

Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <http://orca.cf.ac.uk/111470/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Mullish, Benjamin H., Pechlivanis, Alexandros, Barker, Grace F., Thursz, Mark R., Marchesi, Julian R. and McDonald, Julie A.K. 2018. Functional microbiomics: Evaluation of gut microbiota-bile acid metabolism interactions in health and disease. *Methods* 149 , pp. 49-58.
10.1016/j.ymeth.2018.04.028 file

Publishers page: <http://dx.doi.org/10.1016/j.ymeth.2018.04.028>
<<http://dx.doi.org/10.1016/j.ymeth.2018.04.028>>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



1 **Functional Microbiomics: Evaluation of Gut Microbiota-Bile Acid Metabolism Interactions**
2 **in Health and Disease.**

3 Benjamin H Mullish*¹, Alexandros Pechlivanis¹, Grace F Barker¹, Mark R Thursz¹, Julian R
4 Marchesi^{1,2}, Julie A K McDonald¹.

5 1. Division of Integrative Systems Medicine and Digestive Disease, Faculty of Medicine,
6 Imperial College London, United Kingdom.

7 2. School of Biosciences, Cardiff University, Cardiff, United Kingdom.

8

9 *Corresponding author.

10 Postal address: Division of Integrative Systems Medicine and Digestive Disease
11 Liver Unit, 10th Floor, QEQM Wing
12 St Mary's Hospital Campus, Imperial College London
13 South Wharf Road
14 Paddington, London
15 W2 1NY
16 United Kingdom

17 Email: b.mullish@imperial.ac.uk

18 Telephone: +44(0)203 312 6454

19 Fax: +44(0)207 724 9369

20

21 **Keywords:** microbiota; metabonome; bile; antibiotics; 16S rRNA gene sequencing;
22 qPCR

23

24

25

26

27

28

29

30

31 **Abstract:**

32 There is an ever-increasing recognition that bile acids are not purely simple surfactant
33 molecules that aid in lipid digestion, but are a family of molecules contributing to a diverse
34 range of key systemic functions in the host. It is now also understood that the specific
35 composition of the bile acid *milieu* within the host is related to the expression and activity of
36 bacterially-derived enzymes within the gastrointestinal tract, as such creating a direct link
37 between the physiology of the host and the gut microbiota. Coupled to the knowledge that
38 perturbation of the structure and/or function of the gut microbiota may contribute to the
39 pathogenesis of a range of diseases, there is a high level of interest in the potential for
40 manipulation of the gut microbiota-host bile acid axis as a novel approach to therapeutics.
41 Much of the growing understanding of the biology of this area reflects the recent
42 development and refinement of a range of novel techniques; this study applies a number of
43 those techniques to the analysis of human samples, aiming to illustrate their strengths,
44 drawbacks and biological significance at all stages. Specifically, we used microbial profiling
45 (using 16S rRNA gene sequencing), bile acid profiling (using liquid chromatography-mass
46 spectrometry), *bsh* and *baiCD* qPCR, and a BSH enzyme activity assay to demonstrate
47 differences in the gut microbiota and bile metabolism in stool samples from healthy and
48 antibiotic-exposed individuals.

49

50 **1. Introduction:**

51 **1.1. Overview:**

52 The last few years have been associated with a rapid increase in understanding of the
53 profound contribution of the gut microbiota to the health of the host, as well as its potential
54 roles in the onset and maintenance of a range of diseases. Much initial interest in the gut
55 microbiota has focused on observational studies which defined changes to the structure of
56 the microbiota in different scenarios (e.g. different disease states, impact of diet or
57 antibiotics, etc). However, more recent emphasis has moved away from solely defining the
58 structure of the microbiota, but refocused upon better defining its function, and specifically
59 the many complex routes of communication (including metabolic pathways, immune axes,
60 etc) between the gut microbiota and the host [1]. Given that a key regulator of the
61 composition of the bile acid pool within mammals is the action of bacterially-derived enzymes
62 within the gastrointestinal tract [2], an improved understanding of the close interplay

63 between the gut microbiota and the host's bile acid metabolism is an area of particular
64 interest.

65

66 **1.2. Gut microbiota-bile acid interactions *in vivo*:**

67 Primary bile acids (BA) are synthesised from cholesterol in the liver, where they are
68 conjugated with glycine or taurine. These conjugated bile acids subsequently enter the
69 gallbladder, and are released into the duodenum following the intake of food. Once in the
70 small bowel, the bile acids undertake one of their key physiological roles, the emulsification
71 and solubilisation of dietary lipids. Bile acids will continue along the small intestine, towards
72 the terminal ileum; whilst approximately 95% of bile acids will be reabsorbed via the
73 enterohepatic circulation pathway, the remaining 5% (~400-800 mg per day) are not
74 recovered, and will continue through the distal gut of the terminal ileum and on to the colon
75 [3].

76

77 It is within the small intestine that bile acid modification by the gut microbiota is initiated,
78 driven by enzymes that are produced and secreted by gut microbiota members, but which
79 are not produced by the mammalian host. The first stage of bile acid modification by the gut
80 microbiota is from the enzymes named **bile salt hydrolases (BSHs)**. These enzymes
81 deconjugate the taurine and glycine groups from conjugated bile acids via a hydrolysis
82 reaction, and therefore reform the primary bile acids cholate (CA) and chenodeoxycholate
83 (CDCA). BSHs are found mainly within the bacterial phyla *Firmicutes* and *Bacteroidetes*, but
84 are widely-distributed throughout most major bacterial divisions and archaeal species of the
85 human gut microbiota [4]. At least eight different *bsh* genes exist (see **Supplementary Figure**
86 **1**), with each form having specific properties relating to optimal pH, specificity for taurine- or
87 glycine-conjugated bile acids and gene size [4]. The secondary enzymatic steps are **7- α -**
88 **dehydroxylation**. In these steps, the hydroxyl group of C-7 is removed, thus converting
89 primary bile acids to secondary bile acids. Specifically, in humans, this includes the conversion
90 of cholate to deoxycholate (DCA), and the conversion of chenodeoxycholate to lithocholate
91 (LCA), along with the biosynthesis of other secondary bile acids. 7- α -dehydroxylation is a
92 complex, multi-step process, and only performed by strictly anaerobic bacteria with the bile

93 acid-inducible (*bai*) operon. Based on current microbial genomic annotation, it is estimated
94 that only a very small percentage of gut microbiota members possess 7- α -dehydroxylation
95 activity, with those organisms that do predominantly belonging to the genera *Clostridium*
96 clusters XIVa and XI [5], [6]. Generation of secondary bile acids creates a more hydrophobic
97 bile acid pool, facilitating the elimination of these bile acids within faeces. A range of other
98 gut microbial metabolic actions against bile acids are also described, including the
99 epimerisation of CDCA to synthesise ursodeoxycholic acid, as well as other pathways that
100 result in the generation of iso-, allo- and oxo-/keto-bile acids [2].

101

102 There is now increasing recognition of the diverse roles of bile acids within the host, in
103 particular via their role as endogenous ligands for host cell receptors. These include the
104 nuclear receptor farnesoid X receptor (FXR), and the G protein-coupled plasma membrane
105 bile acid receptor TGR5, all exhibiting varying affinities for different bile acids and their
106 moieties [2]. Bile acids as FXR and TGR5 agonists contribute to a wealth of host physiological
107 processes including the modulation of lipid, glucose and energy homeostasis, as well as the
108 regulation of bile acid synthesis, conjugation and transport. To add to the complexity, there
109 is also evidence that bile acids influence microbiota composition, both via direct and indirect
110 actions [2]. Collectively, the growing evidence for the multiple functions of bile acids within
111 the host – coupled with evidence demonstrating the complex interplay between bile acid
112 metabolism and the gut microbiota – highlights that this axis is a key mechanism by which the
113 gut microbiota directly influences a range of aspects of host physiology.

