nature chemical biology

Article

https://doi.org/10.1038/s41589-023-01357-8

Chemoproteomic identification of a DPP4 homolog in *Bacteroides thetaiotaomicron*

Received: 1 August 2022

Accepted: 8 May 2023

Published online: 22 June 2023

Check for updates

Laura J. Keller[1], Taylor H. Nguyen[2], Lawrence J. Liu[3], Brianna M. Hurysz[3], Markus Lakemeyer[4,11], Matteo Guerra[4,12], Danielle J. Gelsinger[3], Rachael Chanin[5,6], Nhi Ngo[7], Kenneth M. Lum[7], Franco Faucher[8], Phillip Ipock[4], Micah J. Niphakis[7], Ami S. Bhatt[5,6], Anthony J. O'Donoghue[3], Kerwyn Casey Huang[2,9,10] & Matthew Bogyo[4,9]

Serine hydrolases have important roles in signaling and human metabolism, yet little is known about their functions in gut commensal bacteria. Using bioinformatics and chemoproteomics, we identify serine hydrolases in the gut commensal *Bacteroides thetaiotaomicron* that are specific to the Bacteroidetes phylum. Two are predicted homologs of the human dipeptidyl peptidase 4 (hDPP4), a key enzyme that regulates insulin signaling. Our functional studies reveal that BT4193 is a true homolog of hDPP4 that can be inhibited by FDA-approved type 2 diabetes medications targeting hDPP4, while the other is a misannotated proline-specific triaminopeptidase. We demonstrate that BT4193 is important for envelope integrity and that loss of BT4193 reduces *B. thetaiotaomicron* fitness during in vitro growth within a diverse community. However, neither function is dependent on BT4193 proteolytic activity, suggesting a scaffolding or signaling function for this bacterial protease.

The gut microbiota is a diverse microbial community whose dysbiosis has been implicated in a wide variety of conditions. The vast majority of studies aimed at understanding gut microbiota dynamics have focused on global analyses of genes, proteins and/or metabolites using omics methods. While these methods have provided insight into the diversity of gut microbial communities and their metabolic outputs, they provide little information about the functional roles of specific enzymes and mechanisms that regulate communities inside hosts. A better understanding of such regulatory mechanisms is a prerequisite for new avenues to develop drugs that therapeutically modulate the gut microbiota. In addition, it is important to identify enzymes with homologous functions to human proteins, as they represent possible detrimental off-targets that need to be evaluated and considered when developing new therapeutic drugs. While existing approved drugs can have a widespread impact on gut commensal bacteria[1–4], the ability to predict these interactions and prioritize functionally relevant and druggable enzymes produced by commensal microbes in a high-throughput manner remains limited.

Activity-based protein profiling (ABPP) uses chemical probes to covalently modify enzymes in their endogenous environments. Integration with mass spectrometry (MS)-based proteomics allows class-wide identification of enzyme targets and measurement of enzyme activity across physiologically relevant conditions[5]. As a result, functionally relevant enzymes that are likely to be amenable to modulation

[1]Department of Chemical and Systems Biology, Stanford University School of Medicine, Stanford, CA, USA. [2]Department of Bioengineering, Stanford University, Stanford, CA, USA. [3]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, CA, USA. [4]Department of Pathology, Stanford University School of Medicine, Stanford, CA, USA. [5]Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA. [6]Divisions of Hematology and Blood and Marrow Transplantation, Department of Medicine, Stanford University, Stanford, CA, USA. [7]Lundbeck La Jolla Research Center, Inc., San Diego, CA, USA. [8]Department of Chemistry, Stanford University, Stanford, CA, USA. [9]Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA, USA. [10]Chan Zuckerberg Biohub, San Francisco, CA, USA. [11]Present address: Institute of Organic Chemistry and Macromolecular Chemistry, Friedrich-Schiller-University, Jena, Germany. [12]Present address: Department of Biochemical and Cellular Pharmacology, Genentech, San Francisco, CA, USA. e-mail: mbogyo@stanford.edu

by small-molecule drugs can be prioritized for characterization. In applying ABPP to study the gut microbiota, innovations such as the development of new probes[6,7], integration of fluorescence-activated cell sorting[8] and advances in metaproteomics analysis[9] have improved efforts to identify functionally relevant enzymes in native microbiomes such as stool samples. However, the conservation of enzymes across closely related bacteria and limitations in metaproteomics techniques[10] present obstacles to the definitive determination of the function of a specific enzyme using these top-down approaches.

Serine hydrolases are a broad class of proteases, lipases and esterases that have a common catalytic serine residue. Studied extensively in mammalian systems, they have roles in proteolysis, metabolism and cell signaling[11], all of which likely affect the fitness of gut commensals in complex communities. Human serine hydrolases have been targeted with several Food and Drug Administration (FDA)-approved therapeutics to treat diseases such as type 2 diabetes[12], and recent efforts have identified serine hydrolases as new drug targets for bacterial pathogens such as *Mycobacterium tuberculosis*[13–15]. However, little is known about the diversity of serine hydrolases in phylogenetically diverse gut commensal bacteria and their roles in interspecies communication and competition.

One of the most well-characterized serine hydrolases is the human dipeptidyl peptidase 4 (hDPP4), which trims dipeptides from the N-terminus of oligopeptides with a preference for P1 proline or alanine residues. Ubiquitously expressed as a membrane protein that can also be shed to circulate in the bloodstream, this diaminopeptidase cleaves a number of peptide hormones, including glucagon-like peptide 1 (GLP1) and peptide YY, and leads to either degradation of the substrate or alteration of its receptor selectivity[16]. hDPP4-triggered degradation of GLP1, which induces the secretion of insulin, contributes to the regulation of glucose homeostasis[16]. Hence, multiple inhibitors of hDPP4 activity have been approved by the FDA for the treatment of type 2 diabetes[17]. Furthermore, several noncatalytic functions have also been ascribed to hDPP4, mediated through the glycosylation-rich and cysteine-rich regions within the β-propeller domain, including cell adhesion via interactions with extracellular matrix proteins such as collagen and fibronectin and T-cell activation via interactions with adenosine deaminase and caveolin-1 (ref. 18).

The bacterium *Bacteroides thetaiotaomicron* is prevalent and abundant in human gut microbiotas and has served as a model commensal due to its extensive capabilities for digesting complex polysaccharides[19]. Early studies of fecal samples identified *Bacteroides* species as exhibiting high proteolytic activity[20], but the functions of these proteases have yet to be characterized. Recent work has implicated *Bacteroides*-derived proteases as key drivers of inflammatory bowel disease in patients with ulcerative colitis and in animal models of colitis[21]. In particular, dipeptidyl peptidases were identified to be abundant in *Bacteroides* species[21], but the functions of these enzymes remain unclear.

Here we use bioinformatics and a serine-reactive fluorophosphonate (FP) probe to profile serine hydrolases in *B. thetaiotaomicron* and identify six serine hydrolases that were annotated as dipeptidyl peptidases. Bioinformatic predictions indicate that these dipeptidyl peptidases are highly conserved within and restricted to the phylum Bacteroidetes. Two of these dipeptidyl peptidases, BT4193 and BT3254, have been annotated as putative homologs of hDPP4. We show biochemically that BT4193 is a bona fide functional homolog of hDPP4 that can be inhibited by hDPP4-targeted drugs, whereas BT3254 is a misannotated prolyl triaminopeptidase. We demonstrate that BT4193 has a noncatalytic role in the maintenance of envelope integrity and in the establishment of *B. thetaiotaomicron* within multispecies communities in vitro. These findings highlight the importance of the noncatalytic functions of BT4193 and the potential for targeting bacterial enzymes to precisely tune microbiota composition.

Results

Identifying serine hydrolases in gut commensal bacteria

Serine hydrolases have not been extensively studied in gut commensal bacteria, in part because they are difficult to identify bioinformatically due to the paucity of sequence homology and a large variety of protein folds[11]. Thus, we developed a pipeline to predict serine hydrolases by manually curating a list of protein family (Pfam) domains associated with serine hydrolases (Supplementary Fig. 1 and Supplementary Table 1), as defined by the MEROPS peptidase database, ESTHER database and known serine hydrolases identified in mammals[22], plants[23], bacteria[13,24–26] and archaea[27] via ABPP with serine-reactive FP-based probes. Using these serine hydrolase-associated Pfam domains, we bioinformatically predicted the presence of serine hydrolases in humans and 49 representative, phylogenetically diverse gut commensal bacterial strains (Fig. 1a, Supplementary Fig. 2 and Supplementary Tables 2 and 3). In clustering based on the number of serine hydrolases annotated with each Pfam domain, the human proteome appeared as an outgroup, having a distinct set of serine hydrolases driven primarily by the large number of trypsin-like (PF00089) serine hydrolases, and bacterial species clustered consistently by phylum. Furthermore, several types of serine hydrolases were predicted to be found exclusively in gut bacteria, highlighting the importance of characterizing serine hydrolases in these species.

Similar to the human genome[11], predicted serine hydrolases comprise 1–2.5% of each bacterial genome (Extended Data Fig. 1). Interestingly, bacteria from the Bacteroidetes phylum consistently are on the upper end of that range, in part driven by four serine hydrolase-associated Pfam domains that are abundant in the Bacteroidetes phylum (Supplementary Fig. 3a). In a representative species, *B. thetaiotaomicron*, proteins annotated with two of these Bacteroidetes-specific Pfam domains, sialic acid-specific acetylesterase (PF03629) and GDSL-like lipase (PF13472), are also predicted to be carbohydrate-active enzymes (Supplementary Fig. 3b). As one of the hallmarks of bacteria in the Bacteroidetes phylum is their ability to digest complex polysaccharides, it is not surprising that these bacteria have a unique enrichment for carbohydrate-active serine hydrolases. The other two domains in this Bacteroidetes-specific signature are comprised of serine hydrolases with a peptidase S9 (PF00326) or peptidase S41 (PF03572) domain, and this specificity suggests that these peptidases may have unique and vital roles in the Bacteroidetes phylum.

To determine whether the types of enzymes comprising the Bacteroidetes-specific peptidase signature are active and functionally relevant, we performed MS-based ABPP on *B. thetaiotaomicron* VPI-5482 by labeling intact bacteria with an FP-biotin probe. We identified 27 FP-reactive enzymes active during in vitro monoculture growth (Fig. 1b and Supplementary Table 4), representing ~1/3 of the total number of enzymes predicted by Pfam annotation. This discrepancy could be due to growth conditions impacting enzyme expression or activity, or the reactivity of the probe. Identified enzymes are predicted to be localized to every cellular compartment (Supplementary Table 4), so it is likely not due to the permeability of the FP-biotin probe. A protein BLAST (BLAST-P) search against gastrointestinal tract reference genomes from the Human Microbiome Project[28] confirmed that most *B. thetaiotaomicron* serine hydrolases identified by MS–ABPP are conserved and specific to the Bacteroidetes phylum (Fig. 1c and Supplementary Table 5). Six of the identified active serine hydrolases are in the peptidase S9 or peptidase S41 families, including BT3254 and BT4193, which are both annotated as putative homologs of hDPP4 and have the same domain architecture as hDPP4 (Supplementary Fig. 4). Transposon insertions in *BT3254* and *BT4193* caused fitness defects in monocolonized gnotobiotic mice in a pooled transposon-site sequencing assay[29], supporting the relevance of these enzymes for growth in vivo. However, the function of hDPP4-like enzymes in a commensal bacterium residing within a diverse microbiota is unclear.

https://doi.org/10.1038/s41589-023-01357-8

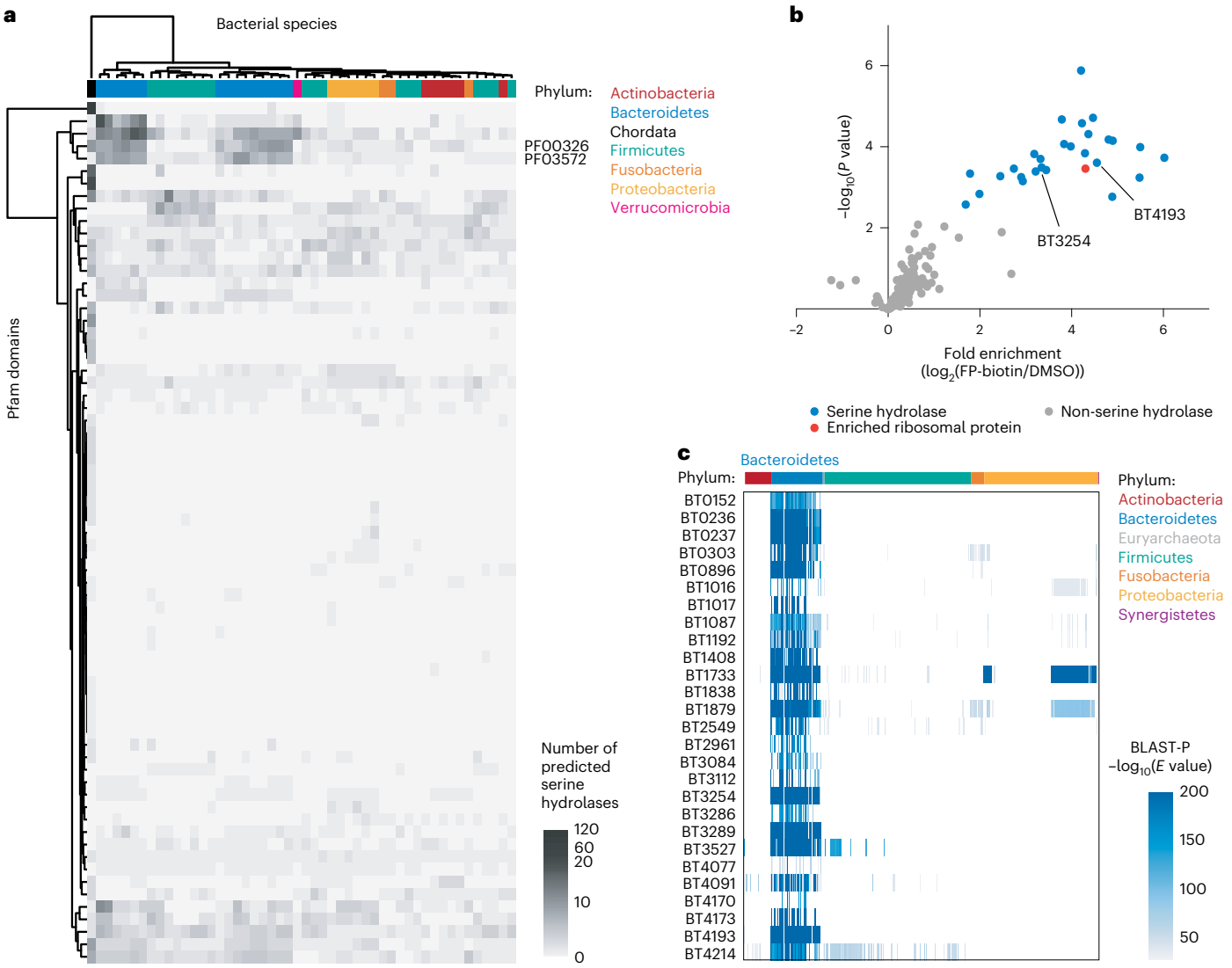


Fig. 1 | Bioinformatics and activity-based protein profiling of *B. thetaiotaomicron* identify serine hydrolases specific to the Bacteroidetes phylum. a, Heatmap of the number of proteins predicted to be serine hydrolases based on annotated Pfam domains in humans (Chordata) and 49 representative gut commensal bacterial species. Species are color-coded by phylum and hierarchically clustered by the number of predicted serine hydrolases with each Pfam domain. Two Bacteroidetes-specific protease Pfam domains, PF00326 (peptidase S9) and PF03572 (peptidase S41), are labeled on the right. **b**, Volcano plot of proteins labeled with FP-biotin relative to dimethyl sulfoxide (DMSO) treatment in *B. thetaiotaomicron* VPI-5482 with three independent replicates. Two-tailed two-sample *t*-test was performed comparing DMSO- and FP-biotin-labeled cells. Significantly enriched hits are colored blue, an enriched ribosomal protein is colored red and hDPP4 homologs, BT4193 and BT3254, are labeled. **c**, Heatmap of homologs of *B. thetaiotaomicron* serine hydrolases across every bacterial strain in the Human Microbiome Project Reference Genomes for the Gastrointestinal Tract database via BLAST-P.

Hence, we sought to confirm whether BT3254 and BT4193 are indeed hDPP4 homologs, determine their sensitivity to current FDA-approved hDPP4 inhibitors and ascertain their physiological functions in *B. thetaiotaomicron*.

BT4193 is a functional homolog of hDPP4

By sequence identity, homologs of the hDPP4 enzyme are predicted to be highly conserved and exclusively found within the Bacteroidetes phylum (Fig. 2a and Supplementary Table 6). Crystallographic studies of hDPP4 have highlighted the following sets of residues key to the activity of the protease: the catalytic triad (Ser630/Asp708/His740), the hydrophobic S1-binding pocket (Tyr631/Val656/Trp659/Tyr662/Try666/Val711) that confers P1 specificity, the double-Glu motif that recognizes the free N-terminus of the substrate (Glu205/Glu206) and the Arg125 residue that hydrogen bonds with Glu205 and the P2 carbonyl group[30]. Based on sequence alignments, both BT4193 and BT3254 have the conserved catalytic triad and Tyr662 and Tyr666 residues that ring-stack with the P1 Pro canonically found in hDPP4 substrates, as well as similar residues forming the rest of the S1-binding pocket (Fig. 2b and Supplementary Figs. 5 and 6). However, only BT4193 has the double-Glu motif and a corresponding Arg residue to recognize and stabilize the free N-terminus of a peptide substrate. Structural homology modeling confirmed the alignment of these residues and suggested that the orientation of the Gln-Glu motif in the predicted structure of BT3254 may not be capable of forming salt bridges with the N-terminus of a substrate (Fig. 2c).

