

# Comprehensive analyses of a large human gut Bacteroidales culture collection reveal species and strain level diversity and evolution

Zhenrun J. Zhang\*, Cody G. Cole†, Michael J. Coyne†, Huaiying Lin†, Nicholas Dylla, Rita C. Smith, Emily Waligurski, Ramanujam Ramaswamy, Che Woodson, Victoria Burgo, Jessica C. Little, David Moran, Amber Rose, Mary McMillin, Emma McSpadden, Anitha Sundararajan, Ashley M. Sidebottom, Eric G. Pamer\*, Laurie E. Comstock\*

† These authors contribute equally.

\* Corresponding authors.

## Authors and Affiliations

### **Duchossois Family Institute, University of Chicago, 900 E. 57th St, Chicago, IL, 60637, USA**

Zhenrun J. Zhang, Huaiying Lin, Michael J. Coyne, Cody G. Cole, Nicholas Dylla, Rita C. Smith, Emily Waligurski, Ramanujam Ramaswamy, Victoria Burgo, Che Woodson, Jessica C. Little, David Moran, Amber Rose, Mary McMillin, Emma McSpadden, Anitha Sundararajan, Ashley M. Sidebottom, Eric G. Pamer, Laurie E. Comstock

### **Department of Medicine, Section of Infectious Diseases & Global Health, University of Chicago Medicine, 5841 South Maryland Ave, Chicago, IL, 60637, USA**

Eric G. Pamer

### **Department of Pathology, University of Chicago Medicine, 5841 South Maryland Ave, Chicago, IL, 60637, USA**

Eric G. Pamer

### **Department of Microbiology, Biological Sciences Division, University of Chicago, 5841 South Maryland Ave, Chicago, IL, 60637, USA**

Zhenrun J. Zhang, Michael J. Coyne, Cody G. Cole, Emily Waligurski, Eric G. Pamer, Laurie E. Comstock

## Summary – 140 words

Species of the Bacteroidales order are among the most abundant and stable bacterial members of the human gut microbiome with diverse impacts on human health. While Bacteroidales strains and species are genomically and functionally diverse, order-wide comparative analyses are lacking. We cultured and sequenced the genomes of 408 Bacteroidales isolates from healthy human donors representing nine genera and 35 species and performed comparative genomic, gene-specific, mobile gene, and metabolomic analyses. Families, genera, and species could be grouped based on many distinctive features. However, we also show extensive DNA transfer between diverse families, allowing for shared traits and strain evolution. Inter- and intra-specific diversity is also apparent in the metabolomic profiling studies. This highly characterized and diverse Bacteroidales culture collection with strain-resolved genomic and metabolomic analyses can serve as a resource to facilitate informed selection of strains for microbiome reconstitution.

# Introduction

Species of the Bacteroidales order are ubiquitous and abundant members of the human colonic microbiome with diverse impacts on host health. While heterogeneity amongst Bacteroidales species was recognized over 90 years ago,<sup>1</sup> with subsequent molecular typing hinting at even greater taxonomic complexity,<sup>2,3</sup> the breadth and depth of Bacteroidales diversity was not fully appreciated until the era of whole genome and deep metagenomic sequencing. Over the past two decades, comparisons of genome sequenced isolates and metagenome-assembled genomes have added new branches to the Bacteroidota phylogenetic tree, with renaming of some taxa and introduction of new names for various genera and species.<sup>4-8</sup> Human gut Bacteroidales species are currently distributed to at least eight families. Of these families, Bacteroidaceae species are by far the best-studied with much less known about members of other families including Odoribacteriaceae, Rickenellaceae and Prevotellaceae.

Bacteroidales species encode many carbohydrate-active enzymes (CAZymes), including glycolytic enzymes, that enable them to break down and utilize a wide range of dietary polysaccharides.<sup>9,10</sup> In the process of catabolizing polysaccharides, some Bacteroidales produce metabolites that include acetate, propionate and succinate, which impact gut mucosal integrity and immune defenses.<sup>11-14</sup> However, studies also hint at species and strain level diversity in the production of various short chain fatty acids.<sup>15,16</sup> In addition to dietary polysaccharides, some Bacteroidales species utilize host mucin glycans,<sup>17</sup> which reduce the thickness of the mucus layer lining the gut mucosa, impacting immune regulation and antimicrobial defenses.<sup>9,18</sup> The spectrum of Bacteroidales strains colonizing the host is impacted by diet, food preparation, and food components.<sup>10</sup> While individual Bacteroidales species generally encode a core constellation of polysaccharide utilization loci (PULs) and metabolize a similar spectrum of polysaccharides,<sup>17</sup> genetic exchange between species<sup>19,20</sup> and loss of some functional PULs, such as those enabling host mucin glycan utilization, can lead to intraspecies diversity in utilizable substrates.<sup>17</sup> We also know little about the glycolytic capabilities of many of the other families of gut Bacteroidales.

Bile acids are prevalent, cholesterol-derived steroidal molecules produced in the liver and introduced into the mammalian intestinal lumen as taurine and glycine conjugates that are sequentially modified by different members of the microbiota. Deconjugation and further modification of bile acids and other steroidal molecules by gut microbes is increasingly recognized to impact the mammalian host<sup>21</sup> and is mediated, in part, by members of the Bacteroidales order.<sup>22-25</sup> For example, a bile salt hydrolase synthesized by *Bacteroides thetaiotaomicron* VPI-5482 impacts host metabolic functions including weight gain and the metabolism of carbohydrates and lipids.<sup>25</sup> A sulfotransferase produced by some Bacteroidales species can sulfonate cholesterol and a subset of secondary bile acids.<sup>22,24</sup> Several Bacteroidales genera and species encode reductases and hydroxysteroid-dehydrogenases that modify secondary bile acids and produce variants that impact lymphocyte differentiation and may also inhibit enteric pathogens.<sup>23</sup> A previous metabolomic study of 59 Bacteroidales strains of three families has been a useful resource to the field, and the current study includes additional

Bacteroidales families and species and additional metabolites further highlighting the extreme species and strain level metabolic heterogeneity.

Another feature of gut Bacteroidales is their rapid evolution due to the exchange of DNA by conjugation. These mobile genetic elements encode numerous properties that affect strain fitness, including PUL that increase utilizable substrates,<sup>19</sup> transport cofactors,<sup>26</sup> Type VI secretion systems<sup>27</sup> and bacteriocins<sup>28,29</sup> that antagonize competing strains, along with protective immunity genes,<sup>30</sup> and antibiotic resistance genes,<sup>31</sup> to name a few. Although extensive conjugal DNA transfer between species within an individual's gut microbiota has been documented,<sup>32,33</sup> there are few studies that examine the degree to which Bacteroidales strains evolve and differentiate in the human gut due to DNA acquisitions.<sup>34</sup> In addition, a better understanding of the phenotypes encoded on these mobile genetic elements and their impact on strain fitness and subsequent impacts on human host phenotypes becomes increasingly important as strains are selected as components of synthetic consortia for microbiome restoration.

Loss of microbiome diversity in a variety of settings is associated with adverse health outcomes and has led to studies of microbiome reconstitution with symbiotic gut bacterial strains or consortia, referred to as live biotherapeutic products (LBPs).<sup>35</sup> As Bacteroidales strains are prevalent colonizers of the human intestine and generate metabolites that contribute to health and disease resistance, identifying those that produce beneficial metabolites is essential. Furthermore, identifying bacterial strains that successfully co-colonize with other species and support microbiome richness and diversity is also an important goal. In this study, we analyzed a large culture collection of diverse Bacteroidales strains of five families and 35 species isolated from healthy human donors to provide an order-wide analysis of their genomic, genetic and metabolomic properties. The substantial accessory genomes harbored by Bacteroidales strains encode a range of functions that, if defined, might be paired with the host recipient's microbiome composition, metabolomic profile and underlying disease status, with the goal of optimizing the effectiveness and health impact of next-generation LBPs.

## Results

### Isolation, identification, and deduplication of Bacteroidales isolates

We collected stool samples from 21 healthy human donors, which were processed for 16S rRNA amplicon sequencing.<sup>36</sup> Among the 21 donors, Bacteroidales represented 12.88% of the bacterial abundance on average (range 3.40% to 46.83%, median 8.32%) (Fig. S1). Fecal samples were diluted, and strains cultured on rich, non-selective media agar plates under anaerobic conditions. Single-colony isolates were sequenced using short-read whole genome sequencing (WGS). The culture collection was not designed to collect the diversity of Bacteroidales strains within each individual, but rather to build a diverse Bacteroidales strain bank for order-wide genomic, genotypic, and metabolomic analyses to facilitate future informed selection of consortia members for microbiome restoration.

To identify isolates of the Bacteroidales order, we performed BLAST searches using the longest 16S rRNA sequence from the WGS against the NCBI RefSeq 16S rRNA database or the Genome Taxonomy Database (GTDB). These analyses identified 408 strains of Bacteroidales (Dataset 1, Bacteroidales collection tab). To correctly speciate these isolates, we used 25 single-copy core genes (SCGs) included in the Anvi'o database<sup>37</sup> (Dataset 1, single copy core genes tab). Multiple sequence alignment of the concatenated sequences of these 25 SCGs from the 408 Bacteroidales genomes and those of Bacteroidales type strains provided clear taxonomic assignments into 35 species of nine distinct genera (Fig 1A, Dataset 1, Bacteroidales collection tab).

As many donor microbiotas contained multiple isolates of the same species, we sought to reduce redundancy by identifying clonal isolates. For this analysis, we performed pairwise comparisons among genomes of all strains of the same species using fastANI<sup>38</sup> for Average Nucleotide Identity (ANI) and nucmer<sup>39</sup> for genomic single nucleotide polymorphisms (SNPs) and insertion-deletions (InDels). Compared to isolates of the same species from different donors, those from the same donor typically have higher ANI and fewer genomic SNPs and InDels (Fig. S2A-C), with clear apparent cutoffs. We used the following criteria as thresholds for concluding isolates were of the same clonal lineage: ANI higher than 99.9%, total genomic SNPs fewer than 1000, and total genomic InDels fewer than 75. These analyses revealed many microbiomes where there were multiple distinct co-colonizing strains of the same species. To confirm that two strains of the same species from a microbiome are either the same or distinct clonal lineages, we performed pangenome analysis of these isolates using Anvi'o.<sup>37</sup> This gene clustering-based approach clearly separated strains, as clonal isolates consistently have more than 95% identical gene clusters whereas isolates of different strains shared 80% or less (Fig. S3). Distinct co-colonizing strains of the same species were identified in nine microbiomes including DFI.3 where the four *B. thetaiotaomicron* isolates segregate into two distinct clonal lineages, and the five *Phocaeicola vulgatus* isolates segregate into three distinct clonal lineages (Dataset 1, de-duplication tab). Similarly, isolates from the MSK.21 microbiome were only of the Prevotellaceae family, including four different *Segatella* species,<sup>5</sup> which is a characteristic of microbiomes from non-industrialized populations. In our strain collection, *Ph. vulgatus* was the most isolated species and also the species found most often to have two or more co-colonizing strains of distinct lineages (in six donor microbiota).

## Phylogeny of gut Bacteroidales based on curated single-copy core genes

We reduced the set of 408 strains by selecting only one strain of the same clonal lineage, which yielded 127 distinct clonal isolates (Dataset 1, de-duplication tab). Next, we built a phylogenetic tree of these 127 unique human gut Bacteroidales strains along with type strains of each of the 35 species using multiple sequence alignment of the concatenated sequences of the 25 SCGs (Fig. 1A, Dataset 1 single copy core genes tab). As expected, families (Bacteroidaceae, Prevotellaceae, Odoribacteraceae, Tannerellaceae and Rikenellaceae) and genera within these families (*Bacteroides*, *Phocaeicola*, *Parabacteroides*, *Segatella*, *Leyella*, *Hoyleseella*, *Butyrivimonas*, *Odoribacter*, and *Alistipes*) were well separated and formed monophyletic clades.

## Genome size and GC content of Bacteroidales isolates

We next analyzed genome size across Bacteroidales isolates (Fig. 1B). Genome sizes varied drastically between species of different families from as small as ~2.8 Mbp (*Alistipes putredinis*) to >7.1 Mbp (*Bacteroides cellulosilyticus* and *Bacteroides ovatus*). In general, the largest genomes are from strains of *Bacteroides* known to have an extensive capability to digest host and dietary polysaccharides and that have the greatest number of carbohydrate active enzymes (CAZymes) and polysaccharide utilization loci (PUL). There was also considerable variation in the genome sizes within strains of the same species, the most dramatic example being genomes of *Ph. vulgatus* that ranged from ~4.57 Mbp to ~5.80 Mbp (Dataset 1, Bacteroidales collection tab). Intraspecies genome size discrepancies highlight the extensive horizontal gene transfer that occurs in these species, and the large pangenes.

Analysis of %GC content revealed a wide range between species of this bacterial order, with *Bacteroides nordii* having the lowest at 40.5%, and *A. communis* the highest at 58.5% (Dataset 1, Bacteroidales collection tab). There are typically narrow distributions of GC content within a species. Moreover, isolates within a genus tend to have similar trends in GC content, with *Alistipes* species having relatively higher GC content (53.5 - 58.5%) and *Phocaeicola* species having a narrow range and relatively lower GC content (42-43%) (Fig. 1C).

## Core genomes of Bacteroidales isolates

We annotated these Bacteroidales genomes using the Prokka pipeline with customization (see Methods).<sup>40</sup> The number of coding regions varied from 2167 to 5671, due to drastic differences in genome size, with an average of 75% of all putative CDS being annotated (range: 65% to 86%). To investigate the core genomes across taxonomic levels, we performed annotation-agnostic clustering of all CDS of the 127 unique Bacteroidales isolates and their type strains. 247 gene clusters were shared across all isolates (order-level core genome)(Fig. 2). At the family level, the core genome size increased to 979 gene clusters for Bacteroidaceae, 842 for Prevotellaceae and 1366 for Odoribacteraceae. Other families included only one genus, so family level analysis was not possible. Genus level analysis was included for all genera with more than one species and further increased the size of the core genomes: 1175 gene clusters for *Bacteroides*, 1438 for *Phocaeicola*, 1862 for *Parabacteroides*, 2045 for *Butyrivibrio*, 992 for *Alistipes*, and 1366 for *Segatella*. Strains of the same species shared on average about 69.6% of core genomes (median 69.5%, range 52.5%-89.8%) (Fig 2).

