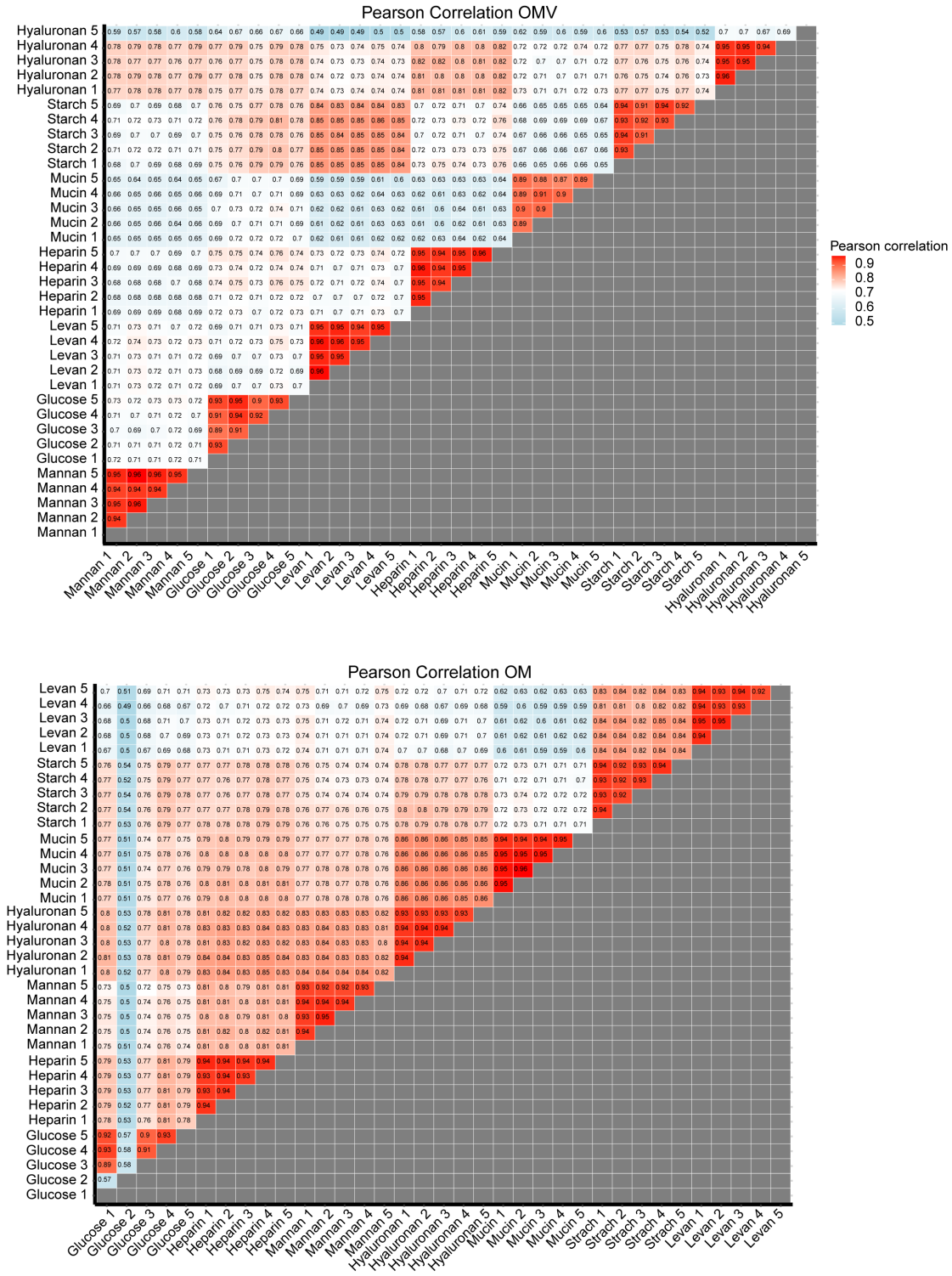

Fig. S8. Pearson Correlation between OMV (upper panel) and OM (lower panel) samples from *Bt* grown in minimal media supplemented with the indicated carbon sources. Five biological replicates were performed for each condition.

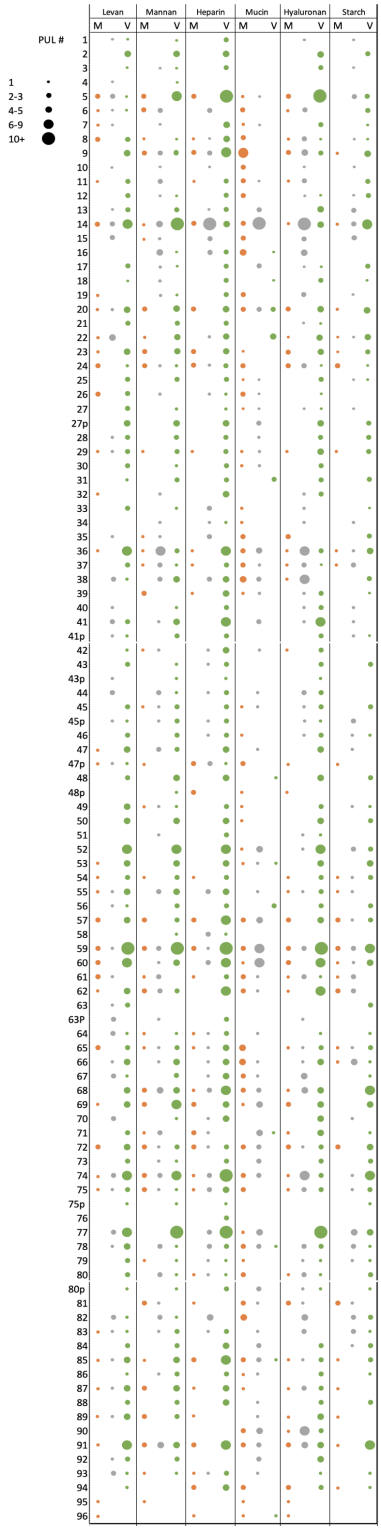

Fig. S9. PUL-encoded proteins were identified and classified as OM-enriched (OMV/OM fold change <-1, M column, colored in orange), OMV-enriched (OMV/OM fold change >1, V column, colored in green) or unclassified (OMV/OM fold change between -1 and 1, colored in gray). Circle size represents the number of identified proteins for each PUL. p means predicted PUL.

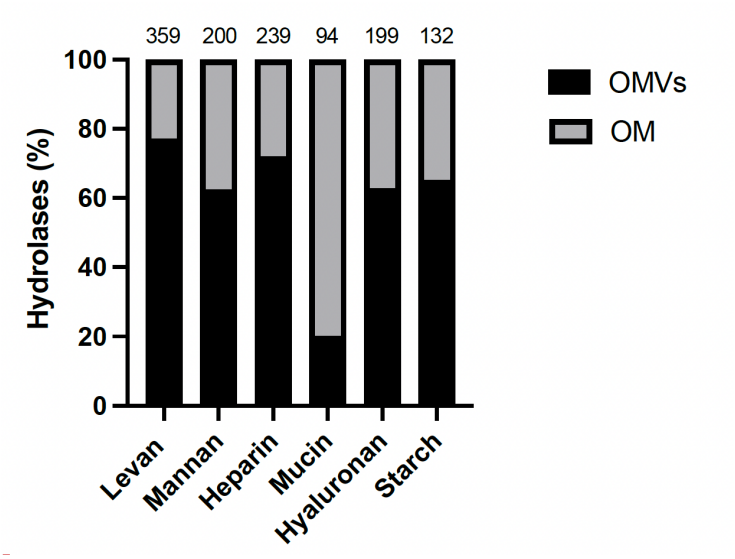

Fig. S10. Hydrolases partitioning between OM and OMV fractions. All predicted hydrolases identified for each condition were classified as OM-enriched (OMV/OM fold change <-1) or OMV-enriched (OMV/OM fold change >1). Numbers at the top of bars indicate the total number of hydrolases identified in each condition.

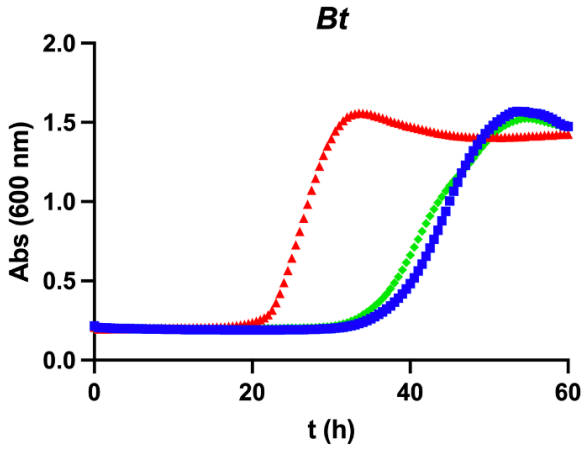

Fig. S11. Growth curves of *Bt* in minimal media with starch. Cultures were not supplemented (blue lines), supplemented with 1 µg/ml of *Bt* OMVs obtained after growth in the same glycan (red lines), or supplemented with 1 µg/ml of *Bt* OMVs obtained after growth in a different glycan (hyaluronan, green lines).

Table S1. Strains, plasmids and oligonucleotides used in this study.

Strains used in this study			
Name	Features	Reference/Source	
B. thetaiotaomicron VPI-5482	Wild-type strain. Erm^S	Jeffrey I. Gordon laboratory	
B. fragilis NCTC 9343/ATCC 25285	Wild-type strain. Erm^S	American Type Culture Collection	
B. ovatus ATCC 8483	Wild-type strain. Erm^S	Juliane Bubeck-Wardenburg laboratory	
P. vulgatus NCTC 11154	Wild-type strain. Erm^S	Juliane Bubeck-Wardenburg laboratory	
E. coli S17-1λpir	*thi pro hsdR hsdM+recA RP4-2-Tc::Mu-Km::Tn7 λpir,* Amp^S	Simon *et. al.*, 1983/Andrew Goodman laboratory	
Plasmids used in this study			
Name	Features	Reference/Source	
pWW3452	AmpR-ermG-RP4/R6K-[P_BfP1E6-RBSlp-LP-GFP- tag-Term]-NBU2, Amp^R Erm^R	Whitaker *et. al.*, 2017	
pWW3515	AmpR-ermG-RP4/R6K-[P_BfP1E6-RBSlp-LP-mCherry-Term]-NBU2, Amp^R Erm^R	Whitaker *et. al.*, 2017	
pWW3867	AmpR-ermG-RP4/R6K-[P_BT1311-RBSphage-GFP- tag-Term]-NBU2, Amp^R Erm^R	Whitaker *et. al.*, 2017	
pWWBoINL-GFP	pWW3452 harboring Inulinase-GFP fusion	This work	
pWWBf1581-GFP	pWW3452 harboring BF_1581-GFP fusion	This work	
pWWBf1581-mCherry	pWW3452 harboring BF_1581-mCherry fusion	This work	
pWWSusG-GFP	pWW3452 harboring SusG-GFP fusion	This work	
pWWBT0418-mCh	pWW3867 harboring BT_0418-mCherry fusion	This work	
pWWBT2844-mCh	pWW3867 harboring BT_2844-mCherry fusion	This work	
pWWBoINLGFP-0418mCh	pWWBoINL-GFP backbone harboring BT_0418-mCherry	This work	
pWWBoINLGFP-BF1581mCh	pWWBoINL-GFP backbone harboring BF_1581-mCherry	This work	
Oligonucleotides used in this study			
Cloned genes in pWW plasmid series by NEBuilder® HiFi DNA Assembly			
Name	Sequence	Template	Description
BoINL F	GGTAATATTAACAATAATTTATTTTCAATGAAGATAAATAAATTCTTAATAAGCGG	*B. ovatus* ATCC 8483 genomic DNA	For cloning into pWW3452
BoINL-GFP R	CTTCGCCTTTACGCATAGATCCTTTCTTAGCGCTTAGATAATGTAATATATTCTTCG	*B. ovatus* ATCC 8483 genomic DNA	For cloning into pWW3452
BF1581 F	TATTAACAATAATTTATTTTCAATGAAAAAAATAAATGCTTTAATTACTAAAATGTGCTT	*B. fragilis* NCTC 9343 genomic DNA	For cloning into pWW3452
BF1581-GFP R	CTCTTCGCCTTTACGCATAGATCCTTTAAAAACTATGCTTGAAGGAAACCAATATTTACC	*B. fragilis* NCTC 9343 genomic DNA	For cloning into pWW3452
BF1581-mCh R	CTTCTTCCCCTTTCGAAACCATAGATCCTTTAAAAACTATGCTTGAAGGAAAC	*B. fragilis* NCTC 9343 genomic DNA	For cloning into pWW3515

SusG F	GGTAATATTAACAATAATTTATTTTCAATGAATAAACATCTCCACTTTTTATC	*B. theta* VPI-5482 genomic DNA	For cloning into pWW3452
SusG-GFP R	CTTCGCCTTTACGCATAGATCCGTTGCCCAACTTGAATACTACAGAGG	*B. theta* VPI-5482 genomic DNA	For cloning into pWW3452
BT0418 F	TCCAAATCTGTTTTTAAAGAATGAAAAAGGGTTTATTGTTTATTTTAATG	*B. theta* VPI-5482 genomic DNA	For cloning into pWW3867
BT0418-mCh R	AGCTCTTCGCCTTTACGCATAGATCCTTCTTCCACGATAGCAAC	*B. theta* VPI-5482 genomic DNA	For cloning into pWW3867
BT2844 F	TCCAAATCTGTTTTTAAAGAATGACAAAGAAGTTGTACTTGC	*B. theta* VPI-5482 genomic DNA	For cloning into pWW3867
BT2844-mCh R	AGCTCTTCGCCTTTACGCATAGATCCTTTGATGATGCTCATGAAGTC	*B. theta* VPI-5482 genomic DNA	For cloning into pWW3867
Tandem-linF	ACAATCAGCCTTACTTGTGCCTG	pWWBoINL-GFP	For tandem cloning
Tandem-linR	TGCAGCCAATGCACAAATGC	pWWBoINL-GFP	For tandem cloning
Tandem cloning prom F	GCATTTGTGCATTGGCTGCAAATTCGTTCCGTCTCGATTCAGATC	pWWBf1581-mCherry/pWWBT0418-mCh	For tandem cloning into pWWBoINL-GFP
Tandem cloning R	CAGGCACAAGTAAGGCTGATTGTTGCAGCCAATGCACAAATGC	pWWBf1581-mCherry/pWWBT0418-mCh	For tandem cloning into pWWBoINL-GFP
pWW3452-linF	ATGCGTAAAGGCGAAGAGCTG	pWW3452 plasmid	For cloning
pWW3515-linF	ATGGTTTCGAAAGGGGAAGAAGATAACATGG	pWW3515 plasmid	For cloning
pWW3452/3515-linR	TGAAAATAAATTATTGTTAATATTACC	pWW3452 plasmid	For cloning

Materials and Methods

Growth assays. Bacterial strains were grown overnight in BHI, washed with PBS, and normalized by OD_{600} of 1. MM with different carbon sources were inoculated to a final OD_{600} of 0.05 and purified OMVs (1 μg/ml final protein concentration) were added when indicated. Growth curves were performed in sterile, round-bottom, polystyrene, 96-well plates in anaerobic and static conditions at 37 °C. OD_{600} values were recorded every 30 min with 10 s shaking before measurements with SmartReader MR9600-T microplate reader. All experiments were performed on 3 independent days with at least 3 wells per strain per condition.

Creation of protein markers fused to GFP and mCherry. OMVs (BT_3968, Bacova_04502, and BF_1581) and OM (BT_0418 and BT_2844) enriched proteins selected as markers were fused to sfGFP and/or mCherry. For this, the genes codifying for the proteins of interest were amplified by PCR using primers listed in table S1. The purified products were cloned into pWW plasmid series upstream of sfGFP/mCherry coding sequences[1] using NEBuilder® HiFi DNA Assembly (NEB). To perform co-expression experiments, OMV and OM fluorescent chimeric markers were cloned into the same integrative plasmid. Constructs were conjugated into *B. thetaiotaomicron* using previously transformed *Escherichia coli* S17-1 λ*pir* as a donor. Strains were plated and *Bacteroides* strains harboring the integrated plasmid were selected in BHI agar plates supplemented with 200 μg/ml gentamicin and 25 μg/ml erythromycin.

Vesicles preparations. OMVs were purified by ultracentrifugation of filtered spent media. Briefly, 50 ml of of *B. thetaiotaomicron* cultures from early stationary phase were centrifuged at 6,500 rpm at 4 °C for 10 minutes. Supernatants were filtered using a 0.22-μm-pore membrane (Millipore) to remove any residual cells. The filtrate was subjected to ultracentrifugation at 200,000 x*g* for 2 h (Optima L-100 XP ultracentrifuge; Beckman Coulter). Supernatants were discarded, and the pellet resuspended in PBS. Purified OMV preparations were lyophilized for MS analysis.

Subcellular fractionation. Total membrane preparations were performed by cell lysis and ultracentrifugation. Cultures from early stationary phase were harvested by centrifugation at 6,500 rpm at 4°C for 10 minutes. The pellets were gently resuspended in a mixture of 50 mM Tris-HCl (pH 8.0), 150 mM NaCl, and 50 mM $MgCl_2$ containing complete EDTA-free protease inhibitor mixture (Roche Applied Science) followed by cell disruption. Centrifugation at 6,500 rpm at 4 °C for 5 minutes was performed to remove unbroken cells. Total membranes were collected by ultracentrifugation at 200,000 x*g* for 1 h at 4 °C, and the soluble fraction was collected from the supernatant. OM and IM were separated by differential extraction with the same buffer supplemented with 1% (v/v) *N*-lauroyl sarcosine and incubated 1 h at room temperature with gentle agitation. The OM fractions were recovered by centrifugation at 200,000 x*g* for 1 h at 4 °C in the pellet fraction, whereas the IM fraction was obtained from the supernatant. OM fractions were lyophilized for MS analysis.

SDS-page and Western blot analyses. Membrane and vesicles fractions were analyzed by standard 10% Tris-glycine SDS-PAGE. For protein marker distribution analyses between different fractions, protein content was quantified using a DC protein assay kit (Bio-Rad), and 10 μg of each fraction was loaded onto an SDS-PAGE gel and stained with Coomasie blue for total protein analysis, or transferred onto a nitrocellulose membrane for

Western blot analysis. Membranes were blocked using Tris-buffered saline (TBS)-based Odyssey blocking solution (LI-COR). Primary antibodies used in this study were rabbit polyclonal anti-His (ThermoFisher), and rabbit polyclonal anti-mCherry (ThermoFisher). Secondary antibodies used were IRDye anti-rabbit 780 (LI-COR). Imaging was performed using an Odyssey CLx scanner (LI-COR).
For SDS-PAGE comparative vesicle production analyses, samples were normalized by OD_{600}. Samples were resuspended in 50 mM Tris-HCl (pH 8.0), 150 mM NaCl, and 50 mM $MgCl_2$ in a final volume defined by the following formulas: for OM fraction, V_{OM} (µl) = 20 µl × OD_{600} × V_c; for vesicle fractions, V_V (µl) = 4 µl × OD_{600} × V_c, where V_c is the starting volume of culture sample (in ml).

Widefield fluorescence microscopy. Bacteria were swabbed from BHI agar plates into MM supplemented with 0.5% (w/v) of glucose and cultured for 20 hs in anaerobic chamber at 37 °C. For GFP and mCherry fluorophores maturation[2–4], cultures were removed from the anaerobic chamber and incubated at 37 °C for 4 h in aerobic conditions. Fifty microliters of bacteria were diluted into 200 µl PBS and 3 µl of the dilution was dotted onto 1% agarose pads in PBS. Excess of liquid was air-dried and 18mm coverslips were added. Images were acquired using a Zeiss Axio Imager M2 upright fluorescence microscope equipped with Plan apo 100x/1.40 N.A. Oil Ph3 M27 objective lens. Images were adjusted and cropped using Fiji[5].
For time lapse fluorescence microscopy, cell culture and fluorophore maturation were performed as previously described. Fifty microliters of bacteria were diluted into 200 µl of pre-warmed MM supplemented with glucose and 3 µl of the dilution was dotted onto 35 mm glass bottom culture dishes (MatTek). Drop was covered with prewarmed 1% agarose pads in MM supplemented with glucose. Images were acquired using a Zeiss Cell Observer Z1 inverted microscope equipped with a temperature-controlled incubation chamber at 37°C. Fluorescence images were acquired every 3.5 min with illumination from a Colibri 7 LED light source (Zeiss) and ORCA-ER digital camera (Hammamatsu Photonics, Japan). A Plan-Apochromat 63x/1.4 N.A. Phase 3 objective and ZEN blue 2.5 software were used for image acquisition.

Protein sample preparation for Mass Spectrometry analyses
Lyophilized protein preparations were solubilized in 100 µl of 5% SDS by boiling them for 10 min at 95 °C. The protein content was assessed by the bicinchoninic acid (BCA) protein assay (Thermo Fisher Scientific) according to manufacturer's instructions. One hundred microgram of each sample were reduced with 10mM DTT for 10 mins at 95°C and alkylated with and alkylated with 40mM IAA in the dark for 1 hour. Reduced and alkylated samples were cleaned up using Micro S-traps (https://protifi.com/pages/s-trap) according to the manufacturer's instructions. Samples were digested overnight with 3 µg of trypsin/Lys-C (1:33 protease/protein ratio) and then collected. Samples were dried down and further cleaned up using C18 Stage tips[6,7] to ensure the removal of any particulate matter.

LC-MS. Prepared purified peptides from each sample were re-suspended in Buffer A* (2% acetonitrile, 0.01% trifluoroacetic acid) and separated using a two-column chromatography setup composed of a PepMap100 C_{18} 20-mm by 75-µm trap and a PepMap C_{18} 500-mm by 75-µm analytical column (Thermo Fisher Scientific). Samples were concentrated onto the trap column at 5 µl/min for 5 min with Buffer A (0.1% formic acid, 2% DMSO) and then infused into an Orbitrap Q-Exactive plus Mass Spectrometer (Thermo Fisher Scientific) or a Orbitrap Fusion Lumos equipped with a FAIMS Pro

interface at 300 nl/minute via the analytical column using a Dionex Ultimate 3000 UPLC (Thermo Fisher Scientific). Ninety-five-minute analytical runs were undertaken by altering the buffer composition from 2% Buffer B (0.1% formic acid, 77.9% acetonitrile, 2% DMSO) to 22% B over 65 min, then from 22% B to 40% B over 10 min, then from 40% B to 80% B over 5 min. The composition was held at 80% B for 5 min, and then dropped to 2% B over 2 min before being held at 2% B for another 8 min. The Q-Exactive plus Mass Spectrometer was operated in a data-dependent mode automatically switching between the acquisition of a single Orbitrap MS scan (375-1400 m/z, maximal injection time of 50 ms, an Automated Gain Control (AGC) set to a maximum of 3×10^6 ions and a resolution of 70k) and up to 15 Orbitrap MS/MS HCD scans of precursors (Stepped NCE of 26%, 28% and 32%, a maximal injection time of 110 ms, an AGC set to a maximum of 2×10^5 ions and a resolution of 35k). The Fusion Lumos Mass Spectrometer was operated in a stepped FAIMS data-dependent mode at two different FAIMS CVs -45 and -65. For each FAIMS CV a single Orbitrap MS scan (300-1600 m/z, maximal injection time of 50 ms, an AGC of maximum of 4×10^5 ions and a resolution of 60k) was acquired every 1.5 seconds followed by Orbitrap MS/MS HCD scans of precursors (NCE 35%, maximal injection time of 100 ms, an AGC set to a maximum of 1.25×10^5 ions and a resolution of 30k).

MS data analysis. Identification and LFQ analysis were accomplished using Max-Quant (v2.0.2.0) [8] using *Bacteroides thetaiotaomicron* VPI-5482 proteome (Uniprot: UP000001414) allowing for oxidation on Methionine. Prior to MaxQuant analysis dataset acquired on the Fusion Lumos were separated into individual FAIMS fractions using the FAIMS MzXML Generator[9]. The LFQ and "Match Between Run" options were enabled to allow comparison between samples. The resulting data files were processed using Perseus (v1.4.0.6)[10] to filter proteins not observed in at least four biological replicates of a single group. ANOVA and Pearson correlation analyses were performed to compare groups. Predicted localization and topology analysis for proteins identified by MS were performed using UniProt[11], PSORT[12], SignalP[13] and PULDB[14].

Movie S1 (separate file). Live OMV formation in *Bt* co-expressing Inulinase-GFP and OmpF-mCherry**.**

Dataset S1 (separate file). Comparative proteomic analysis between OMVs and OM fractions from *Bt* grown in minimal media supplemented with levan.

Dataset S2 (separate file). Comparative proteomic analysis between OMVs and OM fractions from *Bt* grown in minimal media supplemented with mannan.

Dataset S3 (separate file). Comparative proteomic analysis between OMVs and OM fractions from *Bt* grown in minimal media supplemented with heparin.

Dataset S4 (separate file). Comparative proteomic analysis between OMVs and OM fractions from *Bt* grown in minimal media supplemented with mucin.

Dataset S5 (separate file). Comparative proteomic analysis between OMVs and OM fractions from *Bt* grown in minimal media supplemented with hyaluronan.

Dataset S6 (separate file). Comparative proteomic analysis between OMVs and OM fractions from *Bt* grown in minimal media supplemented with starch.

Dataset S7 (separate file). Comparative proteomic analysis between OMVs and OM fractions from *Bt* grown in minimal media supplemented with glucose.

Dataset S8 (separate file). Comparative proteomic analysis between OMVs from all growth conditions showing most enriched proteins in each carbon source.

Dataset S9 (separate file). Comparative proteomic analysis between OMs from all growth conditions showing most enriched proteins in each carbon source.

References

1. Whitaker, W. R., Shepherd, E. S. & Sonnenburg, J. L. Tunable Expression Tools Enable Single-Cell Strain Distinction in the Gut Microbiome. *Cell* **169**, 538-546.e12 (2017).
2. Hebisch, E., Knebel, J., Landsberg, J., Frey, E. & Leisner, M. High Variation of Fluorescence Protein Maturation Times in Closely Related Escherichia coli Strains. *PLoS One* **8**, e75991-(2013).
3. Balleza, E., Kim, J. M. & Cluzel, P. Systematic characterization of maturation time of fluorescent proteins in living cells. *Nat Methods* **15**, 47–51 (2018).
4. Ransom, E. M., Ellermeier, C. D. & Weiss, D. S. Use of mCherry Red Fluorescent Protein for Studies of Protein Localization and Gene Expression in Clostridium difficile . *Appl Environ Microbiol* **81**, 1652–1660 (2015).
5. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676–682 (2012).
6. Rappsilber, J., Ishihama, Y. & Mann, M. Stop and Go Extraction Tips for Matrix-Assisted Laser Desorption/Ionization, Nanoelectrospray, and LC/MS Sample Pretreatment in Proteomics. *Anal Chem* **75**, 663–670 (2003).
7. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc* **2**, 1896–1906 (2007).
8. Prianichnikov, N. *et al.* MaxQuant Software for Ion Mobility Enhanced Shotgun Proteomics*. *Molecular & Cellular Proteomics* **19**, 1058–1069 (2020).
9. Hebert, A. S. *et al.* Comprehensive Single-Shot Proteomics with FAIMS on a Hybrid Orbitrap Mass Spectrometer. *Anal Chem* **90**, 9529–9537 (2018).
10. Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods* **13**, 731–740 (2016).
11. Consortium, T. U. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res* **49**, D480–D489 (2021).
12. Yu, N. Y. *et al.* PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* **26**, 1608–1615 (2010).
13. Almagro Armenteros, J. J. *et al.* SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol* **37**, 420–423 (2019).
14. Terrapon, N. *et al.* PULDB: the expanded database of Polysaccharide Utilization Loci. *Nucleic Acids Res* **46**, D677–D683 (2018).