To determine whether this sequence and structural homology confer functional homology, we used the fluorogenic peptide substrate GlyPro-7-amino-4-methylcoumarin (GP-AMC), which is traditionally used to measure hDPP4-like activity via proteolytic release of

https://doi.org/10.1038/s41589-023-01357-8

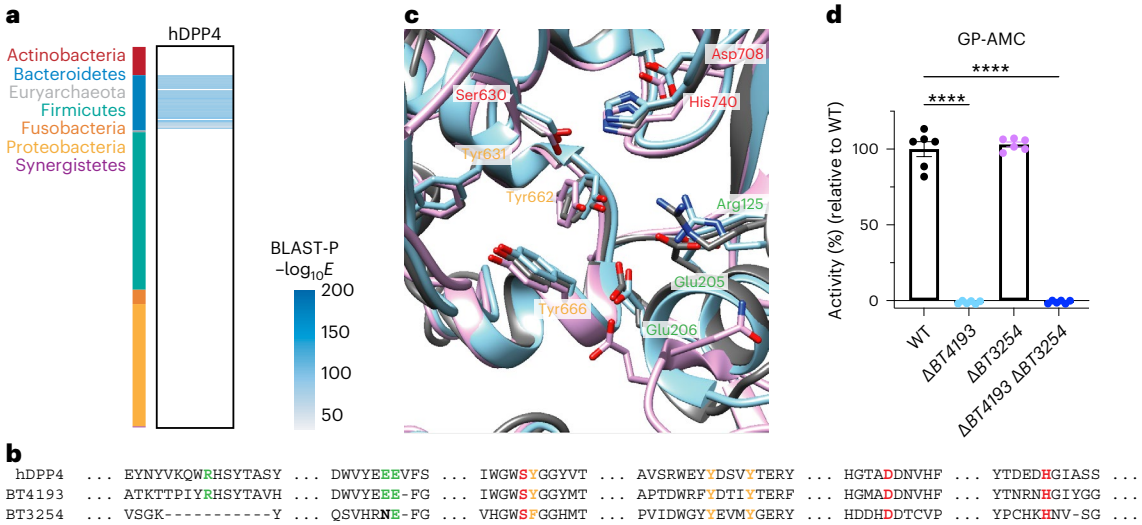


Fig. 2 | BT4193 is a functional homolog of human DPP4 (hDPP4). a, Heatmap of the phylogenetic distribution of hDPP4 homologs across every bacterial strain in the Human Microbiome Project Reference Genomes for the Gastrointestinal Tract database via BLAST-P. **b**, Multiple sequence alignment of hDPP4, BT4193 and BT3254. Residues relevant to hDPP4 function are colored red (catalytic triad, Ser630/Asp708/His740), gold (S1-binding pocket, Tyr631/Tyr662/Try666) or green (free amine-stabilizing residues, Arg125/Glu205/Glu206) along with their corresponding residues in BT4193 and BT3254. **c**, Structural homology modeling alignment of hDPP4 (gray; Protein Data Bank (PDB): 1J2E, chain B) and AlphaFold-predicted structures of BT4193 (blue) and BT3254 (pink). Side chains of residues relevant for the hDPP4 function highlighted in **b** and their corresponding residues in BT4193 and BT3254 are shown, with numbering based on hDPP4. **d**, Quantification of cleavage of canonical hDPP4 fluorogenic peptide substrate GP-AMC in *B. thetaiotaomicron* lysate (ex/em: 380/460 nm; mean ± s.e.m.; $n = 6$ independent replicates). Velocities of substrate cleavage are normalized to WT velocity. Statistical significance was determined using a one-way ANOVA test with post hoc Dunnett's multiple comparisons tests compared with WT ($P < 0.0001$; WT-Δ*BT4193*, $P < 0.0001$; WT-Δ*BT3254*, $P = 0.75$; WT-Δ*BT4193* Δ*BT3254*, $P < 0.0001$; ****$P < 0.0001$).

fluorescent AMC[31]. Wild-type (WT) *B. thetaiotaomicron* lysate cleaved the substrate, suggesting that *B. thetaiotaomicron* has one or more proteases with hDPP4-like activity (Fig. 2d). These results corroborate previously reported evidence of cleavage of a GP-*p*-nitroanilide substrate by *B. thetaiotaomicron*, other species in the Bacteroidetes phylum[32,33], and mouse cecal contents[34]. To ascertain whether the putative hDPP4 homologs, BT4193 and BT3254, contribute to this hDPP4-like activity, we generated single knockouts as well as the double knockout of these genes. GP-AMC processing activity was completely eliminated in the Δ*BT4193* and Δ*BT4193* Δ*BT3254* strains but not in the Δ*BT3254* strain, suggesting that BT4193 is the only functional homolog of hDPP4 with proline diaminopeptidase activity (Fig. 2d). We also recombinantly expressed BT4193 and BT3254 and confirmed that only BT4193 can cleave the GP-AMC substrate (Supplementary Fig. 7).

BT4193 is targeted by hDPP4 inhibitors

As BT4193 is a functional homolog of hDPP4, we wanted to determine if it could be inhibited by compounds targeting the human enzyme, including talabostat and several members of the gliptin class of FDA-approved hypoglycemics for the treatment of type 2 diabetes (Fig. 3a). Pretreatment with two covalent inhibitors, the boronic acid-containing talabostat and the nitrile-containing saxagliptin, specifically and dose-dependently blocked labeling of BT4193 with FP-alkyne (1) as measured by competitive gel-based ABPP (Fig. 3b). Additionally, talabostat and saxagliptin inhibited cleavage of the canonical hDPP4 substrate GP-AMC in WT *B. thetaiotaomicron* lysate, as did the noncovalent inhibitors sitagliptin and linagliptin, although to a lesser extent (Extended Data Fig. 2a). The relative potency of these four inhibitors against BT4193 was confirmed using recombinantly expressed enzyme, with talabostat and saxagliptin having low nanomolar potency against both hDPP4 and BT4193 and sitagliptin and linagliptin being >100- and >1,000-fold weaker inhibitors against BT4193 compared with hDPP4 (Fig. 3c and Extended Data Fig. 2b). None of these inhibitors were active against recombinant BT3254 (Extended Data Fig. 2c).

BT4193 and BT3254 are prolyl aminopeptidases

To comprehensively profile the substrate specificity of BT4193 and BT3254 in an unbiased manner, we applied hDPP4, BT4193 and BT3254 to multiplex substrate profiling by MS (MSP–MS), in which recombinant enzyme is added to a library of synthetic tetradecapeptides whose cleavage is kinetically monitored by liquid chromatography (LC)–MS. Similar to hDPP4, BT4193 primarily trimmed dipeptides from the N-terminus of the substrates, especially after Pro or Ala residues (Fig. 4a–c and Supplementary Table 7). Interestingly, BT3254 also exhibited aminopeptidase activity but trimmed tripeptides with P1 Pro or Ala, suggesting that it is a triaminopeptidase rather than a diaminopeptidase (Fig. 4a–c and Supplementary Table 7).

To confirm these activities, we synthesized a targeted library of fluorogenic peptide substrates that all contained proline in the P1 position and a 7-amino-4-carbamoylmethylcoumarin (ACC) reporter group. The substrates were various lengths and with or without an acetyl cap to determine the necessity of the free N-terminus. In agreement with the results from our MSP–MS screen, we found that BT4193 is a diaminopeptidase like hDPP4 while BT3254 is a triaminopeptidase (Fig. 4d). All three enzymes required substrates to have a free N-terminus, and BT4193 only cleaved dipeptides whereas BT3254 cleaved tripeptides. These substrate preferences were confirmed in vitro with *B. thetaiotaomicron* lysate (Supplementary Fig. 8). The preference for P1 Pro by both enzymes was corroborated with a small P1 library of fluorogenic peptide substrates (Extended Data Fig. 3), which matched the substrate specificity of hDPP4 (ref. 35).

To more comprehensively map substrate specificity at the P2 and P3 positions for each protease, we synthesized a series of positional scanning synthetic combinatorial libraries of P1 Pro fluorogenic peptides, in which each position is systematically scanned through each natural amino acid (excluding cysteine and methionine but including norleucine) and the remaining positions are composed of an isokinetic mixture of those amino acids. Screening the dipeptide library against hDPP4 and BT4193 demonstrated a highly overlapping preference

https://doi.org/10.1038/s41589-023-01357-8

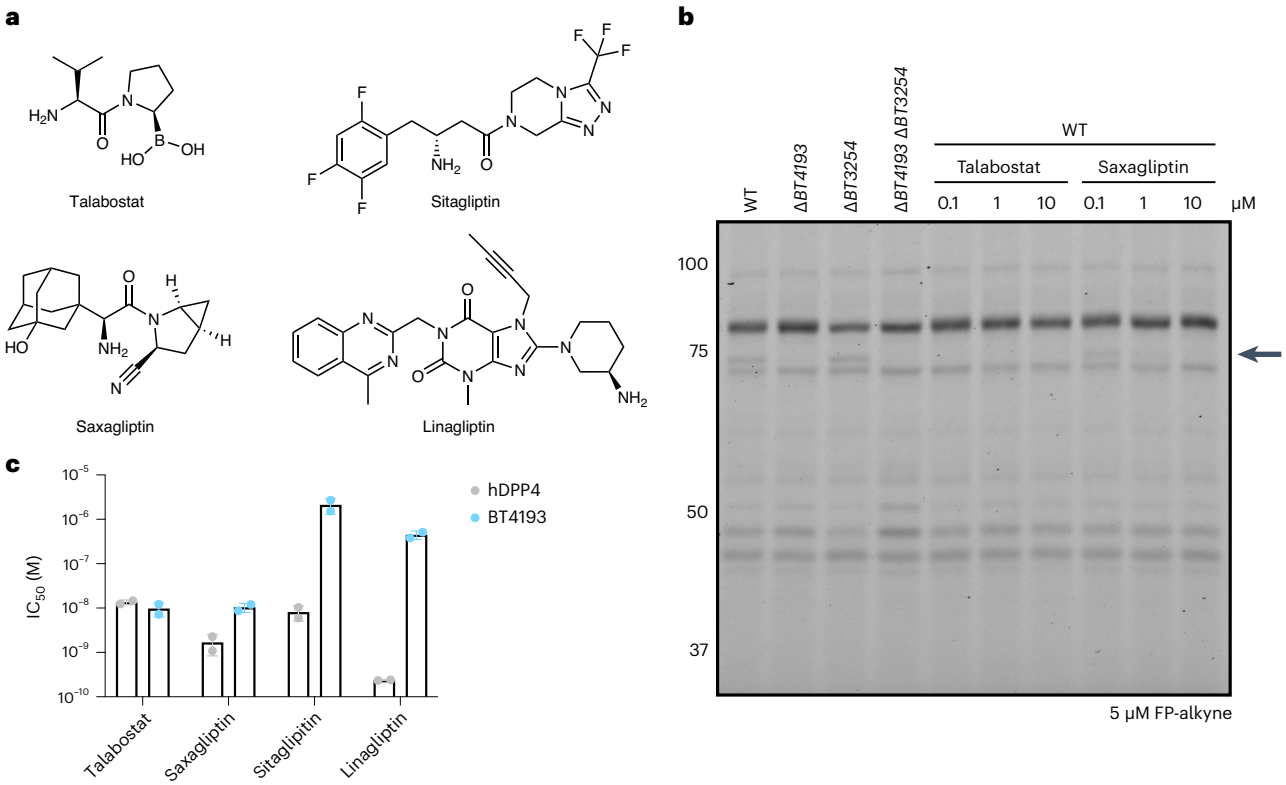


Fig. 3 | Human DPP4 inhibitors target BT4193. a, Structure of hDPP4 inhibitors. **b**, In-gel fluorescence labeling of intact *B. thetaiotaomicron* strains. Pretreatment with talabostat or saxagliptin competes labeling of only BT4193 by FP-alkyne in WT *B. thetaiotaomicron*. FP-alkyne labeling was conjugated with tetramethylrhodamine-azide via Cu(I)-catalyzed azide–alkyne cycloaddition and visualized via SDS–PAGE and in-gel fluorescence. The arrow indicates the location of the band corresponding to BT4193 (predicted molecular weight, 84 kDa). Gel is representative of three independent experiments. **c**, Apparent IC_{50} values of recombinant hDPP4 and BT4193 after treatment with inhibitor for 30 min before measuring activity via GP-AMC cleavage (mean ± s.e.m.; $n = 2$ independent replicates).

for bulky, hydrophobic P2 residues, although most residues could be accommodated by both the bacterial and human enzymes (Fig. 4e and Supplementary Fig. 9). Screening of two tripeptide libraries to scan the P2 and P3 preferences of BT3254 also confirmed that the majority of the residues could be tolerated in either position (Fig. 4f). Together, these results suggest that the specificities of BT4193 and BT3254 are driven primarily by the P1 proline residue and that both enzymes are prolyl aminopeptidases.

Loss of BT4193 causes sensitivity to cell envelope stressors

To characterize the functions of BT4193 and BT3254, we first sought to identify the localization of these proteases in *B. thetaiotaomicron* cells. Both enzymes are predicted to have a type I Sec signal peptide cleavage site without a lipoprotein motif and are predicted to be outer membrane-associated (Supplementary Table 4). Additionally, previous studies suggest that these enzymes are not surface-exposed[36], and homologs of BT4193 and BT3254 in the closely related *Bacteroides fragilis* were not found to be secreted via proteomics[37]. Furthermore, we did not detect the activity of either BT4193 or BT3254 (as indicated by turnover of their substrates AP-ACC and AAP-ACC, respectively) in *B. thetaiotaomicron* culture supernatants (Supplementary Fig. 10). Thus, we conclude that BT4193 and BT3254 are likely localized in the periplasm.

Based on the predicted periplasmic localization, we hypothesized that these enzymes may be important for some aspect of envelope integrity. Previous fitness measurements of a pooled transposon library demonstrated that transposon insertions in *BT4193* have increased susceptibility to a number of cell envelope stressors, including the antibiotic vancomycin and the bile salt deoxycholic acid[38] (Fig. 5a). We confirmed that deletion of *BT4193* results in a substantial growth defect during vancomycin treatment (Fig. 5b,c), and Δ*BT4193* Δ*BT3254* cells exhibited increased sensitivity to deoxycholic acid (Extended Data Fig. 4a). Complementation with WT *BT4193* (Δ*BT4193*::*BT4193*^{WT}) rescued the growth of Δ*BT4193* in the presence of vancomycin (Fig. 5b,c). Interestingly, complementation with the catalytically inactive mutant *BT4193*^{S606A} (Δ*BT4193*::*BT4193*^{S606A}) also rescued growth (Fig. 5b,c and Extended Data Fig. 5). Additionally, Δ*BT4193* and Δ*BT4193* Δ*BT3254* cells exhibited increased susceptibility to the antimicrobial peptide polymyxin B (Fig. 5c and Extended Data Fig. 4b), which destabilizes the outer membrane via binding to lipopolysaccharide molecules[39]. These phenotypes did not reflect a general fitness defect because Δ*BT4193* and Δ*BT4193* Δ*BT3254* cells were similarly sensitive as WT cells to stressors such as pH, ethanol or sodium chloride (Extended Data Fig. 6). Moreover, sensitivity to vancomycin and polymyxin B was not due to a general increase in membrane permeability, as Δ*BT4193* and Δ*BT4193* Δ*BT3254* cells did not exhibit increased uptake of fluorescent vancomycin or propidium iodide (Supplementary Fig. 11). These results suggest that BT4193 has a pivotal role in conferring resistance to various cell envelope stressors through a noncatalytic mechanism.

BT4193 is important for fitness within bacterial communities

Previous studies demonstrated that treatment of a diabetic mouse model with saxagliptin leads to the depletion of bacteria from the Bacteroidetes phylum in the gut[40]. However, this model does not allow for the deconvolution of the effect of inhibiting the host DPP4 enzyme versus bacterial DPP4 enzymes. Thus, to determine whether BT4193

https://doi.org/10.1038/s41589-023-01357-8

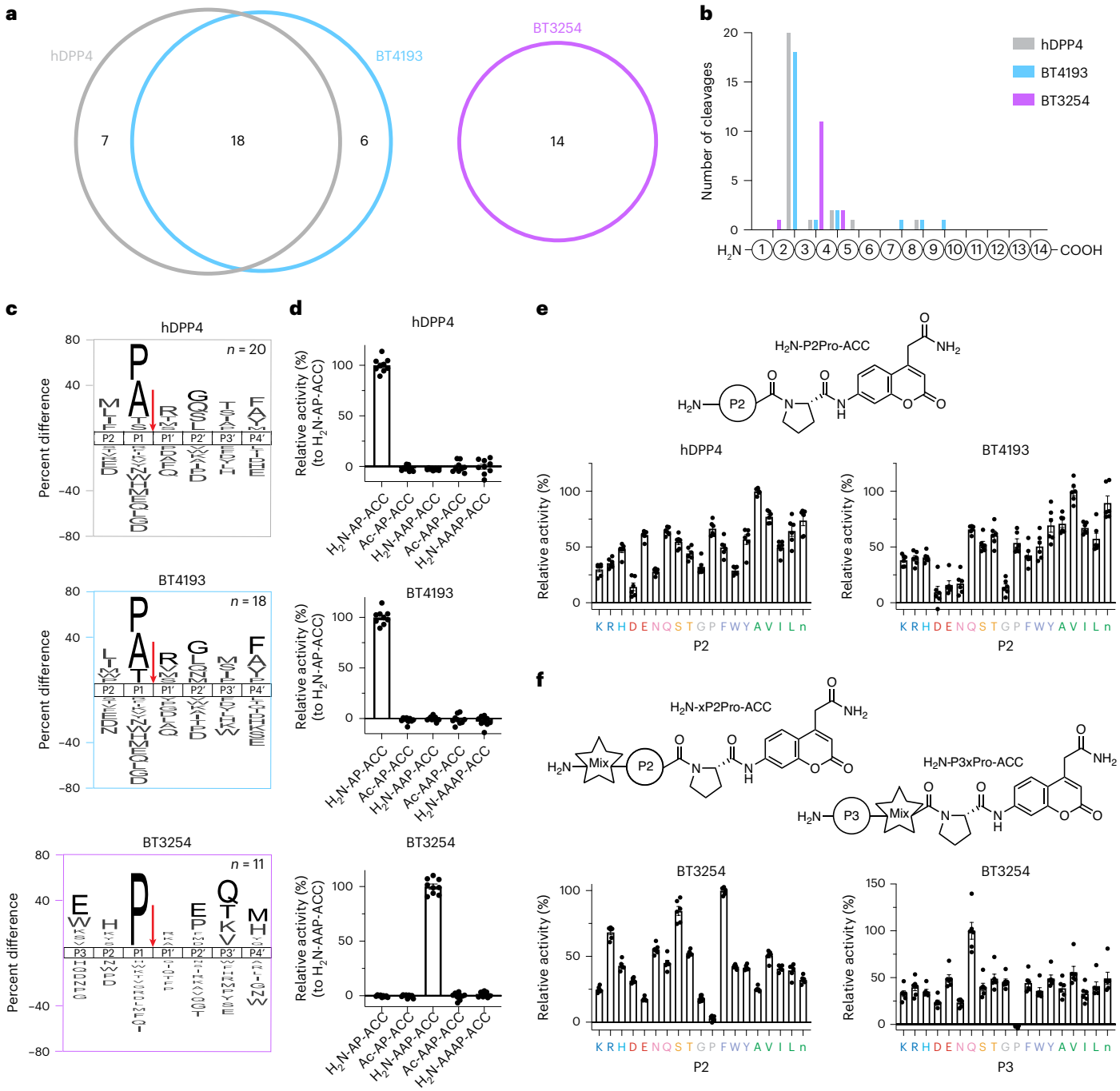


Fig. 4 | Substrate specificity of BT4193 and BT3254 is driven by P1 Pro. **a**–**c**, Substrates cleaved by recombinant hDPP4, BT4193 and BT3254 based on MSP–MS, defined as significant fold-change between 0 h and 4 h of incubation ($n = 4$ independent replicates), quantified as a Venn diagram (**a**) by the site of cleavage within the 14-mer peptide and (**b**) as iceLogo plots of the 20 cleavage sites that occurred at position 2 for hDPP4, 18 cleavage sites that occurred at position 2 for BT4193 and 11 cleavage sites that occurred at position 3 for BT3254 (**c**). For the iceLogo plots, the negative datasets are the 228 cleavage sites at position 2 for hDPP4 and BT4193 or position 3 for BT3254 within the library (amino acids in black are significantly enriched ($P < 0.05$), amino acids in gray have $P < 0.50$; M, norleucine). **d**, Quantification of cleavage of fluorogenic peptide substrates by recombinant hDPP4, BT4193 and BT3254 (ex/em: 355/460 nm; mean ± s.e.m.; $n = 9$ independent replicates). Initial velocities were normalized to the best substrate (H_2N-AP-ACC for hDPP4 and BT4193 and H_2N-AAP-ACC for BT3254). BT4193 is a diaminopeptidase like hDPP4 and BT3254 is a triaminopeptidase. **e**, Quantification of cleavage velocity of a P2Pro-ACC positional scanning library relative to the peptide with the highest turnover for recombinant hDPP4 and BT4193 (ex/em: 355/460 nm; mean ± s.e.m.; $n = 6$ independent replicates). **f**, Quantification of cleavage velocity of xP2Pro-ACC and P3xPro-ACC positional scanning libraries by recombinant BT3254 relative to the peptide in the library with the highest turnover (ex/em: 355/460 nm; mean ± s.e.m.; $n = 6$ independent replicates).

