Ecology and Function of Bacterial Lipases within the Gut Microbiome

Thomas C. A. Hitch[1,*,#], Johannes M. Masson[1,#], Theresa Streidl[1], Thomas Fischöder[2], Ditte Hobbs-Grimmer[3], Debjyoti Ghosh[4], Lothar Elling[2], Nico Jehmlich[4], Martin von Bergen[4,5,6], Lesley Hoyles[7] and Thomas Clavel[1,*]

[1] Functional Microbiome Research Group, Institute of Medical Microbiology, University Hospital of RWTH Aachen, Aachen, Germany

[2] Institute for Biotechnology and Helmhotlz-Institute for Biomedical Engineering, RWTH Aachen University, Aachen, Germany

[3] Hugh Sinclair Unit of Human Nutrition, Department of Food and Nutritional Sciences, University of Reading, Reading, United Kingdom

[4] Department of Molecular Toxicology, Helmholtz-Centre for Environmental Research GmbH (UFZ), Leipzig, Germany

[5] Institute of Biochemistry, Faculty of Biosciences, Pharmacy and Psychology, University of Leipzig, 04103 Leipzig, Germany

[6] German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, 04103 Leipzig, Germany

[7] Department of Biosciences, School of Science & Technology, Nottingham Trent University, Nottingham, United Kingdom

[#]These authors share first authorship

*To whom correspondence should be addressed:

thitch@ukaachen.de (bioinformatics); tclavel@ukaachen.de (all other aspects of the work)

Abstract

Triglycerides are a major dietary component and excessive intake has been linked to many health conditions. While the impact of host lipases on fat degradation and absorption have been well documented, bacterial lipases have been largely ignored. Here, we identified the landscape of microbial lipases within the gut of mice, pigs, and humans. In total, 373 microbial lipases were identified and formed two major clusters based on protein sequences, with the sequences containing highly conserved motifs. Metagenomic and cultivation approaches identified lipase-positive bacteria as diverse, occurring within seven phyla, and abundant within the human gut (9.1% ± 4.6%). Pathway analysis in lipase-positive species identified a low prevalence of fatty acid utilisation and a higher frequency of glycerol metabolism. In vitro testing with stable isotopes confirmed that human gut bacteria can utilise the released glycerol. Overall, this study highlights the role of bacterial lipolytic activity in triglycerides metabolism within the gut.

Keywords: diet, triglycerides, glycerol, fatty acids, gut microbiome, bacterial lipases

Introduction

The gut microbiota is a complex ecosystem that consists of hundreds of microbial species collectively producing millions of proteins [1–5]. Mechanistic understanding of their role in health and disease is made difficult due to this complexity, leading to the use of synthetic communities [6, 7]. Even when the complexity of the microbiota is reduced, mechanistic understanding is hampered by the lack of meaningful annotations for a majority of microbial functions within the gut [8]. One approach for improving functional annotations is the targeted analysis of specific functions, as previously conducted for β-glucuronidases within both the human and murine gut [9, 10] and also for bile acid metabolism [11].

The prevailing knowledge about fat absorption is that it occurs within the small intestine, specifically in the proximal region [12, 13], according to studies conducted in the 1950s [12, 14, 15], 1960s [16] and 1970s [13, 17–19]. However, in light of findings which suggest fat absorption differs greatly depending on fat amount [20]

54 and type provided [21], as well as the potential role of the gut microbiome [22], these assumptions should be
55 re-evaluated and confirmed to provide comprehensive overview of mammalian fat absorption. To date, the role
56 of bacterial lipases on fat degradation and host metabolism has only been investigated a handful of times. Oral
57 administration of a transgenic lipase-producing bacterium was shown to significantly increase fat absorption and
58 improved steatorrhea in pigs that had undergone pancreatic ligation [23]. Carp (*Cyprinus carpio*) fed with LipG1,
59 a bacterial lipase isolated from the gut of carp, significantly increased lipolytic activity and weight gain when fed
60 palm oil, a cheap and industrially advantageous feed [24]. Whilst these studies highlighting both, the health and
61 economic benefits to better understand bacterial lipases and fat degradation, bacterial lipases within the gut are
62 currently understudied compared to host-derived lipases [25–27].
63 In this paper, we studied bacterial lipases present within the gut using a combination of metagenomics,
64 bacterial cultivation, and biochemical methods. This includes identification of the diversity of lipase enzymes,
65 taxonomic diversity of lipase-positive species, and metabolic implications for them.
66

67 # Results

68 **Specific bacterial lipase repertoire within the mammalian gut**

69 While functionally diverse, lipases and esterases share structural and sequence similarity as they are mem-
70 bers of the protein family of alpha/beta hydrolases [28]. To ensure correct identification of lipases within gut
71 microbiota gene catalogues (human, mouse, pig), annotation was conducted against three reference databases.
72 False-positive identification was prevented by selecting proteins with matches in two of the three databases.
73 Multiple conserved structural features exist within lipases, including the alpha-beta fold and motifs such as
74 GxSxG at the active site [29]. Hence filtering was conducted based on the requirement that proteins must con-
75 tain this motif. This reduced the number of total matches by >45% (**Table 1**).
76
77
78
79

80 **Table 1: Annotation of gut bacterial protein catalogues for lipases.** The table indicates the total number of proteins within each host
81 catalogue, in million, followed by the number identified as being lipases throughout the different steps of the workflow.

Database	**Human (9.9 million)** [2]	**Mouse (4.5 million)** [4]	**Pig (7.7 million)** [3]
LED	10,988	5,658	6,490
EC	14	5	20
EGGNOG	239	274	192
Annotation in ≥2 databases	175	250	158
GxSxG motif	116	178	79

82

83 The identified gut lipases were placed into the latest classifcation system for bacterial lipases [30] based
84 on sequence identity using an all-against-all annotation approach (**Figure 1a**). While some gut bacterial lipases
85 were related to family 1.1, 1.2, 1.3, 1.4 and 7, the majority grouped together to form two loosely connected
86 clusters, labelled as Cluster 1 and 2. These two clusters contained lipases originating from the microbiota of
87 humans, mice, and pigs, while some smaller clusters were host-specific.

88 Due to the high sequence identity observed between proteins within Cluster 1 and 2, further sequence
89 analysis was conducted to identify conserved amino acid motifs. Within Cluster 1, the motifs AHSKGG and
90 TTxxTPH both occurred in all 98 member proteins (**Figure 1b**). Both motifs were also highly conserved within
91 Cluster 2, occurring in ≥90% of the 194 member proteins (**Figure 1c**). Although the GxSxG motif was used
92 during identification of these lipases, the conserved version was identified to contain isoleucine and histidine as
93 the variable second and forth positions, respectively. The type proteins of each lipase family/subfamily were
94 scanned for the presence of the novel motifs to identify if they are unique markers of gut lipases. AHSKGG was
95 identified as unique to lipase Cluster 1 and 2 while the TTxxTPH motif was identified within the type proteins for
96 lipase family 1.2 and 1.5.

97 To better understand the variability within this collection of gut bacterial lipases, their features were
98 assessed. The length of each protein was studied in relationship to its host of origin (**Figure 1d**). The mean
99 length was 410 amino acids, ranging from 158 to 706. Interestingly, 40.5% of pig bacterial lipases were less
100 than 300 amino acids long. However, out of the ten longest lipases, four originated from the pig gut, and the
101 other six from humans. The pH preference of each lipase was predicted, with the majority being alkaline (**Figure**
102 **1e**). The percentage varied between hosts, with the murine gut having the highest percentage of alkaline lipases
103 (94.9%) while the porcine gut had the lowest (69.6%).

104 The localisation of enzymes is highly important as secreted enzymes can diffuse away from the cell,
105 while membrane-bound enzymes are limited to substrates within the vicinity of the cells. Therefore, the presence
106 of signal peptides and transmembrane regions was determined. Most bacterial lipases (84.5%) were identified
107 to not contain any known signalling peptide, indicating most are not secreted extracellularly (**Figure 1f**). Of those
108 that did contain a signalling peptide, the most common was the Sec signal peptide (Sec/SPI), followed by the
109 lipoprotein signal peptide (Sec/SPII). No enzymes were observed to contain the Tat signal peptide (Tat/SPI).
110 The majority of lipases were observed to contain transmembrane regions, although the distribution was host-
111 specific (60.8 - 88.8%) (**Figure 1g**). Up to seven transmembrane regions were predicted within a single lipase,
112 although this was observed in only a single lipase from the human gut. The most common number of transmem-
113 brane regions was five (19.0 - 49.4%), followed by one (1.7 - 25.0%).
114

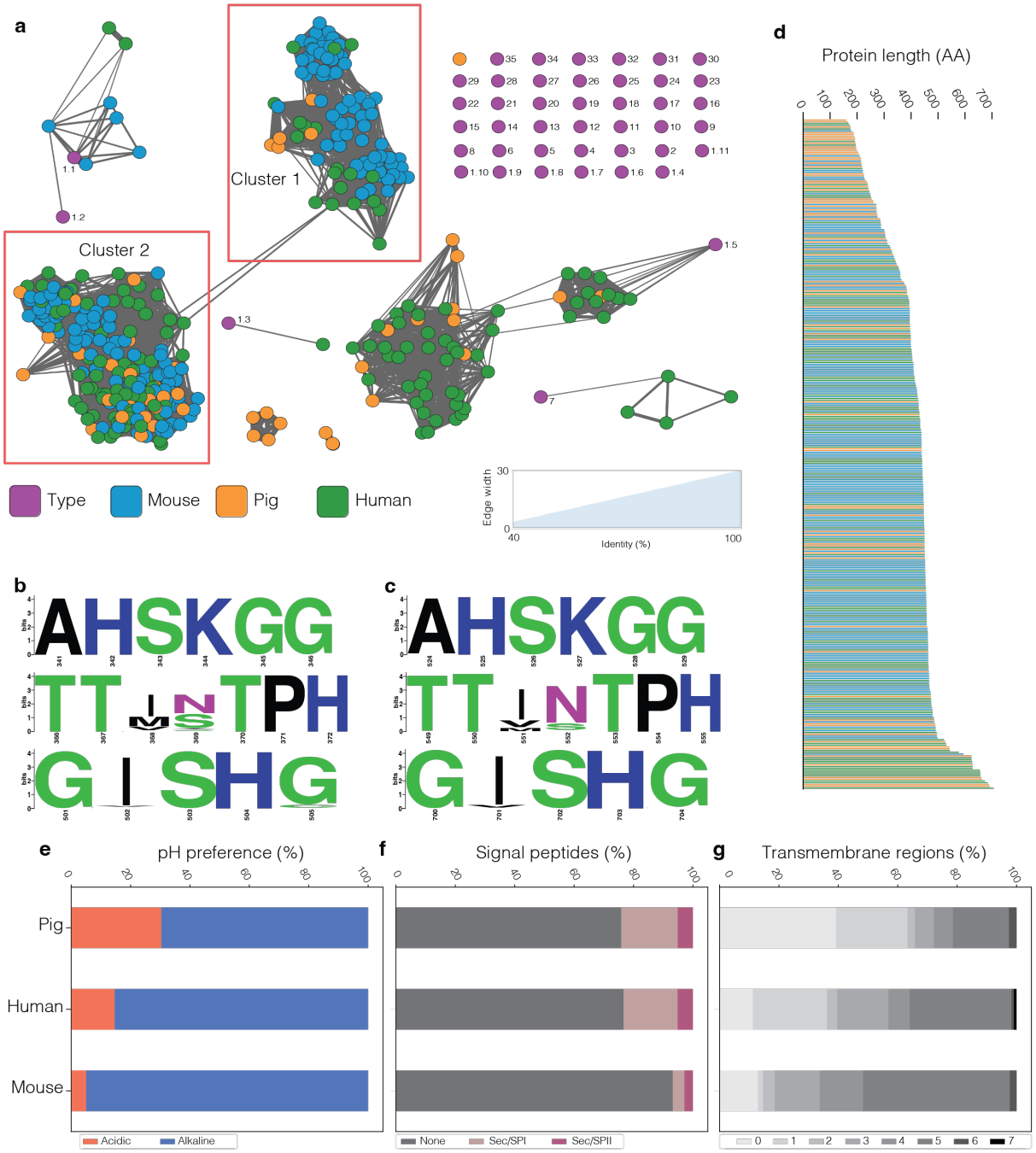


Figure 1: Protein sequence analysis of gut bacterial lipases. a. Network analysis of sequence identity shared between the identified bacterial lipases. Each node represents a single lipase coloured based on its host of origin. Edge lengths and widths are proportional to the sequence identity shared between the two nodes. Cluster 1 and 2 are highlighted with red boxes. **b - c.** The conserved motifs identified within Cluster 1 (**b**) and Cluster 2 (**c**) along with their respective positions in the alignment and coloured based on amino acid features. **d**. Protein length variation across the lipases, coloured based on their host of origin. **e - g.** The features of these enzymes were predicted using AcalPred [31] for determining if enzymes were alkaline or acidic (**e**), SignalP [32] for prediction of signal peptide sequences (**f**) and TMHMM [33] for prediction of transmembrane regions (**g**).