114

115 Two of the most important questions in gut microbiome research are “who is there?” and
116 “what are they doing?”. In the context of bile metabolism we can describe changes in the gut
117 microbiota at several different levels: we can use microbial DNA to define the composition of
118 the gut microbiota and quantify the amount of bile metabolising genes, we can look at the
119 amount of bile metabolising proteins expressed by measuring their enzymatic activity, and
120 we can look at the metabolites being produced by characterising the quantity and
121 composition of bile acid metabolites. In this study we use a set of stool samples from

122 individuals exposed to antibiotics and non-antibiotic-exposed controls to demonstrate how
123 researchers can apply a wide variety of techniques to more fully characterise microbiota-bile
124 interactions in the gut. These techniques include 16S rRNA gene sequencing, liquid
125 chromatography-mass spectrometry-based bile acid profiling, BSH and 7- α -dehydroxylase
126 qPCR, and a BSH enzyme activity assay. In addition, we correlated metataxonomic and
127 metabonomic data to gain a better understanding of the modulation of the bile acid pool by
128 the gut microbiota.

129

130 **2. Material and methods:**

131 **2.1. Study participants:**

132 The study was performed under approval from the UK National Research Ethics Centre
133 (13/LO/1867). Stool samples were collected from a total of eight healthy individuals, and five
134 patients who had recently taken recurrent courses of antibiotics. Antibiotics had been
135 prescribed for a variety of indications, had been used for at least three continuous weeks
136 within the past month, and had last been used between 3 – 6 days prior to sample collection
137 (**Supplementary Table 1**). Healthy individuals had not used antibiotics or been prescribed
138 regular medications for at least six months prior to sample collection. Stool specimens were
139 put on ice within 15 minutes after collection, transferred to the hospital laboratory, and
140 homogenised and aliquoted within 30 minutes. Samples were frozen to and maintained at -
141 80°C prior to analysis.

142

143 **2.2. DNA extraction and 16S rRNA gene sequencing:**

144 DNA was extracted from 250 mg of stool using the PowerLyzer PowerSoil DNA Isolation Kit
145 (Mo Bio, Carlsbad, CA, USA) following manufacturer's instructions, with the addition of a bead
146 beating step for 3 minutes at speed 8 in a Bullet Blender Storm (ChemBio Ltd, St Albans, UK).
147 DNA was stored at -80°C until it was ready to be used.

148

149 Sample libraries were prepared following Illumina's 16S Metagenomic Sequencing Library
150 Preparation Protocol [7] with two modifications. Firstly, the V1-V2 regions of the 16S rRNA

151 gene were amplified using the primers listed in **Table 1**. Additionally, the index PCR reactions
152 were cleaned up and normalised using the SequalPrep Normalization Plate Kit (Life
153 Technologies, Paisley, UK). Sample libraries were quantified using the NEBNext Library Quant
154 Kit for Illumina (New England Biolabs, Hitchin, UK). Sequencing was performed on an Illumina
155 MiSeq platform (Illumina Inc., Saffron Walden, UK) using the MiSeq Reagent Kit v3 (Illumina)
156 and paired-end 300bp chemistry.

157

158 The resulting data was analysed using the Mothur package following the MiSeq SOP Pipeline
159 [8]. The Silva bacterial database was used for sequence alignments (www.arb-silva.de/) and
160 the RDP database reference sequence files were used for classification of sequences using the
161 Wang method [9]. The non-metric multidimensional scaling (NMDS) plot and PERMANOVA
162 p-values were generated using the UniFrac weighted distance matrix generated from Mothur,
163 and analysed using the Vegan library within the R statistical package [10]. Family-level
164 extended error bar plots were generated using the Statistical Analysis of Metagenomic
165 Profiles software package using White's non-parametric t-test with Benjamini-Hochberg FDR
166 [11]. The α diversity (Shannon diversity index, H') and richness (total number of bacterial taxa
167 observed, S_{obs}) were calculated within Mothur and statistical tests (independent t-test and
168 Mann-Whitney U test, respectively) were performed using IBM SPSS Statistics Software
169 version 23. A p-value of 0.05 and a q-value of 0.05 was considered significant.

170

171 **2.3. Inference of gut microbiota function from 16S rRNA gene sequencing data:**

172 To predict the bile-metabolising ability of the microbial communities within the samples, an
173 inferential tool, Piphillin, was applied [12]. This algorithm uses direct nearest-neighbour
174 matching between 16S rRNA gene sequencing datasets and microbial genomic databases to
175 infer the metagenomic content of the samples [12]. In this case, Piphillin was used online
176 [13], using the KEGG May 2017 as reference database, and applying 97% identity cut-off.
177 Inference of gene abundance was assessed for KEGG orthology K01442 (cholyglycine
178 hydrolase, an alternative name for BSH), KEGG orthology K15870 (*baiCD*, a bacterial gene
179 specific to the 7- α -dehydroxylation pathway) and KEGG pathway ko00121 (corresponding to
180 the secondary bile acid biosynthesis pathway).

181

182 **2.4. Ultra performance liquid chromatography-mass spectrometry (UPLC-MS) profiling**
183 **of faecal bile acids:**

184 Faecal samples were lyophilized for 24 hours using a VirTis Benchtop BTP 8ZL freeze dryer
185 (BPS, UK). The dried samples were weighed and bile acids were extracted using a 2:1:1 (vol)
186 mixture of water, acetonitrile and 2-propanol in a Biospec bead beater with 1.0 mm Zirconia
187 beads. After centrifugation (16,000 x g, 20 minutes) the supernatant was filtered using 0.45
188 µm microcentrifuge filters (Costar, Corning).

189
190 Quality control (QC) samples were prepared using equal parts of the faecal filtrates. QC
191 samples were used as an assay performance monitor[14], and as a proxy to remove features
192 with high variation. QC samples were also spiked with mixtures of bile acid standards (55 bile
193 acid standards including 36 non-conjugated, 12 conjugated with taurine, seven conjugated
194 with glycine (Steraloids, Newport, RI, USA)) and were analysed along with the stool samples
195 to determine the chromatographic retention times of bile acids and to aid in metabolite
196 identification.

197
198 Bile acid analysis of faecal extracts was performed using ACQUITY UPLC (Waters Ltd, Elstree,
199 UK) coupled to a Xevo G2 Q-ToF mass spectrometer equipped with an electrospray ionization
200 source operating in negative ion mode (ESI-), using the method described by Sarafian and
201 colleagues [15].

202
203 Waters raw data files were converted to NetCDF format and data were extracted using XCMS
204 (v1.50) package with R (v3.1.1) software. Probabilistic quotient normalisation [16] was used
205 to correct for dilution effects and chromatographic features with coefficient of variation
206 higher than 30% in the QC samples were excluded from further analysis.

207
208 The relative intensities of the features were corrected to the dry weight of the faecal samples.

209
210 **2.5. Integration of 16S rRNA gene sequencing data and bile acid mass spectrometry data:**

211 Correlations between two “omic” datasets acquired from the same set of samples were
212 determined using regularised Canonical Correlation Analysis (rCCA). rCCA modelling of

213 metataxonomic (16S rRNA gene sequencing) and metabonomic (bile acid mass spectrometry)
214 data was employed in the mixOmics library within the R statistical package [17], [18]. The
215 regularisation parameters were determined using the shrinkage method. The rCCA similarity
216 scores between the variables were plotted as heatmaps using the clustered image maps (cim)
217 function. Hierarchical clustering (complete linkage, Euclidean distance) was used to obtain
218 the order of the variables. The correlation circle plot was generated using the plotVar
219 function, which plots strong correlations between variables (plots variables with a correlation
220 above 0.5 outside of the inner circle).