has a role in the establishment of *B. thetaiotaomicron* in a diverse complex bacterial community, we used a synthetic in vitro community of 14 stool-derived gut commensals that can be stably maintained in rich media and reflects the diversity of bacterial families in the host from which the isolates were obtained[41,42] (Fig. 6a). The 14 members comprise several *Bacteroides* species and a *Parabacteroides* species in the

https://doi.org/10.1038/s41589-023-01357-8

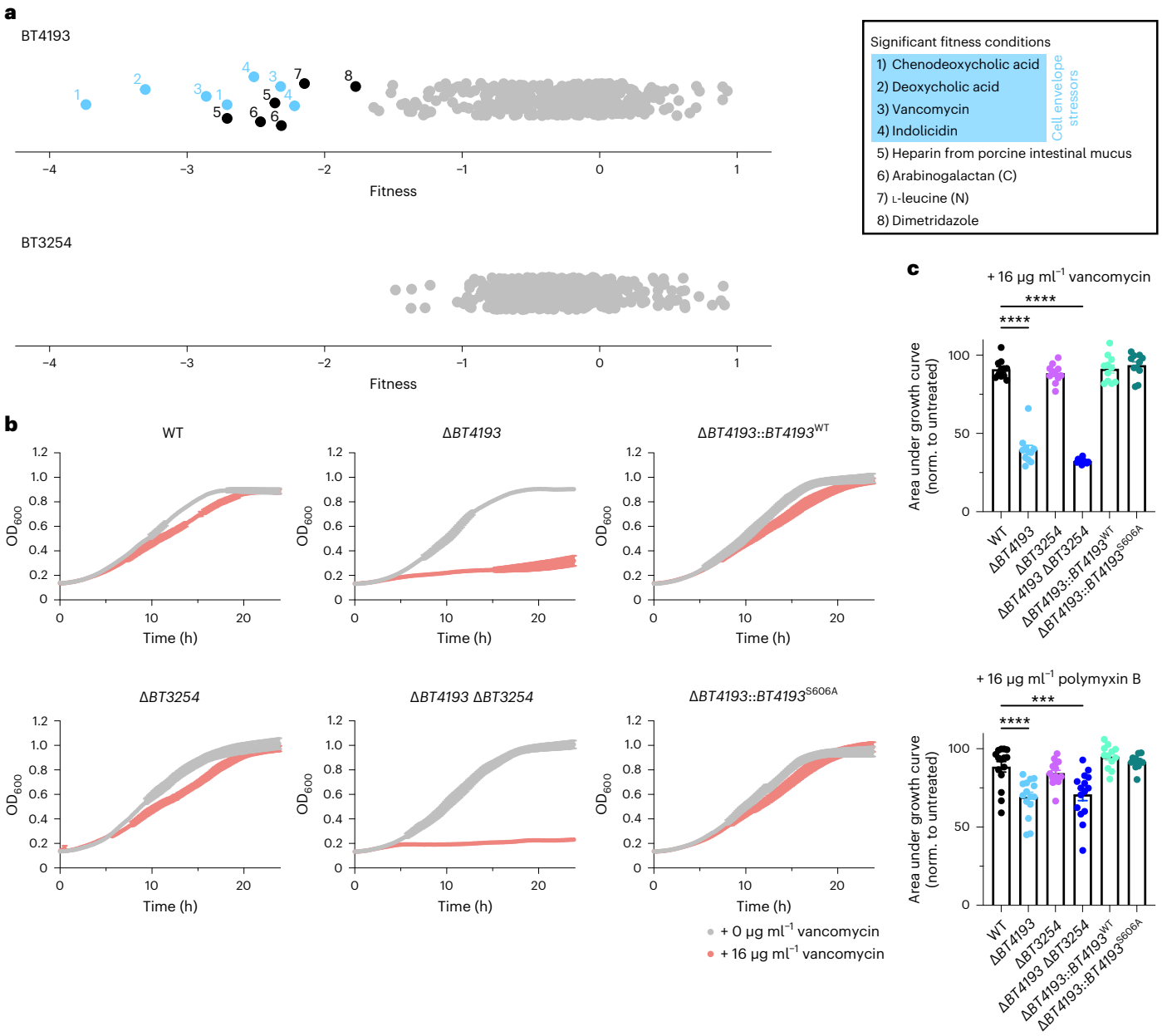


Fig. 5 | BT4193 confers resistance to envelope stressors vancomycin and polymyxin B. a, Fitness scores for BT4193 and BT3254 from screening a genome-wide randomly barcoded transposon insertion *B. thetaiotaomicron* mutant library in >500 conditions[38]. Fitness scores outside the 99.7% confidence interval are considered significant, and significant conditions representing cell envelope stressors are colored blue. **b**, Growth of *B. thetaiotaomicron* strains during vancomycin treatment. **c**, Quantification of the area under the growth curve for treatment with vancomycin and polymyxin B was normalized to the growth curve of untreated cultures. Deletion of *BT4193* substantially impacted fitness in the presence of vancomycin (mean ± s.e.m.; $n = 11$ independent replicates) and polymyxin B (mean ± s.e.m.; $n = 15$ independent replicates for WT, Δ*BT4193*, Δ*BT3254*, Δ*BT4193* Δ*BT3254*; $n = 12$ independent replicates for Δ*BT4193*::*BT4193*^{WT} and Δ*BT4193*::*BT4193*^{S606A}). Statistical significance was determined using a one-way ANOVA test with post hoc Dunnett's multiple comparisons tests compared with WT (vancomycin: $P < 0.0001$; WT-Δ*BT4193*, $P < 0.0001$; WT-Δ*BT3254*, $P = 0.86$; WT-Δ*BT4193* Δ*BT3254*, $P < 0.0001$; WT-Δ*BT4193*::*BT4193*^{WT}, $P = 1.00$; WT-Δ*BT4193*::*BT4193*^{S606A}, $P = 0.85$; polymyxin B: $-P < 0.0001$; WT-Δ*BT4193*, $P < 0.0001$; WT-Δ*BT3254*, $P = 0.74$; WT-Δ*BT4193* Δ*BT3254*, $P = 0.0001$; WT-Δ*BT4193*::*BT4193*^{WT}, $P = 0.42$; WT-Δ*BT4193*::*BT4193*^{S606A}, $P = 0.92$; ***$P < 0.001$, ****$P < 0.0001$).

Bacteroidetes phylum, but not *B. thetaiotaomicron*. The single- and double-knockout strains did not exhibit any growth defects during monoculture growth in the rich brain heart infusion (BHI) medium (Extended Data Fig. 7). When any of our *B. thetaiotaomicron* strains were added individually to the 14-member community, overall community composition was not dramatically altered (Extended Data Fig. 8). However, the relative abundance of the Δ*BT4193* and Δ*BT4193* Δ*BT3254* strains was significantly lower than that of WT (Fig. 6b). To test the extent to which the lower fitness of the knockouts was sensitive to the nutrient environment, we passaged the same 15-member communities in modified Gifu anaerobic medium (mGAM), another rich medium that supports rapid growth of *Bacteroides* species (Extended Data Fig. 7). Similar to BHI, the *ΔBT4193* and *ΔBT4193 ΔBT3254* strains were at lower relative abundance levels than WT in mGAM (Fig. 6c) without any concomitant growth defects in monoculture, indicating that the hDPP4 homolog BT4193 generally has an important role in

https://doi.org/10.1038/s41589-023-01357-8

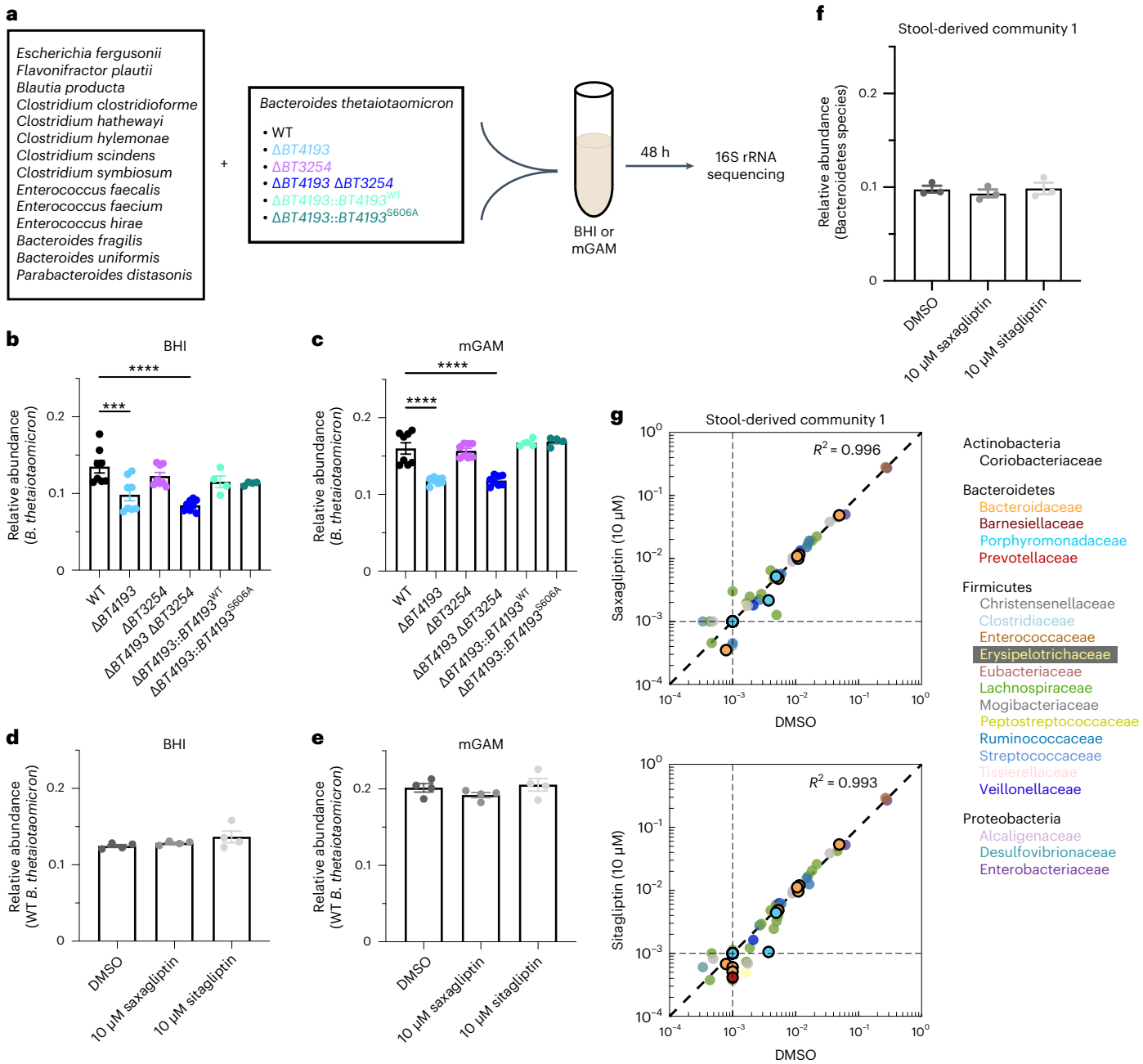


Fig. 6 | BT4193, but not hDPP4-targeting drugs, impacts *B. thetaiotaomicron* fitness during in vitro growth in diverse communities. a, Schematic representation of synthetic community experiments. Fourteen human gut commensal species were cocultured with one of six *B. thetaiotaomicron* strains (WT, Δ*BT4193*, Δ*BT3254*, Δ*BT4193* Δ*BT3254*, Δ*BT4193*::*BT4193*^{WT} or Δ*BT4193*::*BT4193*^{S606A}) for 48 h in BHI or mGAM. The relative abundance of each species was quantified using 16S rRNA sequencing. **b**, Quantification of the relative abundance of *B. thetaiotaomicron* within each synthetic community cocultured in BHI (mean ± s.e.m.; $n = 8$ independent replicates for WT, Δ*BT4193*, Δ*BT3254*, Δ*BT4193* Δ*BT3254*; $n = 4$ independent replicates for Δ*BT4193*::*BT4193*^{WT} and Δ*BT4193*::*BT4193*^{S606A}). Statistical significance was determined using a one-way ANOVA test with post hoc Dunnett's multiple comparisons tests compared with wild type ($P < 0.0001$; WT-Δ*BT4193*, $P = 0.0005$; WT-Δ*BT3254*, $P = 0.48$; WT-Δ*BT4193* Δ*BT3254*, $P < 0.0001$; WT-Δ*BT4193*::*BT4193*^{WT}, $P = 0.23$; WT-Δ*BT4193*::*BT4193*^{S606A}, $P = 0.15$; ***$P < 0.001$; ****$P < 0.0001$). **c**, Quantification of the relative abundance of *B. thetaiotaomicron* within each synthetic community cocultured in mGAM (mean ± s.e.m.; $n = 8$ independent replicates for WT, Δ*BT4193*, Δ*BT3254*, Δ*BT4193* Δ*BT3254*; $n = 4$ independent replicates for Δ*BT4193*::*BT4193*^{WT} and Δ*BT4193*::*BT4193*^{S606A}). Statistical significance was determined using a one-way ANOVA test with post hoc Dunnett's multiple comparisons tests compared with wild type ($P < 0.0001$; WT-Δ*BT4193*, $P < 0.0001$; WT-Δ*BT3254*, $P = 0.98$; WT-Δ*BT4193* Δ*BT3254*, $P < 0.0001$; WT-Δ*BT4193*::*BT4193*^{WT}, $P = 0.74$; WT-Δ*BT4193*::*BT4193*^{S606A}, $P = 0.58$; ****$P < 0.0001$). **d**, Quantification of the relative abundance of WT *B. thetaiotaomicron* in synthetic communities after 48 h of treatment with 10 μM saxagliptin or 10 μM sitagliptin in BHI (mean ± s.e.m.; $n = 4$ independent replicates). Lack of statistical significance was determined using a one-way ANOVA test ($P = 0.2273$). **e**, Quantification of the relative abundance of WT *B. thetaiotaomicron* in synthetic communities after 48 h of treatment with 10 μM saxagliptin or 10 μM sitagliptin in mGAM (mean ± s.e.m.; $n = 4$ independent replicates). Lack of statistical significance was determined using a one-way ANOVA test ($P = 0.31$). **f**, Quantification of the relative abundance of bacteria from the phylum Bacteroidetes in a representative stool-derived community after 48 h of treatment with 10 μM saxagliptin or 10 μM sitagliptin grown in BHI (mean ± s.e.m.; $n = 3$ independent replicates). Lack of statistical significance was determined using a one-way ANOVA test ($P = 0.48$). **g**, Correlation of the relative abundance of each amplicon-sequence variant (ASV) between 48 h of vehicle (DMSO) and drug treatment in a representative stool-derived community grown in BHI. Plotted values for each ASV represent the mean of three independent replicates. Each ASV is colored by family. ASVs within the phylum Bacteroidetes are emphasized with a black outer ring. Dotted lines at 10^{-3} represent the limit of detection for sequencing depth.

the fitness of *B. thetaiotaomicron* specifically within diverse bacterial communities. Much like sensitivity to cell envelope stressors (Fig. 5b,c and Extended Data Fig. 4), strains expressing the catalytically inactive mutant *BT4193*^{S606A} (Δ*BT4193*::*BT4193*^{S606A}) did not exhibit a fitness defect (Fig. 6b,c), and targeting BT4193 with saxagliptin or sitagliptin did not affect the levels of *B. thetaiotaomicron* or the other Bacteroidetes species in the community (Fig. 6d,e and Supplementary Fig. 12). Thus, fitness defects within these synthetic communities are dependent on noncatalytic functions of BT4193.

To determine whether hDPP4-targeting drugs are likely to affect the composition of more diverse communities with a broader range of Bacteroidetes species, we treated eight in vitro communities derived from the stool of individual healthy donors. These stool-derived communities are highly diverse and maintain a stable composition during repeated passaging in BHI and mGAM[42]. Similar to our 15-member synthetic community, treatment with either saxagliptin or sitagliptin for 48 h did not affect the relative abundance of Bacteroidetes species (Fig. 6f and Extended Data Fig. 9) or the overall composition of any of the communities (Fig. 6g and Supplementary Fig. 13). Thus, in vitro community composition and *B. thetaiotaomicron* fitness in these rich media are not dependent on the proteolytic functions of BT4193.

Discussion

Serine hydrolases have critical roles in mammalian pathophysiology, including inflammation, blood clotting and neurotransmitter signaling. As a result, they have been successfully targeted with small-molecule drugs for the treatment of a wide variety of human diseases[12]. However, little is known about the diversity and function of this enzyme superfamily in the phylogenetically divergent bacteria that comprise the human gut microbiota. Our bioinformatic approach should enable the identification of predicted serine hydrolases more broadly in the gut microbiota, and then these predictions can be functionally validated using chemoproteomics. Our analysis suggests that there are several types of bacterial serine hydrolases that are not found in humans (Fig. 1a), indicating that simple protein BLAST analyses based on known serine hydrolases in humans will fail to identify many of these enzymes in bacteria. Furthermore, our findings that BT4193 has functional roles independent of its proteolytic activity indicate the importance of using methods other than simple genetic disruption to define the functions of predicted serine hydrolases.