123

124 **Taxonomic and functional insights into lipase-positive bacteria**

125 Metagenomic assembled genomes (MAGs) allow comprehensive insights into microbial diversity, in-
126 cluding species that are currently uncultured [34–36]. Proteins from 154,723 MAGs previously reconstructed
127 using metagenomes from different human body sites [36] were annotated using the same lipase annotation
128 method as for the gene catalogues. In total, 13,677 lipases were identified within 11,011 of the 154,723 MAGs
129 (7.1%). Most of these proteins were assigned to Cluster 2 (n = 7,604). However, many were also identified as
130 being members of the sub-families within lipase family 1 (n = 1,487) [37]. Of those assigned to Cluster 2, 98.3%
131 contained the three motifs (GxSxG, TTxxTPH and AHSKGG) along with 98.5% of the 653 proteins assigned to
132 Cluster 1.

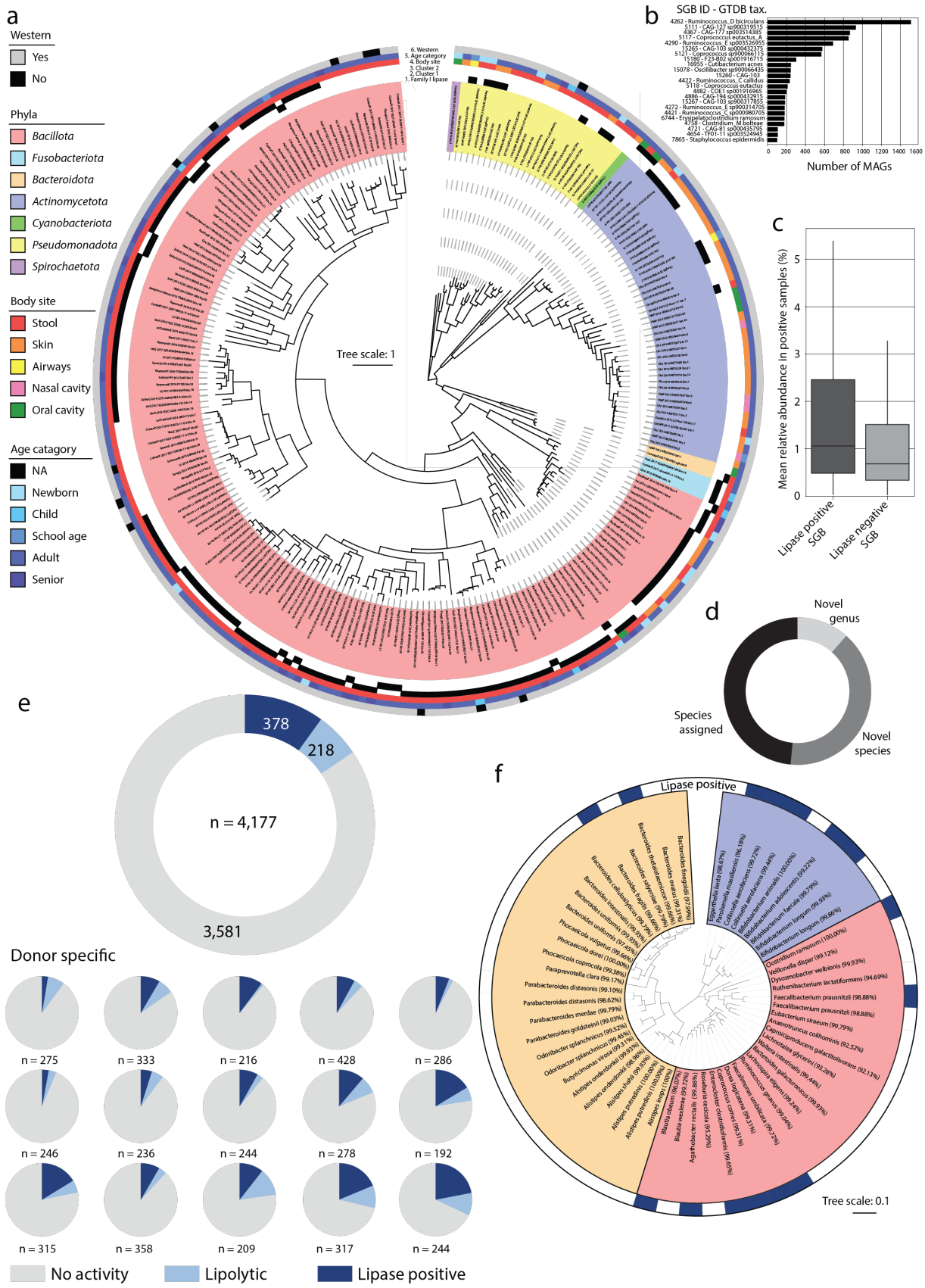
133 To further study the taxonomy of lipase-positive species, the MAGs were grouped based on average
134 nucleotide identity (ANI) values, reducing the 154,723 MAGs into 4,930 species genome bins (SGB). For each
135 SGB, a single representative MAG was selected, of which 235 (4.8%) were identified to contain a lipase (**Figure**
136 **2a**). Taxonomically, these lipase-positive species belong to 37 bacterial families across seven phyla: *Bacillota*,
137 *Actinomycetota*, *Pseudomonadota*, *Fusobacteriota*, *Bacteroidota*, *Spirochaetota,* and *Cyanobacteriota*. Of the
138 most prevalent SGBs (based on the number of MAGs assigned to the bin), only 8 of the top 23 could be assigned
139 to a named species, whilst the rest represented metagenomically inferred species of novel lineages (**Figure**
140 **2b**). The mean relative abundance of lipase-positive species was nearly twice that of the lipase-negative species
141 (1.06% vs 0.68%), suggesting that they are dominant members of human microbiomes (**Figure 2c**). Although
142 dominant, almost 50% of the SGBs were identified to represent unknown species or genera according to GTDB-
143 TK (**Figure 2d**), indicating that many lipase-positive species lack a cultured representative. The majority of the
144 SGB representatives originated from gut samples (n = 176, 74.9%) and represented the major phyla from the
145 gut (*Bacillota*, *Bacteroidota*, *Actinomycetota*, *Pseudomonadota,* and *Fusobacteriota*). The most commonly re-
146 constructed lipase-positive species was *Ruminococcus_D bicirculans*, renamed to *Hominimerdicola aceti* [38],
147 a dominant commensal of the human gut [39] which contains a lipase belonging to Cluster 2 (**Figure 2b**).

148 The second largest body site from which lipase-positive species were identified was the skin (45 SGB,
149 19.1%), of which the majority (29 SGB, 64.4%) belonged to the phylum *Actinomycetota* (**Figure 2a**). Additionally,
150 nine *Bacillota* species belonging to the genus *Staphylococcus* were identified. *Staphylococcus* species have

151 previously been shown to include lipase producers from the skin microbiota [40]. Both *Staphylococcus epider-*
152 *midis* and *Cutibacterium acnes*, common skin colonisers and some of the most prevalent reconstructed species
153 within the utilised dataset (**Figure 2b**), were identified to be lipase-positive, consistent with previous reports
154 about these species [41].

155 In addition to the culture-independent study of MAGs, we determined the lipolytic activity of 4,177 bac-
156 terial colonies from faecal samples from 15 healthy human adults. Of these, 596 (13.5%) were assumed to
157 express either a lipase or esterase based on screening using tributyrin agar (**Figure 2e**). Lipase activity was
158 confirmed by screening the 596 isolates by re-streaking on rhodamine B agar. In total, 378 (8.6%) were lipase-
159 positive, representing between 2.5 and 18% of the isolates from each donor. All donors were found to contain
160 lipase-positive bacteria with a mean prevalence of lipase-positive colonies of 9.1% ± 4.6% (**Supplementary**
161 **Table 1**), confirming lipase activity as a dominant phenotype.

162 To identify the taxonomic diversity of lipase-positive species in the human gut, a curated collection of
163 57 phylogenetically diverse strains representing 31 dominant genera was studied further [42]. Out of these
164 strains, 15 were clearly identified to be lipase-positive based on the rhodamine B agar assay (**Figure 2f**). As
165 this collection contained multiple strains for 6 species (>98.7% 16S rRNA gene sequence identity to the same
166 assigned species), the consistency of lipase activity between strains could be assessed. While strains assigned
167 to *Collinsella aerofaciens, Bifidobacterium longum*, *Alistipes putredinis*, and *Odoribacter splanchnicus* showed
168 consistent results, strain-level variability was observed in *Faecalibacterium prausnitzii,* with one being positive
169 and the other negative.
170


a
Western
Yes
No
Phyla
Bacillota
Fusobacteriota
Bacteroidota
Actinomycetota
Cyanobacteriota
Pseudomonadota
Spirochaetota
Body site
Stool
Skin
Airways
Nasal cavity
Oral cavity
Age catagory
NA
Newborn
Child
School age
Adult
Senior
6. Western
5. Age category
4. Body site
3. Cluster 2
2. Cluster 1
1. Family I lipase
Tree scale: 1
b
SGB ID - GTDB tax.
Number of MAGs
c
Mean relative abundance in positive samples (%)
Lipase positive SGB
Lipase negative SGB
d
Novel genus
Species assigned
Novel species
e
378
218
n = 4,177
3,581
Donor specific
n = 275
n = 333
n = 216
n = 428
n = 286
n = 246
n = 236
n = 244
n = 278
n = 192
n = 315
n = 358
n = 209
n = 317
n = 244
No activity
Lipolytic
Lipase positive
f
Lipase positive
Tree scale: 0.1

171

172 **Figure 2: Culture-independent and -dependant diversity of lipase-positive species. a.** Phylogenomic tree of 235 representative lipase-
173 positive SGBs from human microbiomes generated using PhyloPhlAn [43]. External rings depict important features of SGBs as follows
174 (inner to outer rings): lipase belonging to either family I, Cluster 1 or Cluster 2 (black, encoded; white, absent); metadata associated to the
175 sample the SGB representative was reconstructed from: body site, age of the host, and if from a Western country or not (see colour legends).
176 **b**. The taxonomy and size (number of MAGs) of all lipase-positive SGBs with a size > 100. **c**. Mean relative abundance of lipase-positive
177 and -negative SGBs across 7,783 human gut metagenomes. **d**. Novelty of lipase-positive SGBs according to their GTDB taxonomy. **e.**
178 Occurrence of lipolytic and lipase-positive bacteria across 4,177 isolates (top) or per donor (N = 15) (bottom). **f.** Phylogenetic tree (16S
179 rRNA gene sequences) of a collection of 57 human gut isolates with their lipase activity, as determined by rhodamine B agar testing.
180

181 **Characterization of a novel bacterial lipase from the gut**

182 The lack of experimental validation for bacterial functions has been highlighted as one of the major limita-
183 tions of metagenomic studies [8]. Hence, we utilised cultured isolates and biochemical testing to provide *in vitro*
184 evidence for their functional assignment. Genomes from in-house isolates [44] were searched against the pre-
185 dicted lipases to identify cultured host species. A complete sequence match (100% identity and coverage) was
186 identified between a member of Cluster 2 and a protein within *Clostridium symbiosum* DSM 29356, a prevalent
187 species in the human gut (detected in 28.10% (281/1,000) of human gut metagenomes) [45]. The isolate was
188 cultured in a lipid-rich medium (see methods) for 24h and tested for lipolysis on rhodamine B agar. The fluores-
189 cence detected from the *C. symbiosum* cell solution, compared to the negative control (*Escherichia coli* DSM
190 28618) and the positive control (porcine lipase) confirmed its lipase activity (**Figure 3a**). To determine if this
191 enzymatic activity was cell-bound or excreted, the cell pellet and supernatant were tested separately, with ac-
192 tivity being identified only with the cell pellet (**Figure 3b**). Sequence analysis confirmed that the protein contains
193 the GxSxG, AHSKGG, and TTxxTPH motifs, as well as five transmembrane regions at its N terminus, suggest-
194 ing it is a membrane-bound lipase (**Figure 3c**). The predicted model of the lipase had a low normalised B-factor
195 value for the predicted transmembrane region, indicating its flexibility and therefore supporting its prediction as
196 a transmembrane region (**Figure 3d**).