221

222 **2.6. Abundance and activity of bile-metabolising enzymes:**

223 **2.6.1. Real-time PCR for the quantification of BSH and *baiCD* gene abundance:**

224 qPCR was performed using extracted DNA to quantify gene abundance. Gene abundance was
225 quantified for i) specified groups of *bsh* (using degenerate primer sets previously designed
226 and optimised by our group (**Table 2**)) and ii) *baiCD* (using primers previously described in the
227 literature [19]).

228

229 A total reaction volume of 25µl was used for each reaction, consisting of 20µl master mix and
230 5µl diluted DNA (12.5ng total per reaction). All DNA was diluted in buffer EB (Qiagen, Hilden,
231 Germany). A standard master mix consisting of 5.5µl PCR grade water (Roche, Penzberg,
232 Germany), 12.5µl of 2x SYBR green master mix (ThermoFisher Scientific, Waltham,
233 Massachusetts, USA), 1µl of 10µM forward primer (Eurofins Genomics, Wolverhampton, UK)
234 and 1µl of 10µM reverse primer (Eurofins Genomics) was used. One bacterial strain from the
235 relevant reference group was selected as a standard for each primer set (*bsh* group 1a –
236 *Bacteroides plebius*; *bsh* group 1b – *Bacteroides ovatus*; *bsh* group 3c/e – *Blautia obeum*;
237 *baiCD* – *Clostridium scindens* (DSMZ 5676, Braunschweig, Germany) (**Supplementary**
238 **Methods**). Serial dilutions of each isolate were used to create a standard curve.
239 Thermocycling conditions for each primer set are summarised in **Table 2**. A melt curve stage
240 was performed post-cycling to confirm primer specificity. Products were also visualised using
241 the 2200 TapeStation System (Aligent Technologies, Santa Clara, California, USA) in
242 combination with D1000 Reagents and D100 Screentapes (Aligent Technologies), following
243 the manufacturer's protocol.

244

245 Copy number was calculated from qPCR data using the following formula: gene abundance =
246 (quantity (ng) x 6.022×10^{23} (gene copy number/mol)) / (length of product x 1×10^9 (ng/g) x
247 660 (g/mol)). A mean copy number for each set of triplicates was calculated and divided by
248 the total DNA per reaction to obtain average copy number per ng DNA.

249

250 **2.6.2. Bile salt hydrolase enzyme activity assay:**

251 Faecal water was prepared and total faecal protein quantified using a similar method to that
252 previously-described by Morris and Marchesi [20], but with the addition of bacterial and
253 mammalian protease inhibitor cocktails (G Biosciences, St Louis, MO, USA), as well as DTT to
254 1mM final concentration (Roche, Welwyn Garden City, UK) to minimise enzyme oxidation
255 [21].

256

257 The BSH assay itself was an adaptation of a precipitation-based assay [21]–[23]. The assay
258 was performed in a clear flat-bottomed 96-well microtitre plate and incubated at 37°C at pH
259 5.8 for up to 8 hours. In a total volume of 200µl, 500µg of faecal protein was incubated with
260 sodium phosphate buffer (pH 5.8, final concentration of 0.02mM), and taurodeoxycholic acid
261 (Merck, Damstadt, Germany) (at final concentration 1mM). To prevent evaporation during
262 incubation, wells were overlaid with 50µl of light paraffin oil (0.85g/ml; PanReac AppliChem,
263 Barcelona, Spain) [23]. Samples were assayed in triplicate, with precipitation of insoluble
264 deoxycholic acid monitored by absorbance measurement at 600nm (A_{600}) using a microplate
265 reader (MultiSkan Go, Thermo Scientific, Dartford, UK). Faecal protein incubated with
266 phosphate-buffered saline served as a negative control, and faecal protein incubated with
267 varying concentrations of deoxycholic acid (Merck) was used to establish a standard curve to
268 quantify precipitate formation.

269

270 **2.6.3. Statistical analysis:**

271 A Mann-Whitney U test was used to compare the BSH activity and the BSH and *baiCD* gene
272 abundance data between the antibiotic treated and healthy cohorts. A p-value of <0.05 was
273 considered significant.

274

275 3. Results:**276 3.1. 16S rRNA gene sequencing:**

277 16S rRNA gene sequencing analysis showed patients taking recurrent antibiotics had altered
278 compositions of their gut microbiotas compared to healthy controls (**Figure 1A**, $p < 0.01$,
279 PERMANOVA). Patients taking recurrent antibiotics had lower microbial community diversity
280 (**Figure 1B**, $p < 0.001$, independent t-test) and richness (**Figure 1C**, $p < 0.01$, Mann-Whitney U
281 test) compared to healthy controls. Statistical analysis showed that the altered microbiota in
282 patients taking recurrent antibiotics were due to decreases in the relative abundances of the
283 families *Bacteroidaceae*, *Lachnospiraceae*, *Ruminococcaceae*, and *Oscillospiraceae*, and
284 increases in the relative abundance of the family *Enterobacteriaceae* compared to healthy
285 controls (**Figure 1D**).

286

287 3.2. Inference of gut microbiota function from 16S rRNA gene sequencing data:

288 Results from Piphillin analysis are shown in **Figure 2**. Predicted gene abundance for *bsh* (KEGG
289 orthologue K01442) was significantly reduced in patients who had taken recurrent antibiotics
290 (**Figure 2A**, $p < 0.05$, Mann-Whitney U test). It was not possible to predict gene abundance
291 counts for all samples for *baiCD* (KEGG orthologue K15870) at the cut-off of 97% identity used,
292 implying very low counts. Predicted secondary bile acid biosynthesis (ko00121) trended
293 lower in patients with recurrent antibiotic use compared to controls, but this was not
294 significant (**Figure 2B**, $p = 0.08$).

295

296 3.3. Multivariate statistics analysis of UPLC-MS profiling data:

297 The data table produced by XCMS after normalization to the dry weight of the samples was
298 introduced to SIMCA 14.1 (MKS Umetrics AB). Principal component analysis (PCA) was
299 performed to visualise clustering of samples and assess the quality of the run using the QC
300 samples (**Figure 3A**). Furthermore, supervised OPLS-DA was performed (**Figure 3B**) to reveal
301 the features that were responsible for the discrimination between the recurrent antibiotic-
302 treated and healthy control groups. This feature identification was achieved using the S-plot
303 presented in **Figure 3C**, where feature in the edges of the S-shaped cloud of features were

304 responsible for the separation. Features on top right were higher in the healthy control
305 group, and in bottom left higher in the group treated with recurrent antibiotics. Annotated
306 bile acids are highlighted in the plot.

307

308 Univariate analysis for differences in specific bile acids between healthy participants and
309 people treated with recurrent antibiotics was also performed; data are presented in
310 **Supplementary Figure 2.**

311

312 **3.4. Integration of metataxonomic and metabonomic data:**

313 rCCA modelling was used to determine correlations between metataxonomic (16S rRNA gene
314 sequencing) and **metabonomic** (bile acid mass spectrometry) data (**Figure 4**). We found that
315 correlations between bacterial families and bile acids clustered into three distinct groups
316 (**Figure 4**). Group 1 consisted of correlations where bacterial families were positively
317 associated with conjugated and unconjugated primary bile acids, and negatively correlated
318 with secondary bile acids DCA and LCA. Group 2 consisted of families positively correlated
319 with ursodeoxycholic acid. Group 3 consisted of families positively correlated with secondary
320 bile acids DCA and LCA, and negatively associated with unconjugated primary bile acids CA
321 and CDCA. *Enterobacteriaceae*, which increased in the recurrent antibiotics group, clustered
322 in group 1. *Bacteroidaceae*, *Lachnospiraceae*, *Ruminococcaceae*, and *Oscillospiraceae*, which
323 decreased in the recurrent antibiotics group, clustered in group 3.