The increased sensitivity of Δ*BT4193* cells to vancomycin and the cationic antimicrobial peptide polymyxin B (Fig. 5) supports a role for BT4193 in regulating the integrity of the cell envelope. Polymyxin B is an antimicrobial peptide that acts via pore formation to which WT *B. thetaiotaomicron* is only mildly susceptible[43]. Vancomycin, which prevents cross-linking of nascent peptidoglycan chains in the cell wall, is traditionally thought to target Gram-positive bacteria because Gram-negative bacteria have an outer membrane surrounding the peptidoglycan layer that can act as a barrier to vancomycin and confer resistance[44]. Mutants in *Escherichia coli*[45] and *Caulobacter crescentus*[46] that modulate outer membrane permeability have increased sensitivity to both vancomycin and deoxycholic acid, similar to the phenotypes of *B. thetaiotaomicron* strains lacking *BT4193*. However, we showed that Δ*BT4193* and Δ*BT4193* Δ*BT3254* cells do not exhibit a general increase in uptake of fluorescent vancomycin or propidium iodide (Supplementary Fig. 11). The growth defects of Δ*BT4193* and Δ*BT4193* Δ*BT3254* cells across multiple types of cell envelope stressors provide strong evidence that BT4193 has an important role in maintaining envelope integrity, but through a modality unrelated to general membrane leakiness. Because complementation with either WT or catalytically inactive *BT4193* rescued susceptibility to these stressors (Fig. 5), this function of BT4193 likely contributes to the integrity of the periplasmic space through a noncatalytic function such as protein–protein interactions. Although several critical hDPP4 functions are mediated through protein–protein interactions involving its β-propeller domain[18], the N-linked glycosylation sites and cysteine residues in the glycosylation-rich and cysteine-rich regions of hDPP4 are not conserved in BT4193, limiting our ability to predict protein-binding partners with BT4193. Relative to other Gram-negative bacteria from the Proteobacteria phylum such as *E. coli*, little is known about the architecture of the cell envelope in *Bacteroides* species beyond differences in the lipid A component of lipopolysaccharide molecules[47]. This knowledge gap makes further characterization of Δ*BT4193* phenotypes challenging. Future studies using techniques such as immunoprecipitation–MS may help to identify and validate key interactors for BT4193 in *B. thetaiotaomicron*.

Within a variety of diverse in vitro communities, hDPP4-targeting drugs did not affect the relative abundance of bacteria from the phylum Bacteroidetes, consistent with our finding that *B. thetaiotaomicron*'s niche is independent of the proteolytic activity of BT4193 (Fig. 6). Nonetheless, the presence of BT4193 significantly impacted the establishment of *B. thetaiotaomicron* within communities (Fig. 6). The ability to modulate the *B. thetaiotaomicron* levels through targeting BT4193 has exciting implications for subtle community engineering; current tools for editing community composition such as antibiotics and bacteriophages typically have more drastic effects on community diversity and composition. Our findings highlight the importance of combining ABPP, cellular physiology interrogation and *in vitro* community growth to fully characterize the function of bacterial enzymes within the gut microbiota and define the extent to which drugs that target human proteins impact the gut microbiota.

The role of BT4193 in *B. thetaiotaomicron* physiology and fitness is particularly important considering 10% of patients with diabetes in the US take hDPP4 inhibitors[48]. As a result, any cross-reactivity of these drugs with bacterial enzymes could impact microbial communities in the gut. We assessed the off-target effects of several FDA-approved inhibitors, such as saxagliptin (Onglyza), sitagliptin (Januvia) and linagliptin (Tradjenta), on BT4193 (Fig. 3). The potency of these inhibitors against BT4193 matched their known selectivity for hDPP4 over other human prolyl oligopeptidases[49] and was similar to their efficacy against an hDPP4 homolog in another bacterial species from the phylum Bacteroidetes, *Porphyromonas gingivalis*[50]. The inhibitors sitagliptin and linagliptin derive their selectivity by occupying the S_2 extensive and S_1'/S_2' pockets on hDPP4, respectively[51]. The lack of potency of sitagliptin and linagliptin against BT4193 may be driven by similar differences within its binding pocket. While hDPP4 inhibitors are primarily cleared renally, 20% of saxagliptin[52] and 10% of sitagliptin[53] are eliminated fecally, suggesting that *B. thetaiotaomicron* may be exposed to substantial concentrations of these drugs in the gut. Despite this likely exposure, we showed that treatment with saxagliptin or sitagliptin does not affect the fitness of *B. thetaiotaomicron* or other Bacteroidetes species in synthetic or stool-derived in vitro communities (Fig. 6). These results suggest that the depletion of bacteria from the Bacteroidetes phylum upon saxagliptin treatment seen previously in a diabetic mouse model[40] may be due to inhibition of the host rather than bacterial DPP4 enzymes. One caveat is that the mice in the diabetes model were treated for 8 weeks with saxagliptin[40], as opposed to the single 48-h dose we applied to in vitro communities. Therefore, we cannot rule out that prolonged exposure may have detrimental effects on the gut microbiota of patients with diabetes through direct targeting of BT4193. Regardless, these results highlight the importance of thorough functional analysis of potential off-targets for existing drugs to complement bioinformatic predictions to determine whether unintended impacts on the microbiota are likely.

Although we have identified several critical functions of BT4193, these roles are not dependent on its proteolytic activity. Future efforts to map the endogenous substrates of BT4193 using proteomic methods should provide molecular insight into other functions of BT4193. In addition to potentially having specific peptide or protein targets, both BT4193 and BT3254 may contribute to global protein catabolism. The phylogenetically related oral pathogen *P. gingivalis* is an asaccharolytic

bacterium that relies on a network of peptidases, including homologs of BT4193 and BT3254, to generate dipeptides and tripeptides from exogenous protein that can be more easily imported into the cytoplasm for protein catabolism[54]. Unlike *P. gingivalis*, the hallmark of gut commensal bacteria from the Bacteroidetes phylum is their carbohydrate utilization capabilities. Nevertheless, our chemoproteomic profiling of *B. thetaiotaomicron* identified several other putative dipeptidyl peptidases that may complete a network homologous to that in *P. gingivalis*. Even in the asaccharolytic *P. gingivalis*, fitness defects emerged only in a triple-knockout strain[55]. Thus, there may be compensatory regulation of peptidase activities that obscures defects in nutrient acquisition in the single- and double-deletion strains used in this study. Future investigation will be required to clarify how these enzymes interact with the larger network of dipeptidases in *B. thetaiotaomicron*.

Previous studies identified DPP4 activity in the cecum of germ-free mice, likely from host DPP4 (ref. 56), and FP-reactive hDPP4 can be found in human fecal samples[57]. Interestingly, some of the canonical peptide hormone substrates of hDPP4 are secreted into the gut lumen—peptide YY is primarily secreted luminally[58] and has a role in regulating ion and water absorption[59], and GLP1 has been shown to be secreted apically in ex vivo intestinal tissue[60]. The physiological impact of the luminal secretion of these peptide hormones and hDPP4 is not yet understood, but likely contributes to the multifaceted role of bacterial DPP4 enzymes in the diverse community within the gut lumen.

Overall, our study demonstrates how a combination of bioinformatics and chemoproteomics can identify and prioritize physiologically relevant, druggable enzymes in gut commensal bacteria. This approach led to the successful characterization of two proteases and the discovery of an hDPP4 homolog whose function, at least in part, is not dependent on its proteolytic activity. Moreover, our findings highlight the power of a bottom-up approach to uncover molecular regulators of the dynamics of microbial communities. This approach can be similarly applied to other key gut commensal bacterial species. Finally, it underscores the need to broadly understand the possible off-target effects of existing drugs on bacterial enzymes and how they can impact the gut microbiota.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41589-023-01357-8.

References

1. Maier, L. et al. Extensive impact of non-antibiotic drugs on human gut bacteria. *Nature* **555**, 623–628 (2018).
2. Zimmermann, M., Zimmermann-Kogadeeva, M., Wegmann, R. & Goodman, A. L. Mapping human microbiome drug metabolism by gut bacteria and their genes. *Nature* **570**, 462–467 (2019).
3. Wallace, B. D. et al. Alleviating cancer drug toxicity by inhibiting a bacterial enzyme. *Science* **330**, 831–835 (2010).
4. Maini Rekdal, V., Bess, E. N., Bisanz, J. E., Turnbaugh, P. J. & Balskus, E. P. Discovery and inhibition of an interspecies gut bacterial pathway for Levodopa metabolism. *Science* **364**, eaau6323 (2019).
5. Kidd, D., Liu, Y. & Cravatt, B. F. Profiling serine hydrolase activities in complex proteomes. *Biochemistry* **40**, 4005–4015 (2001).
6. Parasar, B. et al. Chemoproteomic profiling of gut microbiota-associated bile salt hydrolase activity. *ACS Cent. Sci.* **5**, 867–873 (2019).
7. Wu, L. et al. Activity-based probes for functional interrogation of retaining β-glucuronidases. *Nat. Chem. Biol.* **13**, 867–873 (2017).
8. Whidbey, C. et al. A probe-enabled approach for the selective isolation and characterization of functionally active subpopulations in the gut microbiome. *J. Am. Chem. Soc.* **141**, 42–47 (2019).
9. Chatterjee, S. et al. A comprehensive and scalable database search system for metaproteomics. *BMC Genomics* **17**, 642 (2016).
10. Issa Isaac, N., Philippe, D., Nicholas, A., Raoult, D. & Eric, C. Metaproteomics of the human gut microbiota: challenges and contributions to other OMICS. *Clin. Mass Spectrom.* **14**, 18–30 (2019).
11. Simon, G. M. & Cravatt, B. F. Activity-based proteomics of enzyme superfamilies: serine hydrolases as a case study. *J. Biol. Chem.* **285**, 11051–11055 (2010).
12. Bachovchin, D. A. & Cravatt, B. F. The pharmacological landscape and therapeutic potential of serine hydrolases. *Nat. Rev. Drug Discov.* **11**, 52–68 (2012).
13. Ortega, C. et al. Systematic survey of serine hydrolase activity in *Mycobacterium tuberculosis* defines changes associated with persistence. *Cell Chem. Biol.* **23**, 290–298 (2016).
14. Babin, B. M. et al. Identification of covalent inhibitors that disrupt *M. tuberculosis* growth by targeting multiple serine hydrolases involved in lipid metabolism. *Cell Chem. Biol.* **29**, 897–909 (2022).
15. Li, M. et al. Identification of cell wall synthesis inhibitors active against *Mycobacterium tuberculosis* by competitive activity-based protein profiling. *Cell Chem. Biol.* **29**, 883–896 (2022).
16. Deacon, C. F. Physiology and pharmacology of DPP-4 in glucose homeostasis and the treatment of type 2 diabetes. *Front. Endocrinol.* **10**, 80 (2019).
17. Deacon, C. F. Dipeptidyl peptidase-4 inhibitors in the treatment of type 2 diabetes: a comparative review. *Diabetes Obes. Metab.* **13**, 7–18 (2011).
18. Klemann, C., Wagner, L., Stephan, M. & von Hörsten, S. Cut to the chase: a review of CD26/dipeptidyl peptidase-4's (DPP4) entanglement in the immune system. *Clin. Exp. Immunol.* **185**, 1–21 (2016).
19. Sonnenburg, J. L. et al. Glycan foraging in vivo by an intestine-adapted bacterial symbiont. *Science* **307**, 1955–1959 (2005).
20. Macfarlane, G. T., Allison, C., Gibson, S. A. W. & Cummings, J. H. Contribution of the microflora to proteolysis in the human large intestine. *J. Appl. Bacteriol.* **64**, 37–46 (1988).
21. Mills, R. H. et al. Multi-omics analyses of the ulcerative colitis gut microbiome link *Bacteroides vulgatus* proteases with disease severity. *Nat. Microbiol.* **7**, 262–276 (2022).
22. Bachovchin, D. A. et al. Superfamily-wide portrait of serine hydrolase inhibition achieved by library-versus-library screening. *Proc. Natl Acad. Sci. USA* **107**, 20941–20946 (2010).
23. Kaschani, F. et al. Diversity of serine hydrolase activities of unchallenged and Botrytis-infected *Arabidopsis thaliana*. *Mol. Cell. Proteom.* **8**, 1082–1093 (2009).
24. Hatzios, S. K. et al. Chemoproteomic profiling of host and pathogen enzymes active in cholera. *Nat. Chem. Biol.* **12**, 268–274 (2016).
25. Lentz, C. S. et al. Identification of a *S. aureus* virulence factor by activity-based protein profiling (ABPP). *Nat. Chem. Biol.* **14**, 609–617 (2018).
26. Keller, L. J. et al. Characterization of serine hydrolases across clinical isolates of commensal skin bacteria *Staphylococcus epidermidis* using activity-based protein profiling. *ACS Infect. Dis.* **6**, 930–938 (2020).
27. Zweerink, S. et al. Activity-based protein profiling as a robust method for enzyme identification and screening in extremophilic Archaea. *Nat. Commun.* **8**, 15352 (2017).

28. Human Microbiome Project Consortium. et al. Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012).
29. Goodman, A. L. et al. Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell Host Microbe* **6**, 279–289 (2009).
30. Rasmussen, H. B., Branner, S., Wiberg, F. C. & Wagtmann, N. Crystal structure of human dipeptidyl peptidase IV/CD26 in complex with a substrate analog. *Nat. Struct. Biol.* **10**, 19–25 (2003).
31. Hino, M. et al. Glycylprolyl β-naphthylamidase activity in human serum. *Clin. Chim. Acta* **62**, 5–11 (1975).
32. Macfarlane, S. & Macfarlane, G. T. Formation of a dipeptidyl arylamidase by *Bacteroides splanchnicus* NCTC 10825 with specificities towards glycylprolyl-x and valylalanine-x substrates. *J. Med. Microbiol.* **46**, 547–555 (1997).
33. Banbula, A. et al. Emerging family of proline-specific peptidases of *Porphyromonas gingivalis*: purification and characterization of serine dipeptidyl peptidase, a structural and functional homologue of mammalian prolyl dipeptidyl peptidase IV. *Infect. Immun.* **68**, 1176–1182 (2000).
34. Olivares, M. et al. The DPP-4 inhibitor vildagliptin impacts the gut microbiota and prevents disruption of intestinal homeostasis induced by a Western diet in mice. *Diabetologia* **61**, 1838–1848 (2018).
35. Leiting, B. et al. Catalytic properties and inhibition of proline-specific dipeptidyl peptidases II, IV and VII. *Biochem. J.* **371**, 525–532 (2003).
36. Cartmell, A. et al. A surface endogalactanase in *Bacteroides thetaiotaomicron* confers keystone status for arabinogalactan degradation. *Nat. Microbiol.* **3**, 1314–1326 (2018).
37. Wilson, M. M., Anderson, D. E. & Bernstein, H. D. Analysis of the outer membrane proteome and secretome of *Bacteroides fragilis* reveals a multiplicity of secretion mechanisms. *PLoS ONE* **10**, e0117732 (2015).
38. Liu, H. et al. Functional genetics of human gut commensal *Bacteroides thetaiotaomicron* reveals metabolic requirements for growth across environments. *Cell Rep.* **34**, 108789 (2021).
39. Domingues, M. M. et al. Biophysical characterization of polymyxin b interaction with LPS aggregates and membrane model systems. *Biopolymers* **98**, 338–344 (2012).
40. Wang, L., Li, P., Tang, Z., Yan, X. & Feng, B. Structural modulation of the gut microbiota and the relationship with body weight: compared evaluation of liraglutide and saxagliptin treatment. *Sci. Rep.* **6**, 33251 (2016).
41. Aranda-Díaz, A. et al. Establishment and characterization of stable, diverse, fecal-derived in vitro microbial communities that model the intestinal microbiota. *Cell Host Microbe* **30**, 260–272 (2022).
42. Aranda-Díaz, A. et al. Assembly of gut-derived bacterial communities follows 'early-bird' resource utilization dynamics. Preprint at *bioRxiv* https://doi.org/10.1101/2023.01.13.523996 (2023).
43. Cullen, T. W. et al. Antimicrobial peptide resistance mediates resilience of prominent gut commensals during inflammation. *Science* **347**, 170–175 (2015).
44. Miller, S. I. Antibiotic resistance and regulation of the Gram-negative bacterial outer membrane barrier by host innate immune molecules. *mBio* **7**, e01541-16 (2016).
45. Sampson, B. A., Misra, R. & Benson, S. A. Identification and characterization of a new gene of *Escherichia coli* K-12 involved in outer membrane permeability. *Genetics* **122**, 491–501 (1989).
46. Vallet, S.-U. et al. Loss of bacterial cell pole stabilization in *Caulobacter crescentus* sensitizes to outer membrane stress and peptidoglycan-directed antibiotics. *mBio* **11**, e00538-20 (2020).
47. Jacobson, A. N., Choudhury, B. P. & Fischbach, M. A. The biosynthesis of lipooligosaccharide from *Bacteroides thetaiotaomicron*. *mBio* **9**, e02289-17 (2018).
48. Fang, M., Wang, D., Coresh, J. & Selvin, E. Trends in diabetes treatment and control in U.S. adults, 1999–2018. *N. Engl. J. Med.* **384**, 2219–2228 (2021).
49. Waumans, Y., Baerts, L., Kehoe, K., Lambeir, A.-M. & De Meester, I. The dipeptidyl peptidase family, prolyl oligopeptidase, and prolyl carboxypeptidase in the immune system and inflammatory disease, including atherosclerosis. *Front. Immunol.* **6**, 387 (2015).
50. Rea, D. et al. Crystal structure of *Porphyromonas gingivalis* dipeptidyl peptidase 4 and structure-activity relationships based on inhibitor profiling. *Eur. J. Med. Chem.* **139**, 482–491 (2017).
51. Nabeno, M. et al. A comparative study of the binding modes of recently launched dipeptidyl peptidase IV inhibitors in the active site. *Biochem. Biophys. Res. Commun.* **434**, 191–196 (2013).
52. Boulton, D. W. Clinical pharmacokinetics and pharmacodynamics of saxagliptin, a dipeptidyl peptidase-4 inhibitor. *Clin. Pharmacokinet.* **56**, 11–24 (2017).
53. Vincent, S. H. et al. Metabolism and excretion of the dipeptidyl peptidase 4 inhibitor [^{14}C]sitagliptin in humans. *Drug Metab. Dispos.* **35**, 533–538 (2007).
54. Nemoto, T. K. & Ohara-Nemoto, Y. Exopeptidases and gingipains in *Porphyromonas gingivalis* as prerequisites for its amino acid metabolism.*Jpn Dent. Sci. Rev.* **52**, 22–29 (2016).
55. Oda, H., Saiki, K., Tonosaki, M., Yajima, A. & Konishi, K. Participation of the secreted dipeptidyl and tripeptidyl aminopeptidases in asaccharolytic growth of *Porphyromonas gingivalis*. *J. Periodontal. Res.* **44**, 362–367 (2009).
56. Olivares, M. et al. The potential role of the dipeptidyl peptidase-4-like activity from the gut microbiota on the host health. *Front. Microbiol.* **9**, 1900 (2018).
57. Thuy-Boun, P. S. et al. Quantitative metaproteomics and activity-based protein profiling of patient fecal microbiome identifies host and microbial serine-type endopeptidase activity associated with ulcerative colitis. *Mol. Cell. Proteom.* **21**, 100197 (2022).
58. McFadden, D. W., Rudnicki, M., Nussbaum, M. S., Balasubramaniam, A. & Fischer, J. E. Independent release of peptide YY (PYY) into the circulation and ileal lumen of the awake dog. *J. Surg. Res.* **46**, 380–385 (1989).
59. Liu, C. D., Newton, T. R., Zinner, M. J., Ashley, S. W. & McFadden, D. W. Intraluminal peptide YY induces colonic absorption *in vivo*. *Dis. Colon Rectum* **40**, 478–482 (1997).
60. Stevens, L. J. et al. A higher throughput and physiologically relevant two-compartmental human ex vivo intestinal tissue system for studying gastrointestinal processes. *Eur. J. Pharm. Sci.* **137**, 104989 (2019).