197 Structural modelling predicted that removal of the transmembrane region at amino acid 150 of the amino
198 acids sequence would leave a functional and stable enzymatic region (**Figure 3e-g**). Comparison of the pre-
199 dicted structure of the truncated lipase, LipΔ150, against the PDB database using COFACTOR confirmed that
200 the five top matches, based on TM-score, were lipases. Functional annotation of the structure assigned the EC

201 as 3.1.1.3 and the gene ontology term 'triglyceride lipase activity (GO:0004806)'. Ligand binding was also pre-
202 dicted, based on a TM-SCORE of 0.81, of 5LipA bound to octyl-phosphinic acid 1,2-bis-octylcarbamoyloxy-ethyl
203 ester (OCP), a structural homolog for triglycerides which binds irreversibly to the catalytic serine, facilitating
204 crystallography (**Figure 3e**). The AHSKGG, TTxxTPH and GxSxG motifs were all identified to co-localise with
205 OCP within the lipase structure (**Figure 3f**). Co-localisation of the motifs and OCP was studied further by space-
206 filled modelling the motifs and OCP molecule (**Figure 3g**). The conserved nature of these motifs, along with
207 their co-localisation with OCP suggests that they are not only hallmarks of gut lipase clusters but are of functional
208 importance.

209 The LipΔ150 sequence was cloned with a poly-histidine tag into a pET-15b expression vector in *E. coli*
210 BL21(DE3). Once expressed and purified, the enzyme was biochemically characterized to test its predicted
211 function. LipΔ150 was confirmed to be a lipase, as activity was detected against all lengths of fatty acids tested,
212 ranging from C2 to C18 (**Figure 3h**). However, a preference was shown for medium-chain fatty acids under the
213 conditions tested, as the highest concentration of released fatty acids was observed with C8 (133 ± 3 μM), and
214 the least was with C2 (57 ± 5 μM).

215 Due to the use of the GxSxG motif in the prediction of these enzymes, PMSF was used to identify if LipΔ150
216 is a serine-hydrolase. PMSF inhibited 100% of the lipase activity and led to mild precipitation of the protein
217 (**Figure 3i**). Lipolytic activity was also observed to be temperature-dependent, with maximum activity occurring
218 at 45 °C (**Figure 3j**). The possibility of pH altering lipolytic activity was also investigated, with pH of 6 or below
219 preventing activity and maximum activity being observed at pH 9 (**Figure 3k**). As LipΔ150 represents a novel
220 lipase based on its low sequence similarity (<40%) to all existing families, we propose it as the first member of
221 the new lipolytic enzyme family XXXVI [30].

222

223

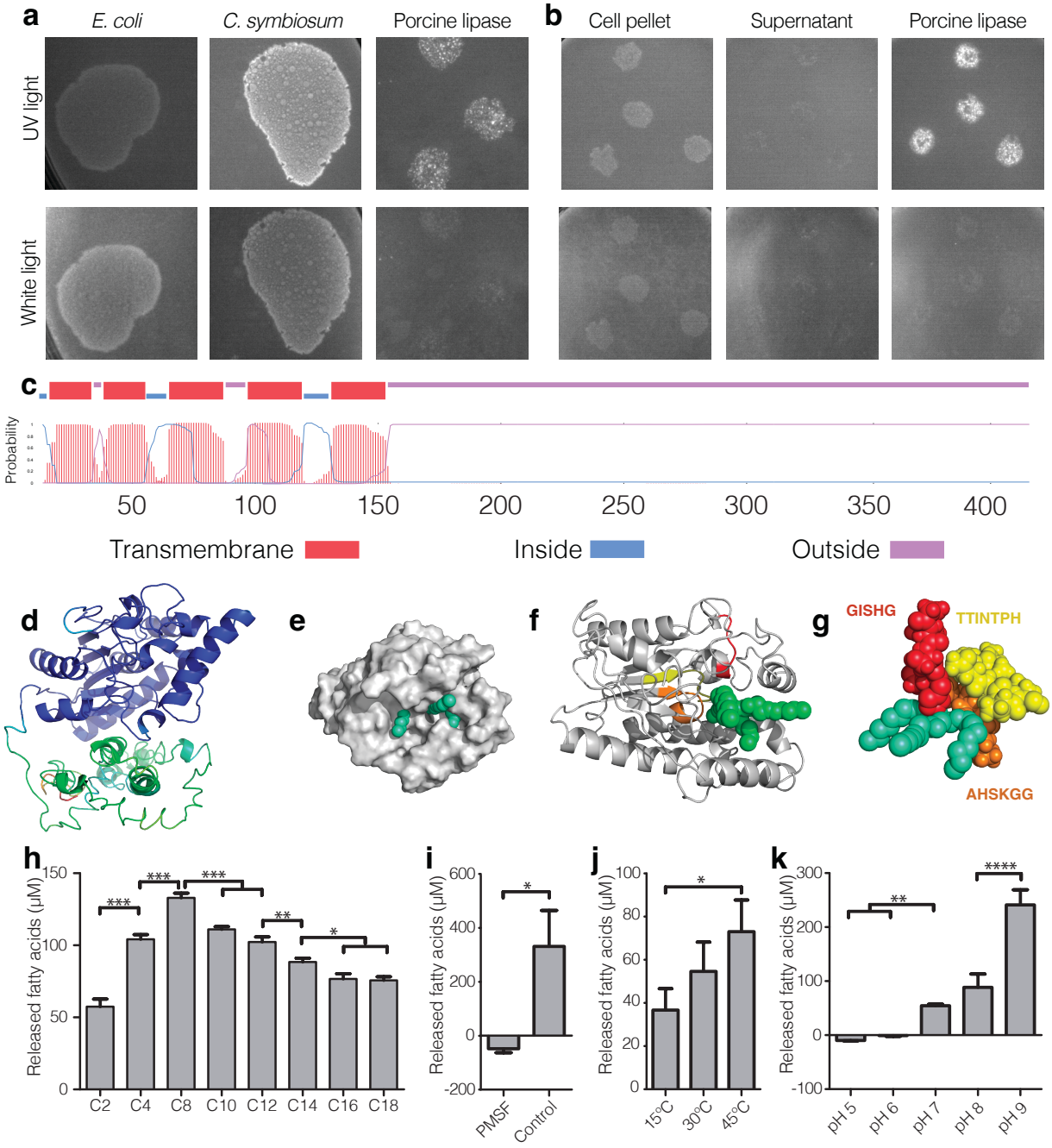


Figure 3: Structural modelling and biochemical characterisation of LipΔ150. a. Rhodamine B agar assay for *C. symbiosum* DSM 29356, *E. coli* DSM 28618 (negative control) and porcine lipase (positive control) viewed under both UV (60 ms) and white light (400 ms). Lipase activity is indicated by fluorescence under UV light, seen as white in the images for *C. symbiosum* and porcine lipase (top). **b**. Rhodamine B agar assay for the cell pellet and supernatant fractions of *C. symbiosum*, cultured for 7 days, along with the porcine lipase-positive control, viewed under both UV (top) and white light (bottom). **c**. Prediction of transmembrane regions via TMHMM [33]. **d**. Structural model of the complete lipase protein coloured according to the B-normalised value, indicative of structural flexibility, from green (low = -1.16) to blue (high = 2.52). **e-g**. Structural modelling of LipΔ150 bound to the substrate OCP (green) on the surface (**e**), using a ribbon

model (**f**) and highlighting only the conserved motifs (**g**). In both **f** and **g**, the conserved motifs are coloured red for GISHG, orange for AHSKGG, and yellow for TTINTPH. **h-k**. Biochemical characterisation of LipΔ150 was conducted using a range of fatty acid lengths (**h**), inhibition via PMSF (**i**) temperature-dependant activity (**j**), and pH (**k**). Triplicates were tested for statistical significance (t-test with Bonferroni multiple test correction) is represented as follows: * < 0.05, ** < 0.01, *** < 0.005.

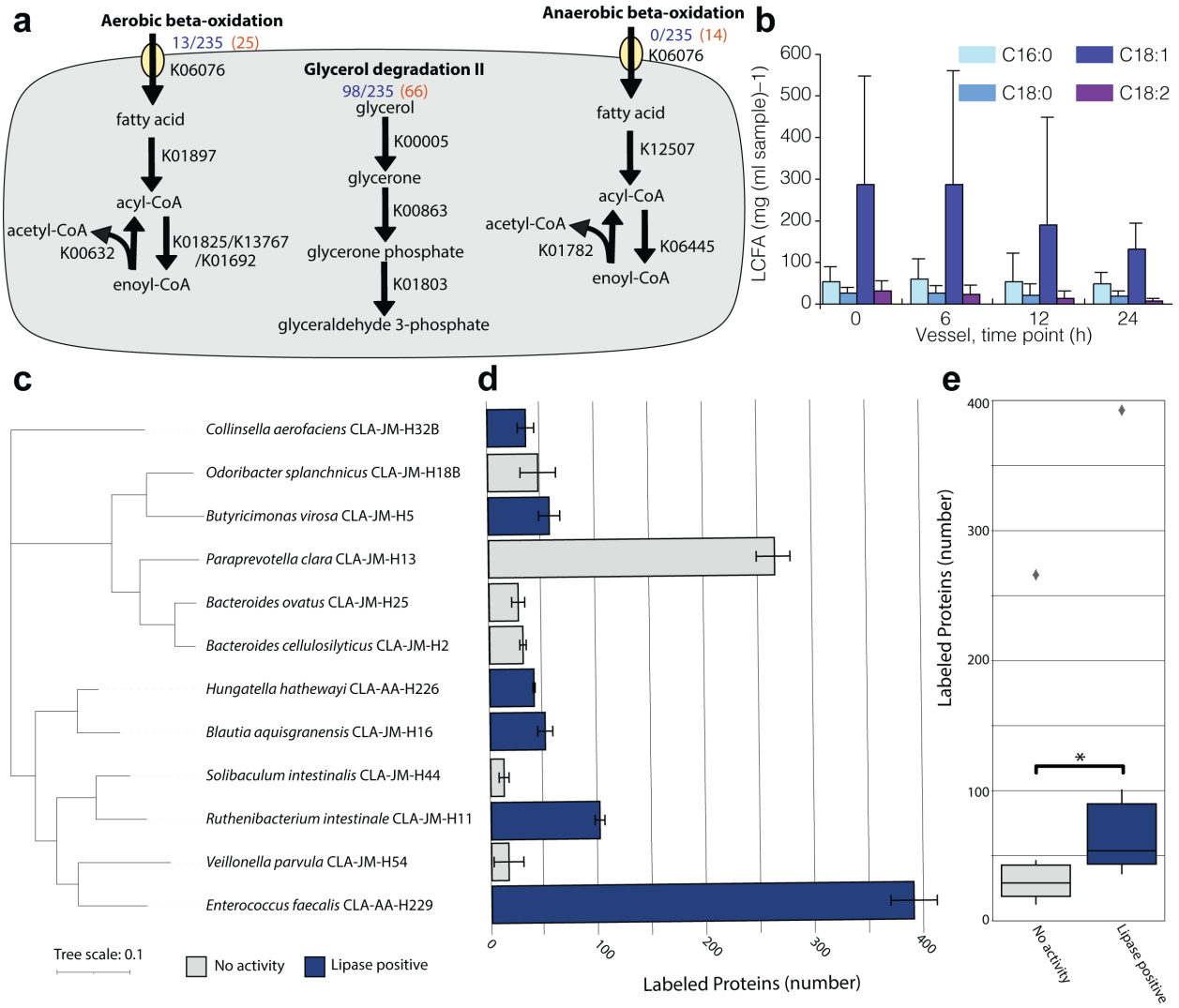
Lipase-positive bacteria preferentially utilise glycerol rather than fatty acids

The presence of lipases within a substantial fraction of gut-colonizing-bacteria suggests a metabolic benefit to triglyceride degradation, such as for energy generation or biomass production [46]. To investigate this further, we examined the metabolic potential of the 235 lipase-positive SGBs (**Figure 4a**). Lipolytic activity releases two products: free fatty acids that can undergo beta-oxidation, and glycerol which can be metabolized to enter glycolysis. Three models of lipid catabolism were studied [47]. Whilst aerobic and anaerobic models of beta-oxidation were examined [48], none of the SGBs contained the anaerobic beta-oxidation pathway and 5.5% of lipase-positive SGBs contained the complete aerobic beta-oxidation pathway. Conversely, 41.7% of SGBs contained the glycerol degradation II pathway. Additionally, when genomes containing incomplete pathways were included, a further 28.1% of SGBs were identified to utilize glycerol.

Due to the high-energy content of long-chain fatty acids, we initially assessed the ability of complex faecal communities to metabolise fatty acids (**Figure 4b**). For this, the faecal samples of five donors were incubated in GMM medium supplemented with 1g olive oil over 24 hours in duplicate. After 24 hours, the amount of $C_{18:1}$ and $C_{18:2}$ within the system was reduced from 149.2 ± 225.4 mg/ml to 69.2 ± 77.5 mg/ml (paired t-test, p=0.26) and 18.5 ± 21.0 mg/ml to 4.7 ± 4.6 mg/ml (paired t-test, p=0.06), respectively, although neither reduction was significant. The saturated fatty acids ($C_{16:0}$, $C_{18:0}$) remained at consistent levels throughout the 24 hours, suggesting the microbiota does not metabolise these substrates.

The genomic evidence suggests the microbiota produces lipases to access the glycerol subunits. To test this hypothesis, we studied the ability of six lipase-positive and six lipase-negative strains to utilise glycerol. We selected phylogenetically diverse species from both groups to ensure the phenotypes are not lineage-dependent (**Figure 4c**). To determine each strain's ability to utilise glycerol, we incubated the strains in YCFA medium supplemented with 1% glycerol, of which 5% was C_{13} isotope labelled. Due to the isotope labelling, strains that utilised glycerol would have a fraction of their proteins labelled; hence proteomics coupled with protein stable isotope probing was used to determine the number of proteins labelled by each strain (**Figure 4d**). Interestingly,

261 all strains were observed to contain labelled proteins, suggesting all could utilise glycerol to some extent. Com-
262 parison of the strains that were lipase-positive vs -negative strains indicated that the positive strains utilised
263 glycerol to a greater extent, leading to a higher number of labelled proteins (**Figure 4e**).
264



265

266 **Figure 4: Glycerol utilisation by lipase-positive bacteria. a.** The occurrence of each module for lipolytic catabolism across lipolytic SGBs
267 is stated in blue; next to this, the number with the incomplete pathway (>50% of the total KOs in the pathway) is in orange. **b**. Degradation
268 of long-chain fatty acids (LCFA) present in GMM and olive oil by faecal microbiota (n = 5) over 24 hours. The detected amount of each
269 LCFA is plotted for each measured time-point during batch fermentation. Each microbiota originated from a different donor. **c**. Phylogenetic
270 (16S rRNA gene sequence) tree of 12 bacterial strains isolated from the human gut, where 6 are lipase-positive (blue bars in panel c). **d**.
271 Number of labelled proteins from each strain after incubation for 24 hours with C_{13}-labelled glycerol. Bars represent the values across three
272 replicates for each strain. **e**. Comparison of the mean number of labelled proteins for each strain between those that are lipase-positive,

and those that are lipase-negative. Outliers are shown as grey diamonds. Statistical significance (Mann-Whitney U) is represented as follows: * < 0.05.

Discussion

The lack of detailed functional assignment is a limitation of metagenomic and genomic studies, limiting our insight into microbial processes. Therefore, we have generated a comprehensive collection of bacterial lipases present within the mammalian gut. The conserved yet unique sequences of these lipases, as compared to those isolated from other environments, was highlighted by the identification of two motifs which are both nearly ubiquitous across all identified bacterial lipases. We hypothesis that these motifs, along with the conserved GxSxG motif, are key to the activity of the bacterial lipases within the gut due to their co-localisation both together and with bound molecules. Of particular note is the high prevalence of alkaline lipases which corroborates previous findings that increasing the pH enhanced the activity of microbial phospholipases from the human gut [49].

Studying the taxonomic range of lipase-positive species identified that lipases are present across the major phyla and families within mammalian body-associated microbiomes. These results show lipolytic activity to be widespread, not limited to a few *Pseudomonadota* as previously suggested [47]. This discrepancy may be explained by the targeted annotation applied within this manuscript compared to the general annotations used to identify KEGG orthologs in previous studies [50].

Many of the most common lipase-positive SGBs were assigned to the genera *Ruminococcus* and *Coprococcus*, both of which are dominant members of the mammalian gut microbiome due to their diverse functional roles [51, 52]. Interestingly, the mean relative abundance of lipase-positive SGBs was twice that of lipase negative species, suggesting that these species are dominant and potentially key members of the ecosystem. The lack of aerobic or anaerobic beta-oxidation from the majority of lipase-positive SGB suggested that the metabolism of fatty acids is not the purpose of their lipolytic activity. In contrast, the majority of lipase-positive species were shown to utilise glycerol significantly more than strains without lipase activity, indicating the release and utilisation of glycerol may enhance the growth of lipase-positive microbes [53].

Direct measurement of lipolytic activity along the gastrointestinal tract can be difficult due to the diverse nature of intestinal content and sample variation. However, biochemical characterisation of a new gut bacterial

300 lipase has provided some hints into their localisation and niche. Activity was observed to decrease with increas-
301 ing detergent concentration, particularly above the critical micelle concentration. In addition to this, lipolytic ac-
302 tivity was completely removed below pH 7, suggesting that bacterial lipases within the gut have evolved to work
303 in an environment with sub-optimal detergent concentrations and at a pH at or above 7. Along the gastrointes-
304 tinal tract, the bile concentration decreases whilst the pH and bacterial load increase, suggesting that the bac-
305 terial lipases have evolved to be active in the distal gut. For validation of this theory as well as to enhance our
306 understanding of bacterial lipase localisation along the GI tract, *in vivo* experiments will be required.
307

308 # Materials and Methods

309 **Databases**

310 **EGGNOG:** EGGNOG (v3) [54] consists of non-supervised orthologous gene clusters formed via Smith-
311 Waterman reciprocal best matches and includes the original COGs [55], which have since been expanded.
312 COG1075 is assigned the function 'triacylglycerol lipase' and contains 1,260 proteins from 633 species. The
313 proteins originating from bacteria (67.5%) were used to form the EGGNOG database.

314 **EC:** Triglyceride lipases are assigned the enzyme commission (EC) notation 3.1.1.3, specifying them
315 as first hydrolases, secondly acting on ester bonds, thirdly carboxylic ester hydrolases, and finally as triglyceride
316 lipases. The UNIPROT database (accessed: December 2017) was searched for genes assigned this EC nota-
317 tion, which returned 162 proteins that formed the EC database.

318 **Lipase engineering database (LED):** This database is a custom database of alpha/beta-hydrolase
319 proteins encompassing six super families made up of 22 homologous families created from 5,278 proteins.
320 Since its creation in 2000, the lipase engineering database has been used multiple times to study the sequence
321 and structural similarity of hydrolases [29].
322

323 **Lipase annotation**

324 Proteins from the pig [3], mouse [4] and human [2] metagenomic gene catalogues were annotated
325 against the EC, EGGNOG and LED databases using DIAMOND BLASTP (v0.9.12.114) [56]. As previous re-

326 search suggests protein families are best captured using 30-40% identity and 35-70% coverage [57], annota-
327 tions were required to cover ≥40% of both the query and subject sequence at a minimum sequence identity of
328 40%. Moreover, proteins needed to have a match within two of the three databases and to contain the GxGxG
329 motif identified to be common in lipases [29].
330
331 **Network analysis and multiple sequence alignment**
332 Identity between the gut lipases and the type proteins for each known lipolytic protein family [30] was
333 identified using DIAMOND with the same cutoffs as for lipase annotation. The protein matches were entered
334 into Cytoscape (v3.8) [58] along with their shared identity for visualisation. The motifs specific for member se-
335 quences of both 'Cluster 1' and 'Cluster 2' (see results section) were identified via cluster-specific alignment
336 with ClustalW [59] within the MEGAX software package [60] and visualised using weblogo [61].
337
338 **Protein modelling**
339 The structure of protein sequences was modelled using the I-TASSER suite [62]. COFACTOR [63] was
340 utilised with TM-ALIGN [64] to compare the predicted structure of the proteins to those within the Protein Data
341 Bank (PDB) [65]. Models were viewed in Pymol (v1.8).
342
343 **Biochemical characterisation**
344 The lipase sequence (WP_003509004.1) from *Clostridium symbiosum* DSM 29356 [44] was modified *in*
345 *silico* by removal of the transmembrane region (first 150 N-terminal amino acids) and addition of a polyhistidine-
346 tag (His_6) at the C-terminus. The new construct (LipΔ150), integrated into a pET-15b vector was purchased from
347 BioCat (Heidelberg, Germany) and transformed into *Escherichia coli* BL21 (DE3). The transformed *E. coli* was
348 cultured in the presence of ampicillin (100 µg/ml) and a low IPTG concentration (0.1 mM) at low temperatures
349 (25 °C) to support a functional folding of the expressed LipΔ150. Cells were harvested after 22 hours by centrif-
350 ugation and frozen at -20 °C.
351 An aliquot of the frozen cells was solved in lysis buffer (50 mM Tris/HCl; 300 mM NaCl; 10 mM imidazole;
352 pH 8) to a concentration of 150 mg/ml. The cells were disrupted by sonication cooled on ice (amplitude 25%;
353 impulse length 3 x 30 s). Cell debris were removed by centrifuging the lysate (16,000 g, 30 min, 4 °C). The

354 supernatant was filtered (0.45 µm) and the target protein purified via immobilized metal-ion affinity chromatog-
355 raphy (IMAC) using a 1 ml HisTrap^{TM} HP column (GE Healthcare, UK). The lysis buffer (50 mM Tris/HCl; 300
356 mM NaCl; 10 mM imidazole; pH 8) was used for equilibration of the column and as washing buffer. Proteins
357 were recovered using the elution buffer (50 mM Tris/HCl; 300 mM NaCl; 300 mM imidazole; pH 8) and collected
358 in 0.5 ml fractions. The protein content during the washing and elution steps was monitored using the Äkta
359 purifier 100 system (GE Healthcare, UK) and double-checked with a Bradford assay. Protein-containing frac-
360 tions from the elution step were pooled and used for further purification.

361 Imidazole was removed using a PD-10 Desalting Column (GE Healthcare, UK) with an imidazole-free equi-
362 libration buffer (50 mM Tris/HCl; 300 mM NaCl; pH 8). Samples of approximately 200 µl were collected, a Brad-
363 ford assay along with the *para*-nitrophenyl (*p*NP) assay (C8) (see below) were performed to identify the lipase
364 fractions, which were then pooled. The final protein concentration was determined by a BSA-Bradford assay (R^2
365 0.99). The purity of the isolated protein was verified with a native polyacrylamide gel electrophoresis.

366 Lipolytic activity was tested utilizing the degradation of *p*NP-bound fatty acid substrates [66–68]. The test
367 was carried out by mixing 83.3 µl of a testing buffer with 16.7 µl protein solution and the absorption measured
368 at 410 nm over a period of 45 min at 37 °C. The standard testing buffer (final concentrations in the 100 µl
369 reaction mix: 50 mM Tris/HCl; 20 mM NaCl; 4.46 mM Tween20; 0.3 mM *p*NP-octanoate (C_8); pH 8) was used
370 during purification and for comparison in every test. A negative control was prepared for every tested condition
371 by adding 16.7 µl of lipase free equilibration buffer (50 mM Tris/HCl; 300 mM NaCl; pH 8) to 83.3 µl of the
372 corresponding testing buffer (*i.e.* containing *p*NP substrates of varying chain length). After 45 min incubation at
373 37 °C, samples (including all controls) were centrifuged (4,000 g, 10 min) and the supernatant measured. For
374 quantification, a standard curve was created using the *p*NP assay with defined concentrations of *p*NP (R^2 =
375 0.99). Results were normalized by subtraction of the negative control values and the standard curve was used
376 to determine the concentration of released fatty acids. Statistical significance was determined using a t-test with
377 Bonferroni multiple test correction adjusted p-values reported.

378

379 **Lipolytic activity assay of bacterial isolates**

380 Hungate tubes with 9 ml of anaerobic brain–heart infusion (BHI) medium were used to cultivate *Escherichia*
381 *coli* DSM 28618 (negative control) and *Clostridium symbiosum* DSM 29356. After 24 hours, 50 µl of each culture
382 was placed onto rhodamine B agar plates [69]. To identify if the lipolytic activity was cell-bound, each isolate

383 was cultured in Hungate tubes with 9 ml BHI broth supplemented with 0.5% cysteine (0.05 g/ml), 0.5% DTT
384 (0.02 g/ml), and 200 µl emulsified olive oil. After 7 days of cultivation, 6 ml of dense culture was centrifuged
385 (10,000 g, 10 min) to collect biomass. The pellet and 50 µl of the supernatant were put onto rhodamine B agar
386 plates. In each experiment, 50 µl of porcine pancreas lipase (Sigma-Aldrich, USA) with a concentration of 20
387 mg/ml was added as positive control to every plate. Afterwards, the plates were incubated for 2 hours at 37 °C.
388 Fluorescence, caused by lipolytic activity was observed under UV light using the GelStudio SA (Analytic Jena,
389 Germany). Curated human gut isolates are all publicly accessible as members of the HiBC [42].
390

391 **Genome analysis of catabolism pathways**

392 Metagenome-assembled genomes (MAGs) were obtained from Pasolli *et al.* (2019) [36], along with
393 metadata for their originating sample, assignment to species genome bins (SGBs), prevalence and relative
394 abundance [36]. Taxonomy was assigned to each MAG using GTDB-TK, release 89 [70, 71]. MAGs were an-
395 notated using PROKKA (v1.13.3) [72] and KEGG orthologs (KOs) [73] identified using the PROKKA2KEGG tool
396 (https://github.com/SilentGene/Bio-py/tree/master/prokka2kegg). Each MAGs KOs were searched against pre-
397 viously published lipolytic catabolism modules [47]. Genome based trees were created using the predicted pro-
398 tein sequences in PhyloPhlAn (v3.0.60) [74] with the high diversity option.
399

400 **16S rRNA gene sequence analysis**

401 Annotation of 16S rRNA gene sequences was done using EZBioCloud [75]. The phylogenetic tree was gener-
402 ated using MegaX [76], with alignment conducted with MUSCLE [77] and a maximum likelihood tree generated
403 using a GTR model with uniform rates of mutation and removal of missing data, bootstrapped via 1000 repli-
404 cates. The tree was then annotated in iToL [78].
405

406 **Large-scale identification of lipase active strains from human faecal samples**

407 Fresh faecal samples were obtained from 15 healthy adults (**Supplementary Table 1**). The collection
408 of anonymised faecal samples used in this study has been described previously [79]. At the laboratory bench,
409 samples were kneaded manually to make the samples homogenous. Sterile, anaerobic PBS (0.1 M, pH 7.2;
410 Oxoid) was added to portions (~5 g) of faeces in mini stomacher bags (Seward) to give 1:9 faeces-to-PBS (w/w).
411 Samples were transferred to an anaerobic cabinet (Don Whitley; 80 % N_2, 10 % H_2, 10 % CO_2) maintained at

412 37 °C. The total amount of time the faecal samples were exposed to air for was ~3 min. Dilution series (10^{-1} to
413 10^{-8}) were prepared from the faecal homogenates in sterile, anaerobic PBS.

414 Isolation was conducted on five agar media (fastidious anaerobe agar containing 5 % laked horse blood
415 (Oxoid), reinforced clostridial agar (Oxoid), phenyl ethyl alcohol agar containing 5 % laked horse blood, Mueller–
416 Rogosa–Sharpe agar (Oxoid) and Maconkey agar (Oxoid)). Plates were incubated for 5 days, then colonies
417 were counted. Each colony from one of the replicates for each plate from which counts were obtained was
418 picked and spotted onto tributyrin agar. Tributyrin agar was prepared by adding 5 g tributyrin agar (Sigma Al-
419 drich) to 250 ml distilled water and 1.25 ml of glyceryl tributyrate (Sigma Aldrich). The mixture was autoclaved
420 at 121 °C, 15 psi for 15 min. After autoclaving the molten agar was homogenized in a 250-ml Ato Mix blender
421 (MSE) for 30 seconds. Rhodamine B agar was prepared in the same way as the tributyrin agar, except that the
422 glyceryl tributyrate was substituted with olive oil and 2.5 ml of filter-sterilized rhodamine B (1 mg ml^{-1} solution
423 prepared in distilled water) was added. Tributyrin plates were examined for zones of clearing at 4, 12, 24 and
424 48 hours. Colonies demonstrating lipolytic activity at any of these times were identified by marking the plates.
425 Lipolytic bacteria were spotted onto the rhodamine B agar to confirm lipase activity as determined by zones of
426 fluorescence detected under a UV light.

427

428 **Glycerol labelled proteomics**

429 Biological triplicates of the selected bacterial strains were added to 10 ml of suitable liquid cultivation
430 medium supplemented with 129.99 mM glycerol (Sigma Aldrich) and 6.84 mM ^{13}C3-labelled glycerol (Eurisotop)
431 (1% (v/v) glycerol of which 5% (v/v) are labelled) and incubated at 37 °C for 24 hours under anaerobic conditions.
432 The cultures were centrifuged at 4,500 g for 10 minutes and the supernatant was removed. After washing the
433 cells with 1 ml of PBS the cell pellets were immediately frozen and stored at -80 °C.

434 Bacterial pellets were resuspended in a lysis buffer consisting of 8 M urea, 2 M thiourea, and 1 mM
435 phenylme-thylsulfonyl fluoride. To disrupt the bacterial cells, bead beating was performed using the FastPrep-
436 24 system (MP Biomedicals, Santa Ana, CA, United States; 5.5 ms, 1 min, 3 cycles), followed by heat shock at
437 90 °C for 10 min with 1,400 rpm (ThermoMixerTM, Eppendorf, Germany) and centrifugation (10,000× g, 10 min).
438 The resulting supernatant was used for protein concentration determination using a Pierce 660 nm Protein As-
439 say (Thermo Fisher Scientific, Waltham, MA, United States). Vivacon 500 columns with a 10 kDA Molecular

440 weight cutoff membrane (Sartorius, Göttingen, Germany) were equilibrated with lysis buffer and centrifuged
441 (14,000× g, 20 °C, 20 min), and the remainder of the centrifugation steps were performed under the same con-
442 ditions. Subsequently, 50 µg of the protein was loaded onto the column and centrifuged. The proteins that ad-
443 hered to the filter were then incubated with 200 µL of 10 mM dithiothreitol in lysis buffer using a Thermo-Mixer
444 (Eppendorf) at 37 °C and 600 rpm for 1 min and then with no agitation for 30 min. After centrifugation, the protein
445 pellets underwent further alkylation and proteolytic cleavage, as described previously [80]. Finally, the resulting
446 peptide lysates were dissolved in 0.1% formic acid (FA) for subsequent mass spectro-metric measurements.

447 For each nanoLC-MS measurement, 1 µg of the peptide lysate was injected into a Vanquish Neo nano-
448 HPLC system (Thermo Fisher Scientific). The peptide lysates were initially trapped on a C18-reverse phase
449 trapping column (Acclaim PepMapTM 100, 75 µm × 2 cm, particle size 3 µM, nanoViper, Thermo Fisher Scien-
450 tific), and then separated on a C18-reverse phase analytical column (Double nanoViper™ PepMap™ Neo,
451 75 µm × 150 mm, particle size 2 µM, Thermo Fisher Scientific). The separation was achieved using a two-step
452 gradient with mobile phases A (0.01% FA in H2O) and B (80% acetonitrile in H2O and 0.01% FA). During the
453 first step of the gradient, which lasted 95 min, the proportion of mobile phase B was increased from 4 to 30%.
454 This was followed by a 40-min period in which the proportion of mobile phase B increased from 30 to 55%. The
455 flow rate during the separation was maintained at 300 nL/min as described [81]. The eluted peptides were ion-
456 ized using a Nanospray Flex™ Ion Source (Thermo Fisher Scientific) and detected using an Orbitrap Exploris™
457 480 mass spectrometer (Thermo Fisher Scientific). The mass spectrometer was operated with the following
458 settings: for MS, the scan range was set to 350–1,550 m/z, the resolution was set to 120,000, the automatic
459 gain control (AGC) target was set to 3′000,000 charges, and the maximum injection time was set to 100 ms. An
460 intensity threshold of 8,000 ions and dynamic exclusion time of 20 s were also applied. The top 10 most intense
461 ions were selected for MS/MS analysis, with an isolation window of 1.4 m/z, resolution of 15,000, AGC target of
462 200,000 ions, and a maximum injection time of 100 ms.

463 Proteins from each strain that incorporated labelled 13C after incubation for 24 hours with 13C-labelled
464 glycerol were identified using Sipros (v4) [82]. The mass spectrometry raw files were extracted into FT1 and
465 FT2 format files using Raxport (https://github.com/thepanlab/Raxport.net) with the mono
466 (https://github.com/mono/mono) framework. The extracted FT2 files of the samples were searched against a
467 target-decoy protein sequence database comprised of the protein sequences available for the strains on Uni-
468 ProtKB and sequences from MAGs for strains whose sequences were not available on UniProt, where the

469 reverse sequences of these target proteins were implemented as decoys. Label-free searches were performed
470 using samples from strains incubated without any 13C labels using Sipros (v4) against the target-decoy data-
471 base, and all identified proteins from the unlabelled searches were subsequently used as references for the
472 labelled searches with Sipros. The false discovery rate (FDR) of peptide identifications was filtered at 0.05 and
473 0.01 for the unlabelled and labelled searches, respectively, by adjusting the score threshold of PSMs. The
474 search required at least two unique peptides for each protein/peptide group to be represented in the results.
475 The isotopic abundance of proteins provided the values of the 'average enrichment levels'. To confirm that the
476 labelled peptides were not representing the natural abundance of 13C, we set the threshold of the average
477 enrichment level to 1.07, the same value as that of the 13C natural abundance [83].
478

479 **Batch fermentation**

480 Fresh faecal samples were obtained from five healthy adult donors of between 25 and 38 years of age
481 (3 females, 2 males). None of the donors had taken antibiotics during the 6 months prior to the study, and none
482 was taking probiotics or prebiotics. Freshly voided faecal samples were processed immediately by diluting them
483 1 in 10 (w/w) in pre-reduced sterile H_2O and homogenizing them in a filter bag in a stomacher (Stomacher 400
484 Lab System; Seward) for 2 min at 'high' speed. For each donor, a batch culture vessel containing 150 ml of gut
485 model medium (GMM) [84] was inoculated with 20 ml of the faecal homogenate and 30 ml of water containing
486 25 µl TWEEN® 80 (Fisher Scientific) and 0 or 1 g of refined/virgin olive oil blend (Carbonell). The water containing
487 the detergent and lipid was steamed for 10 min prior to being added to the batch culture vessel. GMM comprised
488 (g l^{-1} distilled H_2O unless stated otherwise; all reagents were purchased from Sigma Aldrich unless stated oth-
489 erwise): starch from potato, 5.0; pectin from citrus fruit, 2.0; guar, 1.0; porcine gastric mucin (type III), 4.0; xylan
490 from oatspelts, 2.0; arabinogalactan from larchwood, 2.0; inulin (Orafti), 1.0; casein from bovine salts, 3.0; pep-
491 tone water (Oxoid), 5.0; tryptone (Oxoid), 5.0; bile salts no. 3 (Oxoid), 0.4; yeast extract (Oxoid), 4.5;
492 $FeSO_4.7H_2O$, 0.005; NaCl, 4.5; KCl, 4.5; KH_2PO_4, 0.5; $MgSO_4.7H_2O$, 1.25; $CaCl_2.6H_2O$, 0.15; $NaHCO_3$, 1.5; L-
493 cysteine, 0.8; hemin, 0.05 g dissolved in 1 ml of 1 M NaOH. All components of the GMM were combined and
494 autoclaved at 121 °C for 15 min. The final working volume of each batch culture vessel was 200 ml. The pH of
495 each vessel (pH 6.8) was controlled automatically by the addition of 0.5 M HCl or 0.5 M NaOH. pH controllers
496 were supplied by Electrolab. The contents of each vessel were stirred constantly. An anaerobic environment

497 was maintained by constantly sparging the vessels with O_2-free N_2. The temperature of each vessel was main-
498 tained at 37 °C by use of a circulating waterbath connected to each fermentation vessel. Each set of experiments
499 was run for 24 h.

500 Samples (5 ml) were withdrawn from each vessel at 0, 6, 12 and 24 h and immediately placed on ice.
501 Aliquots (2× 2 ml) were centrifuged at 13,000 rpm (MSE MicroCentaur; Sanyo) for 10 min at room temperature.
502 The supernatants were retained for analysis of LCFAs, and stored at -80 °C until processed.
503

504 **Quantification of LCFAs via gas chromatography**

505 For each sample, the supernatants from 1 ml samples were thawed on ice and total fatty acids were
506 extracted as described [85]. $C_{19:0}$ internal standard (stock solution, 200 μg $C_{19:0}$ / ml methanol) was used as the
507 internal standard. Fatty acid methyl esters (FAMEs) were separated and quantified using a gas chromatograph
508 (model 6890; Agilent Technologies UK Ltd, Stockport, UK) equipped with a flame-ionisation detector, quadru-
509 pole mass selective detector (model 5973N), injection port and a 50 m fused silica capillary column (internal
510 diameter 0·25 mm) coated with a 0.2 mm film of cyanopropyl polysiloxane (CP-SIL 88; Varian Analytical Instru-
511 ments, Walton-on-Thames, Surrey, UK). Total FAMEs in a 1 ml sample at a split ratio of 15:1 were determined
512 using a temperature gradient programme (initial temperature 80 °C for 1 min; increased at a rate of 25 °C/min
513 to 160 °C, which was held for 3 min; increased at a rate of 1 °C/min to 190 °C, maintained for 5 min; increased
514 at a rate of 2 °C/min to 230 °C, held for 25 min). Helium was the carrier gas and the apparatus was operated at
515 a constant pressure (20 psi) and flow rate (0·5 ml/min). Injector and detector temperatures were maintained at
516 250 and 275 °C, respectively. Peaks were identified by comparison of retention times with authentic FAME
517 standards obtained from Sigma (Poole, Dorset, UK) and Matreya Inc. (Pleasant Gap, PA, USA).

518 **Acknowledgments**

519 TCAH and TC conceived the study. TCAH conducted all bioinformatic analyses and supervised daily
520 laboratory work. JM, TS, LH, DHG, and TF conducted wet lab experiments. TCAH, JM, LH, DHG, TS, and TC
521 interpreted data. LE, LH, and TC provided guidance and access to essential materials. TCAH and TC secured
522 funding and wrote the paper. All authors reviewed the manuscript and agreed with its final content.

523 TCAH was funded by an internal University Hospital RWTH Aachen START grant titled 'KETOGUT' as
524 well as a RWTH Start-Up grant titled 'ProtoBiome'. Access to the computational resources of the RWTH cluster

were provided under grant; rwth2608. TC received funding from the German Research Foundation (DFG) Project-ID 403224013 – SFB 1382 and Project-ID CL 481/17-1 – SPP 2474. LH's PhD, from which the cultivation (LH, DHG) and batch-culture work (LH) is derived, was funded by GlaxoSmithKline. R. John Wallace and Nest McKain (Rowett Institute of Nutrition and Health, Aberdeen) are thanked for hosting LH and training her in gas chromatography and analyses of data.

Ethical approval and consent to participate

All donors were members of the laboratory who gave free consent for their samples to be used for the purposes of isolating and characterizing lipase-positive bacteria, and for molecular-based analyses and have since been destroyed. At the time the study was conducted (2006/2007) [86], the University of Reading's Research Ethics Committee (UREC) did not require that specific ethical review and approval be given by UREC for the collection of faecal samples from healthy human volunteers to inoculate *in vitro* fermentation systems.

Supplementary Tables

Supplementary Table 1: Large-scale screening of bacterial colonies from human faeces. For each donor five growth medium were used, and for each the total number of colonies, as well as those that were lipolytic and lipase-positive are stated.

References

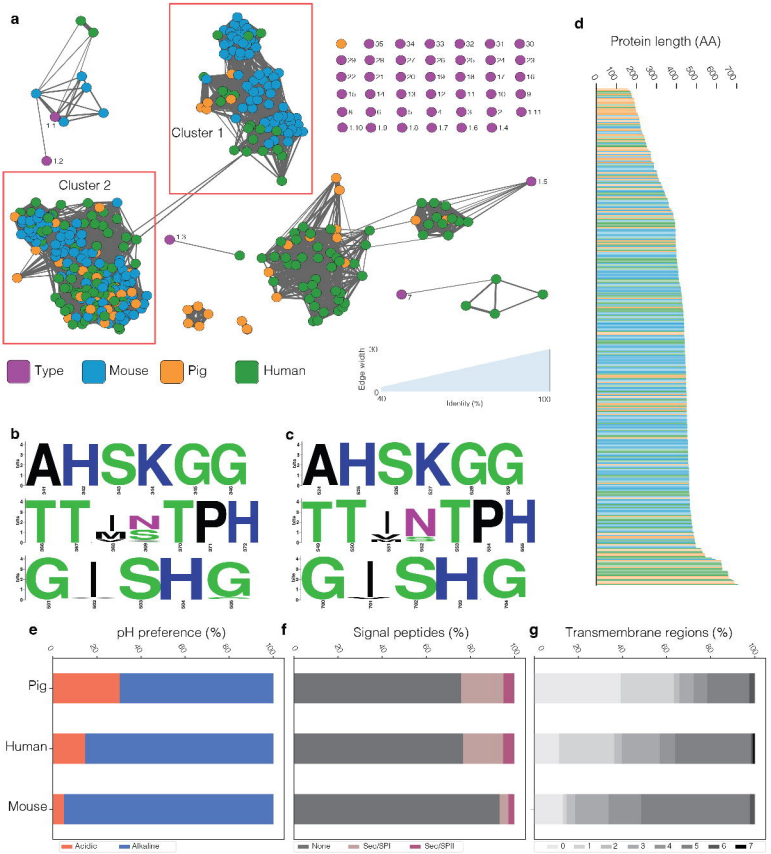
1. Qin J et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 2010;**464**:59–65. https://doi.org/10.1038/nature08821
2. Li J et al. An integrated catalog of reference genes in the human gut microbiome. *Nat Biotechnol* 2014;**32**:834–841. https://doi.org/10.1038/nbt.2942
3. Xiao L et al. A reference gene catalogue of the pig gut microbiome. *Nat Microbiol* 2016;**1**. https://doi.org/10.1038/nmicrobiol.2016.161
4. Lesker TR et al. An Integrated Metagenome Catalog Reveals New Insights into the Murine Gut Microbiome. *Cell Rep* 2020;**30**:2909-2922.e6. https://doi.org/10.1016/j.celrep.2020.02.036
5. Schmitz MA et al. Lineage-specific microbial protein prediction enables large-scale exploration of protein ecology within the human gut. *Nature Communications* 2025;**16**. https://doi.org/10.1038/s41467-025-58442-w
6. Brugiroux S et al. Genome-guided design of a defined mouse microbiota that confers colonization resistance against Salmonella enterica serovar Typhimurium. *Nat Microbiol* 2016;**2**:1–12. https://doi.org/10.1038/nmicrobiol.2016.215
7. Cheng AG et al. Design, construction, and in vivo augmentation of a complex gut microbiome. *Cell* 2022;**185**:3617-3636.e19. https://doi.org/10.1016/j.cell.2022.08.003
8. Thomas AM, Segata N. Multiple levels of the unknown in microbiome research. *BMC Biol* 2019;17–20.

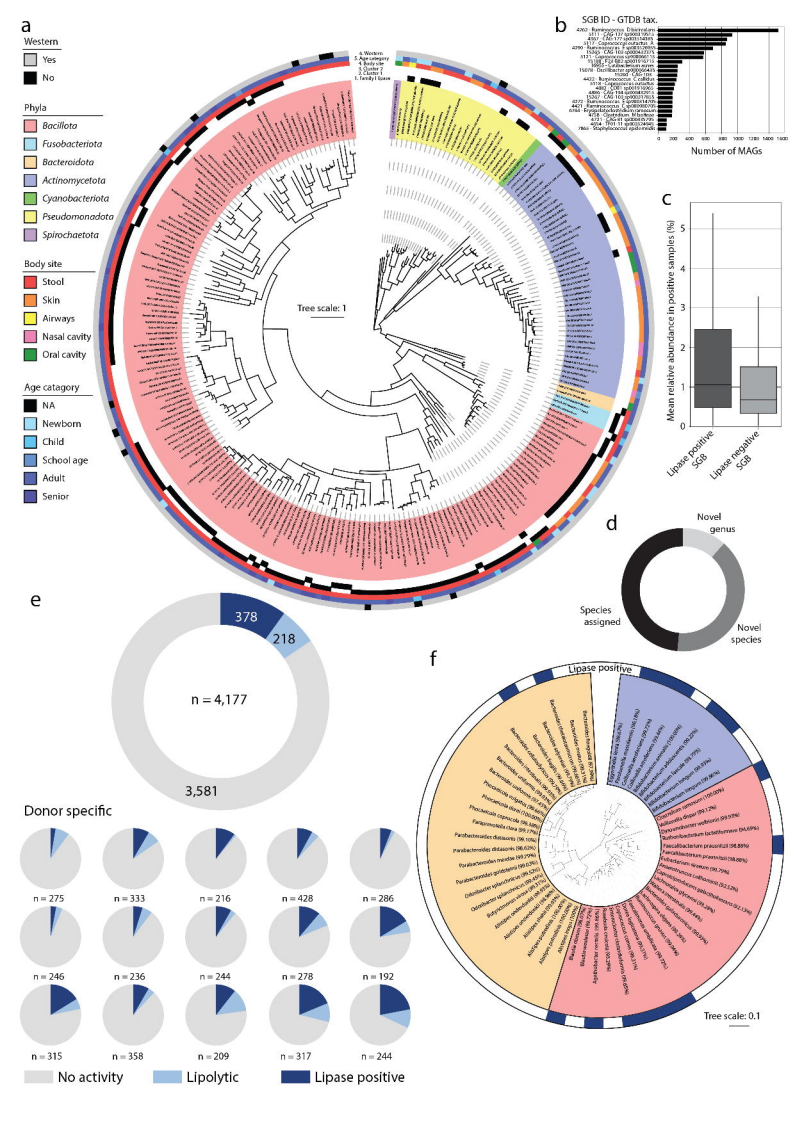
9. Pollet RM et al. An Atlas of β-Glucuronidases in the Human Intestinal Microbiome. *Structure* 2017;**25**:967-977.e5. https://doi.org/10.1016/j.str.2017.05.003
10. Creekmore BC et al. Mouse Gut Microbiome-Encoded β-Glucuronidases Identified Using Metagenome Analysis Guided by Protein Structure. *mSystems* 2019;**4**:1–13. https://doi.org/10.1128/mSystems.00452-19
11. Funabashi M et al. A metabolic pathway for bile acid dehydroxylation by the gut microbiome. *Nature* 2020;**582**:566–570. https://doi.org/10.1038/s41586-020-2396-4
12. Borgström B et al. Studies of intestinal digestion and absorption in the human. *J Clin Invest* 1957;**36**:1521–1536. https://doi.org/10.1172/JCI103549
13. Clark SB, Lawergren B, Martin J V. Regional intestinal absorptive capacities for triolein: an alternative to markers. *American Journal of Physiology* 1973;**225**:574–585. https://doi.org/10.1152/ajplegacy.1973.225.3.574
14. Borgström B. On the Mechanism of the Intestinal Fat Absorption. *Acta Physiol Scand* 1952. https://doi.org/10.1111/j.1748-1716.1954.tb01158.x
15. Borgström B. Determination of pancreatic lipase in human small intestinal content. *Scand J Clin Lab Invest* 1957;**9**:226–228. https://doi.org/10.3109/00365515709079961
16. Aberdeen V, Shepherd PA, Simmonds WJ. CONCURRENT MEASUREMENT, IN UNANAISTHETIZED RATS, OF INTESTINAL TRANSPORT AND FAT ABSORPTION FROM THE LUMEN. *Q J Exp Physiol Cogn Med Sci* 1960.
17. Ammon H V., Phillips SF. Inhibition of colonic water and electrolyte absorption by fatty acids in man. *Gastroenterology* 1973;**65**:744–749. https://doi.org/10.1016/s0016-5085(19)33009-4
18. Snipes RL. Morphological appearance of fat in the epithelial cells of different portions of the intestines in mice. *Experientia* 1977;**33**:1335–1336. https://doi.org/10.1007/BF01920166
19. Wu AL, Clark SB, Holt PR. Transmucosal triglyceride transport rates in proximal and distal rat intestine in vivo. *J Lipid Res* 1975;**16**:251–257.
20. Booth CC, Alldis D, Read AE. Studies on the site of fat absorption. *Gut* 1961;**2**.
21. Kübeck R et al. Dietary fat and gut microbiota interactions determine diet-induced obesity in mice. *Mol Metab* 2016;**5**:1162–1174. https://doi.org/10.1016/j.molmet.2016.10.001
22. Rabot S et al. Germ-free C57BL/6J mice are resistant to high-fat-diet-induced insulin resistance and have altered cholesterol metabolism. *The FASEB Journal* 2010;**24**:4948–4959. https://doi.org/10.1096/fj.10-164921
23. Drouault S et al. Oral treatment with Lactococcus lactis expressing Staphylococcus hyicus lipase enhances lipid digestion in pigs with induced pancreatic insufficiency. *Appl Environ Microbiol* 2002;**68**:3166–3168. https://doi.org/10.1128/AEM.68.6.3166-3168.2002
24. Ran C et al. A novel lipase as aquafeed additive for warm-water aquaculture. *PLoS One* 2015;**10**:1–15. https://doi.org/10.1371/journal.pone.0132049
25. Huang YS et al. In vitro hydrolysis of natural and synthetic γ-linolenic acid-containing triacylglycerols by pancreatic lipase. *J Am Oil Chem Soc* 1995;**72**:625–631. https://doi.org/10.1007/BF02635646
26. Lowe ME. The catalytic site residues and interfacial binding of human pancreatic lipase. *J Biol Chem* 1992;**267**:17069–73.
27. Huggins KW et al. Pancreatic Triglyceride Lipase Deficiency Minimally Affects Dietary Fat Absorption but Dramatically Decreases Dietary Cholesterol Absorption in Mice. *Journal of Biological Chemistry* 2003;**278**:42899–42905. https://doi.org/10.1074/jbc.M303422200
28. Ollis DL et al. The α/β hydrolase fold. *Protein Engineering, Design and Selection* 1992;**5**:197–211. https://doi.org/10.1093/protein/5.3.197
29. Pleiss J et al. Lipase engineering database: Understanding and exploiting sequence-structure-function relationships. *J Mol Catal B Enzym* 2000;**10**:491–508. https://doi.org/10.1016/S1381-1177(00)00092-8
30. Hitch TCA, Clavel T. A proposed update for the classification and description of bacterial lipolytic enzymes. *PeerJ* 2019;**7**. https://doi.org/10.7717/peerj.7249
31. Lin H, Chen W, Ding H. AcalPred: A Sequence-Based Tool for Discriminating between Acidic and Alkaline Enzymes. *PLoS One* 2013;**8**. https://doi.org/10.1371/journal.pone.0075726
32. Almagro Armenteros JJ et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol* 2019;**37**:420–423. https://doi.org/10.1038/s41587-019-0036-z

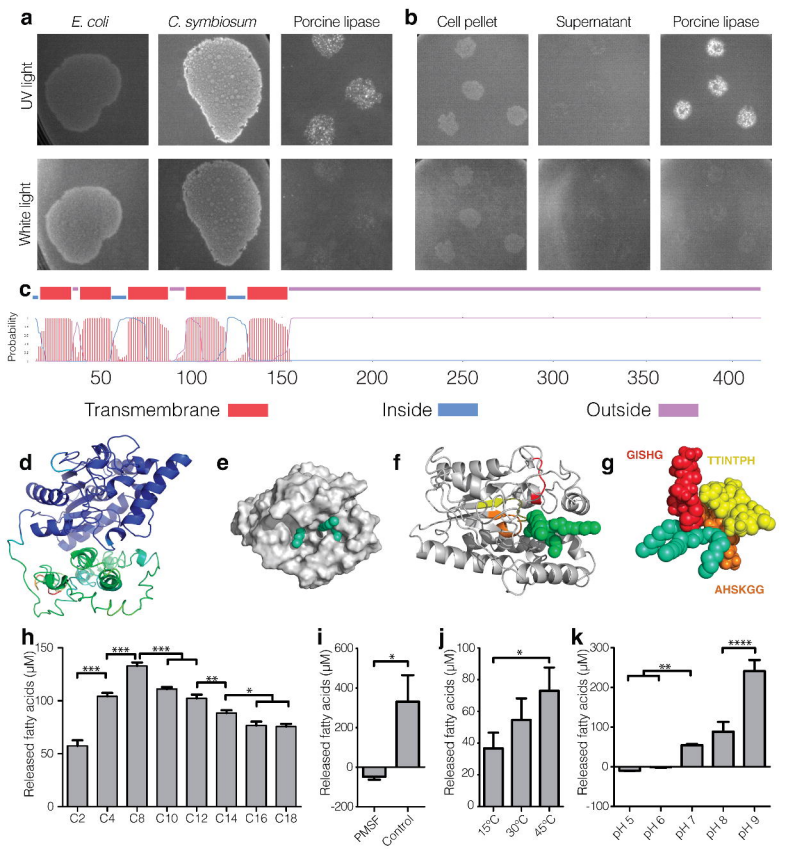
33. Krogh A et al. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J Mol Biol* 2001;**305**:567–580. https://doi.org/10.1006/jmbi.2000.4315
34. Parks DH et al. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol* 2017;**903**:1–10. https://doi.org/10.1038/s41564-017-0012-7
35. Ormerod KL et al. Genomic characterization of the uncultured Bacteroidales family S24-7 inhabiting the guts of homeothermic animals. *Microbiome* 2016;**4**:36. https://doi.org/10.1186/s40168-016-0181-2
36. Pasolli E et al. Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. *Cell* 2019;**176**:649-662.e20. https://doi.org/10.1016/j.cell.2019.01.001
37. Hitch TCA, Clavel T. A proposed update for the classification and description of bacterial lipolytic enzymes. *PeerJ* 2019;**2019**. https://doi.org/10.7717/peerj.7249
38. Hitch TCA et al. Automated analysis of genomic sequences facilitates high-throughput and comprehensive description of bacteria. *ISME Communications* 2021;**1**. https://doi.org/10.1038/s43705-021-00017-z
39. Wegmann U et al. Complete genome of a new Firmicutes species belonging to the dominant human colonic microbiota ('Ruminococcus bicirculans') reveals two chromosomes and a selective capacity to utilize plant glucans. *Environ Microbiol* 2014;**16**:2879–2890. https://doi.org/10.1111/1462-2920.12217
40. Kwaszewska A, Sobiś-Glinkowska M, Szewczyk EM. Cohabitation—relationships of corynebacteria and staphylococci on human skin. *Folia Microbiol (Praha)* 2014;**59**:495–502. https://doi.org/10.1007/s12223-014-0326-2
41. Lebeer S et al. Selective targeting of skin pathobionts and inflammation with topically applied lactobacilli. *Cell Rep Med* 2022;**3**:100521. https://doi.org/10.1016/j.xcrm.2022.100521
42. Hitch TCA et al. HiBC: a publicly available collection of bacterial strains isolated from the human gut. *Nat Commun* 2025;**16**:4203. https://doi.org/10.1038/s41467-025-59229-9
43. Segata N et al. PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nat Commun* 2013;**4**:1–10. https://doi.org/10.1038/ncomms3304
44. Lagkouvardos I et al. The Mouse Intestinal Bacterial Collection (miBC) provides host-specific insight into cultured diversity and functional potential of the gut microbiota. *Nat Microbiol* 2016;**1**:16131. https://doi.org/10.1038/nmicrobiol.2016.131
45. Hitch TCA et al. Automated analysis of genomic sequences facilitates high-throughput and comprehensive description of bacteria. *ISME Communications* 2021;**1**. https://doi.org/10.1038/s43705-021-00017-z
46. Yao J, Rock CO. How bacterial pathogens eat host lipids: Implications for the development of fatty acid synthesis therapeutics. *Journal of Biological Chemistry* 2015;**290**:5940–5946. https://doi.org/10.1074/jbc.R114.636241
47. Vieira-Silva S et al. Species-function relationships shape ecological properties of the human gut microbiome. *Nat Microbiol* 2016;**1**. https://doi.org/10.1038/nmicrobiol.2016.88
48. Campbell JW, Morgan-Kiss RM, Cronan JE. A new Escherichia coli metabolic competency: Growth on fatty acids by a novel anaerobic b-oxidation pathway. *Mol Microbiol* 2003;**47**:793–805. https://doi.org/10.1046/j.1365-2958.2003.03341.x
49. Morotomi M et al. Production of diacylglycerol, an activator of protein kinase C, by human intestinal microflora. *Cancer Res* 1990;**50**:3595–9.
50. Markowitz VM et al. The integrated microbial genomes (IMG) system in 2007: Data content and analysis tool extensions. *Nucleic Acids Res* 2008;**36**:528–533. https://doi.org/10.1093/nar/gkm846
51. Berg Miller ME et al. Diversity and strain specificity of plant cell wall degrading enzymes revealed by the draft genome of Ruminococcus flavefaciens FD-1. *PLoS One* 2009;**4**. https://doi.org/10.1371/journal.pone.0006650
52. Noriega BS et al. Understanding the Impact of Omega-3 Rich Diet on the Gut Microbiota. *Case Rep Med* 2016;**2016**. https://doi.org/10.1155/2016/3089303
53. Xiao L et al. High-fat feeding rather than obesity drives taxonomical and functional changes in the gut microbiota in mice. *Microbiome* 2017;**5**:1–12. https://doi.org/10.1186/S40168-017-0258-6
54. Jensen LJ et al. eggNOG: Automated construction and annotation of orthologous groups of genes. *Nucleic Acids Res* 2008;**36**:D250-4. https://doi.org/10.1093/nar/gkm796