324

325 **3.5. Abundance and activity of bile-metabolising enzymes:**

326 Results from qPCR assays are displayed in **Figure 5**. Recurrent antibiotic use was associated
327 with a significantly reduced abundance of *bsh* genes for all BSH groups tested compared to
328 healthy control participants. Specifically, after recurrent antibiotic use, there was reduced
329 abundance of the genes of *bsh* group 1a gene ($p < 0.01$, Mann-Whitney U test), *bsh* group 1b
330 gene ($p < 0.05$, Mann-Whitney U test), and *bsh* group 3c/e gene ($p < 0.01$, Mann-Whitney U
331 test). *baiCD* gene abundance also significantly reduced after recurrent antibiotic use ($p < 0.05$,
332 Mann-Whitney U test).

333

334 Use of recurrent antibiotics is associated with marked reduction in BSH enzyme activity within
335 faecal samples (**Figure 6**, $p < 0.01$, Mann-Whitney U test).

336

337 **4. Discussion and Conclusions:**

338 In this study, we performed a range of analyses upon stool samples taken from healthy
339 participants and people with recent antibiotic use as a means of demonstrating a range of
340 techniques that may be applied to delineate gut microbiota-host bile acid interactions.

341

342 We found that patients taking recurrent antibiotics had gut microbiotas with reduced
343 proportions of known bile-metabolising enzyme function, including the families
344 *Bacteroidaceae*, *Lachnospiraceae* and *Ruminococcaceae*. Consistent with this, recurrent
345 antibiotic use was associated with enrichment of stool primary bile acids (both conjugated
346 and unconjugated) and loss of secondary bile acids. Correlation analysis showed a distinct
347 clustering of bacterial families and bile acids into three groups, where *Enterobacteriaceae* was
348 positively correlated with unconjugated primary bile acids, and *Bacteroidaceae*,
349 *Lachnospiraceae*, *Ruminococcaeceae* and *Oscillospiraceae* were positively correlated with
350 secondary bile acids. Further analysis demonstrated a loss of BSH gene abundance and
351 enzyme activity within the gut of antibiotic-treated patients, coupled with a loss of 7- α -
352 dehydroxylase *baiCD* gene abundance related to antibiotic use. Most fundamentally, these
353 results emphasise the close and complex interplay between the gut microbiota and bile acid
354 metabolism, and reinforce that any perturbation of the gut microbiota (in this case by
355 antibiotics) may result in marked changes to host physiology. These findings are consistent
356 with other comparable work within this area, including the demonstration that early life
357 antibiotic exposure is associated with a long lasting reduction in bile salt hydrolase function
358 [24]. Furthermore, it has also been recognised that *Clostridium difficile* infection (a
359 gastrointestinal infection occurring predominantly in patients with antibiotic-associated gut
360 dysbiosis) is associated with perturbation of host bile acid profiles, possibly mediated through
361 alteration of gut bile metabolising enzyme functionality [25]–[27].

362

363 We used 16S rRNA gene sequencing to determine the differences in the composition of the
364 gut microbiota between patients taking recurrent antibiotics and healthy controls. We found
365 an increase in the relative abundance of *Enterobacteriaceae* and a decrease in the relative
366 abundance of *Bacteroidaceae*, *Lachnospiraceae*, *Ruminococcaceae*, and *Oscillospiraceae* in
367 the recurrent antibiotic group compared to healthy controls. However, it is important to note
368 that we are reporting changes in the relative abundances of these groups, not the absolute
369 abundances. The total read numbers per sample does not provide information on the total
370 number of 16S rRNA gene copies in the sample [28]. This is especially important in samples
371 where a change in the total bacterial biomass occurs, for example with antibiotic treatment
372 (as is the case in this study). While it is possible that the absolute abundance of
373 *Enterobacteriaceae* increases after recurrent antibiotics, it is also possible that the absolute
374 abundance of *Enterobacteriaceae* has remained unchanged, and there was a decrease in the
375 total biomass due to a decrease in the absolute abundances of *Bacteroidaceae*,
376 *Lachnospiraceae*, *Ruminococcaceae*, and *Oscillospiraceae*. Studies can account for these
377 changes in bacterial biomass by performing 16S rRNA gene qPCR, and weighting their relative
378 abundance data to get a more informative representation of the microbial community
379 composition.

380

381 Whilst 16S rRNA gene sequencing data provides information on the bacterial composition of
382 the sample, it does not provide information regarding the potential functional capabilities of
383 the bacteria and subsequent interactions with the host. Metagenomic sequencing provides
384 information on the collection of genomes in a sample, followed by assembly or mapping to a
385 reference database which allows gene annotation. However, metagenomic sequencing is
386 more expensive than metataxonomics, and the data analysis can be more challenging. In this
387 study, we used Piphillin [12] to indirectly infer the abundance of functional genes as a
388 straightforward and cost-free addition to the study. Piphillin has certain advantages
389 compared to other inferential software tools (including its ease of use, speed of output and
390 the ability to select a reference database of interest [12]), but has not to our knowledge been
391 applied before now for analysis of human gut metataxonomic data. Our intention was to use
392 this method as an exploratory technique, to later confirm with additional methods of analysis
393 (qPCR, LC-MS, and an enzyme assay). The Piphillin results here predicted a reduced *bsh* gene
394 abundance in the recurrent antibiotic group compared to healthy controls, and our qPCR data

395 and enzyme assay were consistent with this. Whilst Piphillin predicted a trend towards
396 reduced secondary bile acid biosynthesis within the recurrent antibiotic group, it was not able
397 to specifically predict *baiCD* gene abundance, and we used qPCR to explore this instead. Our
398 experience here and in other work with inferential algorithms is that whilst they may be a
399 helpful and broadly accurate additional tool to start exploring the function of the microbiota,
400 the current limitations in metagenomic annotation mean that results obtained in this way
401 must be interpreted with caution. However, the constant improvements in metagenome
402 annotation are likely to make such tools ever-more accurate over time.

403

404 Mass spectrometric techniques are the workhorse of bile acids analysis due to their sensitivity
405 and specificity compared to other assays. High resolution time-of-flight mass spectrometry
406 using a soft ionization method (electrospray ionization, ESI) coupled with ultra-performance
407 liquid chromatography is our analytical method of choice as it can provide comprehensive
408 coverage of bile acids and lipids species from complex biological samples needing minimal
409 sample pre-treatment [15]. In our study, we found that antibiotic exposure had a significant
410 impact upon the composition of the bile acid pool, which could have implications on host
411 physiology. In order to develop interventions that target the bile acid metabolic pathway,
412 researchers need to be able to identify specific bacterial taxa responsible for these bile acid
413 conversions.

414

415 One difficulty with 'omics' methodologies is the complexity of the datasets generated, often
416 with very large numbers of variables. Software packages such as mixOmics offer researchers
417 useful exploratory approaches to highlight important correlations between bacteria and
418 metabolites. Integration of metataxonomic and metabonomic data can provide researchers
419 with information on the potential roles of microorganisms within an ecosystem, however it
420 is important to remember that correlation does not equal causation. Strong correlations
421 between bacteria and metabolites must be confirmed with further experiments, such as
422 assays *in vitro* where researchers can assess the direct effects of a substrate/metabolite on
423 the growth or activity of a microorganisms of interest. Examples of assays *in vitro* which may
424 be used include batch cultures, mammalian cell line assays, enzyme assays, etc. It is also
425 important to note that there is no consensus on which data integration method is the best

426 method to integrate metataxonomic and metabonomic data sets, as this is an actively
427 developing field of research.