Methods

Chemicals

BHI medium was purchased from Millipore and BD, and mGAM was purchased from HyServe. GP-AMC was purchased from Santa Cruz Biotechnology. Z-GP-AMC was purchased from Bachem. Recombinant hDPP4 was purchased from Sino Biological. Talabostat mesylate and sitagliptin were purchased from AChemBlock. Saxagliptin and linagliptin were purchased from Carbosynth. Fmoc-protected amino acids (Fmoc-Ala-OH, Fmoc-Asp(tBu)-OH, Fmoc-Glu(tBu)-OH, Fmoc-Phe-OH, Fmoc-Gly-OH, Fmoc-His(Trt)-OH, Fmoc-Ile-OH, Fmoc-Lys(Boc)-OH, Fmoc-Leu-OH, Fmoc-Nle-OH, Fmoc-Asn(Trt)-OH, Fmoc-Pro-OH, Fmoc-Gln(Trt)-OH, Fmoc-Arg(Pbf)-OH, Fmoc-Ser(But)-OH, Fmoc-Thr(But)-OH, Fmoc-Val-OH, Fmoc-Trp(Boc)-OH and Fmoc-Tyr(tBu)-OH) were purchased from Advanced ChemTech, Chem-Impex, AAPPTec and Novabiochem. Vancomycin hydrochloride from *Streptomyces orientalis* was purchased from Sigma-Aldrich. Deoxycholic acid sodium salt was purchased from Fluka. Polymyxin B (sulfate) was purchased from Cayman Chemical. BODIPY FL vancomycin was purchased from Thermo Fisher Scientific. Propidium iodide was purchased from Sigma-Aldrich. Mounting media with DAPI was purchased from Abcam. FP-tetramethylrhodamine (TMR) and FP-biotin have been previously reported[26]. Synthetic methods for solid-phase peptide synthesis of fluorogenic peptides and positional scanning libraries and FP-alkyne (**1**) are included in Supplementary Information.

Bacterial strains and growth conditions

All bacterial strains and plasmids used in this study are listed in Supplementary Table 8. For MS-based ABPP and cloning of genomic DNA, *B. thetaiotaomicron* VPI-5482 was used, and for the remainder of the biological characterization assays, WT *B. thetaiotaomicron* refers to the parent strain *B. thetaiotaomicron* Δ*tdk*. All *B. thetaiotaomicron* strains were grown anaerobically (90% N_2, 5% CO_2 and 5% H_2) at 37 °C in BHI medium, Varel–Bryant minimal medium supplemented with 0.5% glucose (VBG), or modified mGAM. BHI was supplemented with hemin and L-cysteine (BHIS for bacterial genetics and mBHIS for microbiology) or hemin and vitamin K1 (BHIK for bacterial communities). Exact media recipes are detailed in Supplementary Table 9. All *E. coli* strains were grown aerobically at 37 °C in Luria Broth (LB) supplemented with 100 μg ml^{-1} carbenicillin as necessary for *B. thetaiotaomicron* mutagenesis and LB for protein expression.

Clean deletions were generated in *B. thetaiotaomicron* in BHIS, as previously described[61]. Briefly, suicide plasmids containing the flanking regions of *BT3254* and *BT4193* were amplified and assembled into pExchange-*tdk* using a Gibson Assembly Cloning Kit (New England Biolabs). Plasmid inserts were verified by Sanger sequencing, and the resulting plasmids (pRBC9 and pRBC8, respectively) were propagated in *E. coli* DH5α λ*pir*.

To generate *B. thetaiotaomicron* Δ*BT4193* and *B. thetaiotaomicron* Δ*BT3254*, the suicide plasmids pRBC8 and pRBC9, respectively, were conjugated into *B. thetaiotaomicron Δtdk* using S17-1 λ*pir* as the conjugative donor strain. Exconjugants with chromosomally integrated plasmids were recovered on BHIS plates containing 200 μg ml^{-1} gentamycin and 25 μg ml^{-1} erythromycin. Second crossover events were selected using FudR plates (BHIS plates supplemented with 200 μg ml^{-1} 5-fluoro-2-deoxy-uridine). Deletion of the target gene was confirmed by PCR. *B. thetaiotaomicron Δtdk* Δ*BT4193* Δ*BT3254* was generated by conjugating pRBC8 into *B. thetaiotaomicron* Δ*tdk ΔBT3254* and following the selection steps listed above.

Complemented strains were generated by inserting either the WT or catalytically inactive mutant of BT4193 into the intergenic spaces between BT2113-BT2114 and BT3743-BT3744 with the upstream WT promoter region. The plasmids pKI_2 and pKI_1 are the pExchange-*tdk* plasmid containing either the flanking regions for BT2113-BT2114 and BT3743-BT3744, respectively. Plasmids were cut with BamHI, and the amplified $BT4193^{WT}$ or $BT4193^{S606A}$ genes were inserted. For the $BT4193^{WT}$ complementation plasmid (pRBC13), the WT *B. thetaiotaomicron* genome was used as a template for the promoter and coding regions of BT4193. For the $BT4193^{S606A}$ complementation plasmid (pRBC18), the pET28a-BT4193_S606A-His plasmid served as a template for the catalytically inactive *BT4193* gene, whereas the WT genome served as a template for the promoter region. S17-1 λ*pir* harboring complementation plasmids served as the conjugative donor strain. Double crossovers were created as previously described and confirmed via PCR.

Bioinformatic analysis of serine hydrolases

To predict serine hydrolases, a list of serine hydrolase-associated Pfam domains was generated by compiling the Pfam domains associated with each family of serine peptidases in the MEROPS database (https://www.ebi.ac.uk/merops/), each family of alpha/beta-hydrolases in the ESTHER database (https://bioweb.supagro.inrae.fr/ESTHER/general?what=index) and FP-reactive proteins identified by ABPP in humans[22], *Arabidopsis thaliana*[23], *M. tuberculosis*[13], *Staphylococcus aureus*[25], *Staphylococcus epidermidis*[26], *Vibrio cholerae*[24], *Sulfolobus solfataricus* and *Haloferax volcanii*[27]. Each proteome of interest was downloaded from UniProt as a fasta file in March 2022. Proteomes were annotated with Pfam domains using pfam_scan.pl (http://ftp.ebi.ac.uk/pub/databases/Pfam/, November 2021). From each annotated proteome, proteins annotated with serine hydrolase-associated Pfam domains with an *E* value < 0.0001 were considered serine hydrolases. Identified serine hydrolases were then tallied by the Pfam domain, and the heatmap of the number of each type of serine hydrolase was generated using the *pheatmap* package with clustering in R. Predicted serine hydrolases from *B. thetaiotaomicron* were compared with its carbohydrate-active enzymes annotated on the CAZY database (http://www.cazy.org/) using the eulerr package in R.

To assess the conservation of active serine hydrolases identified in *B. thetaiotaomicron* throughout the gut microbiota, the pep.faa file for reference organisms isolated from the gastrointestinal tract as part of the National Institutes of Health (NIH) Human Microbiome Project (https://www.hmpdacc.org/hmp/HMRGD/, retrieved December 2019) was downloaded. This faa file was converted into a BLAST database using the makeblastdb application from National Center for Biotechnology Information BLAST+ v2.7.1. A protein fasta file of *B. thetaiotaomicron* active serine hydrolases was queried against this database using the blastp application with a minimum *E* value of 10^{-30}. These homologs were then sorted by the bacterial strain of origin and query protein sequence, and $-\log_{10}E$ was calculated for each homolog. For representation purposes, all BLAST-P *E* values of 0 were set equal to 10^{-200}. The heatmap of $-\log_{10}E$ was generated using the pheatmap package without clustering in R. This approach was also used to assess the conservation of hDPP4 in the gut microbiota.

To compare human DPP4 to putative *B. thetaiotaomicron* homologs BT4193 and BT3254 and other aminopeptidases, protein sequences were downloaded from UniProt and aligned using the web-based Clustal Omega tool (https://www.ebi.ac.uk/Tools/msa/clustalo/). Phylogenetic trees from multiple sequence alignments were visualized using FigTree v1.4.4. Predicted structures of BT4193 and BT3254 were generated using the web-based AlphaFold[62] Colab tool (https://colab.research.google.com/github/deepmind/alphafold/blob/main/notebooks/AlphaFold.ipynb) and aligned to the crystal structure of hDPP4 (PDB 1J2E) using the MatchMaker function in UCSF Chimera v.1.13.1. Protein domains of hDPP4, BT4193 and BT3254 were predicted using the web-based InterPro tool (https://www.ebi.ac.uk/interpro/), and subcellular localization of the FP-reactive serine hydrolases in *B. thetaiotaomicron* was predicted with the web-based tools SignalP 5.0 (https://services.healthtech.dtu.dk/service.php?SignalP-5.0), CELLO[63] and PSORTb (https://www.psort.org/psortb/index.html).

To identify stressors that cause fitness defects with the disruption of *BT4193* or *BT3254*, fitness scores for both genes were extracted from a chemical genetics dataset screening fitness of a randomly barcoded

transposon insertion *B. thetaiotaomicron* mutant library in >500 conditions[38]. *BT4193* or *BT3254* fitness scores 3 s.d. away from the mean were determined to be significant.

MS-based ABPP sample preparation and analysis

B. thetaiotaomicron VPI-5482 cultures were grown in 30 ml of mBHI in triplicate for 72 h before being aliquoted into three separate tubes and pelleted. For each biological replicate, the pellet was resuspended in 200 µl of 25 µM FP-TMR in PBS or 0.05% DMSO in PBS and incubated anaerobically for 30 min at 37 °C. Then, 2 µl of either 500 µM FP-biotin (final concentration, 5 µM) or DMSO was added to each sample and incubated anaerobically for 30 min at 37 °C. Bacterial pellets were stored at −20 °C before sample preparation.

Treated *B. thetaiotaomicron* cell pellets were thawed, diluted in 200 µl of PBS and lysed by probe sonication. The proteomes were denatured and precipitated using 4:1 MeOH/$CHCl_3$, resuspended in 0.5 ml of 6 M urea in PBS, reduced using tris(2-carboxyethyl)phosphine (10 mM) for 30 min at 37 °C and then alkylated using iodoacetamide (40 mM) for 30 min at room temperature in the dark. The biotinylated proteins were enriched with PBS-washed Pierce Streptavidin Agarose beads (100 µl per sample) by rotating at room temperature for 1.5 h in PBS with 6 ml of 0.2% SDS. The beads were then washed sequentially with 5 ml of 0.2% SDS in PBS (3×), 5 ml PBS (3×) and 5 ml H_2O (3×). On-bead digestion was performed using 2 µg of sequencing-grade trypsin (Promega) in 200 µl of 2 M urea in 100 mM triethylammonium bicarbonate (TEAB) buffer with 2 mM $CaCl_2$ for 12–14 h at 37 °C.

Tryptic digests were desalted using C18 solid-phase extraction (SPE). C18 SPE plates (Thermo Fisher Scientific) were conditioned by the addition of acetonitrile (ACN) to each well and plates were centrifuged at 750*g* for 2 min. This process was repeated with 200 µl of buffer A (0.1% formic acid in water). Samples were diluted with 560 µl of buffer A and triturated with a pipette to mix. A portion (560 µl) of each sample was added to separate wells of the SPE plate, the sample was allowed to slowly load over 5 min and then the plate was centrifuged for 5 min at 200*g*. The remaining sample volume was loaded into the same well of the SPE plate, and the loading step was repeated. Samples were washed twice with buffer A (200 µl, then 100 µl) followed by centrifugation of the SPE plate at 200*g* for 5 min after each wash. Samples were eluted into a clean 96-well plate by the addition of 60 µl of buffer B (70% ACN and 0.1% formic acid in water) and centrifugation at 200*g* for 5 min. Samples were dried by centrifugal vacuum concentrator (50 °C, overnight) and resolubilized in 100 µl of 100 mM TEAB, pH 8.5 containing 30% ACN. Peptides were labeled with the addition of tandem mass tags (TMT) 3 µl per channel, 19.5 µg µl^{-1}) to each sample and incubated at room temperature for 30 min. Hydroxylamine (8 µl, 5% in water) was subsequently added to quench the labeling reaction (15 min), and the TMT-labeled samples were mixed and desalted on C18 using the protocol described above.

Nanoflow LC–MS/MS measurements were performed on an Ultimate 3000 (Thermo Fisher Scientific) interfaced with an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific) via an EASY-Spray source (Thermo Fisher Scientific). Peptides were separated on an EASY-Spray PepMap RSLC C18 column (2 µm particle size, 75 µm × 50 cm; Thermo Fisher Scientific, ES801) heated to 55 °C using a flow rate of 400 nl min^{-1}. The compositions of LC solvents were as follows: solvent A (water and 0.1% formic acid) and solvent B (95% ACN, 5% water and 0.1% formic acid). Peptides were eluted over 4 h using the linear gradient, 2.5–215 min (3–35% solvent B), 215–230 min (25–40% solvent B), 230–231 min (45–70% solvent B), 231–233 min (70–90% solvent B), 233–234 min (5–70% solvent B), 234–236 min (70–90% solvent B) and 236–240 min (3% solvent B).

MS data were acquired using an MS3 data-dependent acquisition method. MS1 profile scans were acquired in the Orbitrap (resolution, 120,000; scan range, 375–1,500 *m/z*; AGC target, 4×10^5 and maximum injection time, 50 ms). Monoisotopic peak determination was set to 'peptide'. Only charge states 2–5 were included. Dynamic exclusion was enabled (repeat count, $n = 1$; exclusion duration, 60 s; mass tolerance: low, 10 ppm and high, 10 ppm, excluding isotopes). An intensity threshold of 5×10^3 was set. Data-dependent MS2 spectra were acquired in centroid mode across a mass range of 400–1,200 *m/z*. Precursor ions were isolated using the quadrupole (isolation window, 0.7 *m/z*), fragmented using collision-induced dissociation (CID) (collision energy, 35%; activation time, 10 ms and activation Q, 0.25) and detected in the ion trap (scan range mode, auto *m/z* normal; scan rate, rapid; AGC target, 1×10^4 and maximum injection time, 50 ms). Data-dependent MS3 spectra were acquired in the Orbitrap (resolution, 50,000; scan range, 100–500 *m/z*; AGC target, 1×10^5 and maximum injection time, 105 ms) following higher-energy collisional dissociation (HCD) activation (collision energy, 65%) using Synchronous Precursor Selection from up to 10 precursors.

Peptide and protein identifications were performed using MaxQuant v.1.6.0.16 (ref. 64) using the search engine Andromeda[65]. Group-specific parameters were set to 'Reporter ion MS3' with 10plex TMT isobaric labels for N-terminal and lysine residue modification selected. Reporter mass tolerance was set to 0.003 Da. The following parameters were used: carbamidomethylation of cysteines as fixed modifications, oxidation of methionine and acetylation of N-terminus as dynamic modifications and trypsin/P as the proteolytic enzyme. Default settings were used for all other parameters. Searches were performed against the UniProt database for *B. thetaiotaomicron* VPI-5482 (Proteome ID: UP000001414, downloaded April 2019) with the following parameters: minimum peptide length, 6; maximum peptide mass, 6,000 Da; minimum peptide length for unspecific search, 6 and maximum peptide length for unspecific search, 40. Identification was performed with at least two unique peptides and quantification only with unique peptides.

Statistical analyses were performed with Perseus v.1.6.0.7 (ref. 66). Putative contaminants, reverse hits and proteins identified by side only were removed. Label-free quantitation intensities were $\log_2$-transformed. Missing values were imputed using a normal distribution (width, 0.3; down-shift, 1.8). *P* values were calculated using a two-sided, two-sample *t*-test. Proteins considered to be enriched serine hydrolases had a minimum fold-change of 3 ($\log_2$ fold-change of 1.58) and minimum *P* value of 0.05 ($-\log_{10}$ of 1.30).

Cloning, expression and purification of recombinant BT4193 and BT3254

All enzymes were purchased from New England BioLabs. A single colony of *B. thetaiotaomicron* VPI-5482 was boiled in 10 µl of water at 95 °C for 5 min to isolate genomic DNA. The genomic sequences for *BT4193* and *BT3254* were amplified and then cloned into the pET28a plasmid with a C-terminal His_6 tag using Gibson assembly following the manufacturer's instructions (for primer sequences, see Supplementary Table 8). For the creation of the catalytically inactive *BT4193* plasmid (pET28a-BT4193^{S606A}-His) used in the generation of the *BT4193*^{S606A} complementation strain, site-directed mutagenesis was performed using the Q5 Site-Directed Mutagenesis Kit (for primer sequences, see Supplementary Table 8). Successful cloning was confirmed by selection on 50 µg ml^{-1} kanamycin followed by Sanger sequencing. The plasmids pET28a-BT4193-His and pET28a-BT3254-His (Supplementary Table 8) were transformed into chemically competent *E. coli* Rosetta 2(DE3) cells with selection using 50 µg ml^{-1} kanamycin and 34 µg ml^{-1} chloramphenicol for expression.