## Diversity of Bacteroidales based on genetic repertoire

To further elucidate the genomic diversity among human gut-derived Bacteroidales, we performed Uniform Manifold Approximation and Projection (UMAP)<sup>41</sup> based on the presence or absence of unique gene clusters from their whole-genome protein coding sequences (Fig. 3, Fig. S4). We found that different families and genera were well separated in UMAP, and in most cases, isolates of the same species and their corresponding type strains clustered together in UMAP and were distinguished from other species (Fig. 3, Fig. S4). *P. vulgatus* isolates occupied

a larger area compared to other species, and some *P. vulgatus* isolates were closer to type strains of other *Phocaeicola* species than to the *P. vulgatus* type strain (Fig. 3 and Fig. S4) with strain distances similar to those of other species of a genus.

## Distribution of CAZymes and PULs among human gut Bacteroidales

The distribution of CAZymes and PUL has been extensively characterized in gut Bacteroidales species as these features indicate a strain's glycan utilizing capabilities. To determine the distribution of CAZymes in this Bacteroidales culture collection, we used Hidden Markov Models from dbCAN database<sup>42</sup> to search for CAZymes and categorized them according to the CAZy database.<sup>43</sup> We detected 305 CAZy families/subfamilies among Bacteroidales isolates with marked variation in the total number of CAZymes in different species (Fig. 4A and Fig. S5A). Most identified CAZymes are of the Glycoside Hydrolases family (GH), followed by Glycosyl Transferases (GT), Polysaccharide Lyases (PL), Carbohydrate Esterases and Carbohydrate-Binding Modules (Fig. 4A). *Bacteroides cellulosilyticus* contained the largest numbers of total and unique CAZy families/subfamilies present, whereas *Alistipes putredinis* contained the least (numbers of total CAZy families/subfamilies in Fig. 4A, and numbers of unique CAZy families/subfamilies in S5B). There are trends in the categories of CAZymes within a genus, with Bacteroidaceae species typically containing the most glycolytic enzymes (GH and PL). The number of CAZymes present in a species also correlated with the diversity of their CAZymes measured by Inverse Simpson Index (Fig. S5A). Glycosyl Transferases are CAZymes that are involved in the synthesis of glycans present as surface molecules (capsular polysaccharides, extracellular polysaccharides, surface layers) or components of Bacteroidales glycoproteins.<sup>44,45</sup> There is also extreme diversity in the number of GT between species with *Alistipes putredinis* encoding only 10 GT and some *P. dorei* strains encoding as many as 80. These differences indicate that the ability to synthesize numerous phase variable surface polysaccharides, which is a trait of the gut Tannerellaceae and Bacteroidaceae,<sup>46</sup> is not a trait shared by all gut Bacteroidales.

UMAP analysis based on the presence or absence of CAZyme families revealed that isolates of the same species tended to cluster together (Fig. 4B). *Bacteroides thetaiotaomicron*, *B. xylanisolvens*, *B. ovatus*, *B. intestinalis* and *B. cellulosilyticus*— species with the most CAZymes— are distant from other clusters, suggesting they possess distinct CAZyme profiles (Fig. 4B). *Alistipes* and Odoribacteraceae species, which encoded the fewest CAZymes, cluster together (Fig. 4B). Analysis of the CAZyme UMAP based on donor showed that isolates from the same individual are scattered around the UMAP (Fig. S5C) supporting that different Bacteroidales species occupy distinct nutritional niches in the microbiome.

Exploitative competition, or the ability to compete for nutrients, often limits co-colonization of strains of the same species in a microbiome. Successful co-colonization may require utilization of distinct glycan sources. We analyzed the CAZymes present in the three distinct *P. vulgatus* strains that co-colonized donor DFI.3 (DFI.3.23, DFI.3.65 and DFI.3.92) and found strain differences in the CAZyme families. GH99 are endo- $\alpha$ -mannosidases that cleave mannoside linkages in *N*-linked glycan chains.<sup>47</sup> GH99 is present in single copy in DFI.3.65, with two copies

in DFI.3.92, but absent in DFI.3.23. Two different GH families, GH10 and GH67 are uniquely present in DFI.3.92. GH10 members are endoxylanases<sup>48</sup> and GH67 members are xylan  $\alpha$ -glucuronidases,<sup>49,50</sup> some of which are upregulated in *Bacteroides* species upon culturing with xylan.<sup>50</sup> These findings suggest that *P. vulgatus* strains within the same donor have unique capabilities to utilize different dietary/host components, exploiting distinct nutritional niches in the same microbiome. We further analyzed the CAZyme profiles of all strains of the same species of distinct clonal lineages co-colonizing individuals and similarly found extensive differences, likely accounting for their ability to co-colonize the same host (Dataset 2). Interestingly, among the CAZyme families that are discordantly present within each comparison, GH46 is the most frequent. GH46 encodes a predicted chitosanase not previously characterized in Bacteroidales. Other frequently discordant GH families include GH117, a neoagaro-oligosaccharide hydrolase family in an agarose metabolic pathway shown to be acquired by a *B. uniformis* strain,<sup>51</sup> and GH96, an  $\alpha$ -agarase family previously uncharacterized in Bacteroidales.

In Bacteroidales, the majority of glycolytic enzymes are encoded in PULs. Although PULs often contain genes encoding surface and periplasmic glycolytic enzymes, as well as regulatory genes,<sup>52</sup> the minimum requirement for a locus to be considered a PUL is a set of *susC/susD* genes, encoding the outer membrane/surface glycan import machinery, and a gene encoding a glycolytic enzyme.<sup>52</sup> We used PULpy<sup>53</sup> to identify PULs in our Bacteroidales strain collection. *Bacteroides cellulosilyticus* strains contained the most PULs on average (67-72 PULs), followed closely by *B. ovatus* isolates (66-71 PULs)(Fig. 4C), consistent with these strains containing the most CAZymes (Fig. 4A, Fig. S5B). *Alistipes putredinis* contained only one *susC/susD* pair without a nearby GH or PL encoding gene. In general, *Bacteroides* species contained more PULs whereas *Alistipes* and *Odoribacter* species contain fewer (Fig. 4C), suggesting limited use of dietary and host glycans by these latter species.

## Bile acid metabolism capacities among human gut Bacteroidales

Gut bacteria produce enzymes that modify host-generated primary bile acids into secondary bile acids, which have numerous effects on both the host and the microbiota. Many Bacteroidales produce bile salt hydrolases (BSH) that deconjugate the taurine or glycine moiety which enables further modification by other microbiota-encoded enzymes.<sup>21,25,54,55</sup> Using three distinct BSH sequences as queries, we found that most gut Bacteroidales of our collection have at least one BSH ortholog with the exception of species of Prevotellaceae, most *Odoribacter splanchnicus* and the single *B. finegoldii* isolate (Dataset 3). Therefore, this important deconjugation reaction is performed by most gut Bacteroidales. A recent study showed that centenarians have increased levels of fecal lithocholic acid (LCA) including iso-LCA, allo-LCA, 3-oxo-LCA, 3-oxoallo-LCA and isoallo-LCA.<sup>23</sup> We searched our Bacteroidales collection for 5 $\alpha$ -reductase (5-AR) and 5 $\beta$ -reductase (5-BR) that produce 3-oxoallo-LCA and 3-oxo-LCA, respectively (Dataset 3).<sup>23</sup> Of the 127 strains, the majority contain orthologs of both 5AR and 5BR with the exception of all Prevotellaceae, *Bacteroides nordii*, *O. splanchnicus* and one outlier *B. uniformis* strain, all of which lacked both proteins. In addition, *Parabacteroides distasonis* isolates contain only an ortholog of 5BR, but not 5AR. The genes encoding 5AR and 5BR are adjacent as previously described (Dataset 3).<sup>23</sup> Lastly, we searched for hydroxysteroid



dehydrogenases (HSDH). With the exception of the Prevotellaceae, *Bacteroides nordii*, *O. splanchnicus* and the same outlier *B. uniformis* strain, most Bacteroidales contain at least one ortholog of 3 $\beta$ HSDH, which produces isoalloLCA from 3-oxoallo-LCA. These genes are adjacent to the 5AR and or 5BR genes as described.<sup>23</sup> 7 $\alpha$ HSDH is present in only some *Bacteroides* species and one *O. splanchnicus* isolate (Dataset 3). Collectively, these data indicate that most Bacteroidales strains encode enzymes that can modify bile acids to produce a variety of secondary bile acids, with species and strain level diversity. It remains unclear if the Prevotellaceae have bile acid conversion capacities, but our data show that they do not have close ortholog to any of the known Bacteroidales proteins with these functions.

## Metabolome of Human Gut Bacteroidales

Human gut Bacteroidales produce a wide range of metabolites that affect other microbes and impact host phenotypes.<sup>11-13,56,57</sup> We performed metabolomic analyses of culture supernatants of Bacteroidales strains grown in BHIS medium. We quantified the short chain fatty acids (SCFAs) acetate, propionate, butyrate and the dicarboxylic acid succinate. In addition, a panel of 29 metabolites was analyzed semiquantitatively, enabling us to compare relative metabolite amounts to baseline levels present in the medium to determine production versus consumption. The metabolomes of 111 strains that grew well in BHIS were included.

Bacteroidales strains produce acetate, succinate and propionate, with some exceptions. For example, *B. stercoris* MSK.20.72 did not produce acetate, produced only small amounts of propionate and consumed succinate. Butyrate was produced by Odoribacteraceae isolates (two *Odoribacter splanchnicus* and one *Butyrimonas synergistica*) (Fig. 5A, Fig. S6A). These *O. splanchnicus* isolates contain both Acetyl-CoA-Type and Lysine-Type complete butyrate production pathways. The genomes of both *Alistipes putredinis* strains also encode complete pathways for butyrate production, however, when grown in BHIS, they produced only very low amounts of butyrate (0.055-0.14 mM compared to 0.54-18 mM from *O. splanchnicus*) (Fig. S6A). Most notable in these analyses are numerous examples of extreme intra-species diversity in the production of these metabolic end products.

The relative production versus consumption of metabolites, such as branched-chain and aryl acids, amino acids, medium-chain fatty acids, among Bacteroidales strains was also determined and demonstrated, for some metabolites, significant inter- and intra-species variation (Fig. 5B).

One of the most variably produced metabolites in our panel is phenylacetate. Gut microbiota-derived phenylacetate is the critical precursor for host production of phenylacetylglutamine, which is associated with both development of atherosclerotic cardiovascular disease, major adverse cardiovascular events, and stroke risk.<sup>58-60</sup> Phenylacetate is reported to be produced from phenylalanine via two different pathways by some gut microbes, including *Bacteroides* species,<sup>57</sup> and involves specific oxidoreductases. We observed that in some species, such as *B. uniformis* and *B. thetaiotaomicron*, the numbers of FAD-dependent oxidoreductases encoded in their genomes correlates with phenylacetate production (Fig. S6B, S6C).

Hierarchical clustering of metabolomic profiles did not precisely follow phylogeny (Fig. S6D), consistent with previous findings.<sup>16</sup> Procrustean Approach to Cophylogeny (PACo) test did not detect a significant correlation between phylogeny and metabolomic profiles ( $P = 0.129$ ) (Fig. S6E). Thus, our metabolomic results suggest that while genomic analyses can in many cases determine whether production or consumption of specific metabolites occurs, inter- and intra-species diversity of metabolomes cannot, at this time, be fully explained by genomic composition, highlighting the need for broader and deeper analyses of bacterial metabolite production or consumption and gene transcription under varying culture conditions.

## Analysis of genes encoding antagonistic proteins/machinery

Bacteroidales species, like most species that live in communities, produce many antagonistic weapons to thwart competitors. To date, five distinct antibacterial toxin families (BSAPs, BfUbb, BcpT, bacteroidetocins, and Fab1) and three distinct Type VI secretion systems (T6SSs) architectures (GA1, GA2 and GA3) have been identified. Many of these antibacterial toxins have been shown to limit competing species in the mammalian gut.<sup>61-66</sup> We queried the genomes for genes encoding BSAP-1 – BSAP-4,<sup>61,67,68</sup> involved in intra-species antagonism in *B. fragilis*, *B. uniformis*, *Ph. vulgatus* and *Ph. dorei*; BcpT,<sup>28</sup> which is restricted largely to *Ph. vulgatus* and *Ph. dorei*; bacteroidetocin A and B, which are widely distributed among gut Bacteroidaceae;<sup>29</sup> and Fab1<sup>62</sup> and BfUbb,<sup>69</sup> which are restricted to *B. fragilis*. As expected, a subset of strains of the expected species contains the genes encoding each of these toxins with the exception of BfUbb, which was not encoded in any of the 7 *B. fragilis* genomes in our collection, consistent with a previous reported prevalence of 14% among sequenced *B. fragilis* strains.<sup>69</sup> There were two other notable findings from this analysis. The first is that *bcpT* and the 9.1 kb plasmid that encodes it were found not only in *Ph. vulgatus* and *Ph. dorei* as expected, but also in *B. caccae*, a species not previously shown to harbor this mobile genetic element. The second interesting finding is that BcpT and Bd-B, both encoded on mobile genetic elements (MGEs), are present in some isolates of the same clonal lineage but not all (see below), indicating that these mobile elements were acquired or lost by some strains in those individuals' gut microbiome.

To investigate T6SS distribution, we performed BLAST searches using concatenated T6SS consensus sequences of each of the three GA types and the five GA2 subtypes.<sup>33,70</sup> We detected at least one T6SS in most genera except *Butyrivibrio*, *Segatella* and *Hoylesella*, which extends Bacteroidales species containing T6SS loci to *Odoribacter* and *Alistipes*. However, in the *Odoribacter spanchnicus* isolates of DFI.1, 14 genes the GA1 T6SS region are present in the reverse orientation, resulting in the disruption *tssD* and *tssC* for which partial genes are contained on each side of the inverted region (Fig S6A) leading to a non-functional T6SS. Of the 127 unique isolates analyzed, GA1 is present in 20 genomes, GA2b in 14 genomes, GA2a in 5 genomes, GA2c in 1 genome, and GA3 in 4 genomes. GA1 and GA2 T6SSs loci are present on integrative and conjugative elements (ICE) and we found them to be variably present in strains of the same clonal lineage (see below), demonstrating acquisition of these antagonistic weapons by strains in the microbiota of these human subjects. Our previous analyses had revealed GA2 ICE exclusion wherein we had not previously identified strains with both a GA2

and GA3 T6SS, and rarely found isolates containing both a GA2 and GA1 T6SS loci. Of those genomes that did contain both GA1 and GA2 ICE, we previously identified a defect in one of the T6SS loci or in a gene adjacent to the GA1 T6SS locus designated *mghA*.<sup>33</sup> Three strains in our collection, *B. uniformis* DFI.1.121, *B. thetaiotaomicron* DFI.1.148 and *B. thetaiotaomicron* DFI.2.27, contain both GA2a and GA1 T6SS loci in their genomes (Dataset 1, T6SS tab), and in each, the *mghA* gene is similarly interrupted by the insertion of a putative transposase-like element encoding a reverse transcriptase (Fig. S6B). Mhga is a protein containing both an *N*<sup>6</sup>-adenine methylase domain and an SNF2 helicase domain, which is similar to protective genes in the phage-defense system (DISARM).<sup>71</sup> The most surprising result of these analyses was the identification of a *B. fragilis* strains with all three GA types (GA1, GA2a, and GA3) (Fig. S7; Dataset 1, T6SS tab). In this strain, the *mghA* gene of the GA1 ICE was also interrupted, but with a different transposable element (Fig S7).