428

429 *bsh* qPCR primer sets were designed to quantify the differences in *bsh* gene abundance in our
430 samples. We found a statistically significant decrease in *bsh* group 1a, group 1b, and group
431 G3c/e gene abundance, together with a significant reduction in that of *baiCD*, associated with
432 antibiotic use. Even though these primers were optimised by us to target a select group of
433 BSH-producing bacteria and were confirmed to not cross-react between groups, the bacterial
434 strains used from each group during the optimisation stage were subject to availability.
435 Therefore, it is reasonable to suggest that, due to their degenerate nature, the primers could
436 also target the *bsh* gene in other bacterial species within a group which were not tested during
437 the optimisation stage, thereby potentially providing a more comprehensive assessment of
438 *bsh* gene abundance within the faecal samples. DNA sequencing would be required to
439 categorically confirm the BSH-producing bacterial species targeted by these primer sets. We
440 also performed qPCR of the *baiCD* operon; whilst this operon is not found in all bacteria with
441 7- α -dehydroxylating ability, it is present within the two bacterial species with high activity of
442 this enzyme, *Clostridium scindens* and *Clostridium hiranonis*, and most strains of these species
443 will be amplified by this PCR [19]. Furthermore, *Clostridium scindens* is particularly of interest
444 within this context, since its loss from the gut microbiota in association with antibiotic use has
445 been associated with altered gut bile acid metabolism and a potential vulnerability to
446 *Clostridium difficile* infection [27]. Whilst this qPCR will not amplify certain bacteria with low
447 secondary bile acid biosynthesis functionality (including *Clostridium leptum* and *Clostridium*
448 *sordeii*), good correlation has been noted between *baiCD* PCR assay results and 7- α -
449 dehydroxylase activity in an *in vitro* assay, demonstrating that this is still a highly useful assay
450 [19].

451

452 Whilst qPCR of bacterial genes is useful, similar to metataxonomic data, there are concerns
453 that what is being assessed relates to which bacterial genes are present, rather than if those
454 genes are being actively transcribed and the resultant functional effects. As such,
455 metatranscriptomics – the sequencing of RNA from within a microbial community – is of great
456 interest for its ability to more directly establish gene transcription and therefore microbiota
457 functionality. However, there remain certain practical difficulties in undertaking such studies,

458 including the considerable cost, the computational complexity, and the difficulties in high-
459 quality RNA extraction and sequencing given its relative instability compared to DNA.
460 Furthermore, whilst protocols have been described that aim to simplify collection of stool and
461 preserve samples for subsequent streamlined combined metagenomic and
462 metatranscriptomic analysis (e.g. via the addition of ethanol or RNAlater to samples) [29], the
463 implications of these preservatives upon the quality of the metabolic profile obtained from
464 the sample remain undefined.

465

466 The quantification of gene abundance using qPCR data, metagenomic data, and Piphillin data
467 cannot categorically confirm gene expression and functionality *in vivo*. Therefore, we
468 developed an enzyme activity assay to measure the amount of BSH activity in each sample
469 through substantial adaptation of a plate-based precipitation assay [21]. Other groups have
470 used a ninhydrin assay to measure BSH activity [23], [30]; however, these studies used pure
471 bacterial strains, and in our experience, this assay is not sensitive enough to detect BSH
472 activity within faecal water. Whilst BSH activity does not require strict anaerobic conditions,
473 7- α -dehydroxylation does [33], complicating development of a similar activity assay.
474 However, an assay applying thin layer chromatography and radiolabelled cholic acid to human
475 caecal aspirate or stool obtained after enema use to assess 7- α -dehydroxylase activity has
476 been described [33], [32].

477

478 In this study, we compared healthy people with patients taking antibiotics, and did not match
479 the participants for other demographics. There are a variety of variables that have been
480 shown to influence the composition and/or functionality of the gut microbiota, which (in
481 addition to antibiotics/ microbial infections) include diet, age, surgery, stress, BMI, and
482 pregnancy[34]–[36]. As such, we are unable to say if the differences seen between our groups
483 related purely to antibiotic use, or if there was a contribution from other factors. Where
484 studies compare healthy and diseased groups in attempting to generate novel hypotheses
485 regarding the contribution of gut microbiota-bile acid interactions to the disease process,
486 regard for these factors must be taken to ensure that control groups are appropriate.

487

488 Future challenges regarding methodology within this area remain. The relationship between
489 the gut microbiota, bile acid metabolism and the host is complex and bidirectional, and

490 methodologies that further delineate this relationship are required. Development of
491 standardised pipelines for analysing these complex datasets – coupled with more
492 standardised methods for integration of different data sets – are key immediate challenges.
493 At present, whilst there is growing sophistication in our ability to define and correlate gut
494 microbial and bile acid profiles, there is little work (particularly within humans) that has linked
495 this back to systemic host effects. Given the growing recognition that bile acids are signalling
496 molecules with complex systemic effects upon the host, it is clearly of interest and importance
497 to be able to link microbial and bile acid interplay to host physiological function, in relation to
498 health and disease.

499

500 **Acknowledgements:** Infrastructure support was provided by the National Institute for
501 Health Research (NIHR) Imperial Biomedical Research Centre (BRC).

502

503 **Funding:** BHM is the recipient of a Medical Research Council Clinical Research Training
504 Fellowship (grant reference: MR/R000875/1). The Division receives financial support from
505 the National Institute of Health Research (NIHR) Imperial Biomedical Research Centre (BRC)
506 based at Imperial College Healthcare NHS Trust and Imperial College London.

507

508 **5. References:**

509 [1] E. Holmes, J. V. Li, J. R. Marchesi, and J. K. Nicholson, "Gut microbiota composition
510 and activity in relation to host metabolic phenotype and disease risk," *Cell*
511 *Metabolism*, vol. 16, no. 5. Cell Press, pp. 559–564, 07-Nov-2012.

512 [2] A. Wahlströ, S. I. Sayin, H.-U. Marschall, and F. Bäckhed, "Intestinal Crosstalk
513 between Bile Acids and Microbiota and Its Impact on Host Metabolism," *Cell Metab.*,
514 vol. 24, pp. 41–50, 2016.

515 [3] A. F. Hofmann, "The Continuing Importance of Bile Acids in Liver and Intestinal
516 Disease," *Arch. Intern. Med.*, vol. 159, no. 22, p. 2647, Dec. 1999.

517 [4] B. V Jones, M. Begley, C. Hill, C. G. M. Gahan, and J. R. Marchesi, "Functional and
518 comparative metagenomic analysis of bile salt hydrolase activity in the human gut
519 microbiome.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 105, no. 36, pp. 13580–5, Sep. 2008.

- 520 [5] M. Kitahara, F. Takamine, T. Imamura, and Y. Benno, "Assignment of Eubacterium sp.
521 VPI 12708 and related strains with high bile acid 7 α -dehydroxylating activity to
522 Clostridium scindens and proposal of Clostridium hylemonae sp. nov., isolated from
523 human faeces," *Int. J. Syst. Evol. Microbiol.*, vol. 50, no. 3, pp. 971–978, May 2000.
- 524 [6] J. M. Ridlon, D.-J. Kang, and P. B. Hylemon, "Bile salt biotransformations by human
525 intestinal bacteria," *J. Lipid Res.*, vol. 47, no. 2, pp. 241–259, Feb. 2006.
- 526 [7] Illumina, "16S Metagenomic Sequencing Library Preparation." [Online]. Available:
527 [https://support.illumina.com/downloads/16s_metagenomic_sequencing_library_pre](https://support.illumina.com/downloads/16s_metagenomic_sequencing_library_preparation.html)
528 [paration.html](https://support.illumina.com/downloads/16s_metagenomic_sequencing_library_preparation.html). [Accessed: 27-Nov-2017].
- 529 [8] J. J. Kozich, S. L. Westcott, N. T. Baxter, S. K. Highlander, and P. D. Schloss,
530 "Development of a dual-index sequencing strategy and curation pipeline for analyzing
531 amplicon sequence data on the MiSeq Illumina sequencing platform.," *Appl. Environ.*
532 *Microbiol.*, vol. 79, no. 17, pp. 5112–20, Sep. 2013.
- 533 [9] Q. Wang, G. M. Garrity, J. M. Tiedje, and J. R. Cole, "Naive Bayesian Classifier for
534 Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy," *Appl.*
535 *Environ. Microbiol.*, vol. 73, no. 16, pp. 5261–5267, Aug. 2007.
- 536 [10] "R: The R Project for Statistical Computing." [Online]. Available: [https://www.r-](https://www.r-project.org/)
537 [project.org/](https://www.r-project.org/). [Accessed: 27-Nov-2017].
- 538 [11] D. H. Parks, G. W. Tyson, P. Hugenholtz, and R. G. Beiko, "STAMP: statistical analysis
539 of taxonomic and functional profiles," *Bioinformatics*, vol. 30, no. 21, pp. 3123–3124,
540 Nov. 2014.
- 541 [12] S. Iwai *et al.*, "Piphillin: Improved Prediction of Metagenomic Content by Direct
542 Inference from Human Microbiomes," *PLoS One*, vol. 11, no. 11, p. e0166104, Nov.
543 2016.
- 544 [13] Second Genome Therapeutics, "Piphillin." [Online]. Available:
545 <http://secondgenome.com/solutions/resources/data-analysis-tools/piphillin/>.
546 [Accessed: 14-Feb-2018].
- 547 [14] T. Sangster, H. Major, R. Plumb, A. J. Wilson, and I. D. Wilson, "A pragmatic and

- 548 readily implemented quality control strategy for HPLC-MS and GC-MS-based
549 metabonomic analysis," *Analyst*, vol. 131, no. 10, p. 1075, Oct. 2006.
- 550 [15] M. H. Sarafian *et al.*, "Bile Acid Profiling and Quantification in Biofluids Using Ultra-
551 Performance Liquid Chromatography Tandem Mass Spectrometry," *Anal. Chem.*, vol.
552 87, no. 19, pp. 9662–9670, Oct. 2015.
- 553 [16] K. A. Veselkov *et al.*, "Optimized Preprocessing of Ultra-Performance Liquid
554 Chromatography/Mass Spectrometry Urinary Metabolic Profiles for Improved
555 Information Recovery," *Anal. Chem.*, vol. 83, no. 15, pp. 5864–5872, Aug. 2011.
- 556 [17] K.-A. Lê Cao, I. González, and S. Déjean, "integrOmics: an R package to unravel
557 relationships between two omics datasets," *Bioinformatics*, vol. 25, no. 21, pp. 2855–
558 2856, Nov. 2009.
- 559 [18] I. Gonzalez, S. Déjean, P. Martin, and A. Baccini, "CCA : An R Package to Extend
560 Canonical Correlation Analysis," *J. Stat. Softw.*, vol. 23, no. 12, pp. 1–14, Jan. 2008.
- 561 [19] J. E. Wells, K. B. Williams, T. R. Whitehead, D. M. Heuman, and P. B. Hylemon,
562 "Development and application of a polymerase chain reaction assay for the detection
563 and enumeration of bile acid 7 α -dehydroxylating bacteria in human feces.," *Clin.*
564 *Chim. Acta.*, vol. 331, no. 1–2, pp. 127–34, May 2003.
- 565 [20] L. S. Morris and J. R. Marchesi, "Assessing the impact of long term frozen storage of
566 faecal samples on protein concentration and protease activity," *J. Microbiol.*
567 *Methods*, vol. 123, pp. 31–38, Apr. 2016.
- 568 [21] K. Smith, X. Zeng, J. Lin, S. Chincholkar, and M. Mohnl, "Discovery of Bile Salt
569 Hydrolase Inhibitors Using an Efficient High-Throughput Screening System," *PLoS One*,
570 vol. 9, no. 1, p. e85344, Jan. 2014.
- 571 [22] W. H. Ling, R. Korpela, H. Mykkänen, S. Salminen, and O. Hänninen, "Lactobacillus
572 strain GG supplementation decreases colonic hydrolytic and reductive enzyme
573 activities in healthy female adults.," *J. Nutr.*, vol. 124, no. 1, pp. 18–23, Jan. 1994.
- 574 [23] H. Tanaka, H. Hashiba, J. Kok, and I. Mierau, "Bile salt hydrolase of *Bifidobacterium*
575 *longum*-biochemical and genetic characterization.," *Appl. Environ. Microbiol.*, vol. 66,

- 576 no. 6, pp. 2502–12, Jun. 2000.
- 577 [24] K. Korpela *et al.*, “Intestinal microbiome is related to lifetime antibiotic use in Finnish
578 pre-school children,” *Nat. Commun.*, vol. 7, p. 10410, Jan. 2016.
- 579 [25] A. R. Weingarden *et al.*, “Microbiota transplantation restores normal fecal bile acid
580 composition in recurrent *Clostridium difficile* infection,” *AJP Gastrointest. Liver
581 Physiol.*, vol. 306, no. 4, pp. G310–G319, Feb. 2014.
- 582 [26] J. R. Allegretti *et al.*, “Recurrent *Clostridium difficile* infection associates with distinct
583 bile acid and microbiome profiles,” *Aliment. Pharmacol. Ther.*, vol. 43, no. 11, pp.
584 1142–53, Jun. 2016.
- 585 [27] C. G. Buffie *et al.*, “Precision microbiome reconstitution restores bile acid mediated
586 resistance to *Clostridium difficile*,” *Nature*, vol. 517, no. 7533, pp. 205–208, Oct.
587 2014.
- 588 [28] G. B. Gloor and G. Reid, “Compositional analysis: a valid approach to analyze
589 microbiome high-throughput sequencing data,” *Can. J. Microbiol.*, vol. 62, no. 8, pp.
590 692–703, Aug. 2016.
- 591 [29] E. A. Franzosa *et al.*, “Relating the metatranscriptome and metagenome of the human
592 gut,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 111, no. 22, pp. E2329–38, Jun. 2014.
- 593 [30] Z. Wang, X. Zeng, Y. Mo, K. Smith, Y. Guo, and J. Lin, “Identification and
594 characterization of a bile salt hydrolase from *Lactobacillus salivarius* for development
595 of novel alternatives to antibiotic growth promoters,” *Appl. Environ. Microbiol.*, vol.
596 78, no. 24, pp. 8795–802, Dec. 2012.
- 597 [31] C. Degirolamo, S. Rainaldi, F. Bovenga, S. Murzilli, and A. Moschetta, “Microbiota
598 modification with probiotics induces hepatic bile acid synthesis via downregulation of
599 the Fxr-Fgf15 axis in mice,” *Cell Rep.*, vol. 7, no. 1, pp. 12–8, Apr. 2014.
- 600 [32] L. A. Thomas, M. J. Veysey, G. French, P. B. Hylemon, G. M. Murphy, and R. H.
601 Dowling, “Bile acid metabolism by fresh human colonic contents: a comparison of
602 caecal versus faecal samples,” *Gut*, vol. 49, no. 6, pp. 835–42, Dec. 2001.
- 603 [33] L. A. Thomas, A. King, G. L. French, G. M. Murphy, and R. H. Dowling, “Cholyglycine

604 hydrolase and 7 α -dehydroxylase optimum assay conditions in vitro and caecal
605 enzyme activities ex vivo," *Clin. Chim. Acta*, vol. 268, no. 268, pp. 61–72, 1997.

606 [34] A. W. DuPont and H. L. DuPont, "The intestinal microbiota and chronic disorders of
607 the gut," *Nat. Rev. Gastroenterol. Hepatol.*, vol. 8, no. 9, pp. 523–31, Aug. 2011.

608 [35] E. K. Costello, K. Stagaman, L. Dethlefsen, B. J. M. Bohannan, and D. A. Relman, "The
609 application of ecological theory toward an understanding of the human
610 microbiome.," *Science*, vol. 336, no. 6086, pp. 1255–62, Jun. 2012.

611 [36] M. Dave, P. D. Higgins, S. Middha, and K. P. Rioux, "The human gut microbiome:
612 current knowledge, challenges, and future directions.," *Transl. Res.*, vol. 160, no. 4,
613 pp. 246–57, Oct. 2012.

614

615

616 **Figure legends:**

617 **Figure 1:** Antibiotics alters the gut microbiota composition in patients taking recurrent
618 antibiotics compared to healthy controls. (A) Nonmetric multidimensional scaling (NMDS)
619 plot showing the difference in gut microbiota composition of patients taking recurrent
620 antibiotics and healthy controls ($p < 0.01$, PERMANOVA). (B) α diversity was decreased in
621 patients taking recurrent antibiotics compared to healthy controls (** $p < 0.001$,
622 independent t-test). (C) Richness (total number of bacterial taxa observed) was decreased in
623 patients taking recurrent antibiotics compared to healthy controls (** $p < 0.01$, Mann-
624 Whitney U test). (D) Extended error bar plot comparing the differences in the mean
625 proportions of significantly altered families and the difference in the proportions of the means
626 (White's non-parametric t-test with Benjamini-Hochberg FDR). Plot only shows families where
627 the difference between the proportions was greater than 1%.

628

629 **Figure 2:** Inference of bile-metabolising function from 16S data using Piphillin. (A) Bile salt
630 hydrolase KEGG orthologue counts (K01442) (* $p < 0.05$, Mann-Whitney U test). (B) Secondary
631 bile acid biosynthesis KEGG orthologue counts (ko00121) ($p > 0.05$, Mann-Whitney U test).

632

633 **Figure 3:** Multivariate analysis of UPLC-MS bile acid profiling data. (A) PCA scores plot (B)
634 OPLS-DA scores plot (C) OPLS-DA S-plot, showing the contribution of bile acids to the
635 separation of the two groups. AB: recurrent antibiotic treated patients; HC: healthy controls;
636 QC: quality controls.

637

638 **Figure 4:** Regularized CCA (rCCA) modelling of metataxonomic (16S rRNA gene sequencing
639 data, family-level) and metabonomic data (bile acid data). (A) The representation of units for
640 the first two canonical variates showing the correlations between variables in patients
641 receiving recurrent antibiotics and healthy controls. (B) Correlation circle plot showing strong
642 correlations between metataxonomic and metabonomic data (plot only shows variables with
643 a correlation above 0.5). Variables projected in the same direction from the origin have a
644 strong positive correlation, and variables projected in opposite directions from the origin have
645 strong negative correlations. Variables that are at a farther distance from the origin have a
646 stronger correlation. (C) Heatmaps of the rCCA similarity scores between metataxonomic and
647 metabolomic data. Bacterial families outlined in black boxes clustered according to
648 correlations with distinct groups of bile acids.

649

650 **Figure 5:** qPCR to quantify gene abundance of bile metabolising genes. (A) *bsh* group 1a gene
651 (** $p < 0.01$, Mann-Whitney U test); (B) *bsh* group 1b gene (* $p < 0.05$, Mann-Whitney U test);
652 (C) *bsh* group 3c/e (** $p < 0.01$, Mann-Whitney U test); (D) *baiCD* gene ($p < 0.05$, Mann-
653 Whitney U test).

654

655 **Figure 6:** Bile salt hydrolase (BSH) enzyme activity assay. Taurodeoxycholic acid was used as
656 the substrate for the enzyme assay, and results are therefore expressed as rate of deoxycholic
657 acid formation (* $p < 0.05$, Mann-Whitney U test).

658 **Tables:**

659

660 **Table 1.** Primers used for 16S rRNA gene sequencing on the Illumina MiSeq. The forward
 661 primer mix was composed of four different forward primers, mixed at a ratio of 4:1:1:1 (28F-
 662 YM:28F-Borrellia:28FChloroflex:28F-Bifdo). Bases in bold are the MiSeq adapter sequences.

Primer name	Primer sequence
28F-YM (forward primer)	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGAGTTTGATYMTGGCTCAG
28F-Borrellia (forward primer)	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGAGTTTGATCCTGGCTTAG
28FChlorofle x (forward primer)	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGAATTTGATCTTGGTTCAG
28F-Bifdo (forward primer)	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGGTTTCGATTCTGGCTCAG
388R (reverse primer)	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGCTGCCTCCCGTAGGAG T

663

664

665

666

667

668

669

670

671

672

673

674 **Table 2.** Primers sequence and PCR conditions for *bsh* and *baiCD* qPCR.

675

676

677

678

679

680

681

682

683

684

685

686

687

688

689

690

691

692

693

694

695

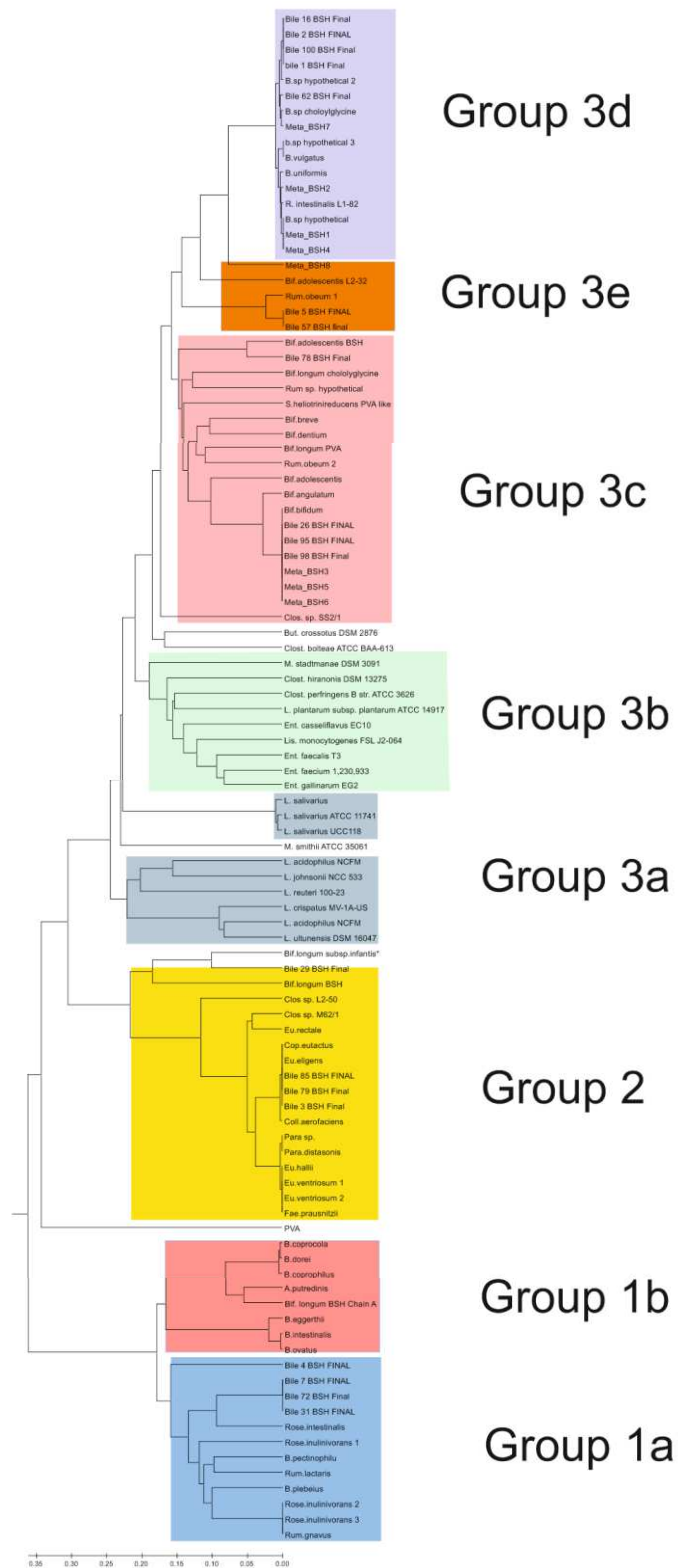
Group	Primer Sequence (5'-3')	F/R	Cycling Conditions	Expected Product Size (bp)
1a	CACATATTGTGGCACGAACAATH GAR TGGGG	F	95°C for 10 min, (95°C for 15 sec, 55°C for 1 min) x 40 cycles	570
	CTGTGCCCGGATACAGATTAACR TAR TTRTT	R		
1b	CGGCGTTCCGCATTTYTAYGARA A	F	95°C for 10 min, (95°C for 15 sec, 55°C for 1 min) x 40 cycles	318
	GTTCAATGCCAATCGGAATATCR AAR TTRTT	R		
3c/e	TTTTGGCCGAACACTGGAYTAYG ARTT	F	95°C for 5 min, (95°C for 15 sec, 54°C for 30 sec, 72 for 10 min) x 40 cycles	774
	TCAACGGAGCCCAGAATATGRA ARA AYTG	R		
<i>baiCD</i>	GGWTTCCAGCCRCAGATGTTCTT TG	F	94°C for 2 min, (94°C for 20 sec, 52°C for 30 sec, 69°C for 90 sec) x 35 cycles, 68°C for 10 min	1300
	GAATTCCGGGTTTCATGAACATT CTKCKAAG	R		

Supplementary Material:

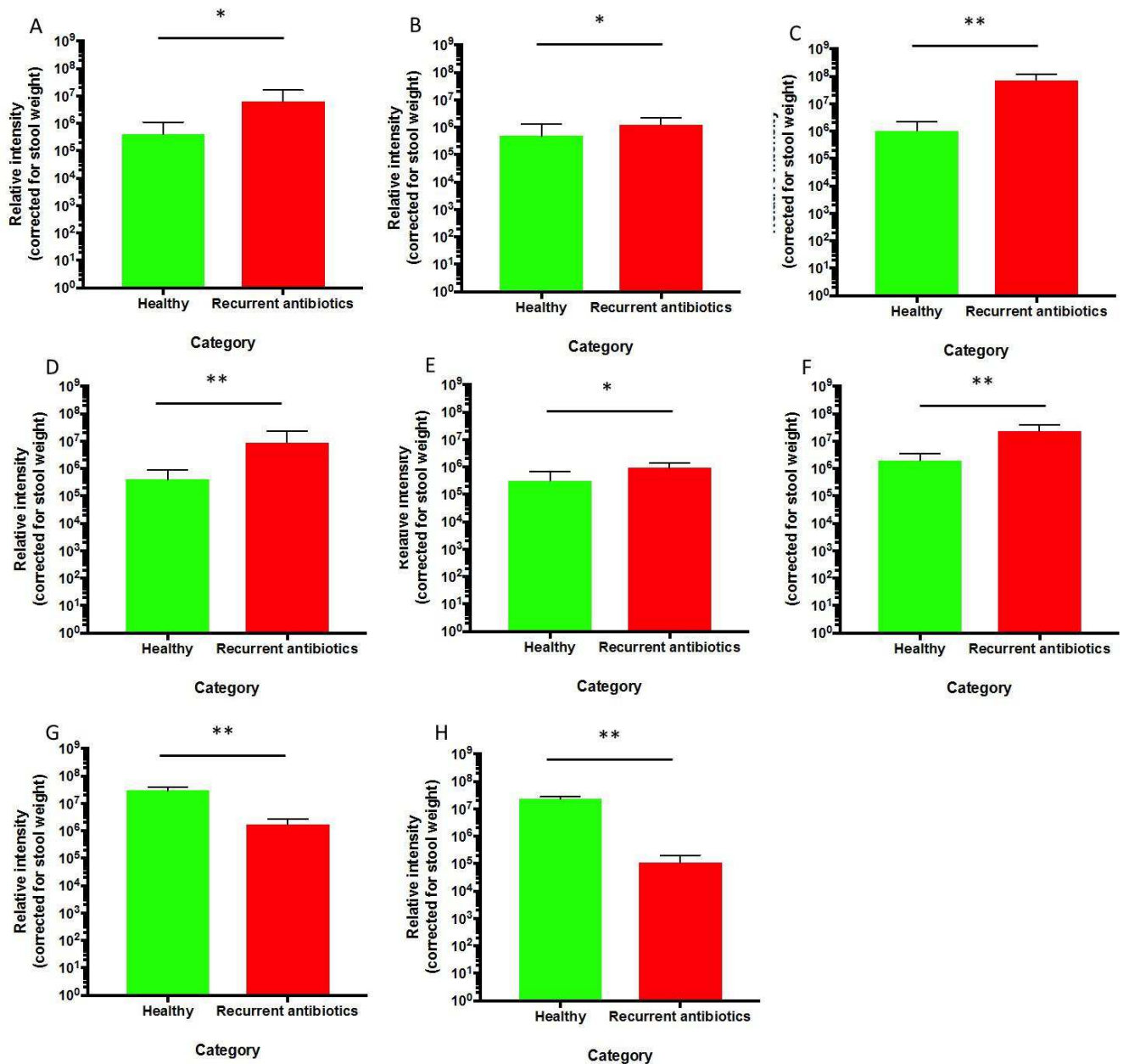
Supplementary Methods: Isolation of bacteria used as standards for *bsh* gene qPCR:

Bacteroides plebius, *Bacteroides ovatus*, and *Blautia obeum* were previously isolated from the stool of a healthy 29 year-old male donor. *Bacteroides plebius* was isolated from Fastidious Anaerobe Agar plates (Acumedia, USA) with 5% horse blood (VWR, USA), *Bacteroides ovatus* was isolated from nutrient agar plates (Sigma-Aldrich, USA), and *Blautia obeum* was isolated from de Man, Rogosa and Sharpe agar plates (Sigma-Aldrich).

DNA extraction was performed on the isolates using the EZNA Isolation Kit Bacterial DNA (Omega, USA) with the addition of a bead beating using the Bullet Blender Storm (speed 8 for 3 min). A ~900 bp region of the 16S gene was amplified using previously published primers [1] and DNA was sequenced at Macrogen Europe and isolates were identified by performing a standard nucleotide BLAST of the 16S rRNA sequences (NCBI).



Supplementary Figure 1: Bile salt hydrolase grouping map.



Supplementary Figure 2: Univariate analysis of differences in specific bile acids between healthy participants and people treated with recurrent antibiotics. (A) Taurocholic acid (* p < 0.05, Mann-Whitney U test); (B) Glycocholic acid (* p < 0.05); (C) Cholic acid (** p < 0.01); (D) Taurochenodeoxycholic acid (** p < 0.01); (E) Glycochenodeoxycholic acid (* p < 0.05); (F) Chenodeoxycholic acid (** p < 0.01); (G) Deoxycholic acid (** p < 0.01); (H) Lithocholic acid (** p < 0.01).

Patient characteristics	Value
Sex	Male (<i>n</i> = 2), female (<i>n</i> = 3)
Age	63+/- 14 years
Prior antibiotics	Penicillins (<i>n</i> =3), cephalosporins (<i>n</i> =1), fluoroquinolones (<i>n</i> =2)
Significant co-morbidities	Benign prostatic hypertrophy (<i>n</i> =1), diverticulitis (<i>n</i> =1), urinary tract infection (<i>n</i> =1), inguinal hernia repair within past three months (<i>n</i> =1), suspected lung carcinoma (<i>n</i> =1)

Supplementary Table 1: Key clinical characteristics of patients included within study.

Supplementary Material Additional References:

- [1] E. Ben-Dov, O. H. Shapiro, N. Siboni, and A. Kushmaro, "Advantage of using inosine at the 3' termini of 16S rRNA gene universal primers for the study of microbial diversity.," *Appl. Environ. Microbiol.*, vol. 72, no. 11, pp. 6902–6, Nov. 2006.