For recombinant BT4193, an overnight culture of transformed cells was inoculated into 1 l of LB supplemented with 50 µg ml^{-1} kanamycin and 34 µg ml^{-1} chloramphenicol and grown with shaking until the optical density at 600 nm (OD_{600}) = 0.6. After the culture was cold-shocked on ice for 20 min, 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to the culture to induce protein expression, and the culture was incubated for 20 h at 19 °C. The bacterial pellet was flash-frozen in

liquid nitrogen (LN_2) and stored at −80 °C. The pellet was lysed in 30 ml of lysis buffer (20 mM *N*-2-hydroxyethylpiperazine-*N*-2-ethane sulfonic acid (HEPES), 10 mM imidazole, 150 mM NaCl, pH 8) supplemented with 750 mM trehalose by probe sonication on ice (3 min at 30% power in 1 s bursts; 1 min at 60% power in 1 s bursts; 3 min at 30% power in 1 s bursts and 1 min at 60% power in 1 s bursts). Cell debris was pelleted at 38,400*g* for 20 min at 4 °C, and the lysate was transferred to a new tube and centrifuged at 12,000*g* for 15 min at 4 °C. The lysate was then clarified sequentially through a 5-µm and then a 1-µm filter before purification. The clarified lysate was injected into an ÄKTApurifier FPLC system (GE Healthcare) and separated with two in-parallel 1-ml HisTrap FF crude columns (GE Healthcare). The column was washed with lysis buffer supplemented with 10 mM imidazole, then 20 mM imidazole and then 40 mM imidazole. The protein was eluted with lysis buffer supplemented with 300 mM imidazole, and fractions containing protein (as determined by the ultraviolet (UV) trace and SDS–PAGE) were combined. The protein was further purified using a HiPrep 16/60 Sephacryl S-300 HR column (Cytiva Life Sciences) using lysis buffer, and fractions containing protein (as determined by the UV trace and SDS–PAGE) were combined.

For recombinant BT3254, an overnight culture of transformed cells was inoculated into 1 l of LB medium supplemented with 1% glucose (to suppress leaky expression), 50 µg ml^{-1} kanamycin and 34 µg ml^{-1} chloramphenicol and grown with shaking until OD_{600} reached 0.8. After the culture was cold-shocked on ice for 20 min, 0.5 mM IPTG was added and the culture was incubated for 20 h at 19 °C. The bacterial pellet was flash-frozen in LN_2 and stored at −80 °C until purification. The pellet was lysed in lysis buffer (20 mM HEPES, 10 mM imidazole, 150 mM NaCl, pH 8) supplemented with 750 mM trehalose by probe sonication on ice (3 min at 30% power in 1 s bursts, 1 min at 60% power in 1 s bursts, 3 min at 30% power in 1 s bursts and 1 min at 60% power in 1 s bursts). Cell debris was pelleted at 15,000*g* for 30 min at 4 °C, but overexpressed protein remained in the insoluble fraction, as determined by SDS–PAGE. To purify the unfolded protein, the pellet was solubilized in denaturing buffer (100 mM Na_2PO_4, 10 mM Tris, 8 M urea, pH 8) for 1 h at 25 °C with rotating and then centrifuged at 15,000*g* for 20 min at 25 °C. The lysate was incubated with 1 ml of nickel-nitrilotriacetic acid (Ni-NTA) agarose (Qiagen) for 1 h at 25 °C with rotation. The resin was washed three times in denaturing buffer at pH 6.3, and protein was eluted three times in denaturing buffer at pH 5.9 and then three times in denaturing buffer at pH 4.5. Elution fractions were combined and pH-adjusted to pH 7.5 before dialysis into 20 mM HEPES, 150 mM NaCl, 2 mM reduced glutathione, 200 µM oxidized glutathione, pH 8 for 72 h, with the dialysis buffer being replaced once.

Enzyme kinetics

For each assay, 20 µl of lysate or recombinant enzyme were combined with 5 µl of 50 µM fluorogenic peptide substrate (final concentration, 10 µM, 0.2% DMSO) in a 384-well plate. Fluorescence (ex/em, 380/460 nm for AMC substrates; ex/em, 355/460 nm for ACC substrates) was measured every minute for 30 min at 37 °C on a Cytation 3 plate reader (BioTek). Initial velocities were quantified by calculating the slope of the linear phase of the reaction as relative fluorescent units per min. All assays were performed with three biological replicates over multiple experiments.

To measure in vitro enzyme activity, *B. thetaiotaomicron* WT and Δ*BT4193*, Δ*BT3254*, Δ*BT4193* Δ*BT3254*, Δ*BT4193*::*BT4193*^{WT} and Δ*BT4193*::*BT4193*^{S606A} strains were diluted into mBHIS and grown to late exponential phase in three parallel cultures. Bacteria were pelleted by centrifugation and lysed in 100 mM HEPES (pH 7.5) by probe sonication at 15% power (3 × 20 s, with 2 min on ice in between). Cell debris was pelleted by centrifugation, and lysate concentration was normalized to 0.25 µg $µl^{-1}$ with 100 mM HEPES, pH 7.5 (final concentration, 5 µg per well). Each replicate was measured in technical triplicate, and initial velocities were averaged. To measure the activity of recombinant hDPP4 (Sino Biological), BT4193 and BT3254, enzyme was diluted to 1.25 nM in 100 mM HEPES, pH 7.5 (final concentration, 1 nM).

For calculation of half maximal inhibitory concentration (IC_{50}) values, recombinant enzyme was diluted to 1.25 nM in 20 mM HEPES (pH 7.5) and incubated with inhibitor, diluted 1:100 from DMSO stocks, for 30 min at 37 °C before monitoring activity. Initial velocities were normalized to DMSO-pretreated enzyme (100%) or buffer only (0%).

Competitive gel-based ABPP

B. thetaiotaomicron WT and Δ*BT4193*, Δ*BT3254* and Δ*BT4193* Δ*BT3254* knockout strains were diluted into mBHIS and grown to a late exponential phase. For each sample, 4 µl of DMSO or 100× stocks of covalent hDPP4 inhibitor (10 µM, 100 µM or 1 mM stocks of talabostat or saxagliptin in DMSO) were added to 400 µl of bacterial culture and incubated anaerobically for 1 h at 37 °C. To each sample, 4 µl of 500 µM FP-alkyne (final concentration, 5 µM) were added and incubated anaerobically for 30 min at 37 °C. Bacterial pellets were pelleted by centrifugation and lysed in PBS by probe sonication at 15% power (3 × 20 s, with 2 min on ice in between). Cell lysate was clarified by centrifugation, and lysate concentration was normalized. TMR–azide was conjugated to FP-modified enzymes using Cu(I)-catalyzed azide–alkyne cycloaddition (5 mM BTTAA, 1 mM $CuSO_4$, 15 mM sodium ascorbate and 20 mM azide–TMR for 30 min at 37 °C), and labeled enzymes were separated via SDS–PAGE. In-gel fluorescence was visualized using the Cy3 channel on Typhoon 9410 Imager (Amersham Biosciences).

MSP–MS

Recombinant hDPP4, BT4193 and BT3254 (4 nM, 4 nM and 10 nM, respectively) were incubated with a mixture of 228 14-mer peptides (0.5 µM for each peptide) in 20 mM HEPES, pH 7.5 at room temperature. The reaction was quenched with 8 M urea at 15 min, 1 h and 4 h for hDPP4 and BT4193 and 4 h only for BT3254. A control assay consisted of each enzyme mixed with 8 M urea before the addition of the peptide library. Assays were conducted in quadruplicate. Samples were then acidified by the addition of 40 µl of 2% trifluoroacetic acid (TFA), desalted using spin tips consisting of octadecyl C18 (CDS Analytical), evaporated to dryness in a vacuum centrifuge and placed at −80 °C. Samples were resuspended in 40 µl of 0.1% TFA, and 4 µl of each sample were used for LC–MS/MS.

MSP–MS assay samples were subjected to LC–MS/MS performed on a Q-Exactive Mass Spectrometer (Thermo Fisher Scientific) equipped with an Ultimate 3000 high-performance liquid chromatography (HPLC) (Thermo Fisher Scientific). Peptides were separated by reverse phase chromatography on a C18 column (1.7 µm bead size, 75 µm × 20 cm, 65 °C) at a flow rate of 300 nl min^{-1} using solvent A (0.1% FA in water) and solvent B (0.1% FA in ACN). LC separation was performed using a 65-min linear gradient of 5–40% solvent B followed by a 5-min linear gradient of 40–85% solvent B and staying at 85% solvent B for 10 min. The flow rate was increased to 400 nl min^{-1}, and solvent B was increased to 90% for 10 min. Survey scans were recorded over a 200–2,000 *m/z* range (70,000 resolutions at 200 *m/z*, AGC target 1×10^6, 75 ms maximum). MS/MS was performed in data-dependent acquisition mode with HCD fragmentation (30 normalized collision energy) on the 10 most intense precursor ions (17,500 resolutions at 200 *m/z*, AGC target 5×10^4, 120 ms maximum, dynamic exclusion 15 s).

MS/MS data analysis was performed using PEAKS v.8.5 (Bioinformatics Solutions). MS2 data were searched with the 228-member 14-mer library sequence as a database, and a decoy search was conducted with peptide amino acid sequences in reverse order. A precursor tolerance of 20 ppm and 0.01 Da for MS2 fragments was defined. No protease digestion was specified. Data were filtered to 1% peptide- and protein-level false discovery rates with the target-decoy strategy. Peptides were quantified with label-free quantification, and data were normalized by TIC. Outliers from replicates were removed by Dixon's *Q* test. Missing and zero values were imputed with normally distributed

random numbers under the limit of detection, as determined by the lowest 5% of peptide intensities. A one-way analysis of variance (ANOVA) was performed for peptide data of the 0 (control), 15 min, 1 h and 4 h incubation conditions for hDPP4 and BT4193, and a *t*-test was performed for 0 and 4 h incubation conditions for BT3254. Those with $P < 0.05$ were considered for further analysis. Criteria for identification of cleaved peptide products included those with intensity scores of eightfold or more above the quenched inactive enzymes (time, 0 min), evaluated by $\log_2$(active/inactivated enzyme ratio) for each peptide product with $P < 0.01$ by two-tailed *t*-test. Data were analyzed using the eulerr package in R and iceLogo software[67].

Growth analyses

B. thetaiotaomicron WT, Δ*BT4193*, Δ*BT3254*, Δ*BT4193* Δ*BT3254*, Δ*BT4193*::*BT4193*^{WT} and Δ*BT4193*::*BT4193*^{S606A} strains were grown in mBHIS until the late exponential phase and washed via centrifugation with VBG. Cultures were normalized to an OD_{600} of 1 and diluted 1:100 in three to four wells into VBG, supplemented with vancomycin hydrochloride ($n = 3$), deoxycholic acid ($n = 3$), polymyxin B ($n = 3$), ethanol ($n = 2$) or sodium chloride ($n = 2$). To test the pH-dependence of growth, the pH of the VBG medium was adjusted using 10 N hydrochloric acid and sterile-filtered. Anaerobic growth was monitored by measuring absorbance (OD_{600}) every 10 min for 24 h at 37 °C using a Cytation 3 plate reader (BioTek) or every 3 min for 24 h using a Stratus plate reader (Cerillo). Quantification of the area under the growth curve was analyzed using the bayestestR package in R.

Fluorescent vancomycin imaging and analysis

B. thetaiotaomicron WT, Δ*BT4193*, Δ*BT3254*, Δ*BT4193* Δ*BT3254*, Δ*BT4193*::*BT4193*^{WT} and Δ*BT4193*::*BT4193*^{S606A} strains were diluted into VBG and grown for 5 h into the mid-exponential phase. BODIPY FL vancomycin was added to each culture to a final concentration of 8 μg ml^{-1}, and cultures were incubated anaerobically for 30 min at 37 °C. Bacteria were pelleted via centrifugation and diluted 2.5-fold into 20 mM HEPES, pH 7.5. For imaging, 1 μl of bacteria was added to a 1% agarose pad and allowed to dry before a coverslip was placed on the pad. Phase-contrast and fluorescence images were acquired with a Ti-E inverted epifluorescence microscope (Nikon Instruments) using a ×100 (numerical aperture 1.40) oil immersion objective and a Neo 5.5 sCMOS camera (Andor Technology) using μManager v1.4. To calculate the fluorescence intensity of each cell, the MATLAB image processing package Morphometrics[68] was used to segment cell contours in phase-contrast images. For ~1,000 contours in each condition, the mean background-subtracted fluorescence was calculated over all pixels inside the contour.

Propidium iodide imaging and analysis

B. thetaiotaomicron WT, Δ*BT4193*, Δ*BT3254*, Δ*BT4193* Δ*BT3254*, Δ*BT4193*::*BT4193*^{WT} and Δ*BT4193*::*BT4193*^{S606A} strains were diluted into VBG and grown for 5 h into the mid-exponential phase. Cultures were normalized to an OD_{600} of 1 and washed twice via centrifugation in 0.85% NaCl in water. Bacteria were incubated in the dark in 3 μM propidium iodide in 0.85% NaCl for 10 min at 25 °C. Bacteria were washed twice via centrifugation in 0.85% NaCl and incubated in 2% paraformaldehyde in water for 15 min at 25 °C. Bacteria were washed twice via centrifugation in water and diluted 1:1,000 before mounting (10 μl) on a slide with mounting media with DAPI.

Confocal microscopy images were acquired on a Zeiss LSM700 confocal microscope using a ×63 oil-emersion objective. DAPI was excited with the 405 nm laser line, and propidium iodide was excited with the 555 nm laser line. For each of the four independent experiments, laser power, pinhole size and gain were set at the beginning of the acquisition period and maintained throughout the session. For each image, regions were selected based on the DAPI stain, and the focal plane was chosen to maximize propidium iodide intensity. Confocal images were analyzed with ImageJ v1.51g and the macro FluoQ Macro V3-97 (ref. 69). Briefly, the macro first operated a background subtraction via the built-in subtraction function. Next, a median filter was applied to each image. To exclude low-value pixels, the images were then thresholded manually with the ImageJ built-in threshold function and the regions of interest were segmented automatically in the PI channel. After segmentation, the mean fluorescence intensity in the region of interest was calculated and used for further analysis.

Community assembly experiments and 16S rRNA gene sequencing analysis

To generate the 15-member synthetic communities, all bacterial isolates were struck on BHI blood agar plates (5% defibrinated horse blood in 1.5% (wt/vol) agar). The resulting colonies were inoculated into 1 ml of BHI, BHIK or mGAM and were grown in a 96-deep well plate at 37 °C for 48 h. Synthetic communities were assembled from stationary phase cultures of isolates mixed at equal OD_{600}, and 1 μl of the mixture was inoculated into 200 μl of BHI or mGAM growth medium. Stool-derived communities were generated by inoculating frozen stool samples from each of eight healthy donors in either 1 ml of BHI or 1 ml mGAM, and communities were serially diluted 1:200 for seven 48-h passages[42]. These stool-derived communities were inoculated from frozen glycerol stocks into 3 ml of the medium in which they were derived (BHI or mGAM) and serially diluted 1:200 for three 48-h passages to ensure that composition reached a steady state before measurements.

Experiments with both the synthetic and stool-derived communities were performed in clear, flat-bottomed 96-well plates (Greiner Bio-One) sealed with breathable AeraSeals (Excel Scientific). All reagents were prereduced by incubation in an anaerobic chamber for at least 24 h. For drug conditions, saxagliptin and sitagliptin were added at a final concentration of 10 μM, and 0.1% (vol/vol) DMSO was added as vehicle control. Sealed plates were placed in a loosely closed bag to minimize evaporation and incubated anaerobically at 37 °C without shaking for 48 h. Plates were then resealed with foil seals and stored at −80 °C for sequencing preparation. Amplicon sequencing data were obtained and processed as previously described[41]. Relative abundances were determined to a minimum threshold of 10^{-3}, reflecting the typical depth of sequencing.

Software

Graphs and statistical analyses were generated using RStudio, GraphPad Prism 9 and Matlab 2020a (Mathworks). Code is available upon request. Data are plotted as mean ± s.e.m. For the comparison of WT to knockout strains, one-way ANOVA tests were performed, with post hoc Dunnett's multiple comparisons tests with WT as the control where appropriate (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$ and ****$P < 0.0001$).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Selection of serine hydrolase-associated Pfam domains was performed with the MEROPS (https://www.ebi.ac.uk/merops/) and ESTHER (https://bioweb.supagro.inrae.fr/ESTHER/general?what=index) databases. Proteomes for bioinformatic prediction of serine hydrolases were downloaded from UniProt, and accession codes for each proteome are included in Supplementary Table 2. Proteomes were annotated with Pfam domains using pfam_scan.pl (http://ftp.ebi.ac.uk/pub/databases/Pfam/). Proteomes of reference isolates from the NIH Human Microbiome Project gastrointestinal tract were downloaded from https://www.hmpdacc.org/hmp/HMRGD/. Carbohydrate-active enzymes were identified using the CAZyme database (http://www.cazy.org/). Raw proteomics data for this study have been deposited to the ProteomeXchange Consortium via the PRIDE partner

repository with the dataset identifier PXD035963. Raw MSP–MS data can be obtained through massive.ucsd.edu under the dataset identifier numbers MSV000091339 (hDPP4), MSV000091338 (BT4193) and MSV000089969 (BT3254). Raw 16S rRNA data for community assembly experiments can be obtained at https://doi.org/10.25740/rn970zy4428. All other data are available in the source data provided with this paper.

Code availability

Code for bioinformatic analyses of serine hydrolases, analysis of fitness data, quantification of area under the growth curves and plotting of Venn diagrams can be obtained at https://doi.org/10.5281/zenodo.7835000. Code for analyzing community assembly experiments can be obtained at https://doi.org/10.5281/zenodo.7830074.

References

61. Zhu, W. et al. Xenosiderophore utilization promotes *Bacteroides thetaiotaomicron* resilience during colitis. *Cell Host Microbe* **27**, 376–388 (2020).
62. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
63. Yu, C.-S., Lin, C.-J. & Hwang, J.-K. Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. *Protein Sci.* **13**, 1402–1406 (2004).
64. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
65. Cox, J. et al. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10**, 1794–1805 (2011).
66. Tyanova, S. et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731–740 (2016).
67. Colaert, N., Helsens, K., Martens, L., Vandekerckhove, J. & Gevaert, K. Improved visualization of protein consensus sequences by iceLogo. *Nat. Methods* **6**, 786–787 (2009).
68. Ursell, T. et al. Rapid, precise quantification of bacterial cellular dimensions across a genomic-scale knockout library. *BMC Biol.* **15**, 17 (2017).
69. Stein, F., Kress, M., Reither, S., Piljić, A. & Schultz, C. FluoQ: a tool for rapid analysis of multiparameter fluorescence imaging data applied to oscillatory events. *ACS Chem. Biol.* **8**, 1862–1868 (2013).