## Horizontal DNA transfer between co-resident species

We and others have previously demonstrated extensive transfer of conjugative elements between Bacteroidales species in a person's gut microbiota.<sup>27,32,72</sup> However, most of these transfers were detected within and between species of Bacteroidaceae and *Parabacteroides*. As our culture collection contains diverse Bacteroidales species of five different families, we had the opportunity to identify DNA transfer among these diverse families. We used previously established criteria of 99.99% DNA identity over a minimum of 5 kbp to detect elements transferred between species within an ecosystem. Using the entire set of 408 isolate genomes, we detected extensive transfer of DNA between co-resident species, which included mobile and conjugative plasmids, integrative and conjugative elements (ICE), lysogenic phage and other mobile genetic elements (Dataset 4). As mentioned above, these transfers included the GA1 and GA2 T6SS loci that are contained on ICEs. In the gut microbiome of donor DFI.1, a GA1 ICE transferred among seven different species of three families: Bacteroidaceae, Tannerellaceae, and Rikenellaceae. We also detected transfer of other genetic elements between Bacteroidaceae and *Odoribacter*, *Butyrivibrio* and *Alistipes* (Dataset 4). The community of donor MSK.21 contained five different Prevotellaceae species, from which we detected DNA transfers among *Leyella stercorea* and *Segatella brunsvicensis*. Figure 6A illustrates five different elements that were transferred between different species in the microbiota of donor DFI.9. The full list of these genetic regions and their putative annotations are included for each ecosystem (Dataset 4). These analyses also highlight the ubiquity of lysogenic phage transferring between species in the human gut. We identified 20 phages that transferred between Bacteroidales species in this collection — one as large as 102 kb — and all of them belong to the Cauldoviricetes group which are contractile tail phages. In donor DFI.1, *B. cellulosilyticus*, *O. splanchnicus*, and *Parabacteroides merdae*, comprising three different Bacteroidales families, all contained the identical prophage in their genomes, demonstrating a broad host range.

We also detected numerous plasmids transferred among species. This size cutoff imposed in this analysis excluded pBI143, a ubiquitous mobile plasmid of ~2.8 kb that was recently described as one of the most abundant genetic elements of the human gut microbiome.<sup>73</sup> We

found pBI143 in strains of 16 of the 21 ecosystems and 40 of the 127 clonal lineages, including 14 species. In addition, we found another small plasmid of ~5 kbp that was present in strains of seven communities. The largest plasmid detected was 105 kbp and is a variant of pMMCAT, a large conjugative plasmid carrying fimbriae genes and an extracellular polysaccharide biosynthesis locus.<sup>33</sup> We previously found pMMCAT to be one of the most highly transferred elements with a conserved architecture.<sup>33</sup> In our collection, we found evidence of transfer of pMMCAT-like plasmids in at least six communities.

## Strain evolution within the gut microbiota

Although we know that Bacteroidales strains evolve in the human gut due to mutation and DNA transfers mediated by conjugation and phage,<sup>19,27-32,34,72,74</sup> tracking the evolution of a strain due to these DNA acquisitions or losses requires the analysis of numerous clonal strains of the same lineage from an individual. During our strain de-duplication process, we grouped strains into “clonal” isolates, i.e., those likely derived from a single strain acquisition by an individual. Seventy of the 127 de-duplicated strains contained more than one isolate of the same clonal group (Dataset 1, de-duplication tab). Within these clonal groups we searched for DNA regions of at least 5 kbp that were present in only a subset of clonal strains, but not all. These analyses revealed, in many cases, the divergence and evolution of strains due to the presence of various MGE including plasmids, ICE, and lysogenic phage. In fact, 45 of these 70 clonal lineages revealed strain evolution and divergence based on the presence or absence of MGE (Dataset 5). In Figure 6B we illustrate the genetic differences in the four *B. thetaiotaomicron* clonal isolates of DFI.6, showing the location of a 29,213 bp MGE that inserted just downstream of *dnaK* in one strain, and the insertion of a 10,545 bp phage in a gene of unknown function (DUF3822) in three of the four strains. All four strains also contain four plasmids, and three of the four strains contain an additional plasmid of 44,031 bp. Even greater divergence was detected in *Ph. vulgatus* isolates of clonal lineage 2 of the gut community of donor MSK.19, where we detected eight different *Ph. vulgatus* “strains” that evolved from a single ancestral strain acquired by that individual (Fig 6C). In addition to prophage, the genetic elements variably present in these eight evolved strains encode proteins involved in nutrient acquisition, restriction and modification enzymes, fimbriae, proteases, and antibacterial defense proteins to name a few (Dataset 5). In other clonal lineages, strains variably contain elements encoding T6SSs, antibacterial proteins and peptides, and proteins conferring antibiotic resistance. These data reveal the rapid and extensive evolution and divergence of clonal strains of Bacteroidales in the human gut due to horizontal gene transfer and the numerous phenotypic capabilities that likely accompany these acquisitions.

## Discussion

Species belonging to the Bacteroidales order constitute a major fraction of the lower intestinal tract microbiota and contribute to the host’s physiology and immune defenses. Here, we characterize a large and diverse Bacteroidales culture collection combining genome sequencing and extensive genomic and genetic analyses, with metabolomic analyses of individual

Bacteroidales isolates. To our knowledge, this is the first Bacteroidales stain collection that is fully sequenced, characterized at the genomic, genetic, and metabolomic levels, with individual strains available to the research community.

The characterization of this strain bank has reinforced some properties of this order of bacteria and revealed new findings. For example, it was shown, first by culture-based analyses<sup>75</sup> and more recently by single cell analyses<sup>34</sup> that two or more strains of distinct clonal lineages of a species can co-colonize the human gut. The genetic and phenotypic differences that support such a co-occurrence have not been well-described. Our strain bank has captured numerous gut microbiomes in which multiple strains of distinct lineages of species were isolated, including multiple instances of co-colonizing *Ph. vulgatus*, *B. thetaiotaomicron*, and *B. uniformis* strains. In all cases, we find that co-colonizing strains have distinct accessory genomes including genes for PULs and other genes that likely support their co-colonization. Among our collection, *Ph. vulgatus* is also notable in having the greatest variation in genome size, size of the accessory genome, and wide distribution on the UMAP plot. Although some of this variation may be because our collection contains more *Ph. vulgatus* than other species, this diversity likely contributes to the frequent presence of multiple *Ph. vulgatus* strains of distinct lineages in a single gut ecosystem.

Not only do distinct lineages of a species co-colonize, but so to do numerous isolates of the same clonal lineage that have diverged due to the acquisition or loss of genetic elements. The properties conferred by the genetic cargo of these MGE are beginning to be characterized and include the utilization of dietary polysaccharides,<sup>19,20</sup> cofactor import,<sup>26</sup> antagonistic weapons,<sup>27-29</sup> protection from T6SS mediated antagonism,<sup>30</sup> transcriptional regulators that influence expression of chromosomal genes,<sup>76</sup> biofilm formation, and antibiotic resistance.<sup>77,78</sup> It can be anticipated that introduction of new Bacteroidales strains for therapeutic purposes will evolve differently in each ecosystem into which they are introduced and will also shape the Bacteroidales strains preexisting in or acquired by each ecosystem.

Metabolomic profiling of Bacteroidales isolates revealed significant inter- and intra-species diversity. While butyrate production was restricted to a small subset of species encoding genes for a complete butyrate production pathway, it remains unclear whether these symbiotic strains are significant contributors to the overall butyrate pool in the colon. Along similar lines, while we identified a correlation between phenylacetate production levels with the number of FAD-dependent oxidoreductases encoded in Bacteroidales genomes, further study is required to determine the impact of these strains and the production of phenylacetate on the host. Our finding that the metabolomic output of strains did not consistently align with phylogeny, though interesting, likely results from the limited media conditions under which we analyzed strain metabolomes. It is likely that media compositions that more closely reflect the *in vivo* environment that Bacteroidales strains encounter, including the presence of other microbial species, would provide insights into the regulation of metabolite production.

Bacteroidales species have been associated with disease susceptibility, on the one hand, and health benefits, on the other hand. The broad range of genomic and metabolomic features associated with individual strains within a given Bacteroidales species suggests that strains to

be used for therapeutic purposes should be selected carefully.<sup>79-81</sup> Our strain-resolution Bacteroidales repository provides detailed description and characterization of these human gut isolates, which are available to the scientific community (both in NCBI RefSeq database and at the DFI Symbiotic Biobank website <https://dfi.cri.uchicago.edu/biobank/>) and will facilitate development of targeted therapeutics and defined synthetic microbial consortia.<sup>35,82</sup>

## Declaration of interests

E.G.P. serves on the advisory board of Diversigen; is an inventor on patent applications WPO2015179437A1, titled “Methods and compositions for reducing *Clostridium difficile* infection,” and WO2017091753A1, titled “Methods and compositions for reducing vancomycin-resistant enterococci infection or colonization”; and receives royalties from Seres Therapeutics, Inc. The other authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this manuscript.

## Acknowledgment

We thank members of the E.G.P. laboratory who have provided valuable feedback on this manuscript. Z.J.Z was supported by The GI Research Foundation early career grant. This work was supported by US National Institutes of Health grants P01 CA023766 to E.G.P. and by the Duchossois Family Institute of the University of Chicago. We also acknowledge the important role of the late Eric Littmann in initiating this study of the Bacteroidales order.

# Figure Legends

**Figure 1. Phylogeny and genomic characterization of the 127 unique gut Bacteroidales isolates.** **A.** Phylogenetic tree of unique gut Bacteroidales isolates and type strains based on concatenated single-copy core gene sequences. *Escherichia coli* serves as an outgroup. Scale bar indicates phylogenetic distance. 127 isolates (circles), type strains (triangles). **B.** Genomic size of human gut Bacteroidales isolates (circles) and type strains (triangles), arranged by decreasing average genome size of species from top to bottom. **C.** Genomic %GC content of human gut Bacteroidales isolates (circles) and type strains (triangles), arranged by decreasing average %GC content of species from top to bottom.

**Figure 2. Core genomes of Bacteroidales isolates at various taxonomic levels.** The number of core gene clusters shared by isolates grouped at different taxonomic ranks is plotted. Color indicates core gene clusters in different taxonomic ranks. Bacteroidales collection isolates (circles) or type strains (triangles).

**Figure 3. UMAP analysis of the genomic diversity of Bacteroidales isolates.** UMAP plot based on the presence or absence of gene clusters in whole-genome protein coding sequences of all Bacteroidales human isolates and type strains in our collection. Bacteroidales collection isolates (circles) or type strains (triangles).

**Figure 4. CAZyme and PUL analyses of Bacteroidales isolates.** **A.** Total numbers of CAZymes in each Bacteroidales isolate, arranged by species from most (top) to least (bottom). CAZyme families are categorized into five major classes according to CAZy database, each panel represents one class. **B.** UMAP plot based on the presence or absence of CAZyme families in all Bacteroidales human isolates and type strains used here. Color indicates species identity. Bacteroidales collection isolates (circles) or type strains (triangles). **C.** Number of PULs in all Bacteroidales isolates and type strains, ranked by their average PUL numbers from most (top) to least (bottom).

**Figure 5. Metabolomic profiles of Bacteroidales isolates.** **A.** Quantitative metabolomics on SCFAs among Bacteroidales isolates. Each column represents one SCFA. Numbers in wells indicate relative concentration changes in mM compared to BHIS medium. Color squares on the left indicate species identity. **B.** Semi-quantitative metabolomics on 25 metabolites among Bacteroidales isolates. Each column represents one metabolite. Columns are grouped by hierarchical clustering. Color scale indicates  $\log_2$ (fold change) of metabolite concentration relative to BHIS medium. Color squares on the left indicate species identity.

**Figure 6. Strain evolution in the human gut microbiota due to DNA transfer events.** **A.** Phage/genetic elements that transferred between the various species of the DFI.9 microbiome. **B.** Divergence and evolution due to DNA acquisitions of four *B. thetaiotaomicron* isolates of DFI.6 of the same clonal lineage. Sites of insertion of the elements in each of the genomes is shown. The MGE is 29,213 bp, the phage is 10,545 bp and all strain contain plasmids of 98,230 bp, 4,971 bp, 4,148 bp and 2,750 bp, and three of the four strains also contain a 44,031 bp

plasmid shown in pink. **C.** Extensive strain divergence of a clonal lineage of *P. vulgatus* of ecosystem MSK.19 due to differential DNA acquisitions/losses. Eight distinct strains are differentiated based on the 14 transferred elements listed across the top. Strains that are the same in regard to the presence or absence of these elements are shaded with the same color. Refer to Dataset 4 for details regarding the genes of these regions. Red X indicates presence of the element and black O indicates the element is absent in that strain. Number in parenthesis refers to the region listed in Dataset 4, and this Dataset also contains details of the genes of these regions. AID – acquired interbacterial defense element. LCP – large conjugative plasmid.

## Supplementary Figure Legends

**Figure S1. Gut microbial composition of human donors.** Bar plot of relative abundance of bacterial taxa according to 16S rRNA amplicon sequencing. Each column is a human donor fecal sample.

**Figure S2. De-duplication of Bacteroidales isolates.** Violin plots showing pairwise differences of Average Nucleotide Identity (**A**), total genomic SNPs (**B**) and total genomic insertions-deletions (**C**) between all isolates of same species. Each dot represents one pairwise comparison of two isolates of same species.

**Figure S3. Pan-genome comparison of strains of the same species from the same donor.** Pangenome comparison circular plot among *Bacteroides thetaiotaomicron* strains from donor DFI.2 (**A**), *B. thetaiotaomicron* strains from donor DFI.3 (**B**), *Phocaeicola vulgatus* strains from donor DFI.3 (**C**), *Bacteroides uniformis* strains from donor DFI.6 (**D**), *Ph. vulgatus* strains from donor DFI.6 (**E**), *Ph. vulgatus* strains from donor DFI.9 (**F**), *Ph. vulgatus* strains from donor MSK.9 (**G**), *Ph. vulgatus* strains from donor MSK.17 (**H**), *Ph. vulgatus* strains from donor MSK.19 (**I**), *B. uniformis* strains from donor MSK.20 (**J**), and *Segatella copri* strains from donor MSK.21 (**K**). Corresponding type strains are also included in red tracks. Light shading indicates absence of that gene cluster.

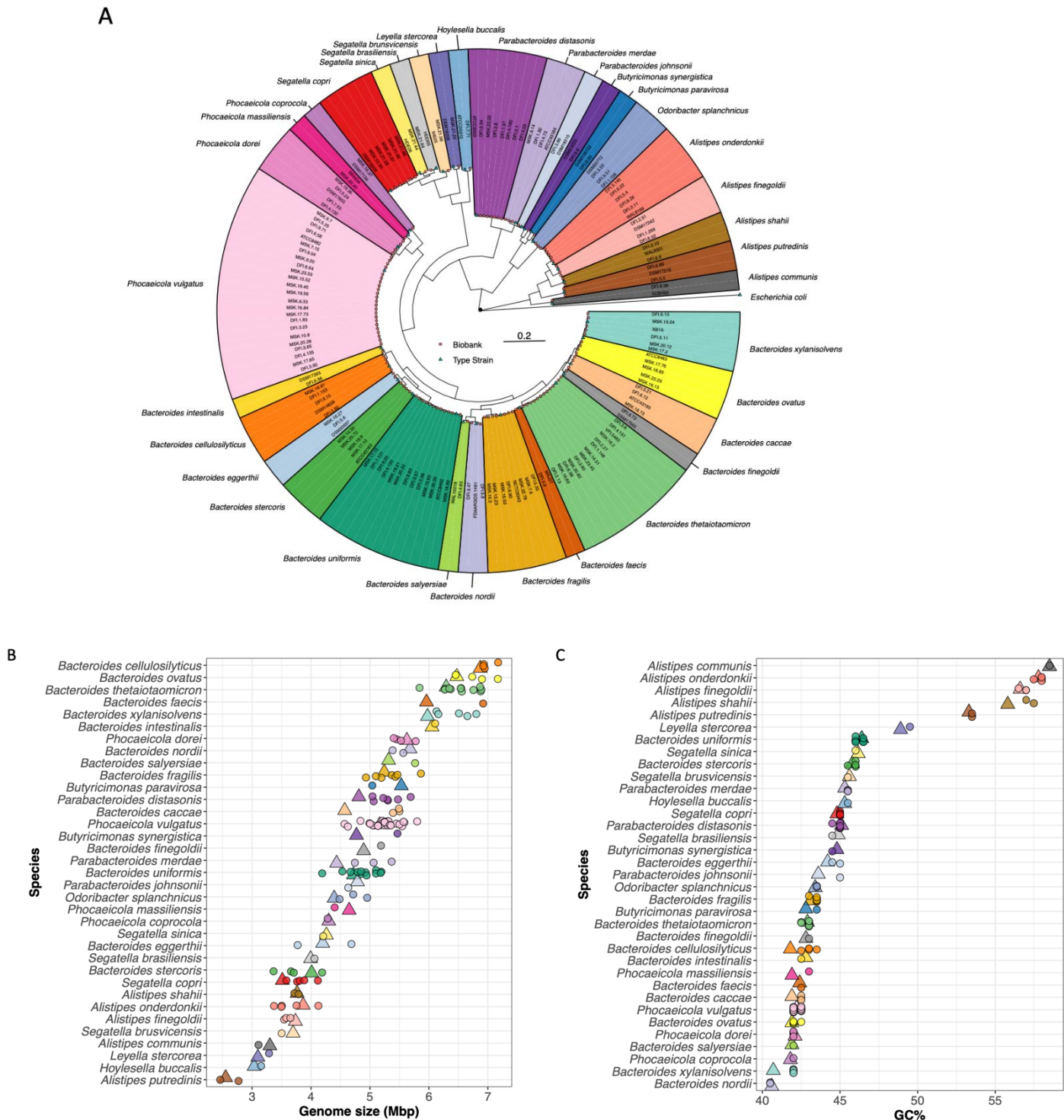
**Figure S4. UMAP analysis of the genomic diversity of Bacteroidales isolates within the genera or family.** Whole-genome gene cluster UMAP based on presence or absence of gene clusters in Bacteroidales genera and families. Triangle indicates type strains, whereas circles indicate human isolates included in our collection.

**Figure S5. Additional CAZyme and PUL analyses of Bacteroidales isolates.** **A.** Number of CAZymes in Bacteroidales isolates and their Inverse Simpson Indices. **B.** Numbers of unique CAZyme families in each Bacteroidales isolate, arranged by species from most (top) to least (bottom). CAZyme families are categorized into five major classes according to CAZy database, each panel representing one class. **C.** UMAP plot based on the presence or absence of CAZyme families in all Bacteroidales isolates and type strains (as in Figure 4B). Color indicates donor identity.



**Figure S6. Inter- and intra-specific diversity of metabolome among Bacteroidales isolates.** **A.** Butyrate concentration changes after culturing in BHIS among Bacteroidales isolates with complete or incomplete/absent butyrate production pathways in their genomes. Color indicates species identity. **B.** Relative concentration fold change of phenylacetate after culturing in BHIS among *Bacteroides thetaiotaomicron* isolates compared to numbers of FAD-dependent oxidoreductase (OR) in their genomes. **C.** Relative concentration fold change of phenylacetate after culturing in BHIS among *Bacteroides uniformis* isolates compared to numbers of FAD-dependent oxidoreductase (OR) in their genomes. **D.** Comparison between single-copy core gene dendrogram and hierarchical clustering dendrogram of metabolomic profiles of Bacteroidales isolates. Lines connect the same isolate on two dendrograms. **E.** Procrustes superimposition of PCoA plots of single-copy core genes (blue) and metabolomic profiles (orange). Lines connect the same isolate on two PCoA plots.

**Figure S7. Genetic features of select Type VI secretion system loci among Bacteroidales isolates.** **A.** A dysfunctional GA1 T6SS locus from *Odoribacter splanchnicus* where 14 of the *tss* genes are inverted resulting in partial *tssD* and *tssC* genes at the inversion termini. **B.** Open reading frame maps of the GA1 and flanking regions showing the *mghA* gene (green) interrupted by a 2463 bp insertion encoding a transposase-like gene in strains also containing a GA2 locus. **C.** The T6SS genetic regions of a *B. fragilis* strains containing all three T6SS genetic architectures (GA1-GA3). The *mghA* gene near the GA1 locus is also interrupted but by a much larger insertion.



**Figure 1. Phylogeny and genomic characterization of the 127 unique gut Bacteroidales isolates.**

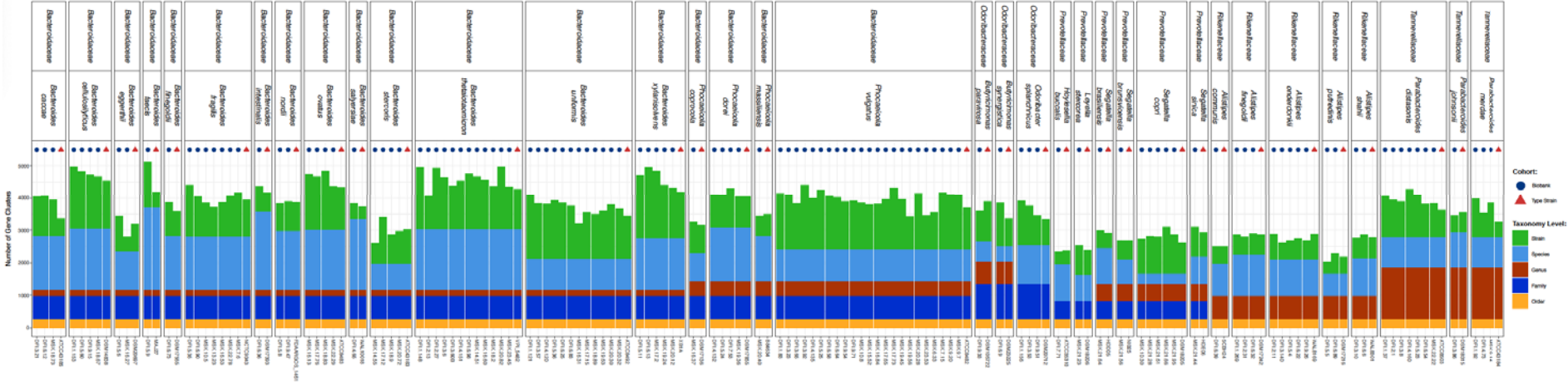


Figure 2. Core genomes of Bacteroidales isolates at various taxonomic levels.

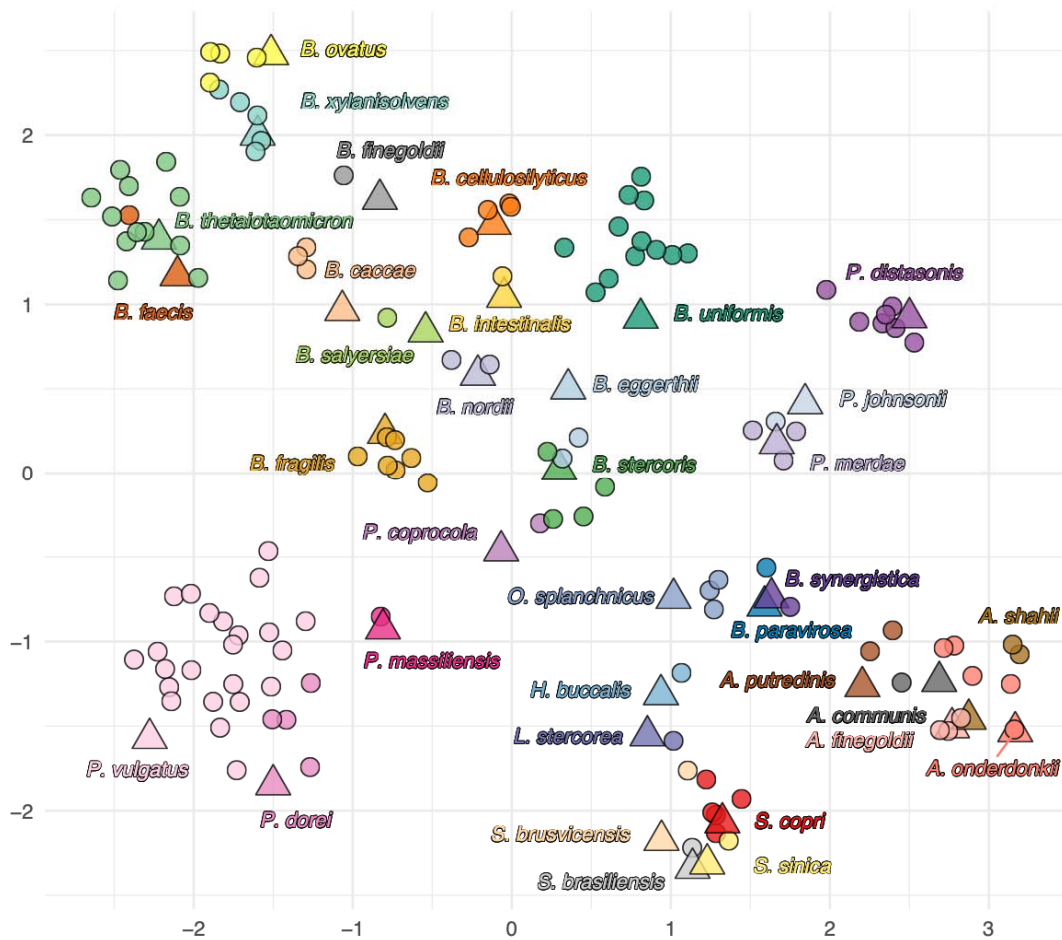


Figure 3. UMAP analysis of the genomic diversity of Bacteroidales isolates

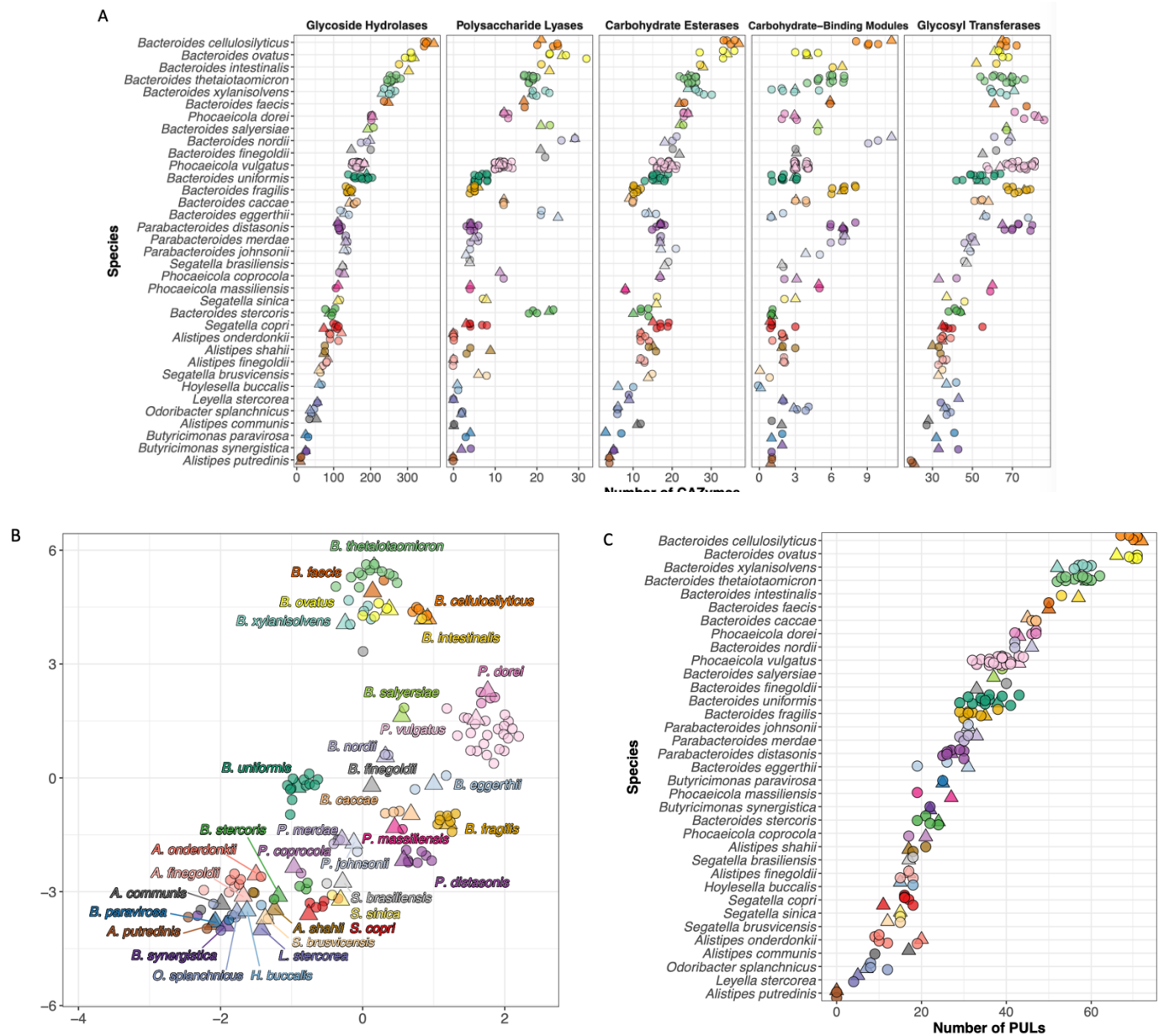
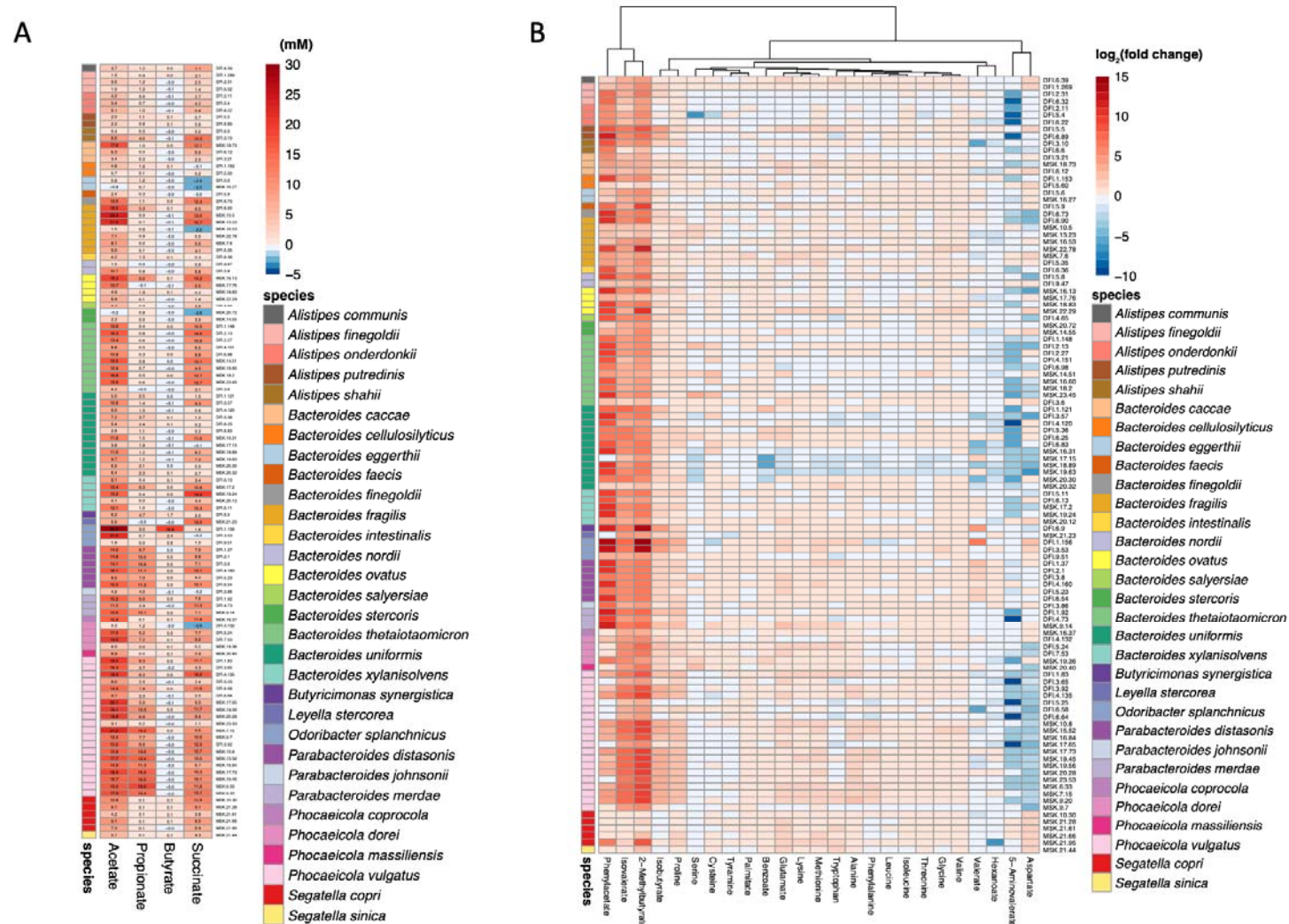


Figure 4. CAZyme and PUL analyses of Bacteroidales isolates.





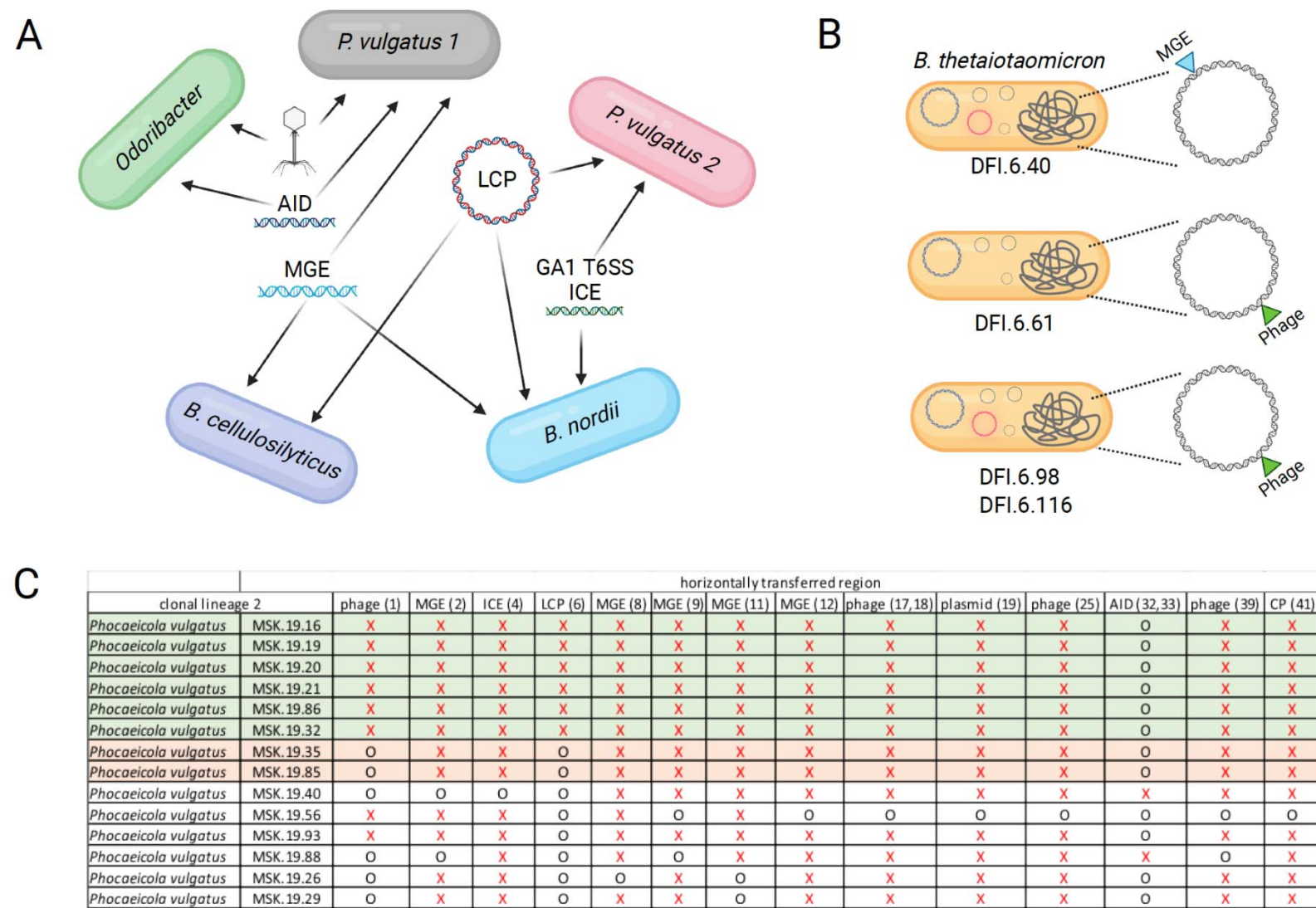
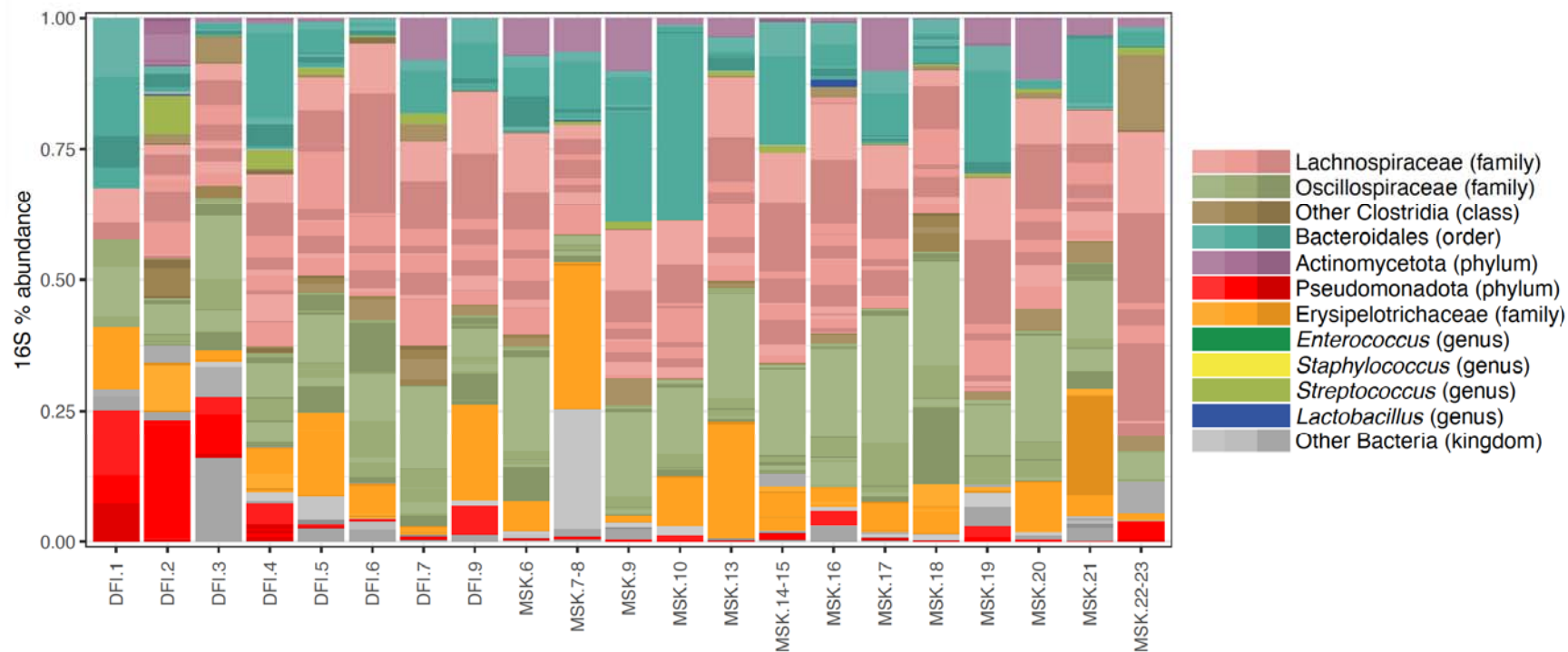
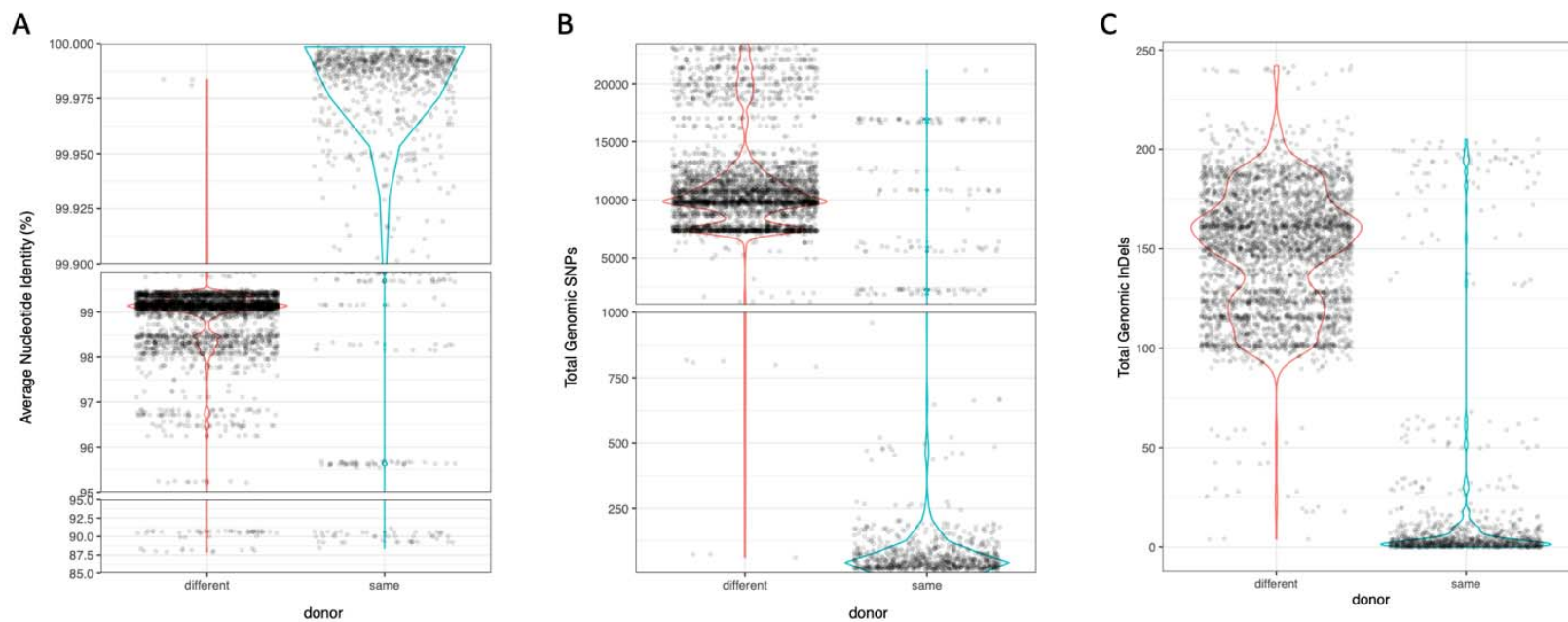


Figure 6. Strain evolution in the human gut microbiota due to DNA transfer events.



**Figure S1. Gut microbial composition of human donors.**





**Figure S2. De-duplication of Bacteroidales isolates.**

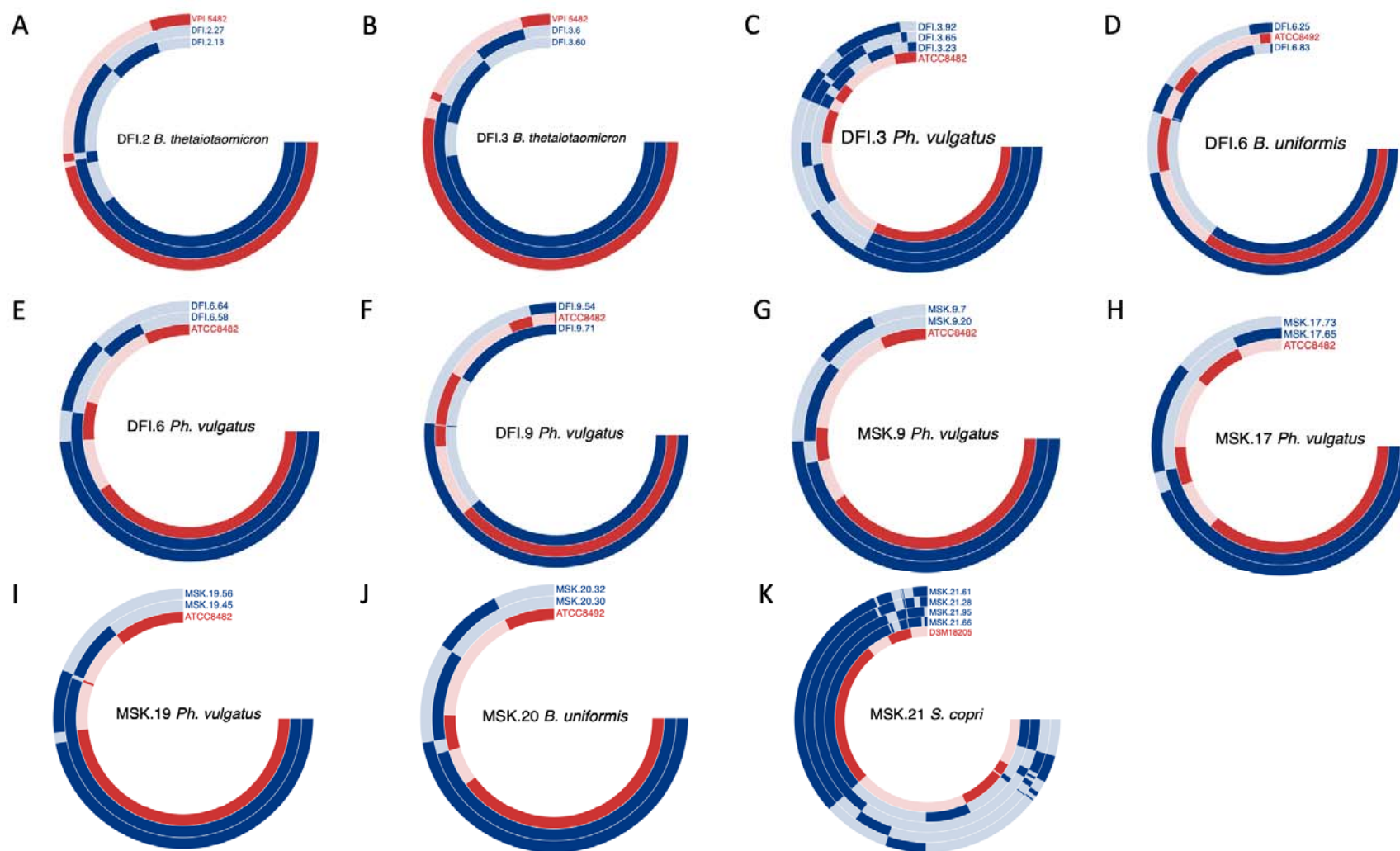


Figure S3. Pan-genome comparison of multiple strains from the same donor.

**Figure S4. UMAP analysis of genomic diversity of Bacteroidales isolates within genera or family.**

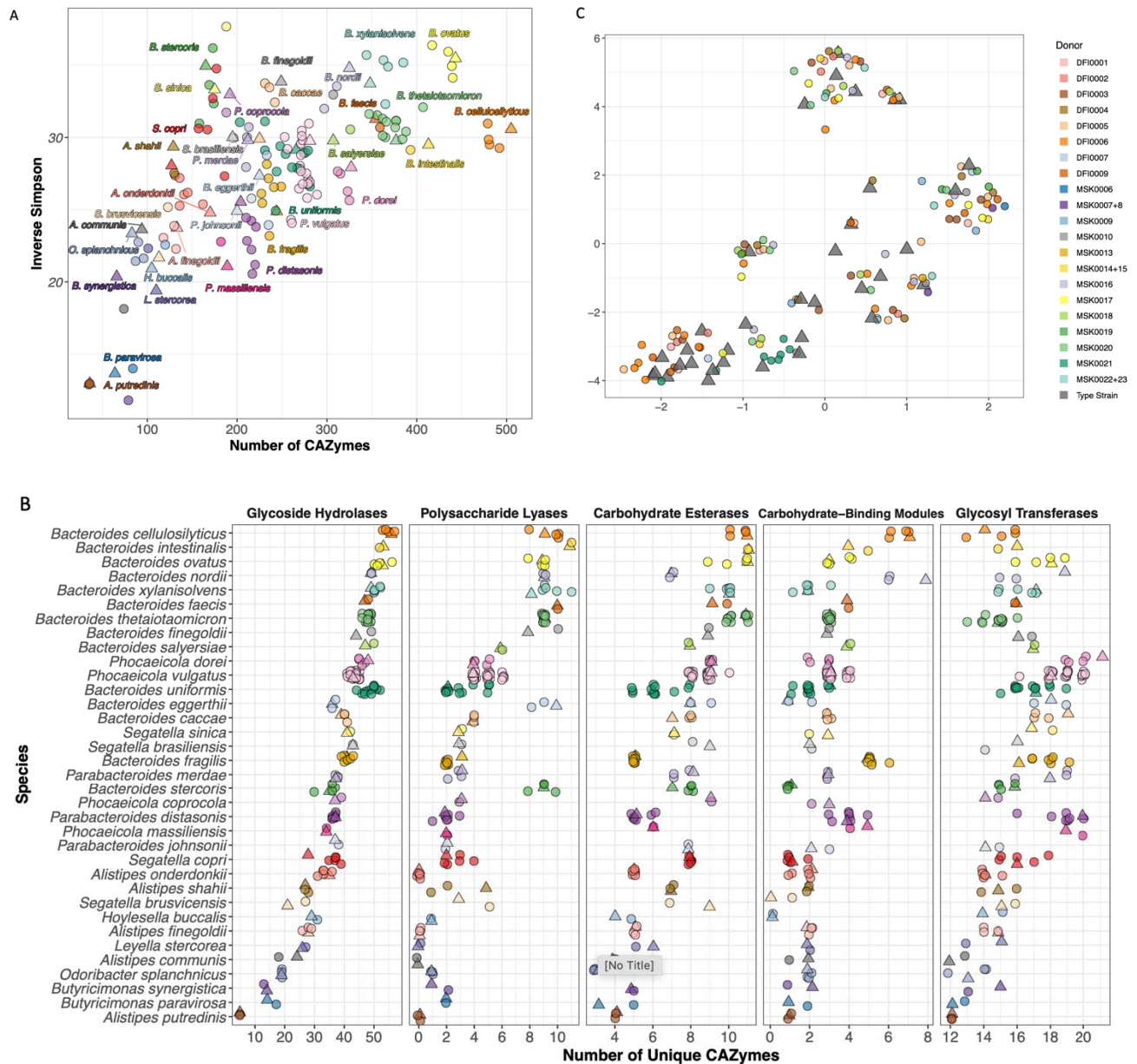


Figure S5. Additional CAZyme and PUL analyses of Bacteroidales isolates.

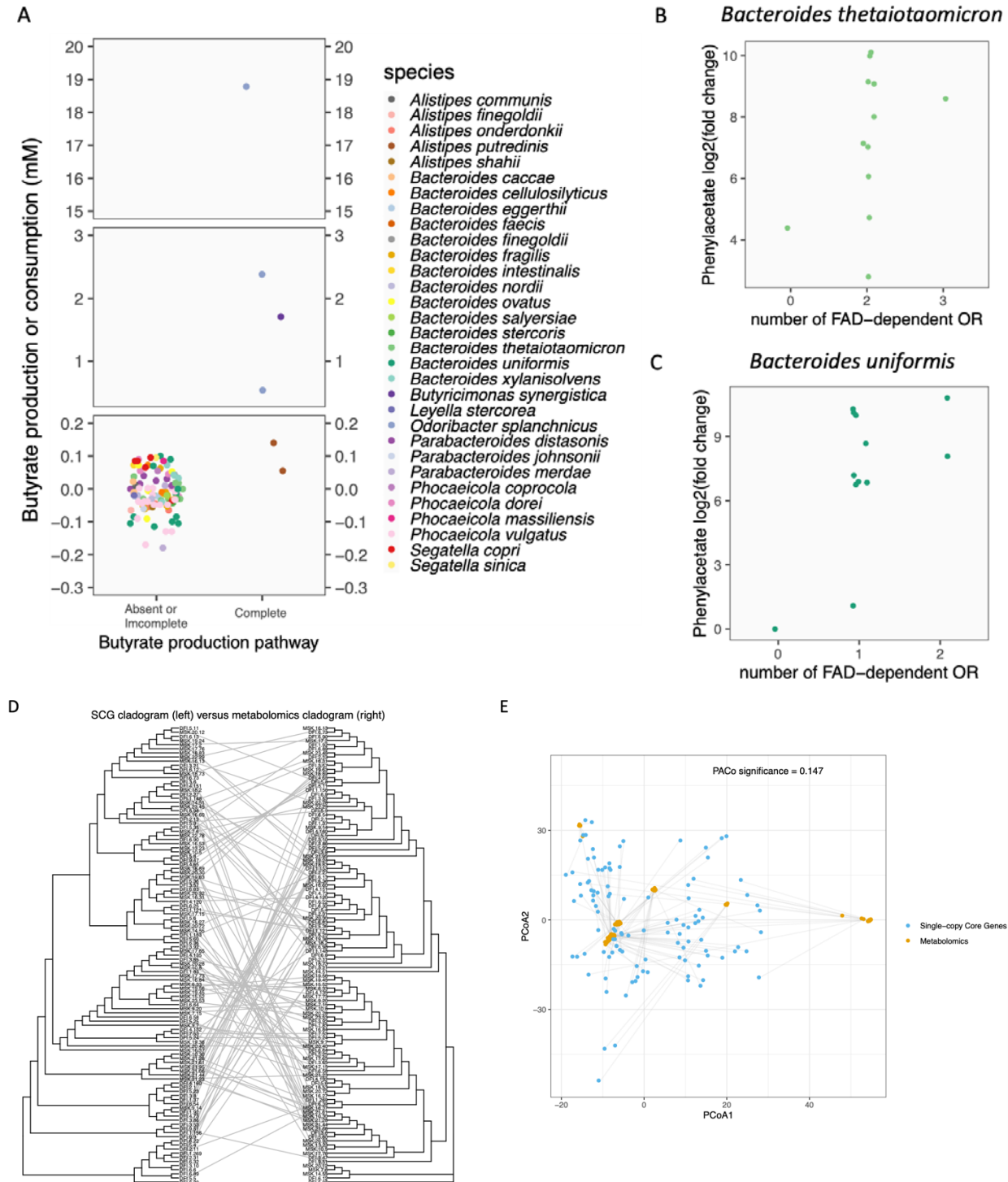
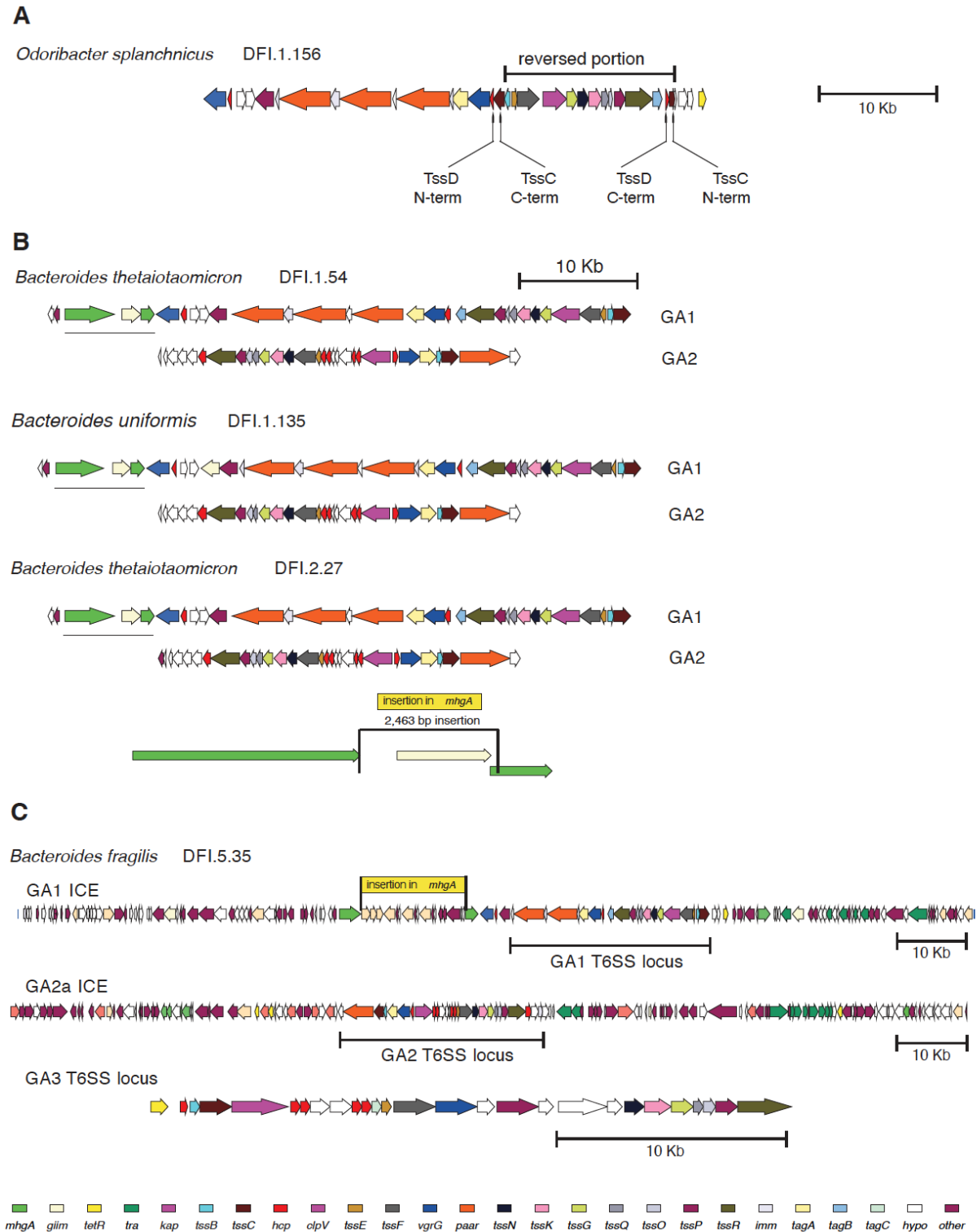


Figure S6. Inter- and intra-specific diversity of metabolome among Bacteroidales isolates.





**Figure S7. Genetic features of select Type VI secretion system loci among Bacteroidales isolates.**

# References

1. Eggerth, A.H., and Gagnon, B.H. (1933). The Bacteroides of human feces. *J Bacteriol* 25, 389-413. 10.1128/jb.25.4.389-413.1933.
2. Johnson, J.L. (1980). Specific strains of Bacteroides species in human fecal flora as measured by deoxyribonucleic acid homology. *Applied and environmental microbiology* 39, 407-413. 10.1128/aem.39.2.407-413.1980.
3. Salyers, A.A. (1984). Bacteroides of the human lower intestinal tract. *Annual Review of Microbiology* 38, 293-313. 10.1146/annurev.mi.38.100184.001453.
4. Hitch, T.C.A., Bisdorf, K., Afrizal, A., Riedel, T., Overmann, J., Strowig, T., and Clavel, T. (2022). A taxonomic note on the genus Prevotella: Description of four novel genera and emended description of the genera Hallella and Xylanibacter. *Systematic and Applied Microbiology* 45, 126354. 10.1016/j.syapm.2022.126354.
5. Blanco-Míguez, A., Gálvez, E.J.C., Pasolli, E., De Filippis, F., Amend, L., Huang, K.D., Manghi, P., Lesker, T.-R., Riedel, T., Cova, L., et al. (2023). Extension of the Segatella copri complex to 13 species with distinct large extrachromosomal elements and associations with host conditions. *Cell host & microbe* 31, 1804-1819.e1809. 10.1016/j.chom.2023.09.013.
6. Garcia-Lopez, M., Meier-Kolthoff, J.P., Tindall, B.J., Gronow, S., Woyke, T., Kyrpides, N.C., Hahnke, R.L., and Goker, M. (2019). Analysis of 1,000 type-strain genomes improves taxonomic classification of Bacteroidetes. *Front Microbiol* 10, 2083. 10.3389/fmicb.2019.02083.
7. Sakamoto, M., and Benno, Y. (2006). Reclassification of Bacteroides distasonis, Bacteroides goldsteinii and Bacteroides merdae as Parabacteroides distasonis gen. nov., comb. nov., Parabacteroides goldsteinii comb. nov. and Parabacteroides merdae comb. nov. *Int J Syst Evol Microbiol* 56, 1599-1605. 10.1099/ijs.0.64192-0.
8. Rautio, M., Eerola, E., Vaisanen-Tunkelrott, M.L., Molitoris, D., Lawson, P., Collins, M.D., and Jousimies-Somer, H. (2003). Reclassification of Bacteroides putredinis (Weinberg et al., 1937) in a new genus Alistipes gen. nov., as Alistipes putredinis comb. nov., and description of Alistipes finegoldii sp. nov., from human sources. *Syst Appl Microbiol* 26, 182-188. 10.1078/072320203322346029.
9. Porter, N.T., and Martens, E.C. (2017). The critical roles of polysaccharides in gut microbial ecology and physiology. *Annual Review of Microbiology* 71, 349-369. 10.1146/annurev-micro-102215-095316.

10. Glowacki, R.W.P., and Martens, E.C. (2020). If you eat it, or secrete it, they will grow: the expanding list of nutrients utilized by human gut bacteria. *J Bacteriol* 203, e00481-00420. 10.1128/JB.00481-20.
11. Furusawa, Y., Obata, Y., Fukuda, S., Endo, T.A., Nakato, G., Takahashi, D., Nakanishi, Y., Uetake, C., Kato, K., Kato, T., et al. (2013). Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. *Nature* 504, 446-450. 10.1038/nature12721.
12. Arpaia, N., Campbell, C., Fan, X., Dikiy, S., van der Veeken, J., deRoos, P., Liu, H., Cross, J.R., Pfeffer, K., Coffey, P.J., and Rudensky, A.Y. (2013). Metabolites produced by commensal bacteria promote peripheral regulatory T-cell generation. *Nature* 504, 451-455. 10.1038/nature12726.
13. Smith, P.M., Howitt, M.R., Panikov, N., Michaud, M., Gallini, C.A., Bohlooly, Y.M., Glickman, J.N., and Garrett, W.S. (2013). The microbial metabolites, short-chain fatty acids, regulate colonic Treg cell homeostasis. *Science* 341, 569-573. 10.1126/science.1241165.
14. van der Hee, B., and Wells, J.M. (2021). Microbial regulation of host physiology by short-chain fatty acids. *Trends Microbiol* 29, 700-712. 10.1016/j.tim.2021.02.001.
15. Werner, H., Rintelen, G., and Kunstek-Santos, H. (1975). [A new butyric acid-producing bacteroides species: *B. splanchnicus* n. sp. (author's transl)]. *Zentralbl Bakteriol Orig A* 231, 133-144.
16. Han, S., Van Treuren, W., Fischer, C.R., Merrill, B.D., DeFelice, B.C., Sanchez, J.M., Higginbottom, S.K., Guthrie, L., Fall, L.A., Dodd, D., et al. (2021). A metabolomics pipeline for the mechanistic interrogation of the gut microbiome. *Nature* 595, 415-420. 10.1038/s41586-021-03707-9.
17. Pudlo, N.A., Urs, K., Crawford, R., Pirani, A., Atherly, T., Jimenez, R., Terrapon, N., Henrissat, B., Peterson, D., Ziemer, C., et al. (2022). Phenotypic and genomic diversification in complex carbohydrate-degrading human gut bacteria. *mSystems* 7, e0094721-e0094721. 10.1128/msystems.00947-21.
18. Hayase, E., Hayase, T., Jamal, M.A., Miyama, T., Chang, C.-C., Ortega, M.R., Ahmed, S.S., Karmouch, J.L., Sanchez, C.A., Brown, A.N., et al. (2022). Mucus-degrading *Bacteroides* link carbapenems to aggravated graft-versus-host disease. *Cell* 185, 3705-3719.e3714. 10.1016/j.cell.2022.09.007.
19. Hehemann, J.H., Kelly, A.G., Pudlo, N.A., Martens, E.C., and Boraston, A.B. (2012). Bacteria of the human gut microbiome catabolize red seaweed glycans with carbohydrate-active enzyme updates from extrinsic microbes. *Proc Natl Acad Sci U S A* 109, 19786-19791. 10.1073/pnas.1211002109.



20. Pudlo, N.A., Pereira, G.V., Parnami, J., Cid, M., Markert, S., Tingley, J.P., Unfried, F., Ali, A., Varghese, N.J., Kim, K.S., et al. (2022). Diverse events have transferred genes for edible seaweed digestion from marine to human gut bacteria. *Cell Host Microbe* 30, 314-328 e311. 10.1016/j.chom.2022.02.001.
21. Devlin, A.S., and Fischbach, M.A. (2015). A biosynthetic pathway for a prominent class of microbiota-derived bile acids. *Nature chemical biology* 11, 685-690. 10.1038/nchembio.1864.
22. Yao, L., D'Agostino, G.D., Park, J., Hang, S., Adhikari, A.A., Zhang, Y., Li, W., Avila-Pacheco, J., Bae, S., Clish, C.B., et al. (2022). A biosynthetic pathway for the selective sulfonation of steroidal metabolites by human gut bacteria. *Nature microbiology* 7, 1404-1418. 10.1038/s41564-022-01176-y.
23. Sato, Y., Atarashi, K., Plichta, D.R., Arai, Y., Sasajima, S., Kearney, S.M., Suda, W., Takeshita, K., Sasaki, T., Okamoto, S., et al. (2021). Novel bile acid biosynthetic pathways are enriched in the microbiome of centenarians. *Nature* 599, 458-464. 10.1038/s41586-021-03832-5.
24. Le, H.H., Lee, M.T., Besler, K.R., Comrie, J.M.C., and Johnson, E.L. (2022). Characterization of interactions of dietary cholesterol with the murine and human gut microbiome. *Nat Microbiol* 7, 1390-1403. 10.1038/s41564-022-01195-9.
25. Yao, L., Seaton, S.C., Ndousse-Fetter, S., Adhikari, A.A., DiBenedetto, N., Mina, A.I., Banks, A.S., Bry, L., and Devlin, A.S. (2018). A selective gut bacterial bile salt hydrolase alters host metabolism. *Elife* 7, e37182. 10.7554/eLife.37182.
26. Frye, K.A., Piamthai, V., Hsiao, A., and Degnan, P.H. (2021). Mobilization of vitamin B(12) transporters alters competitive dynamics in a human gut microbe. *Cell Rep* 37, 110164. 10.1016/j.celrep.2021.110164.
27. Coyne, M.J., Roelofs, K.G., and Comstock, L.E. (2016). Type VI secretion systems of human gut Bacteroidales segregate into three genetic architectures, two of which are contained on mobile genetic elements. *BMC Genomics* 17, 58. 10.1186/s12864-016-2377-z.
28. Evans, J.C., McEneaney, V.L., Coyne, M.J., Caldwell, E.P., Sheahan, M.L., Von, S.S., Coyne, E.M., Tweten, R.K., and Comstock, L.E. (2022). A proteolytically activated antimicrobial toxin encoded on a mobile plasmid of Bacteroidales induces a protective response. *Nat Commun* 13, 4258. 10.1038/s41467-022-31925-w.
29. Coyne, M.J., Bechon, N., Matano, L.M., McEneaney, V.L., Chatzidaki-Livanis, M., and Comstock, L.E. (2019). A family of anti-Bacteroidales peptide toxins widespread in the human gut microbiota. *Nat Commun* 10, 3460. 10.1038/s41467-019-11494-1.

30. Ross, B.D., Verster, A.J., Radey, M.C., Schmidtke, D.T., Pope, C.E., Hoffman, L.R., Hajjar, A.M., Peterson, S.B., Borenstein, E., and Mougous, J.D. (2019). Human gut bacteria contain acquired interbacterial defence systems. *Nature* 575, 224-228. 10.1038/s41586-019-1708-z.
31. Whittle, G., Shoemaker, N.B., and Salyers, A.A. (2002). The role of *Bacteroides* conjugative transposons in the dissemination of antibiotic resistance genes. *Cell Mol Life Sci* 59, 2044-2054. 10.1007/s000180200004.
32. Coyne, M.J., Zitomersky, N.L., McGuire, A.M., Earl, A.M., and Comstock, L.E. (2014). Evidence of extensive DNA transfer between bacteroidales species within the human gut. *mBio* 5, e01305-01314. 10.1128/mBio.01305-14.
33. García-Bayona, L., Coyne, M.J., and Comstock, L.E. (2021). Mobile Type VI secretion system loci of the gut Bacteroidales display extensive intra-ecosystem transfer, multi-species spread and geographical clustering. *PLoS Genet* 17, e1009541. 10.1371/journal.pgen.1009541.
34. Zheng, W., Zhao, S., Yin, Y., Zhang, H., Needham, D.M., Evans, E.D., Dai, C.L., Lu, P.J., Alm, E.J., and Weitz, D.A. (2022). High-throughput, single-microbe genomics with strain resolution, applied to a human gut microbiome. *Science* 376, eabm1483. 10.1126/science.abm1483.
35. Oliveira, R.A., and Pamer, E.G. (2023). Assembling symbiotic bacterial species into live therapeutic consortia that reconstitute microbiome functions. *Cell Host Microbe* 31, 472-484. 10.1016/j.chom.2023.03.002.
36. Sorbara, M.T., Littmann, E.R., Fontana, E., Moody, T.U., Kohout, C.E., Gjonbalaj, M., Eaton, V., Seok, R., Leiner, I.M., and Pamer, E.G. (2020). Functional and genomic variation between human-derived isolates of Lachnospiraceae reveals inter- and intra-species diversity. *Cell Host Microbe* 28, 134-146 e134. 10.1016/j.chom.2020.05.005.
37. Eren, A.M., Kiefl, E., Shaiber, A., Veseli, I., Miller, S.E., Schechter, M.S., Fink, I., Pan, J.N., Yousef, M., Fogarty, E.C., et al. (2021). Community-led, integrated, reproducible multi-omics with anvi'o. *Nat Microbiol* 6, 3-6. 10.1038/s41564-020-00834-3.
38. Jain, C., Rodriguez, R.L., Phillippy, A.M., Konstantinidis, K.T., and Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* 9, 5114. 10.1038/s41467-018-07641-9.
39. Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., and Salzberg, S.L. (2004). Versatile and open software for comparing large genomes. *Genome Biol* 5, R12. 10.1186/gb-2004-5-2-r12.

40. Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068-2069. 10.1093/bioinformatics/btu153.
41. McInnes, L., Healy, J., Saul, N., and Großberger, L. (2018). UMAP: Uniform Manifold Approximation and Projection. *Journal of Open Source Software* 3. 10.21105/joss.00861.
42. Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., and Xu, Y. (2012). dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research* 40, W445-W451. 10.1093/nar/gks479.
43. Drula, E., Garron, M.-L., Dogan, S., Lombard, V., Henrissat, B., and Terrapon, N. (2021). The carbohydrate-active enzyme database: functions and literature. *Nucleic acids research* 50, D571-D577. 10.1093/nar/gkab1045.
44. Fletcher, C.M., Coyne, M.J., Villa, O.F., Chatzidaki-Livanis, M., and Comstock, L.E. (2009). A general O-glycosylation system important to the physiology of a major human intestinal symbiont. *Cell* 137, 321-331. 10.1016/j.cell.2009.02.041.
45. Fletcher, C.M., Coyne, M.J., Bentley, D.L., Villa, O.F., and Comstock, L.E. (2007). Phase-variable expression of a family of glycoproteins imparts a dynamic surface to a symbiont in its human intestinal ecosystem. *Proc Natl Acad Sci U S A* 104, 2413-2418. 10.1073/pnas.0608797104.
46. Coyne, M.J., and Comstock, L.E. (2008). Niche-specific features of the intestinal bacteroidales. *J Bacteriol* 190, 736-742. 10.1128/JB.01559-07.
47. Thompson, A.J., Williams, R.J., Hakki, Z., Alonzi, D.S., Wennekes, T., Gloster, T.M., Songsrirote, K., Thomas-Oates, J.E., Wrodnigg, T.M., Spreitz, J., et al. (2012). Structural and mechanistic insight into N-glycan processing by endo-alpha-mannosidase. *Proc Natl Acad Sci U S A* 109, 781-786. 10.1073/pnas.1111482109.
48. Zhang, M., Chekan, J.R., Dodd, D., Hong, P.Y., Radlinski, L., Revindran, V., Nair, S.K., Mackie, R.I., and Cann, I. (2014). Xylan utilization in human gut commensal bacteria is orchestrated by unique modular organization of polysaccharide-degrading enzymes. *Proc Natl Acad Sci U S A* 111, E3708-3717. 10.1073/pnas.1406156111.
49. Rogowski, A., Briggs, J.A., Mortimer, J.C., Tryfona, T., Terrapon, N., Lowe, E.C., Basle, A., Morland, C., Day, A.M., Zheng, H., et al. (2015). Glycan complexity dictates microbial resource allocation in the large intestine. *Nat Commun* 6, 7481. 10.1038/ncomms8481.
50. Wang, K., Pereira, G.V., Cavalcante, J.J., Zhang, M., Mackie, R., and Cann, I. (2016). *Bacteroides intestinalis* DSM 17393, a member of the human colonic

- microbiome, upregulates multiple endoxylanases during growth on xylan. *Sci Rep* 6, 34360. 10.1038/srep34360.
51. Pluvinage, B., Grondin, J.M., Amundsen, C., Klassen, L., Moote, P.E., Xiao, Y., Thomas, D., Pudlo, N.A., Anele, A., Martens, E.C., et al. (2018). Molecular basis of an agarose metabolic pathway acquired by a human intestinal symbiont. *Nat Commun* 9, 1043. 10.1038/s41467-018-03366-x.
52. Xu, J., Mahowald, M.A., Ley, R.E., Lozupone, C.A., Hamady, M., Martens, E.C., Henrissat, B., Coutinho, P.M., Minx, P., Latreille, P., et al. (2007). Evolution of symbiotic bacteria in the distal human intestine. *PLoS Biol* 5, e156. 10.1371/journal.pbio.0050156.
53. Stewart, R.D., Auffret, M.D., Roehe, R., and Watson, M. (2018). Open prediction of polysaccharide utilisation loci (PUL) in 5414 public Bacteroidetes genomes using PULpy. *bioRxiv*, 421024. 10.1101/421024.
54. Sato, Y., Atarashi, K., Plichta, D.R., Arai, Y., Sasajima, S., Kearney, S.M., Suda, W., Takeshita, K., Sasaki, T., Okamoto, S., et al. (2021). Novel bile acid biosynthetic pathways are enriched in the microbiome of centenarians. *Nature* 599, 458-464. 10.1038/s41586-021-03832-5.
55. Song, X., Sun, X., Oh, S.F., Wu, M., Zhang, Y., Zheng, W., Geva-Zatorsky, N., Jupp, R., Mathis, D., Benoist, C., and Kasper, D.L. (2020). Microbial bile acid metabolites modulate gut RORgamma(+) regulatory T cell homeostasis. *Nature* 577, 410-415. 10.1038/s41586-019-1865-0.
56. Louis, P., and Flint, H.J. (2017). Formation of propionate and butyrate by the human colonic microbiota. *Environ Microbiol* 19, 29-41. 10.1111/1462-2920.13589.
57. Zhu, Y., Dwidar, M., Nemet, I., Buffa, J.A., Sangwan, N., Li, X.S., Anderson, J.T., Romano, K.A., Fu, X., Funabashi, M., et al. (2023). Two distinct gut microbial pathways contribute to meta-organismal production of phenylacetylglutamine with links to cardiovascular disease. *Cell Host Microbe* 31, 18-32 e19. 10.1016/j.chom.2022.11.015.
58. Yu, F., Li, X., Feng, X., Wei, M., Luo, Y., Zhao, T., Xiao, B., and Xia, J. (2021). Phenylacetylglutamine, a novel biomarker in acute ischemic stroke. *Front Cardiovasc Med* 8, 798765. 10.3389/fcvm.2021.798765.
59. Yu, F., Feng, X., Li, X., Luo, Y., Wei, M., Zhao, T., and Xia, J. (2021). Gut-Derived Metabolite Phenylacetylglutamine and White Matter Hyperintensities in Patients With Acute Ischemic Stroke. *Front Aging Neurosci* 13, 675158. 10.3389/fnagi.2021.675158.

60. Nemet, I., Saha, P.P., Gupta, N., Zhu, W., Romano, K.A., Skye, S.M., Cajka, T., Mohan, M.L., Li, L., Wu, Y., et al. (2020). A cardiovascular disease-linked gut microbial metabolite acts via adrenergic receptors. *Cell* 180, 862-877 e822. 10.1016/j.cell.2020.02.016.
61. Roelofs, K.G., Coyne, M.J., Gentyala, R.R., Chatzidaki-Livanis, M., and Comstock, L.E. (2016). Bacteroidales secreted antimicrobial proteins target surface molecules necessary for gut colonization and mediate competition in vivo. *mBio* 7. 10.1128/mBio.01055-16.
62. Bao, Y., Verdegaaal, A.A., Anderson, B.W., Barry, N.A., He, J., Gao, X., and Goodman, A.L. (2021). A common pathway for activation of host-targeting and bacteria-targeting toxins in human intestinal bacteria. *mBio* 12, e0065621. 10.1128/mBio.00656-21.
63. Jiang, K., Li, W., Tong, M., Xu, J., Chen, Z., Yang, Y., Zang, Y., Jiao, X., Liu, C., Lim, B., et al. (2023). *Bacteroides fragilis* ubiquitin homologue drives intraspecies bacterial competition in the gut microbiome. *Nat Microbiol.* 10.1038/s41564-023-01541-5.
64. Wexler, A.G., Bao, Y., Whitney, J.C., Bobay, L.M., Xavier, J.B., Schofield, W.B., Barry, N.A., Russell, A.B., Tran, B.Q., Goo, Y.A., et al. (2016). Human symbionts inject and neutralize antibacterial toxins to persist in the gut. *Proc Natl Acad Sci U S A* 113, 3639-3644. 10.1073/pnas.1525637113.
65. Hecht, A.L., Casterline, B.W., Earley, Z.M., Goo, Y.A., Goodlett, D.R., and Bubeck Wardenburg, J. (2016). Strain competition restricts colonization of an enteric pathogen and prevents colitis. *EMBO Rep* 17, 1281-1291. 10.15252/embr.201642282.
66. Chatzidaki-Livanis, M., Geva-Zatorsky, N., and Comstock, L.E. (2016). *Bacteroides fragilis* type VI secretion systems use novel effector and immunity proteins to antagonize human gut Bacteroidales species. *Proc Natl Acad Sci U S A* 113, 3627-3632. 10.1073/pnas.1522510113.
67. Shumaker, A.M., Laclare McEneaney, V., Coyne, M.J., Silver, P.A., and Comstock, L.E. (2019). Identification of a fifth antibacterial toxin produced by a single *Bacteroides fragilis* strain. *J Bacteriol* 201. 10.1128/JB.00577-18.
68. Chatzidaki-Livanis, M., Coyne, M.J., and Comstock, L.E. (2014). An antimicrobial protein of the gut symbiont *Bacteroides fragilis* with a MACPF domain of host immune proteins. *Mol Microbiol* 94, 1361-1374. 10.1111/mmi.12839.
69. Chatzidaki-Livanis, M., Coyne, M.J., Roelofs, K.G., Gentyala, R.R., Caldwell, J.M., and Comstock, L.E. (2017). Gut symbiont *Bacteroides fragilis* secretes a

- eukaryotic-like ubiquitin protein that mediates intraspecies antagonism. *mBio* 8. 10.1128/mBio.01902-17.
70. Coyne, M.J., Roelofs, K.G., and Comstock, L.E. (2016). Type VI secretion systems of human gut Bacteroidales segregate into three genetic architectures, two of which are contained on mobile genetic elements. *BMC Genomics* 17, 58. 10.1186/s12864-016-2377-z.
71. Ofir, G., Melamed, S., Sberro, H., Mukamel, Z., Silverman, S., Yaakov, G., Doron, S., and Sorek, R. (2018). DISARM is a widespread bacterial defence system with broad anti-phage activities. *Nature Microbiology* 3, 90-98. 10.1038/s41564-017-0051-0.
72. Zhao, S., Lieberman, T.D., Poyet, M., Kauffman, K.M., Gibbons, S.M., Groussin, M., Xavier, R.J., and Alm, E.J. (2019). Adaptive evolution within gut microbiomes of healthy people. *Cell Host Microbe* 25, 656-667 e658. 10.1016/j.chom.2019.03.007.
73. Fogarty, E.C., Schechter, M.S., Lolans, K., Sheahan, M.L., Veseli, I., Moore, R.M., Kiefl, E., Moody, T., Rice, P.A., Yu, M.K., et al. (2024). A cryptic plasmid is among the most numerous genetic elements in the human gut. *Cell* 187, 1206-1222 e1216. 10.1016/j.cell.2024.01.039.
74. Groussin, M., Poyet, M., Sistiaga, A., Kearney, S.M., Moniz, K., Noel, M., Hooker, J., Gibbons, S.M., Segurel, L., Froment, A., et al. (2021). Elevated rates of horizontal gene transfer in the industrialized human microbiome. *Cell* 184, 2053-2067 e2018. 10.1016/j.cell.2021.02.052.
75. Zitomersky, N.L., Coyne, M.J., and Comstock, L.E. (2011). Longitudinal analysis of the prevalence, maintenance, and IgA response to species of the order Bacteroidales in the human gut. *Infect Immun* 79, 2012-2020. 10.1128/IAI.01348-10.
76. Sheahan, M.L., Coyne, M.J., Flores, K., Garcia-Bayona, L., Chatzidaki-Livanis, M., Sundararajan, A., Holst, A.Q., Barquera, B., and Comstock, L.E. (2023). A ubiquitous mobile genetic element disarms a bacterial antagonist of the gut microbiota. *bioRxiv*. 10.1101/2023.08.25.553775.
77. Husain, F., Veeranagouda, Y., Boente, R., Tang, K., Mulato, G., and Wexler, H.M. (2014). The Ellis Island Effect: A novel mobile element in a multi-drug resistant *Bacteroides fragilis* clinical isolate includes a mosaic of resistance genes from Gram-positive bacteria. *Mob Genet Elements* 4, e29801. 10.4161/mge.29801.
78. Wang, Y., Shoemaker, N.B., and Salyers, A.A. (2004). Regulation of a *Bacteroides* operon that controls excision and transfer of the conjugative



- transposon CTnDOT. *J Bacteriol* 186, 2548-2557. 10.1128/JB.186.9.2548-2557.2004.
79. Park, S.Y., Rao, C., Coyte, K.Z., Kuziel, G.A., Zhang, Y., Huang, W., Franzosa, E.A., Weng, J.K., Huttenhower, C., and Rakoff-Nahoum, S. (2022). Strain-level fitness in the gut microbiome is an emergent property of glycans and a single metabolite. *Cell* 185, 513-529 e521. 10.1016/j.cell.2022.01.002.
80. Dapa, T., Ramiro, R.S., Pedro, M.F., Gordo, I., and Xavier, K.B. (2022). Diet leaves a genetic signature in a keystone member of the gut microbiota. *Cell Host Microbe* 30, 183-199 e110. 10.1016/j.chom.2022.01.002.
81. Geva-Zatorsky, N., Sefik, E., Kua, L., Pasman, L., Tan, T.G., Ortiz-Lopez, A., Yanortsang, T.B., Yang, L., Jupp, R., Mathis, D., et al. (2017). Mining the human gut microbiota for immunomodulatory organisms. *Cell* 168, 928-943 e911. 10.1016/j.cell.2017.01.022.
82. Zhang, Z.J., Lehmann, C.J., Cole, C.G., and Pamer, E.G. (2022). Translating microbiome research from and to the clinic. *Annu Rev Microbiol* 76, 435-460. 10.1146/annurev-micro-041020-022206.