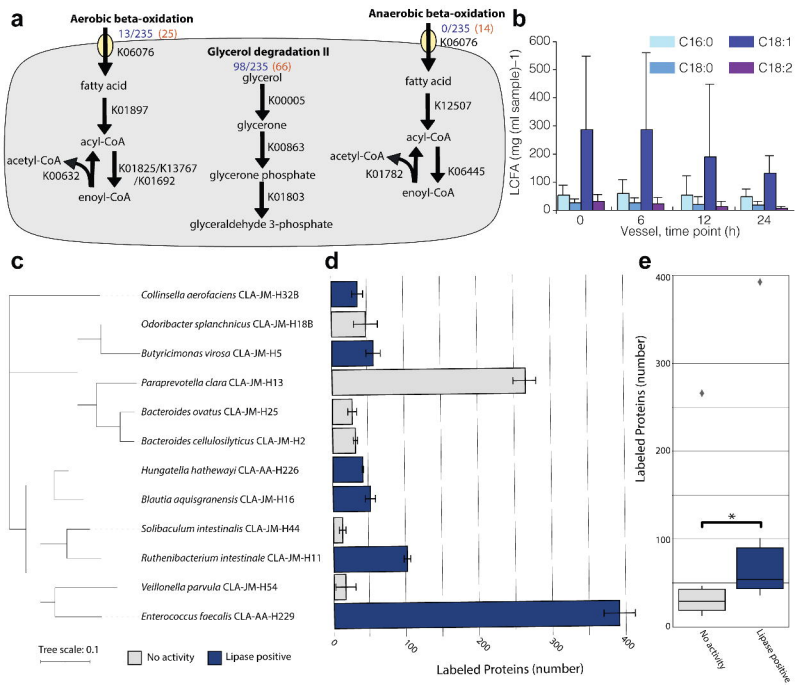
55. Tatusov RL. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res* 2000;**28**:33–36. https://doi.org/10.1093/nar/28.1.33
56. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 2014;**12**:59–60. https://doi.org/10.1038/nmeth.3176
57. Veeramachaneni V, Makałowski W. Visualizing sequence similarity of protein families. *Genome Res* 2004;**14**:1160–1169. https://doi.org/10.1101/gr.2079204
58. Shannon P et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;**13**:2498–2504. https://doi.org/10.1101/gr.1239303.metabolite
59. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994;**22**:4673–4680. https://doi.org/10.1093/nar/22.22.4673
60. Kumar S et al. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 2018;**35**:1547–1549. https://doi.org/10.1093/molbev/msy096
61. Crooks G et al. WebLogo: a sequence logo generator. *Genome Res* 2004;**14**:1188–1190. https://doi.org/10.1101/gr.849004.1
62. Yang J et al. The I-TASSER suite: Protein structure and function prediction. *Nat Methods* 2014;**12**:7–8. https://doi.org/10.1038/nmeth.3213
63. Roy A, Yang J, Zhang Y. COFACTOR: An accurate comparative algorithm for structure-based protein function annotation. *Nucleic Acids Res* 2012;**40**:471–477. https://doi.org/10.1093/nar/gks372
64. Zhang Y, Skolnick J. TM-align: A protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res* 2005;**33**:2302–2309. https://doi.org/10.1093/nar/gki524
65. Berman HM et al. The Protein Data Bank. *Nucleic Acids Res* 2000;**28**:235–242. https://doi.org/10.1093/nar/28.1.235
66. Peng Y et al. Accurately Determining Esterase Activity via the Isosbestic Point of p-Nitrophenol. *Biore-sources* 2016;**11**. https://doi.org/10.15376/biores.11.4.10099-10111
67. Zhang Y et al. Self-assembly of lipase hybrid nanoflowers with bifunctional Ca2+ for improved activity and stability. *Enzyme Microb Technol* 2020;**132**:109408. https://doi.org/10.1016/j.enzmictec.2019.109408
68. de Oliveira BH et al. Biocatalytic benefits of immobilized Fusarium sp. (GFC) lipase from solid state fermentation on free lipase from submerged fermentation. *Ind Crops Prod* 2020;**147**:112235. https://doi.org/10.1016/j.indcrop.2020.112235
69. Kouker G, Jaeger KE. Specific and sensitive plate assay for bacterial lipases. *Appl Environ Microbiol* 1987;**53**:211–213. https://doi.org/10.1128/aem.53.1.211-213.1987
70. Parks DH et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *bioRxiv* 2018;256800. https://doi.org/10.1101/256800
71. Chaumeil P-A et al. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 2019;**36**:1925–1927. https://doi.org/10.1093/bioinformatics/btz848
72. Seemann T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* 2014;**30**:2068–2069. https://doi.org/10.1093/bioinformatics/btu153
73. Ogata H et al. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 1999;**27**:29–34. https://doi.org/10.1093/nar/27.1.29
74. Asnicar F et al. Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. *Nat Commun* 2020;**11**:1–10. https://doi.org/10.1038/s41467-020-16366-7
75. Yoon SH et al. Introducing EzBioCloud: A taxonomically united database of 16S rRNA gene se-quences and whole-genome assemblies. *Int J Syst Evol Microbiol* 2017;**67**:1613–1617. https://doi.org/10.1099/ijsem.0.001755
76. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* 2016;**33**:1870–1874. https://doi.org/10.1093/molbev/msw054
77. Edgar RC. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004;**32**:1792–1797. https://doi.org/10.1093/nar/gkh340
78. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* 2019;**47**:W256–W259. https://doi.org/10.1093/nar/gkz239

79. Hoyles L, McCartney AL. What do we mean when we refer to Bacteroidetes populations in the human gastrointestinal microbiota? *FEMS Microbiol Lett* 2009;**299**:175–183. https://doi.org/10.1111/j.1574-6968.2009.01741.x
80. Wiśniewski JR et al. Universal sample preparation method for proteome analysis. *Nat Methods* 2009;**6**:359–362. https://doi.org/10.1038/nmeth.1322
81. Castañeda-Monsalve V et al. High-throughput screening of the effects of 90 xenobiotics on the simplified human gut microbiota model (SIHUMIx): a metaproteomic and metabolomic study. *Front Microbiol* 2024;**15**. https://doi.org/10.3389/fmicb.2024.1349367
82. Guo X et al. Sipros Ensemble improves database searching and filtering for complex metaproteomics. *Bioinformatics* 2018;**34**:795–802. https://doi.org/10.1093/bioinformatics/btx601
83. Coplen TB, Shrestha Y. Isotope-abundance variations and atomic weights of selected elements: 2016 (IUPAC Technical Report). *Pure and Applied Chemistry* 2016;**88**:1203–1224. https://doi.org/10.1515/pac-2016-0302
84. Macfarlane GT, Macfarlane S, Gibson GR. Validation of a Three-Stage Compound Continuous Culture System for Investigating the Effect of Retention Time on the Ecology and Metabolism of Bacteria in the Human Colon. *Microb Ecol* 1998;**35**:180–187. https://doi.org/10.1007/s002489900072
85. Morales ER et al. Ricinoleic acid inhibits methanogenesis and fatty acid biohydrogenation in ruminal digesta from sheep and in bacterial cultures1,2. *J Anim Sci* 2012;**90**:4943–4950. https://doi.org/10.2527/jas.2011-4670
86. Hoyles L et al. Metabolic retroconversion of trimethylamine N-oxide and the gut microbiota. *Microbiome* 2018;**6**:73. https://doi.org/10.1186/s40168-018-0461-0


a
Cluster 1
Cluster 2
Type
Mouse
Pig
Human
Edge width
Identity (%)
b
c
d
Protein length (AA)
e
pH preference (%)
Pig
Human
Mouse
Acidic
Alkaline
f
Signal peptides (%)
None
Sec/SPI
Sec/SPII
g
Transmembrane regions (%)


a
Western
Yes
No
Phyla
Bacillota
Fusobacteriota
Bacteroidota
Actinomycetota
Cyanobacteriota
Pseudomonadota
Spirochaetota
Body site
Stool
Skin
Airways
Nasal cavity
Oral cavity
Age catagory
NA
Newborn
Child
School age
Adult
Senior
6. Western
5. Age category
4. Phylum
3. Cluster 2
2. Cluster 1
1. Family Lipase
Tree scale: 1
b
SGB ID - GTDB tax.
Number of MAGs
c
Mean relative abundance in positive samples (%)
Lipase positive SGB
Lipase negative SGB
d
Novel genus
Species assigned
Novel species
e
378
218
n = 4,177
3,581
Donor specific
n = 275
n = 333
n = 216
n = 428
n = 286
n = 246
n = 236
n = 244
n = 278
n = 192
n = 315
n = 358
n = 209
n = 317
n = 244
No activity
Lipolytic
Lipase positive
f
Lipase positive
Tree scale: 0.1


a
E. coli
C. symbiosum
Porcine lipase
UV light
White light
b
Cell pellet
Supernatant
Porcine lipase
c
Probability
50
100
150
200
250
300
350
400
Transmembrane
Inside
Outside
d
e
f
g
GISHG
TTINTPH
AHSKGG
h
Released fatty acids (μM)
C2
C4
C8
C10
C12
C14
C16
C18
i
Released fatty acids (μM)
PMSF
Control
j
Released fatty acids (μM)
15°C
30°C
45°C
k
Released fatty acids (μM)
pH 5
pH 6
pH 7
pH 8
pH 9


a
Aerobic beta-oxidation
13/235 (25)
K06076
fatty acid
K01897
acyl-CoA
acetyl-CoA
K00632
K01825/K13767
/K01692
enoyl-CoA
Glycerol degradation II
98/235 (66)
glycerol
K00005
glycerone
K00863
glycerone phosphate
K01803
glyceraldehyde 3-phosphate
Anaerobic beta-oxidation
0/235 (14)
K06076
fatty acid
K12507
acyl-CoA
acetyl-CoA
K01782
K06445
enoyl-CoA
b
C16:0
C18:1
C18:0
C18:2
LCFA (mg (ml sample)−1)
Vessel, time point (h)
c
Collinsella aerofaciens CLA-JM-H32B
Odoribacter splanchnicus CLA-JM-H18B
Butyricimonas virosa CLA-JM-H5
Paraprevotella clara CLA-JM-H13
Bacteroides ovatus CLA-JM-H25
Bacteroides cellulosilyticus CLA-JM-H2
Hungatella hathewayi CLA-AA-H226
Blautia aquisgranensis CLA-JM-H16
Solibaculum intestinalis CLA-JM-H44
Ruthenibacterium intestinale CLA-JM-H11
Veillonella parvula CLA-JM-H54
Enterococcus faecalis CLA-AA-H229
Tree scale: 0.1
No activity
Lipase positive
d
Labeled Proteins (number)
e
Labeled Proteins (number)
No activity
Lipase positive