Acknowledgements

We thank the laboratories of J. Sonnenburg and M. Howitt at Stanford University for the use of their equipment and W. Zhu at Vanderbilt University Medical Center for gifting the plasmids pKI_1 and pKI_2. L.J.K. was supported by the Stanford ChEM-H Chemistry/Biology Interface Predoctoral Training Program (T32 GM120007), a Stanford Molecular Pharmacology Training Grant (T32 GM113854), and a Stanford Graduate Fellowship. T.H.N. was supported by a National Science Foundation Graduate Research Fellowship (DGE-1656518). B.M.H. was supported by the UCSD Graduate Training Program in Cellular and Molecular Pharmacology through an institutional training grant from the National Institute of General Medical Sciences (T32 GM007752). M.L. was supported by Deutsche Forschungsgemeinschaft (DFG) for funding under the Walter Benjamin Program. R.C. was supported by the NIH Training (T32 HG000044) and a Stand Up 2 Cancer grant (to A.S.B.). F.F. was supported by a National Science Foundation Graduate Research Fellowship (DGE-1656518), a Stanford ChEM-H O'Leary-Thiry Graduate Fellowship, and Stanford's Enhancing Diversity in Graduate Education Doctoral Fellowship Program. This work was supported by the NIH (grants R01 EB026332 and R01 EB026285 (to M.B.), R01 DK131005 (to A.J.O.), R01 AI148623 and R01 AI143757 (to A.S.B.) and RM GM135102 and R01 AI147023 (to K.C.H.)) and National Science Foundation (grant EF-2125383 (to K.C.H.)). K.C.H. is a Chan Zuckerberg Biohub Investigator.

Author contributions

L.J.K. and M.B. conceived and designed the study. L.J.K. ran bioinformatic analyses, synthesized fluorogenic peptide substrates, purified recombinant enzymes and performed enzyme kinetics, gel-based ABPP and microbiology experiments. T.H.N. performed community assembly and fluorescent vancomycin microscopy experiments. L.J.L., B.M.H. and D.J.G. performed MSP–MS experiments. N.N., K.M.L. and M.J.N. performed MS-based ABPP experiments. R.C. performed *B. thetaiotaomicron* genetics. F.F. and P.I. synthesized FP-alkyne. P.I. assisted with bioinformatic analyses. L.J.K., T.H.N., L.J.L., B.M.H., M.L., M.G., D.J.G., P.I., A.J.O., K.C.H. and M.B. analyzed and interpreted data. L.J.K., T.H.N., K.C.H. and M.B. prepared the figures and wrote and edited the paper. All authors reviewed and revised the paper.

Competing interests

The authors declare no competing interests.

Additional information

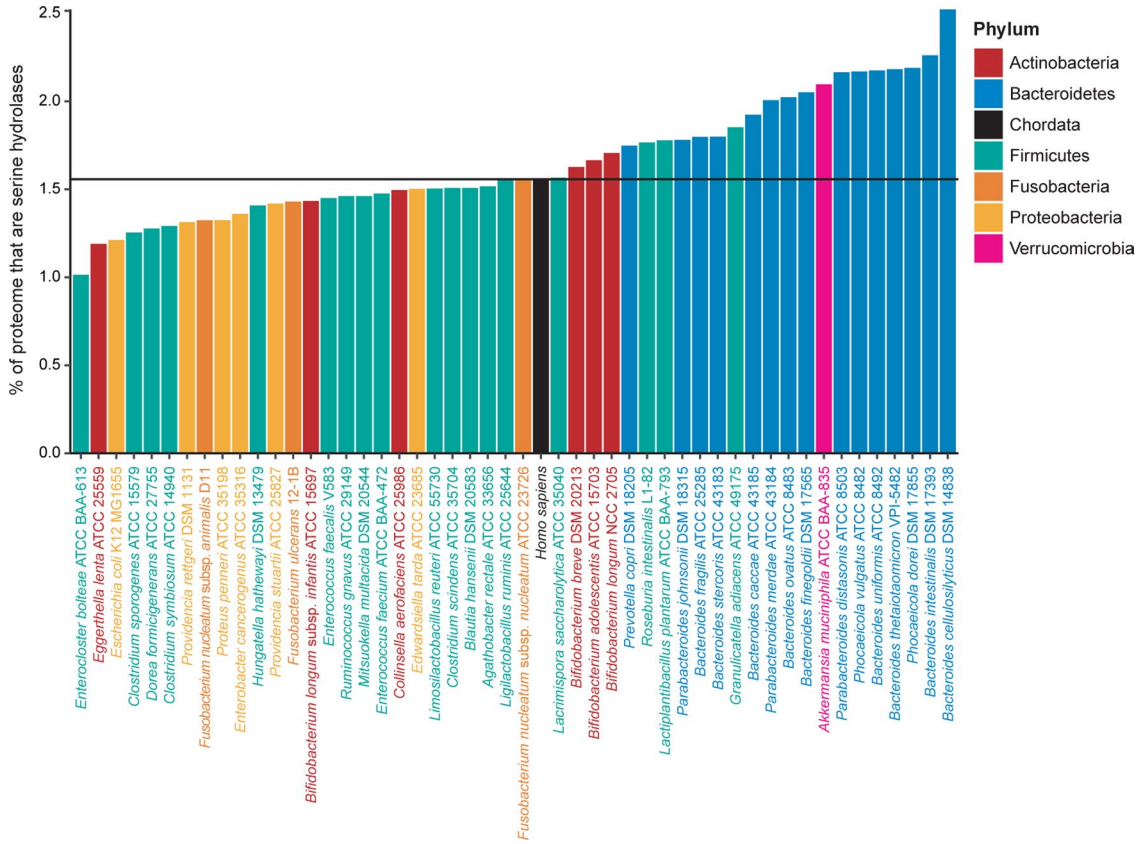
Extended data is available for this paper at https://doi.org/10.1038/s41589-023-01357-8.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41589-023-01357-8.

Correspondence and requests for materials should be addressed to Matthew Bogyo.

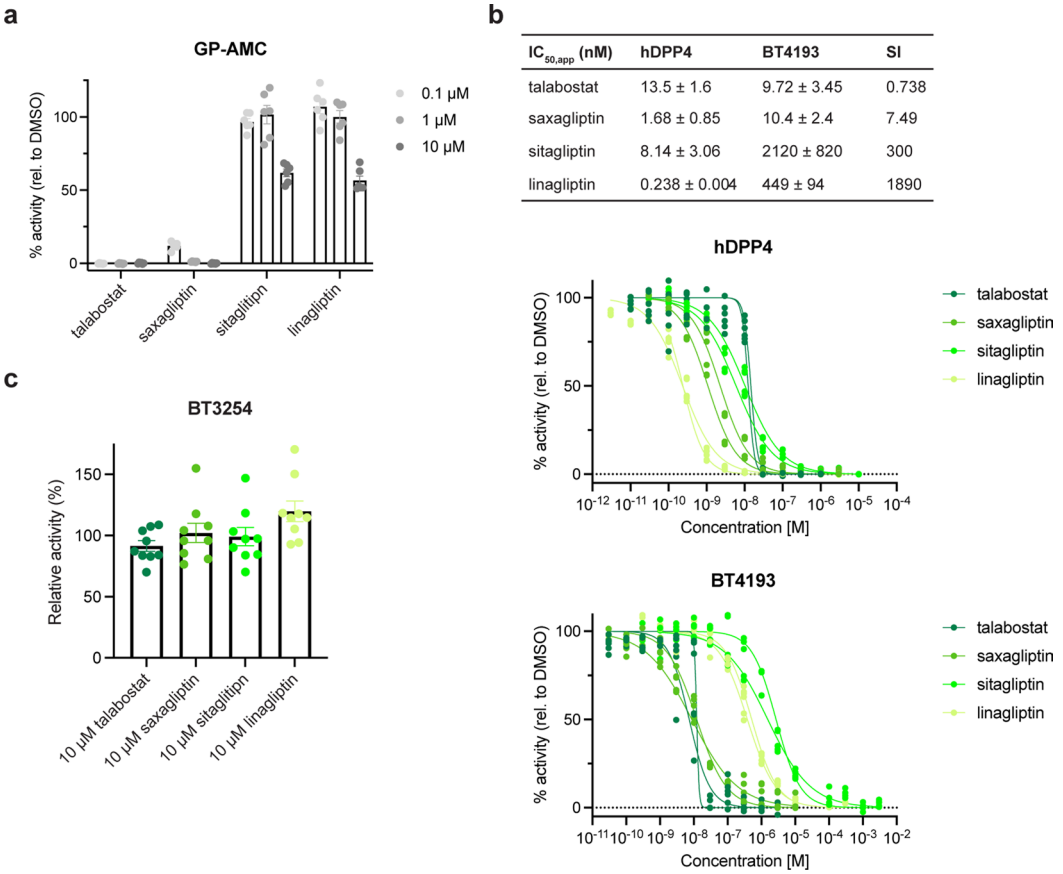
Peer review information *Nature Chemical Biology* thanks Michael Zimmermann and the other anonymous reviewers for their contribution to the peer review of this work.

https://doi.org/10.1038/s41589-023-01357-8



Extended Data Fig. 1 | Bacteria from the phylum Bacteroidetes have a larger fraction of their proteome predicted to be serine hydrolases. Percentage of the total proteome based on the number of proteins that are predicted serine hydrolases, colored by phylum. The horizontal line corresponds to the percentage of predicted human serine hydrolases as a reference.

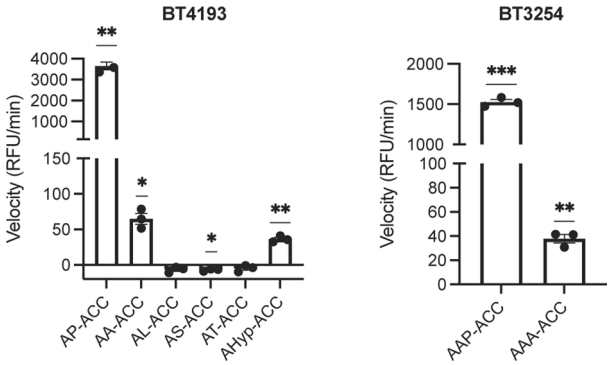
https://doi.org/10.1038/s41589-023-01357-8



$IC_{50,app}$ (nM)	hDPP4	BT4193	SI
talabostat	13.5 ± 1.6	9.72 ± 3.45	0.738
saxagliptin	1.68 ± 0.85	10.4 ± 2.4	7.49
sitagliptin	8.14 ± 3.06	2120 ± 820	300
linagliptin	0.238 ± 0.004	449 ± 94	1890

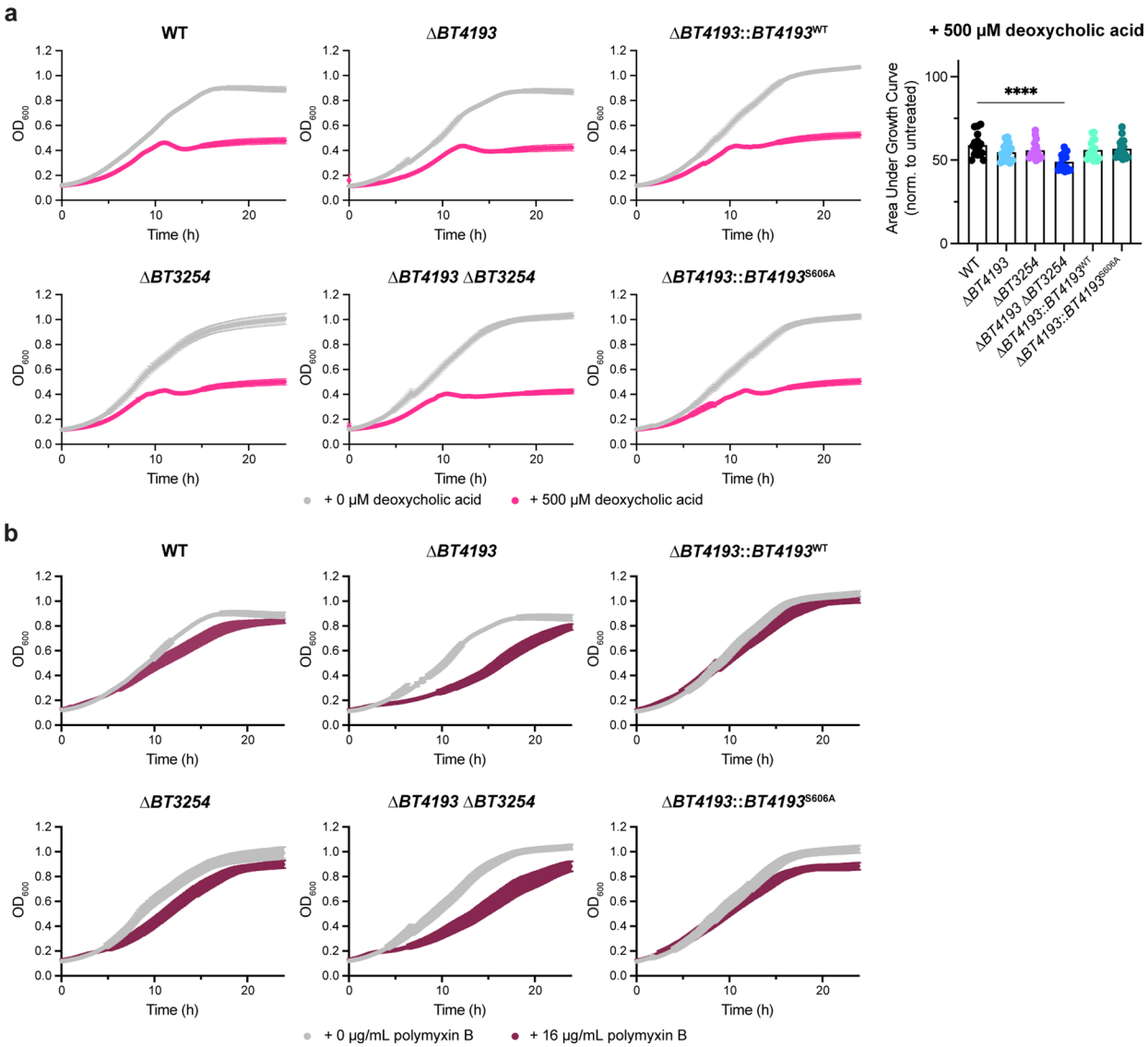
Extended Data Fig. 2 | hDPP4 inhibitors act on BT4193 but not BT3254. a, Quantification of percent inhibition of GP-AMC cleavage in wild type (WT) *B. thetaiotaomicron* lysate (ex/em: 380/460 nm) after pretreatment with inhibitor for 30 min at 37 °C. Activity was normalized to DMSO treatment (mean ± SEM; $n = 6$ independent replicates). **b**, Apparent IC_{50} values of recombinant hDPP4 and BT4193 after treatment with inhibitor for 30 min prior to measuring activity via GP-AMC cleavage and the corresponding curves. Activity was normalized to DMSO treatment (100%) and no enzyme controls (0%), and fit with dose-dependent four parameter inhibition. Selectivity index (SI) is the ratio of BT4193 to hDPP4 apparent IC_{50} values (mean ± SD; $n = 2$ independent experiments, each calculated with 3 independent replicates). **c**, Quantification of inhibition of recombinant BT3254 by hDPP4 inhibitors after 30 min of pretreatment. Velocities were normalized to DMSO pretreatment. Data represent the mean ± SEM of 9 independent replicates.

https://doi.org/10.1038/s41589-023-01357-8



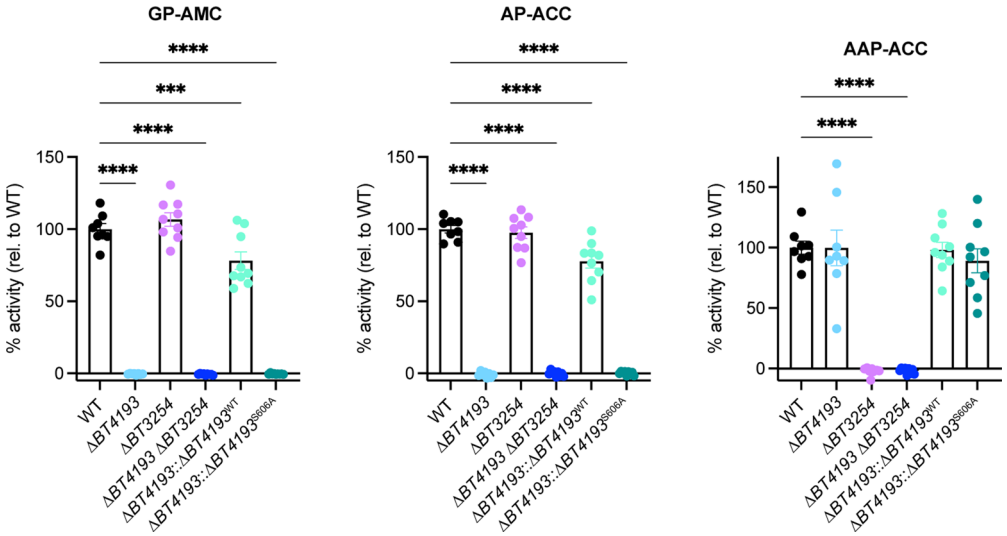
Extended Data Fig. 3 | Recombinant BT4193 and BT3254 prefer P1 Pro residues. Quantification of cleavage velocity of di- and tripeptide fluorogenic peptides with ACC-containing substrates (ex/em: 355/460 nm) by recombinantly expressed and purified BT4193 and BT3254. Hydroxyproline is abbreviated as Hyp. Data represent the mean ± SEM of one representative experiment with 3 independent replicates. Statistical significance was determined using a two-tailed one-sample *t*-test compared with 0 (BT4193: AP-ACC, $p = 0.0026$; AA-ACC, $p = 0.014$; AL-ACC, $p = 0.10$; AS-ACC, $p = 0.019$; AT-ACC, $p = 0.20$; AHyp-ACC, $p = 0.0049$; BT3254: AAP-ACC, $p = 0.0004$; AAA-ACC, $p = 0.0085$; $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$).

https://doi.org/10.1038/s41589-023-01357-8



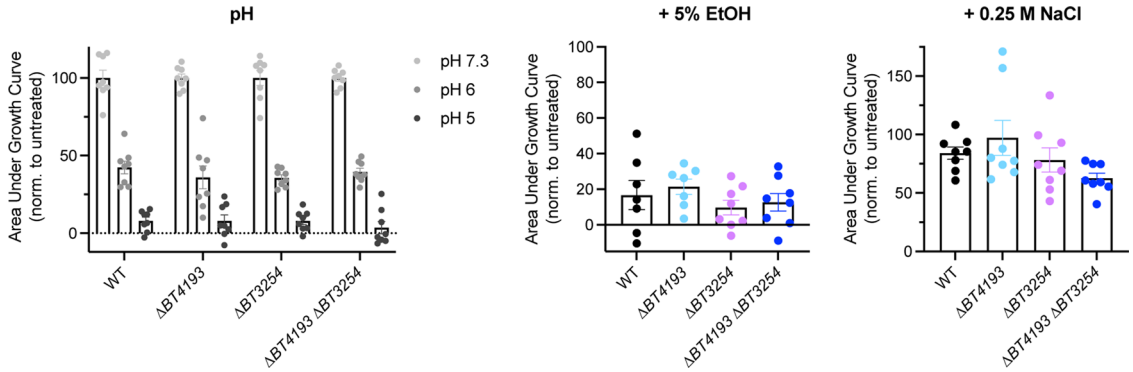
Extended Data Fig. 4 | BT4193 confers resistance to deoxycholic acid and polymyxin B. a, Growth of *B. thetaiotaomicron* strains during deoxycholic acid treatment, and quantification of area under the growth curve normalized to untreated bacteria. The Δ*BT4193* Δ*BT3254* strain exhibited decreased fitness in the presence of deoxycholic acid (mean ± SEM; 16 independent replicates). Statistical significance was determined using a one-way ANOVA test with post hoc Dunnett's multiple comparisons tests compared with wild type ($p = 0.0002$; WT-Δ*BT4193*, $p = 0.14$; WT-Δ*BT3254*, $p = 0.42$; WT-Δ*BT4193* Δ*BT3254*, $p < 0.0001$; WT-Δ*BT4193*::*BT4193*^{WT}, $p = 0.45$; WT-Δ*BT4193*::*BT4193*^{S606A}, $p = 0.72$; ****$P < 0.0001$). **b**, Growth of *B. thetaiotaomicron* strains during polymyxin B treatment.

https://doi.org/10.1038/s41589-023-01357-8



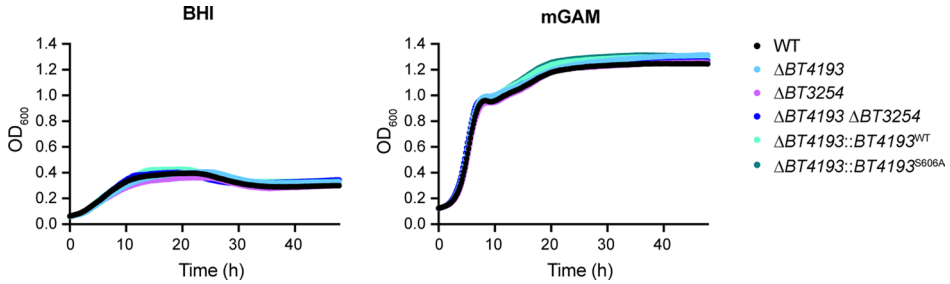
Extended Data Fig. 5 | Complementation with catalytically inactive *BT4193* does not rescue loss of DPP4 activity in *B. thetaiotaomicron* lysate. Quantification of initial velocities of fluorogenic peptide substrate cleavage (ex/em: 380/460 nm for AMC substrates; ex/em: 355/460 nm for ACC substrates) in lysate generated from wild type (WT), knockout, and complemented *B. thetaiotaomicron* strains. Data represent the mean ± SEM of 8 independent replicates for WT and Δ*BT4193* and 9 independent replicates for Δ*BT3254*, Δ*BT4193* Δ*BT3254*, Δ*BT4193*::*BT4193*^{WT}, and Δ*BT4193*::*BT4193*^{S606A}. Statistical significance was determined using a one-way ANOVA test with post hoc Dunnett's multiple comparisons tests compared with wild type (GP-AMC: $p < 0.0001$; WT-Δ*BT4193*, $p < 0.0001$; WT-Δ*BT3254*, $p = 0.55$; WT-Δ*BT4193* Δ*BT3254*, $p < 0.0001$; WT-Δ*BT4193*::*BT4193*^{WT}, $p = 0.0004$; WT-Δ*BT4193*::*BT4193*^{S606A}, $p < 0.0001$; AP-ACC: $p < 0.0001$; WT-Δ*BT4193*, $p < 0.0001$; WT-Δ*BT3254*, $p = 0.96$; WT-Δ*BT4193* Δ*BT3254*, $p < 0.0001$; WT-Δ*BT4193*::*BT4193*^{WT}, $p < 0.0001$; WT-Δ*BT4193*::*BT4193*^{S606A}, $p < 0.0001$; AAP-ACC: $p < 0.0001$; WT-Δ*BT4193*, $p = 1.00$; WT-Δ*BT3254*, $p < 0.0001$; WT-Δ*BT4193* Δ*BT3254*, $p < 0.0001$; WT-Δ*BT4193*::*BT4193*^{WT}, $p = 1.00$; WT-Δ*BT4193*::*BT4193*^{S606A}, $p = 0.77$; ***$P < 0.001$; ****$P < 0.0001$).

https://doi.org/10.1038/s41589-023-01357-8



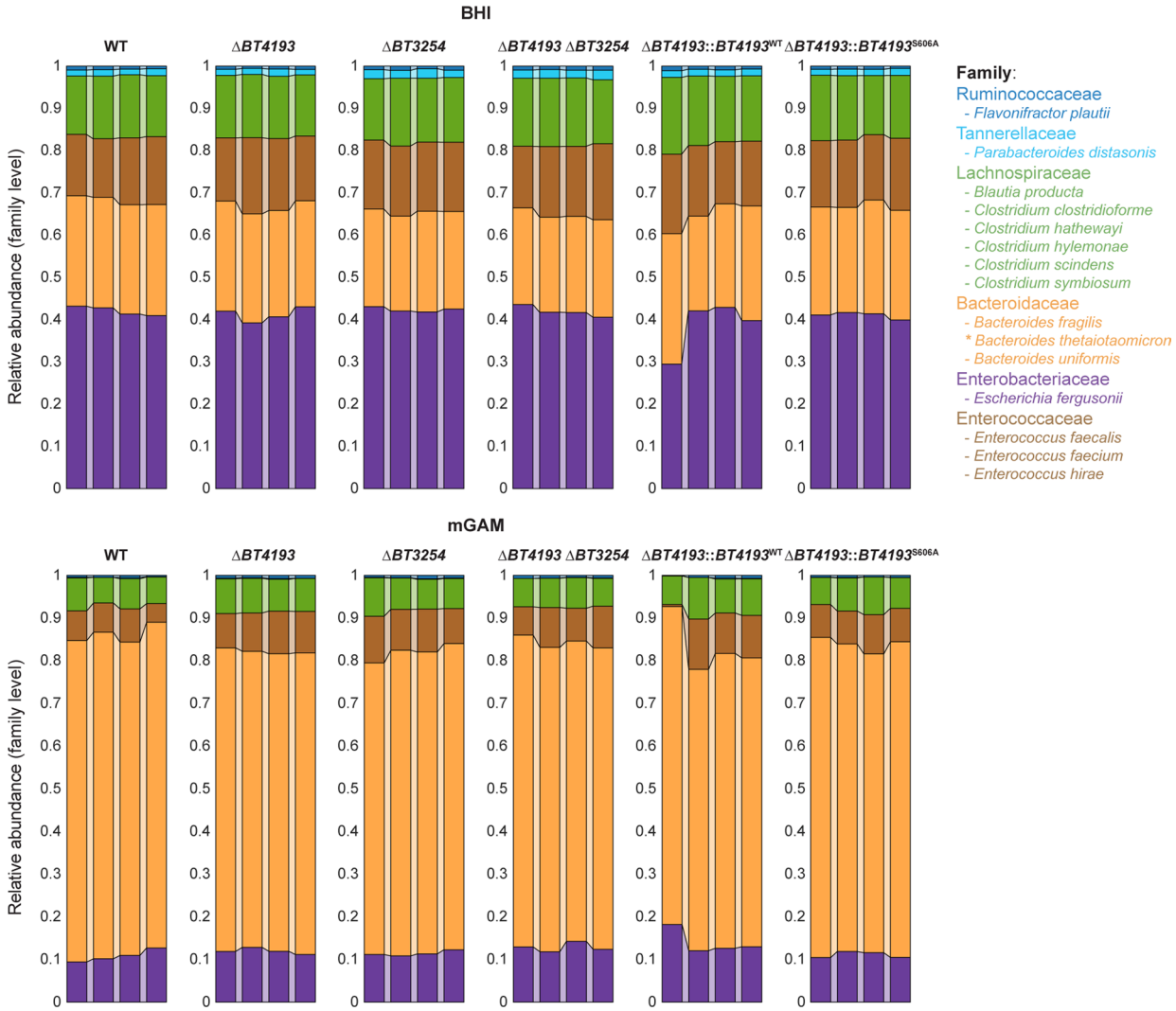
Extended Data Fig. 6 | Δ*BT4193* strains are not universally susceptible to stressors. Quantification of the area under the growth curve of *B. thetaiotaomicron* strains under pH, ethanol, or sodium chloride stress, normalized to untreated bacteria. Deletion of *BT4193* did not impact fitness in response to pH, ethanol treatment, or sodium chloride treatment (mean ± SEM; $n = 8$ independent replicates for pH 7.3, pH 6, pH 5, 0.25 M NaCl; $n = 7$ independent replicates for 5% EtOH WT, 5% EtOH Δ*BT4193*; $n = 8$ independent replicates for 5% EtOH Δ*BT3254*, 5% EtOH Δ*BT4193* Δ*BT3254*). Lack of statistical significance was determined using a one-way ANOVA test (pH 7.3, $p = 1.00$; pH 6, $p = 0.69$; pH 5, $p = 0.71$; 5% EtOH, $p = 0.48$; 0.25 M NaCl, $p = 0.11$).

https://doi.org/10.1038/s41589-023-01357-8



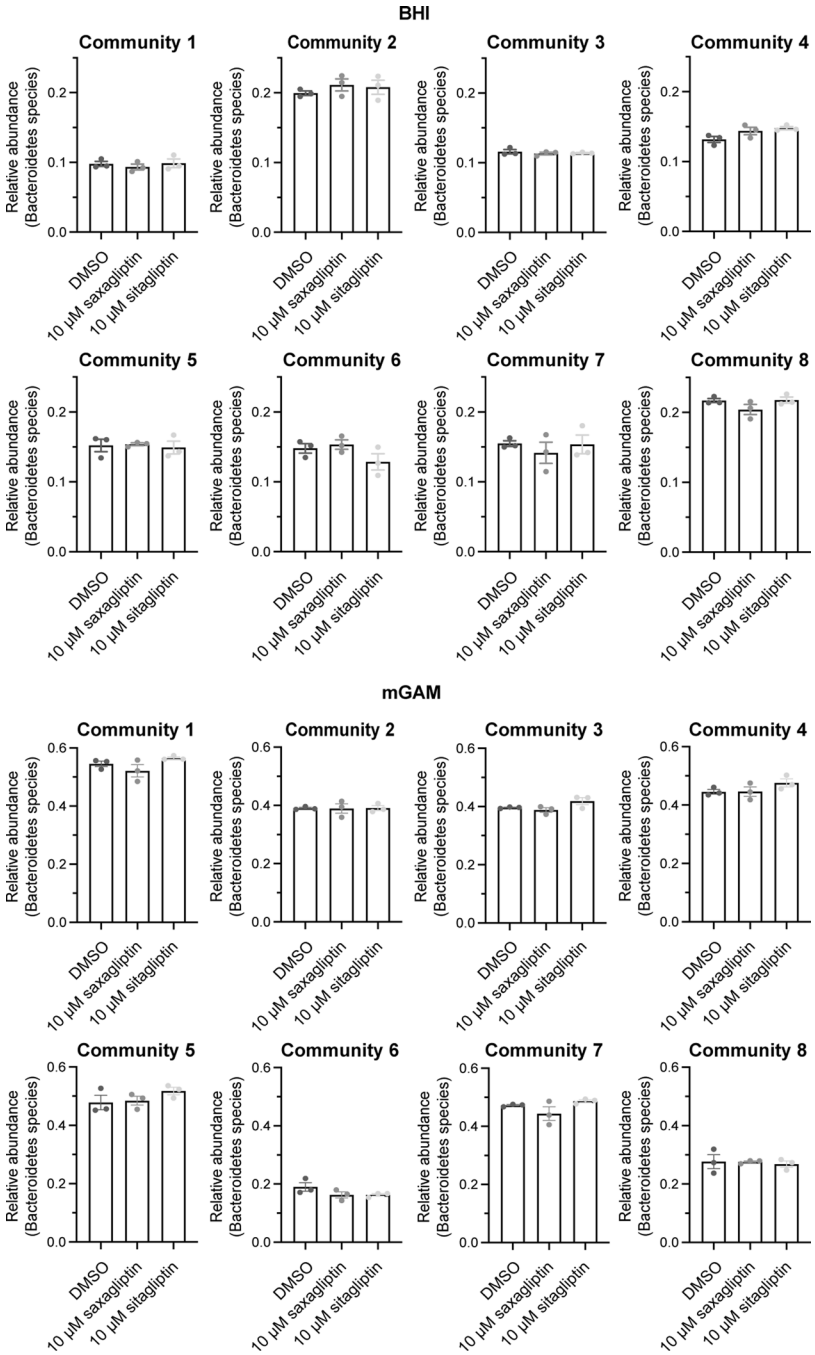
Extended Data Fig. 7 | Loss of *BT4193* does not affect monoculture growth in rich media. Representative growth of *B. thetaiotaomicron* strains in BHI or mGAM (mean ± SEM; $n = 8$ independent replicates).

https://doi.org/10.1038/s41589-023-01357-8



Extended Data Fig. 8 | Overall community structure was not affected by deletion of *BT4193*. Quantification of relative abundance of bacteria within the 15-member synthetic communities at the family level after 48 h of growth in BHI or mGAM. Each group of four bars represents replicates for a community with each strain and medium combination.

https://doi.org/10.1038/s41589-023-01357-8



Extended Data Fig. 9 | Treatment with hDPP4-targeting drugs does not affect the fitness of bacteria from the phylum Bacteroidetes in stool-derived communities. Quantification of the relative abundance of bacteria from the phylum Bacteroidetes in eight stool-derived communities after 48 h treatment with 10 μM saxagliptin or 10 μM sitagliptin in BHI or mGAM (mean ± SEM; $n = 3$ independent replicates). Lack of statistical significance was determined using a one-way ANOVA test (BHI: Community 1, $p = 0.71$; Community 2, $p = 0.58$; Community 3, $p = 0.69$; Community 4, $p = 0.077$; Community 5, $p = 0.91$; Community 6, $p = 0.19$; Community 7, $p = 0.70$; Community 8, $p = 0.19$; mGAM: Community 1, $p = 0.15$; Community 2, $p = 1.00$; Community 3, $p = 0.10$; Community 4, $p = 0.25$; Community 5, $p = 0.33$; Community 6, $p = 0.20$; Community 7, $p = 0.16$; Community 8, $p = 0.91$).

nature research

Corresponding author(s): Matthew Bogyo

Last updated by author(s): Apr 17, 2023

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Collection of raw proteomics data was performed using an UltiMate 3000 - Orbitrap Fusion Lumos system (Thermo Scientific). Collection of in-gel fluorescence images was performed using a Typhoon 9410 Imager (Amersham). Collection of MSP-MS raw data was performed using an UltiMate 3000 - Q-Exactive system (Thermo Scientific). Collection of enzyme activity was performed using a Cytation 3 plate reader (BioTek). Collection of bacterial growth curves was performed using either a Cytation 3 plate reader (BioTek), an Epoch 2 plate reader (BioTek), or a Stratus plate reader (Cerillo). Collection of microscopy images was performed using a Ti-E inverted epifluorescence microscope (Nikon) with a Neo 5.5 sCMOS camera (Andor Technology) or a Zeiss LSM700 confocal microscope.
Data analysis	Multiple sequence alignments were performed using the web tool Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) and FigTree v1.4.4. Predicted protein structures were performed using AlphaFold Colab (https://colab.research.google.com/github/deepmind/alphafold/blob/main/notebooks/AlphaFold.ipynb) and aligned using UCSF Chimera v1.13.1. Protein domains were predicted using the web-based InterPro tool (https://www.ebi.ac.uk/interpro/), and subcellular localization of the FP-reactive serine hydrolases were predicted with the web-based tools SignalP 5.0 (https://services.healthtech.dtu.dk/service.php?SignalP-5.0), CELLO8, and PSORTb (https://www.psort.org/psortb/index.html). Proteomics data analyses were performed using MaxQuant v1.6.0.16 and Perseus v1.6.0.7. MSP-MS data analyses were performed using PEAKS v8.5. Microscopy data analyses were performed using the Matlab image processing package Morphometrics or the ImageJ macro FluoQ Macro V3-97. Sequencing samples were demultiplexed with Qiime v1.9.1 and further analyzed with custom Matlab (Mathworks) scripts. Graphs, data processing, and statistical analyses were generated using RStudio, GraphPad Prism 9, and Matlab 2020a (Mathworks). Code for bioinformatic analyses of serine hydrolases, analysis of fitness data, quantification of area under the growth curves, and plotting of Venn diagrams can be obtained at https://doi.org/10.5281/zenodo.7835000. Code for analyzing community assembly experiments can be obtained at https://doi.org/10.5281/zenodo.7830074.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Selection of serine hydrolase-associated Pfam domains was performed with the MEROPS (https://www.ebi.ac.uk/merops/) and ESTHER (https://bioweb.supagro.inrae.fr/ESTHER/general?what=index) databases. Proteomes for bioinformatic prediction of serine hydrolases were downloaded from UniProt, and accession codes for each proteome are included in Supplementary Table 2. Proteomes were annotated with Pfam domains using pfam_scan.pl (http://ftp.ebi.ac.uk/pub/databases/Pfam/). Proteomes of reference isolates from the NIH Human Microbiome Project gastrointestinal tract were downloaded from https://www.hmpdacc.org/hmp/HMRGD/. Carbohydrate-active enzymes were identified using the CAZyme database (http://www.cazy.org/). Raw proteomics data for this study have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD035963. Raw MSP-MS data can be obtained through massive.ucsd.edu under the dataset identifier numbers MSV000091339 (hDPP4), MSV000091338 (BT4193), and MSV000089969 (BT3254). Raw 16S rRNA data for community assembly experiments can be obtained at https://doi.org/10.25740/rn970zy4428. All other manuscript data are available in the source data provided with this paper.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We did not make any prior statistical calculations to determine sample size. Sample sizes of at least 3 were used, as they are the minimum requirement to allow for statistical analysis.
Data exclusions	In Extended Data Figure 2a, 2 technical replicates in four different conditions were excluded due to pipetting error. In Figure 5b and Extended Data Figure 4, 1 replicate was excluded from plotting the growth curves of 5 different conditions due to a minor technical artifact, but the data were still used for quantification of area under the growth curve. In Extended Data Figure 6, 1 replicate across 2 different conditions was excluded from the quantification of area under the growth curve due to a major technical artifact.
Replication	Each experiment was successfully replicated across at least two independent experiments.
Randomization	Randomization was not relevant to this study, as each sample was treated with the same set of conditions.
Blinding	Experiments were not blinded because each experiment was performed and analyzed by a single scientist.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging