

1 Low Gut Ruminococcaceae Levels are Associated with Occurrence of
2 Antibiotic-associated Diarrhea

3 Xiaoqiong Gu^{1,2}, Jean XY Sim³, Wei Lin Lee^{1,2}, Liang Cui^{1,2}, Yvonne FZ Chan³, Ega Danu
4 Chang^{1,2}, Yii Ean Teh³, An-Ni Zhang⁴, Federica Armas^{1,2}, Franciscus Chandra^{1,2}, Hongjie
5 Chen^{1,2}, Shijie Zhao⁴, Zhanyi Lee^{1,2}, Janelle R. Thompson^{5,6}, Eng Eong Ooi^{7,8,9}, Jenny G.
6 Low^{3,7,8}, Eric J. Alm^{1,2,4,10,11,*}, Shirin Kalimuddin^{3,7,*}

7

- 8 1. Antimicrobial Resistance Interdisciplinary Research Group, Singapore-MIT Alliance
9 for Research and Technology, 1 Create Way, Singapore 138602, Singapore
- 10 2. Campus for Research Excellence and Technological Enterprise (CREATE),
11 Singapore 138602
- 12 3. Department of Infectious Diseases, Singapore General Hospital, 20 College Road,
13 Singapore 169856, Singapore
- 14 4. Department of Biological Engineering, Massachusetts Institute of Technology, 21
15 Ames Street, Cambridge, MA 02142, USA
- 16 5. Singapore Centre for Environmental Life Sciences Engineering, Nanyang
17 Technological University, 60 Nanyang Drive, Singapore 637551, Singapore
- 18 6. Asian School of the Environment, Nanyang Technological University, 62 Nanyang
19 Drive, Singapore 637459, Singapore
- 20 7. Program in Emerging Infectious Diseases, Duke-NUS Medical School, 8 College
21 Road Singapore 169857, Singapore
- 22 8. Viral Research and Experimental Medicine Center, SingHealth Duke-NUS Academic
23 Medical Centre (ViREMiCS), 20 College Road, Singapore 169856, Singapore

- 24 9. Saw Swee Hock School of Public Health, National University of Singapore, 12
25 Science Drive 2, Singapore 117549, Singapore
- 26 10. Center for Microbiome Informatics and Therapeutics, Massachusetts Institute of
27 Technology, Building E25-321, Cambridge, MA 02139, USA
- 28 11. Broad Institute of MIT and Harvard, 415 Main Street, Cambridge, MA 02142, USA

29 * Corresponding Authors

30 Correspondence to:

31 Shirin Kalimuddin

32 Department of Infectious Diseases, Singapore General Hospital

33 Academia Level 3, 20 College Road

34 Singapore 169856

35 Tel: +65 6321 3479; Email: shirin.kalimuddin@singhealth.com.sg

36 Eric J Alm

37 Department of Biological Engineering, Massachusetts Institute of Technology

38 Cambridge, Massachusetts 02139, United States

39 Tel: +1 617 253 27726; Email: ejalm@mit.edu

40 Conflict of Interest Statement: E.J.A is a co-founder and shareholder of Finch Therapeutics, a
41 company that specializes in microbiome-targeted therapeutics. All the other authors have
42 declared that no conflict of interest exists.

43

44 ABSTRACT

45 Patients receiving antibiotics often suffer from antibiotic-associated diarrhea (AAD). AAD is
46 of clinical significance as it can result in premature antibiotic discontinuation and suboptimal
47 treatment of infection. The drivers of AAD however, remain poorly understood. We sought
48 to understand if differences in the gut microbiome, both at baseline and during antibiotic
49 administration, would influence the development of AAD. We administered a 3-day course of
50 oral amoxicillin-clavulanate to 30 healthy adult volunteers, and performed a detailed
51 interrogation of their stool microbiome at baseline and up to 4-weeks post antibiotic
52 administration, using 16S rRNA gene sequencing. Lower levels of Ruminococcaceae were
53 significantly and consistently observed from baseline till Day 7 in participants who developed
54 AAD. The probability of AAD could be predicted based on qPCR-derived levels of
55 *Faecalibacterium prausnitzii*, the most dominant species within the Ruminococcaceae family.
56 Overall, participants who developed AAD experienced a greater decrease in microbial
57 diversity during antibiotic dosing. Our findings suggest that a lack of gut Ruminococcaceae at
58 baseline influences development of AAD. In addition, quantification of *F. prausnitzii* in stool
59 prior to antibiotic administration may help identify patients at risk of AAD, and aid clinicians
60 in devising individualised treatment regimens to minimise such adverse effects.

61 **KEYWORDS:** Antibiotic-associated diarrhea (AAD), amoxicillin-clavulanate, gut
62 microbiome, Ruminococcaceae, *F. prausnitzii*

63

64 INTRODUCTION

65 Antibiotic-associated diarrhea (AAD) occurs in a significant proportion of patients, and is
66 particularly associated with use of broad-spectrum antibiotics (1). AAD may sometimes be
67 severe enough to result in premature discontinuation of antibiotics, and can in turn result in
68 suboptimal treatment of infection. AAD has also been shown to prolong hospital stay,
69 increase risk of other infections and lead to higher overall healthcare costs (2). Therefore,
70 AAD is of significant clinical importance, and a better understanding of its underlying
71 mechanisms and drivers is needed in order to devise therapeutic strategies to minimise its
72 occurrence.

73 While it is well known that antibiotics disrupt and alter the diversity of microorganisms
74 within the gut (gut microbiome), it is less clear how AAD develops (3). One hypothesis is
75 that AAD results from the overgrowth of toxigenic bacteria, such as *Clostridium difficile*,
76 which are resistant to the administered antibiotic (4). However, *C. difficile* diarrhea only
77 accounts for 15-25% of all cases of AAD (5). Other possible pathogens which have been
78 implicated include *Clostridium perfringens*, *Staphylococcus aureus*, *Escherichia coli*,
79 *Pseudomonas aeruginosa*, and *Klebsiella pneumoniae* (3, 6), supported by studies in murine
80 models demonstrating overgrowth of these pathogens in mice with AAD (6, 7). Another
81 proposed mechanism of AAD is the loss of functional and beneficial gut microbes with
82 critical metabolic activities, resulting in reduced carbohydrate fermentation and short-chain
83 fatty acids (SCFAs) that are important for colonic health (8, 9). Despite these mechanistic
84 explanations for AAD, many gaps in knowledge still remain. Firstly, the etiopathogenesis of
85 non-*C. difficile* AAD is poorly defined. Secondly, although much work has been done in
86 animal models to study the link between gut microbiota alteration and AAD (6, 7), studies in
87 humans are lacking. This is of consequence as there are significant differences between the
88 gut microbiome of humans and animals. Thirdly, while post-antibiotic treatment changes

89 could explain AAD through increases of pathogenic bacteria or decreases in beneficial gut
90 microbes, it remains unknown whether the baseline microbiome composition prior to
91 antibiotic administration may confer predisposition to AAD.

92 We hypothesized that baseline differences in the gut microbiome prior to antibiotic
93 administration could account for why certain patients developed AAD, and that these
94 baseline differences would in turn modulate changes in the gut microbiome composition post
95 antibiotic-administration. To test our hypothesis, we conducted an experimental medicine
96 study in 30 healthy adult volunteers. We administered a 3-day course of oral amoxicillin-
97 clavulanate to study participants. Amoxicillin-clavulanate, a broad-spectrum antibiotic with
98 activity against Gram-positive and Gram-negative organisms, including anaerobes, was
99 chosen as it is one of the most widely-prescribed antibiotics (10), and is associated with a
100 high incidence of AAD (11, 12). We monitored individuals for occurrence of AAD while on
101 amoxicillin-clavulanate. Using 16S rRNA gene sequencing, we tracked dynamic changes in
102 the composition and diversity of the gut microbiome at baseline and up to four-weeks post
103 antibiotic administration, in order to identify differences between individuals who developed
104 AAD (AAD group) and those who did not (non-AAD group).

105 RESULTS

106 **Gut Ruminococcaceae levels differentiate AAD and non-AAD groups both at baseline** 107 **and post-antibiotic treatment**

108 30 healthy adult volunteers were orally administered 1g of amoxicillin-clavulanate twice
109 daily for 3 days, a dose commonly used in clinical practice. Individuals enrolled had a mean
110 age of 30.3 ± 6.2 years and a mean body mass index of 24.8 ± 3.4 kg m⁻², with an equal male
111 to female ratio (1:1) (Table S1). Faecal samples were collected prior to (day 0, baseline),
112 during (days 1, 2, 3) and after antibiotic treatment (days 7, 14 and 28) (Figure 1A). Nucleic

113 acids were extracted from faecal samples and subjected to 16S rRNA gene sequencing. Using
114 droplet digital PCR, we confirmed that *C. difficile* toxic TcdA and TcdB genes were not
115 detected in any of the faecal samples at baseline and during-antibiotic treatment (days 1-3) ,
116 ruling out *C. difficile* colitis as a cause of diarrhea (13). For this study, we *a priori* defined
117 AAD as at least 1 episode of Bristol Stool Scale type 6 or 7 on either days 1, 2 or 3 (Figure
118 1B). Based on this definition, there were 13 individuals who developed AAD (AAD group)
119 and 17 individuals who did not (non-AAD group) (Table S1-2). There were no differences in
120 demographics between the two groups (Table S1). In order to ensure the safety and well-
121 being of the study volunteers, the study protocol mandated that amoxicillin-clavulanate
122 would be discontinued early if an individual experienced 3 or more episodes of watery stool
123 in a 24-hour period - this occurred in 4 of 13 individuals in the AAD group (Table S1, S3).
124 One individual in the non-AAD group developed severe vomiting after a single dose of
125 amoxicillin-clavulanate, resulting in antibiotic discontinuation on day 1. In all 5 individuals
126 however, faecal samples were collected and sequenced as per protocol (Table S1).

127 We first examined whether the composition of the gut microbiome at baseline would
128 influence development of AAD. Aggregation of microbial sequence types at the taxonomic
129 levels of "family" and "genus" revealed notable dynamics in the AAD compared to non-AAD
130 groups (Figure 2A, Figure S1A). We observed that Enterobacteriaceae blooms (days 1-3)
131 were more common in the AAD group (76.9% vs 29.4% occurrence frequency) (Figure S1B),
132 with a much higher magnitude (mean 59.1% vs 21.0%, median 67.9% [IQR 33.7-78.7] vs
133 20.0% [IQR 17.4-25.2], Bonferroni-corrected $p < 0.0001$) (Figure S1C). Within the
134 Enterobacteriaceae family, these were assigned to the genus *Escherichia-Shigella* (Figure
135 S1D).

136 Among the major taxonomies, we looked for features present at baseline that could
137 differentiate between the AAD and non-AAD groups. Significantly, we found that

138 Ruminococcaceae levels were distinctly different between the two groups, both at baseline
139 and post-antibiotic treatment. At baseline, the AAD group had a lower proportion of
140 Ruminococcaceae (mean 8.4% vs 14.4%, median 7.9% [IQR 4.2-11.9] vs 14.2% [IQR 11.6-
141 17.7], Bonferroni-corrected $p=0.02$, $n=30$) (Figure 2B). The AAD group also experienced a
142 consistently lower proportion of Ruminococcaceae compared to the non-AAD group till day
143 7. On average, across the duration of the study, the AAD group had a lower proportion of
144 Ruminococcaceae (mean 7.5% vs 15.3%, median 5.5 [IQR 3.0-10.8] vs 15.6 [IQR 11.8-19.2],
145 Bonferroni-corrected $p=2.1e-14$, $n=197$). These differences were not observed with any of the
146 other major taxonomies.

147 We identified potential drivers of AAD at the genus level within the Ruminococcaceae
148 family across the duration of the study, with *Faecalibacterium* being the most abundant
149 (mean 67.1%, median 66.6 [IQR 56.8-77.7]), followed by *Subdoligranulum* (mean 11.7%,
150 median 10.4 [IQR 6.5-14.5]) and *Ruminococcus* (7.7%, median 6.5 [IQR 3.1-10.1]). We
151 observed that the genera *Faecalibacterium*, *Subdoligranulum* and *Ruminococcus* were
152 significantly less abundant in the AAD group across most days between day 0 to day 7.

153 **Amoxicillin-clavulanate causes greater gut microbiome diversity loss and community** 154 **disturbance in the AAD group compared to the non-AAD group**

155 We next determined the extent to which the gut microbiota was disrupted during antibiotic
156 treatment and the timescale of recovery, in terms of the abundance of microbial amplicon
157 sequence variants (ASVs) and composition. By day 3, the faecal microbiota from the AAD
158 group was distinctly different from microbiomes sampled at the other timepoints and from
159 the non-AAD group, forming a separate cluster along the first and second axes of a principal-
160 coordinate analysis (PCoA) (Figure 3A). We quantified the microbial diversity within each
161 individual at a given time point (α diversity) and the differences between each individual's

162 baseline and post-treatment gut microbiota (β diversity) (Figure 3B). We observed a greater
163 decrease in diversity in the AAD group, compared to the non-AAD group on days 2 and/or 3
164 (Figure 3B, $P < 0.05$, Bonferroni-corrected Mann-Whitney U test). Analysis of the relative
165 abundance of bacterial taxonomic groups at the phylum level supported our finding that the
166 AAD group was more severely impacted than the non-AAD group (Figure 3C). This
167 difference in diversity between the two groups was driven by a sharp increase of
168 Proteobacteria, and a decrease of Firmicutes and Actinobacteria in the AAD group, as
169 compared to the non-AAD group on days 2 and 3 (Figure 3C). Individuals in both groups
170 returned to their baseline taxonomy and diversity by day 7, as shown by permutational
171 multivariate analysis of variance (PERMANOVA) (Figure 3A).

172 **Predicting the risk of AAD based on the relative abundance of Ruminococcaceae at** 173 **baseline**

174 While post-treatment changes in the gut microbiome could explain AAD, predicting which
175 patients are at higher risk of developing AAD would be clinically useful, to enable
176 personalization of antibiotic prescription to minimize the incidence of AAD. Through
177 hierarchical clustering of microbiome composition at baseline, the majority of non-AAD
178 individuals (Figure 4A, blue labels) were grouped into a single cluster (cluster 1), while AAD
179 individuals (Figure 4A, red labels) were separated into various clusters (non-cluster 1). This
180 clustering suggests that some features could be a potential indicator to identify individuals at
181 risk of AAD. Moreover, it also suggests the presence of common characteristics amongst the
182 non-AAD group, but not for the AAD group. Our findings suggest that individuals who went
183 on to develop AAD could be differentiated from those who did not, by the relative abundance
184 of Ruminococcaceae at baseline. This was evident from PCoA based on Bray-Curtis
185 dissimilarities (Figure S2). Hence, we sought to determine if we could predict the risk of
186 AAD based on the relative abundance of Ruminococcaceae at the baseline (D0). Upon

187 ranking the subjects based on their relative-abundance of Ruminococcaceae, we found that
188 there was a clear separation between the AAD and non-AAD groups at the extremes of
189 relative abundance (<0.05 or >0.16) (Figure 4B). Next we quantified absolute values of one
190 species under Ruminococcaceae (*Faecalibacterium prausnitzii* [*F. prausnitzii*], the most
191 abundant Ruminococcaceae species, using a qPCR assay (14). We found that the gene copies
192 of *F. prausnitzii* followed a similar trend with 16S rRNA gene relative abundance of
193 Ruminococcaceae (Spearman's $\rho = 0.85$, $p = 3.0e-9$) (Figure 4C-D). We calculated the risk of
194 developing AAD based on the absolute abundance of *F. prausnitzii* derived using qPCR at
195 baseline. Lower relative abundance of *F. prausnitzii* at baseline were predictive of risk of
196 AAD. The probability of developing AAD was above 0.7 if *F. prausnitzii* levels were less
197 than 2.4×10^7 GC/ μ L, and below 0.3 if *F. prausnitzii* levels were above 8.0×10^7 GC/ μ L
198 (Figure 4E).

199 DISCUSSION

200 Our study has shed important insights on differences in gut microbiome responses between
201 individuals who developed AAD and those who did not. We found that individuals who
202 developed AAD experienced greater gut microbiome community changes, accompanied by
203 lower diversity and a greater disturbance in abundance across taxonomies. Individuals in the
204 AAD group experienced a sharp increase in Proteobacteria, belonging to the genus
205 Escherichia-Shigella. Looking for taxonomic signatures that could differentiate between the
206 two groups, we found that amongst all bacterial families, gut Ruminococcaceae levels were
207 significantly and consistently different between the two groups. Ruminococcaceae levels
208 were lower both at baseline prior to antibiotic treatment, and up till day 7 post-dose. In fact,
209 ranking of gut Ruminococcaceae levels, or simply *F. prausnitzii* (the most dominant species

210 within the Ruminococcaceae family), prior to antibiotic treatment was able to indicate if an
211 individual would develop AAD upon treatment with amoxicillin-clavulanate.

212 Ruminococcaceae is a group of strictly anaerobic bacteria that is present in the colonic
213 mucosal biofilm of healthy individuals (15). Decreased abundance of Ruminococcaceae has
214 been implicated in a number of inflammatory bowel diseases, including ulcerative colitis and
215 Crohn's disease (16–18), inflammatory diseases such as hepatic encephalopathy (19) and has
216 also been associated with *C. difficile* infection and *C. difficile*-negative nosocomial diarrhea
217 (20). Ruminococcaceae plays an important role in the maintenance of gut health through its
218 ability to produce butyrate and other SCFAs. These SCFAs are essential carbon and energy
219 sources to colonic enterocytes (21), in the absence of which, functional disorders of the
220 colonic mucosa may occur – which may manifest in the form of osmotic diarrhea (9). Indeed,
221 supplementation of butyrate and other SCFAs has been shown to reduce colonic
222 inflammation and improve diarrhea in conditions such as inflammatory bowel diseases,
223 irritable bowel syndrome, and diverticulitis (22–24). We thus posit that a lack of
224 Ruminococcaceae resulting in decreased SCFA production may be driving the development
225 of AAD in our cohort of otherwise healthy individuals who received amoxicillin-clavulanate.

226 To date, many others have studied the role of probiotic bacteria such as *Lactobacillus*,
227 *Bifidobacterium*, *Clostridium*, *Bacillus* and *Lactococcus* in the development and prevention
228 of AAD (25). In a systematic review and a meta-analysis, administration of such probiotics
229 were associated with a reduced incidence of AAD (26, 27). However, a clinical trial
230 performed on a cohort of 3000 older patients concluded that *Lactobacillus* and
231 *Bifidobacterium* probiotic administration was not found to be effective in preventing *C.*
232 *difficile* - associated diarrhea (1). Likewise, *Lactobacillus reuteri* was not effective in
233 preventing AAD in a cohort of 250 children (28, 29). There remains a need to search for
234 probiotic candidates that will be effective against AAD. Our findings suggest that certain

235 species within the Ruminococcaceae family may be useful as probiotics to prevent AAD, but
236 this will need to be further evaluated in randomised-controlled clinical trials.

237 We have identified clear differences in baseline gut microbial composition that may enable
238 pre-identification of individuals at higher risk of developing AAD. Although our findings at
239 present are only applicable in the context of AAD caused by amoxicillin-clavulanate, our
240 study provides a framework to identify potential drivers of AAD caused by other classes of
241 antibiotics. We acknowledge that a limitation of our study is the relatively small sample size,
242 which may have reduced the statistical power given the variability of inter-individual
243 differences in the gut microbiome. Despite this, we were still able to observe clear and
244 significant differences between the AAD and non-AAD groups.

245 Our findings provide evidence for the first time that baseline differences in the individual's
246 gut microbial composition can influence the risk of developing AAD with certain antibiotics,
247 and would guide the development of point-of-care diagnostics. Being able to pre-identify
248 individuals at increased risk of AAD would aid clinicians in devising an individualised
249 antibiotic regime best suited to the patient that is least likely to result in premature antibiotic
250 discontinuation and suboptimal treatment of infection - An example of such a clinical
251 workflow is presented in Figure 5. In addition, the use of Ruminococcaceae as a prebiotic to
252 prevent AAD in patients who receive amoxicillin-clavulanate also warrants further
253 exploration. Overall, our study provides insights into how the gut microbiome influences
254 development of AAD, and opens a window of opportunity for further research in this area.

255

256

257 METHODS

258 Study design and participants

259 From Aug 2019 to Jan 2020, 30 healthy adult volunteers who fulfilled the pre-determined
260 inclusion and exclusion criteria were enrolled into the study. Individuals were eligible for the
261 study if they were: 1) Aged between 21-40 years, 2) Willing and able to provide written
262 informed consent, and 3) Agreeable to abstain from probiotics and/or prebiotics during the
263 study period. Individuals were excluded if they met any of the following criteria: 1) Presence
264 of underlying chronic medical illness, 2) History of *C. difficile* diarrhea, 3) Inflammatory
265 bowel disease or any other chronic gastrointestinal tract illness, 4) Allergy to beta-lactam
266 antibiotics, 5) Acute infection in the preceding 7 days, 6) Were pregnant or breastfeeding,
267 and/or 7) Receipt of antibiotics in the past 3 months.

268 Upon enrolment, individuals received oral amoxicillin-clavulanate at a dose of 1g (875 mg of
269 amoxicillin trihydrate and 125 mg of potassium clavulanate) twice a day for 3 days, i.e. a
270 total of 6 doses. Study drug compliance was assessed via daily phone calls and pill count at
271 each study visit. Individuals were instructed to maintain consistent dietary habits and abstain
272 from pre- or probiotics throughout the duration of the study.

273 Metadata and sample collection

274 Baseline demographics were recorded. Individuals were provided with a standardised diary to
275 record frequency of bowel opening and any adverse events. Individuals were followed-up for
276 28 days (screening, baseline [prior to antibiotic administration], and days 1, 2, 3, 7, 14 and
277 28). At each study visit, frequency of bowel opening in the past 24 hours was recorded and a
278 faecal sample passed on the day of the study visit was collected. Faecal samples were
279 collected prior to (day 0, baseline), during (days 1, 2, 3) and after antibiotic treatment (days 7,

280 14 and 28), over a time span of 4 weeks. Faecal samples were collected using a disposable
281 commode 2-piece specimen collector (MEDLINE, USA) and stored immediately in -20°C
282 freezers prior to being transported to the laboratory which was off-site. The Bristol Stool
283 Scale of each faecal sample was assessed by the same study team member each time to
284 ensure consistency.

285 **DNA extraction, library construction and Illumina 16S rRNA Sequencing**

286 DNA was extracted from approximately 300 - 600 mg of faecal samples using DNeasy®
287 PowerSoil® Pro Kit (Qiagen, Germany, Cat# 47016) following the manufacturer's
288 instructions. DNA purity and quantity of extracted DNA was determined by Nanodrop®
289 Spectrophotometer ND-1000 (Thermo Fisher Scientific, USA) before sending to the
290 NovogeneAIT Genomics Singapore for sequencing. Bacterial 16S V4 region was amplified
291 with the Earth Microbiome Project recommended primer pairs 515F
292 (GTGCCAGCMGCCGCGGTAA) and 806R (GGACTACHVGGGTWTCTAAT) (30) using
293 Phusion® High-Fidelity PCR Master Mix (New England Biolabs, Cat# M0531L). The size of
294 the amplicon was checked using 1 % agarose gel electrophoresis. The amplicon with correct
295 size was purified from agarose gel using Qiagen Gel Extraction Kit (Qiagen, Germany) and
296 proceeded to library preparation using NEBNext Ultra DNA Library Prep Kit for Illumina®
297 (New England Biolabs, Cat# E7370L) following manufacturer's instructions. Unique indexes
298 were added to each sample. The library was quantified using qPCR. The sequencing libraries
299 were normalized and pooled at equimolar concentration before performed on NovaSeq-6000
300 (Illumina, USA) to generate 250 bp paired-end raw reads.

301 **16S rRNA gene sequencing datasets pre-processing analysis**

302 Paired-end raw reads were assigned to a sample by their unique barcode, and the barcode and
303 primer sequence were then truncated. Paired-end reads were merged using FLASH (V1.2.7)

304 (31) to merge pairs of reads when the original DNA fragments are shorter than twice of the
305 reads length. The obtained splicing sequences were called raw tags. Quality filtering were
306 then performed on the raw tags under specific filtering conditions of QIIME (V1.7.0) (32)
307 quality control process. After filtering, high-quality clean tags were obtained. The tags were
308 compared with the reference database (Gold database,
309 http://drive5.com/uchime/uchime_download.html) using UCHIME algorithm (33) to detect
310 chimeric sequences, and then the chimeric sequences were removed to obtain the Effective
311 Tags finally. Sequence analysis and processing of paired-end demultiplexed sequences were
312 performed in Quantitative Insights into Microbial Ecology pipeline (QIIME 2, v 2020.6) (34).
313 Demultiplexed sequences were imported using QIIME 2 “Fastq manifest” format by mapping
314 the sample identifiers to absolute file paths containing sequence information for each sample.
315 PairedEndFastqManifestPhred33V2 format was used. Interactive quality plot and a summary
316 distribution of sequence qualities at each base pair position in the sequence data was
317 visualized to determine input parameters for denoising. DADA2 was used to denoise,
318 dereplicate and filter chimaeras in paired-end sequences to identify all amplicon sequence
319 variants (ASVs), equivalent to 100% Operational Taxonomy Unit (OTUs) (35, 36). Forward
320 and reverse reads were truncated at 200 bases to retain high quality bases respectively. In
321 total, we characterized an average of $150,087 \pm 14,526$ (mean \pm SD) 16S rRNA sequences for
322 197 samples.

323 **Microbiome composition and diversity**

324 The alpha diversity metrics (Shannon entropy) and beta diversity metrics (Jensen-Shannon
325 Distance) from ASVs were generated via q2-diversity plugin. A sampling depth of 96,935
326 sequences per sample was used and optimal alpha-rarefaction curves were achieved.
327 Taxonomy assignment to ASVs was performed using q2-feature-classifier plugin using a pre-
328 trained Naive Bayes classifier against the reference 515F/806R region of sequences in Silva

329 138 at 99% OTUs (37). We computed ASV pairwise distances using the Pearson correlation
330 (ASV abundances across 30 subjects at baseline). The resulting distance matrix was
331 subsequently inputted into a hierarchical clustering function ('fcluster'). The linkage
332 approach was set as 'average'. The colour threshold was set to '0.4'. Principal coordinate
333 analysis (PCoA) was performed using the 'scipy' package in python based on the ASV-level
334 Bray-Curtis dissimilarities between the composition of baseline samples. PERMANOVA
335 analysis was calculated using the 'skbio' package in python based on the ASV-level Bray-
336 Curtis dissimilarities and 9999 permutations.

337 **Calculating predictive probability of developing AAD from baseline abundance**

338 To calculate the predictive probability of developing AAD from *F. prausnitzii* baseline
339 abundance, the absolute abundance of *F. prausnitzii* was min-max normalized to their
340 transformed value between 0 and 1. The predictive probability of developing AAD was
341 calculated using kernel density estimation with a Gaussian distribution kernel. We assumed
342 that each sample had an independent and identical distribution with a mean at its
343 concentration and a standard deviation, which is the hyperparameter of this model. We
344 aggregated the distribution of every sample from the AAD group and obtained the AAD
345 probability density function (PDF). The same approach was applied to calculating the non-
346 AAD PDF. The probability of developing AAD was calculated as $\frac{P_{AAD}}{P_{AAD} + P_{non-AAD}}$.

347 **Molecular ddPCR assay for *C. difficile* toxin genes TcdA and TcdB quantification**

348 We excluded *C. difficile* infection in all individuals via PCR of toxin genes TcdA and TcdB
349 in stool at baseline (day 0) and during-antibiotic treatment (days 1-3). Briefly, droplet digital
350 PCR was performed on both 10x and 100x diluted DNA extracted from all baseline and
351 during-antibiotic treatment (days 1-3) samples on Bio-Rad droplet digital PCR system (Bio-
352 Rad, California). Reaction mixtures of 22 μ L were prepared with 11 μ L of ddPCR supermix

353 for probes (no dUTP) (Bio-Rad, Cat# 1863024), 0.2 μ M of forward, reverse primers each,
354 0.2 μ M of fluorescent probes and 2.2 μ L of template DNA (13). After that, 20 μ L of the
355 reaction mixtures were transferred into the cassette with 70 μ L of droplet generator oil for
356 probes (Bio-Rad, Cat# 1863005) for droplet generation. 10,000 - 20,000 droplets were
357 generated from 20 μ L of each reaction mix with the QX200 droplet generator (Bio-Rad).
358 Droplet-partitioned samples were transferred to a Twin-tec PCR 96-well plate (Eppendorf),
359 sealed and amplified in the thermal cycler under the Bio-Rad recommended thermal cycling
360 protocol (95°C for 10 min, followed by 40 cycles of 94°C for 30s and 60°C for 1 min, ending
361 with 98°C for 1 min with a ramp rate of 2°C/s). The amplified samples were immediately
362 transferred to the QX200 reader (Bio-Rad) and read in the FAM channel. Analysis of the
363 ddPCR data was performed using QuantaSoft software (Bio-Rad). gBlocks Gene Fragments
364 (Integrated DNA Technologies, Iowa) were designed with DNA sequences corresponding to
365 the amplification regions of the primer-probe sets as positive control (Table S5).

366 **Molecular qPCR assay for *Faecalibacterium prausnitzii* and 16S rRNA gene** 367 **characterization**

368 To determine the absolute concentration of *F. prausnitzii* and 16S rRNA gene, qPCR was
369 performed on 100x diluted DNA extracted from all baseline samples (14, 38). Reaction
370 mixtures of 10 μ L of extracted stool DNA were prepared in triplicates with 5 μ L of qPCR
371 SsoAdvanced Universal SYBR Green Supermix (Biorad, Cat# 1725271), 0.25 μ M of
372 forward, reverse primers each and 1 μ L of template DNA. The reactions are set up using
373 electronic pipettes (Eppendorf), sealed and amplified in the Bio-Rad CFX384 real-time PCR
374 thermal cycler under the Bio-Rad recommended thermal cycling protocol (98°C for 3 min for
375 polymerase activation and DNA denaturation, followed by 40 cycles of 95°C for 15 s and
376 60°C for 30 s). All baseline samples for both assays are set up in one 384 plate to avoid inter-
377 plate variation. No template controls were included for both assays. Melt-curve analysis was

378 performed from 65 °C - 95 °C in 0.5 °C increments at 5 sec/step. gBlocks Gene Fragments
379 (Integrated DNA Technologies, Iowa) were designed with DNA sequences corresponding to
380 the amplification regions of the primer-probe sets (Table S5). 10-fold serial dilutions were
381 performed on the gBlocks for a range of 1 to 1×10^7 copies/ μ L, representing the range of the
382 standard curve for quantification. gBlocks Gene Fragments containing *F. prausnitzii* and 16S
383 rRNA gene were quantified and calibrated using ddPCR to determine the absolute gene
384 copies.

385 *F. prausnitzii* gene copies per PCR reaction were normalized to the 16S rRNA gene. The
386 standard curve for *F. prausnitzii* was $y=37.342-3.674x$ (efficiency = 85.1%) and for 16S
387 rRNA gene was $y=38.230-3.693x$ (efficiency = 85.4%). To adjust the *F. prausnitzii* for each
388 sample due to nucleic acids extracted from uneven biomass of stool samples, we first
389 calculated the deviation of the 16S rRNA gene from the median of the 16S rRNA gene in all
390 baseline samples, i.e., deviation factor = $\frac{\text{Sample concentration}}{\text{Median concentration}}$. We then divided the *F.*
391 *prausnitzii* concentrations by this deviation. The final absolute concentration of *F. prausnitzii*
392 (GC/ μ L per PCR reaction) = The adjusted absolute concentration of *F. prausnitzii* (GC/ μ L
393 per PCR reaction) * 100 (the dilution factor).

394 **Quantification and Calibration of gBlocks Gene Fragments**

395 gBlocks Gene Fragments containing *F. prausnitzii* and 16S rRNA gene were quantified and
396 calibrated using ddPCR to determine the absolute gene copies. Reaction mixture of 22 μ L of
397 gBlock Gene Fragments were prepared in duplicates with 11 μ L of ddPCR EvaGreen
398 Supermix (Biorad, Cat# 1864034), 0.2 μ M of forward, reverse primers each and 2.2 μ L of
399 template standard DNA. The reactions are set up in the Bio-Rad CFX96 real-time PCR
400 thermal cycler under the Bio-Rad recommended thermal cycling protocol (95°C for 5 min for
401 polymerase activation, followed by 40 cycles of 95°C for 30 s for DNA denaturation and

402 60°C for 1 min for annealing/extension, 4°C for 5 min and 90 °C for 5 min for signal
403 stabilization). During the process, the ramp rate was set as 2°C/s.

404

405

406 **Data Availability**

407 The sequencing datasets generated during this study are available at European Nucleotide
408 Archive (ENA): PRJEB46061. Deposited data includes FASTQ files for the 197 16S rRNA
409 amplicon sequences, with adaptors removed and filtered for good quality. The code to
410 reproduce all of the analysis and figures in this paper is available at
411 https://github.com/XiaoqiongGu/augmentin_16S/Final_Figure.

412 **Statistical analysis**

413 Statistical significance of bacterial diversity and bacterial composition aggregated at ‘family’
414 and ‘genus’ level between the AAD and non-AAD groups was calculated using two-sided
415 Mann-Whitney U-test. Multiple hypothesis testing in each dataset with Bonferroni correction
416 was used with Mann-whitney test and ‘two-sided’ method in ‘statannot’ package. All
417 statistical analysis was performed using Python v3.6. The statistical tests used, the definition
418 of statistical significance and multiple tests corrections were indicated in the figure legends.

419 **Study Approval**

420 This study was approved by the SingHealth Centralised Institutional Review Board (Ref:
421 2019/2377). Written informed consent was obtained from all participants prior to inclusion in
422 the study.

423 **AUTHOR CONTRIBUTIONS**

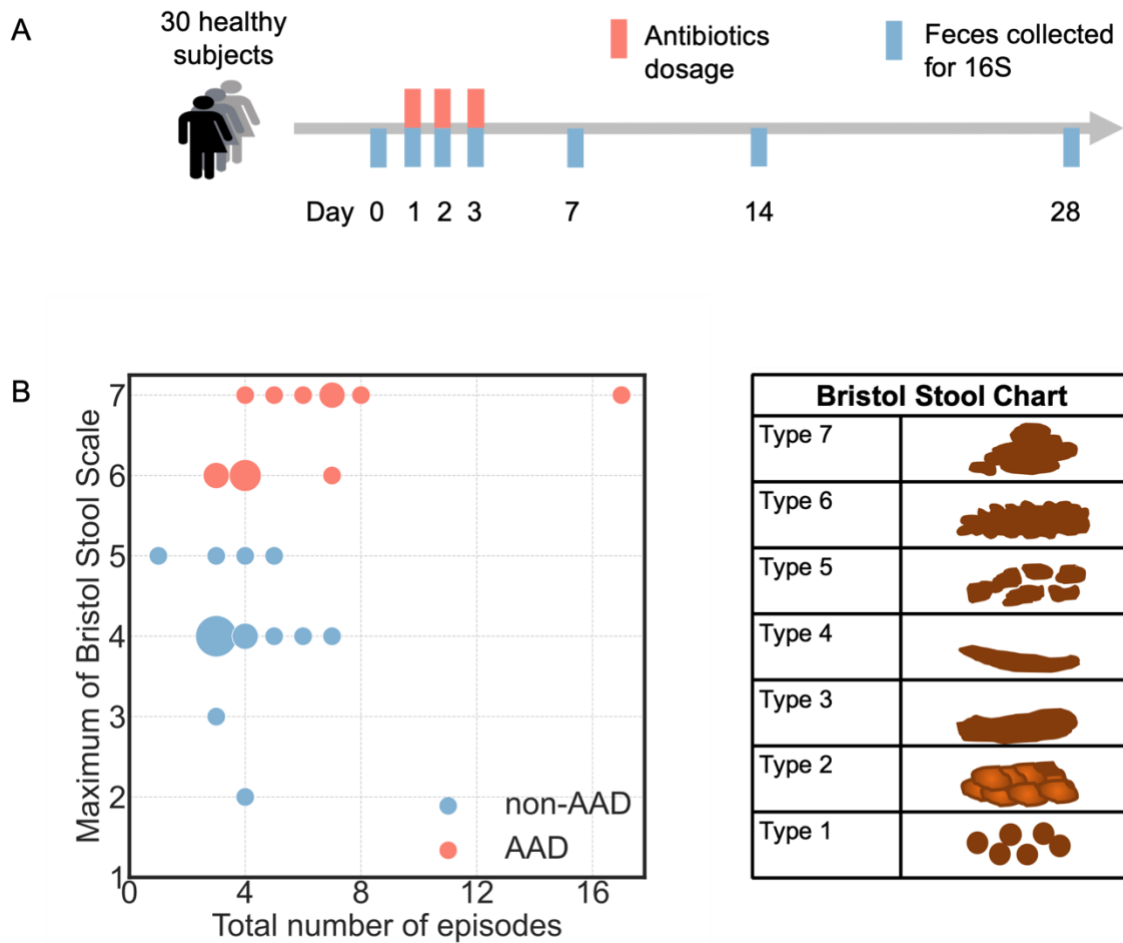
424 Clinical portion of the study and subject enrolment, S.K., J.X.Y.S., Y.E.T. and Y.F.Z.C.;
425 Conceptualization and study design, S.K., E.J.A., E.E.O. and J.G.L.; Stool sample pre-
426 processing, X.G., E.D.C. and L.C.; Molecular qPCR/ddPCR analysis, X.G. and Z.L.;
427 Sequencing data analysis, X.G., E.J.A., W.L.L., F.A., F.C., A.N.Z., H.C. and E.D.C.; Writing
428 - Original Draft, X.G., S.K., W.L.L, E.J.A., F.A. and F.C.; Writing - Review & Editing, X.G.,
429 S.K., W.L.L, E.J.A., J.S.X.Y., L.C., F.A., F.C., Y.F.Z.C., A.N.Z., Y.E.T., H.C., S.Z., J.R.T.,
430 E.E.O., and J.G.L. All authors agreed to submit the manuscript, read and approved the final
431 draft and take full responsibility of its content, including the accuracy of the data.

432 ACKNOWLEDGMENTS

433 We thank the Singapore General Hospital Clinical Trials Research Centre and all study
434 volunteers. We thank the SMART AMR core team (Tse Mien Tan, Peiyong Ho, Megan
435 Mcbee, and Farzad Olfat) for their support with logistics and lab supplies; Hafiz Ismail
436 (SMART) for his assistance in quantifying *C. difficile* in the baseline faecal samples;
437 Mathilde Poyet (MIT) and Charmaine Ng (NUS) for their guidance in the stool samples
438 collection and nucleic acids extraction. This study was funded by a SingHealth Academic
439 Medicine Research Grant (AM-CT003-2018) and the National Research Foundation,
440 Singapore, under its Campus for Research Excellence and Technological Enterprise
441 (CREATE) program funding to the Singapore-MIT Alliance for Research and Technology
442 (SMART) Antimicrobial Resistance Interdisciplinary Research Group (AMR IRG).

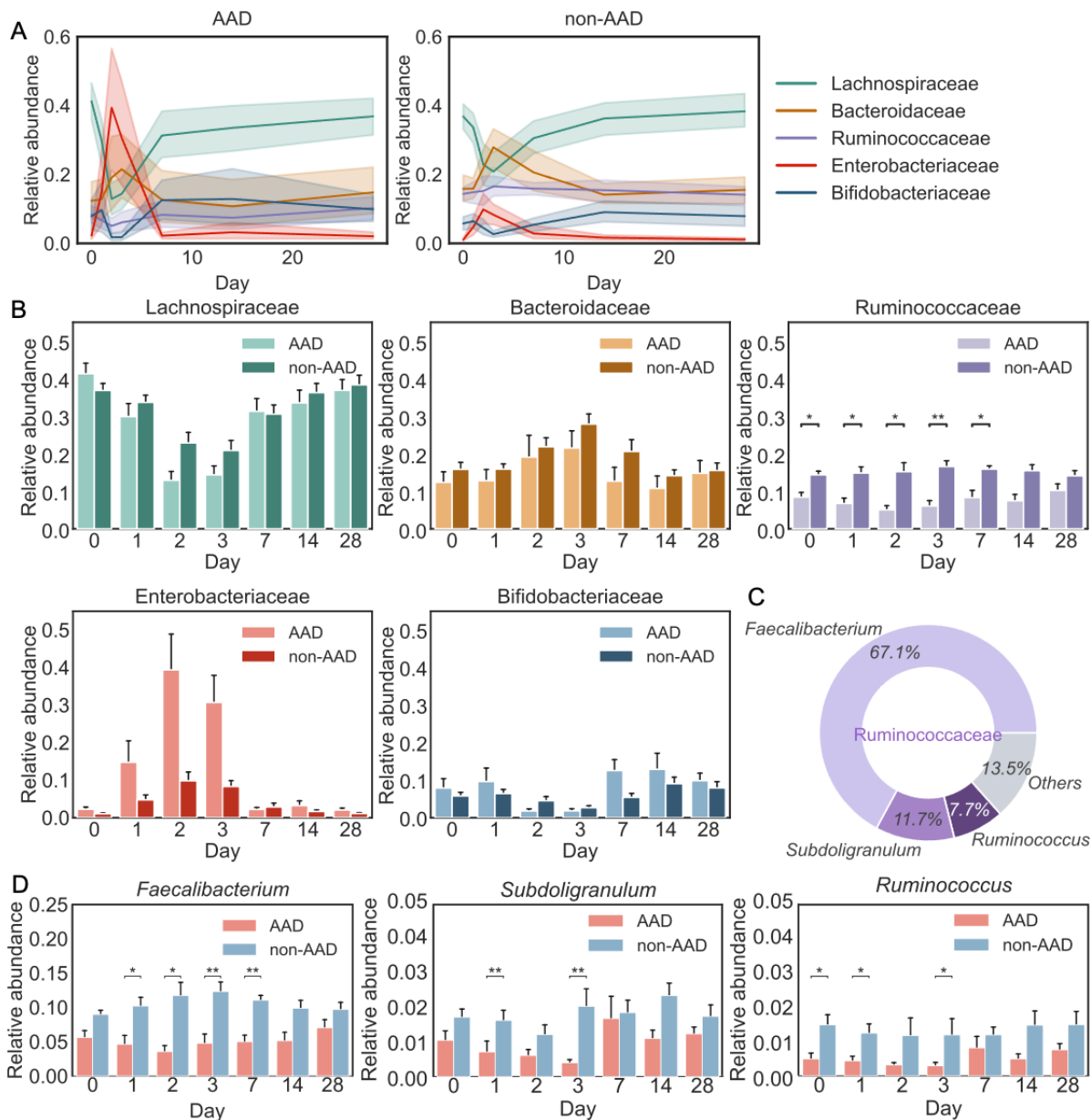
443

444 FIGURES

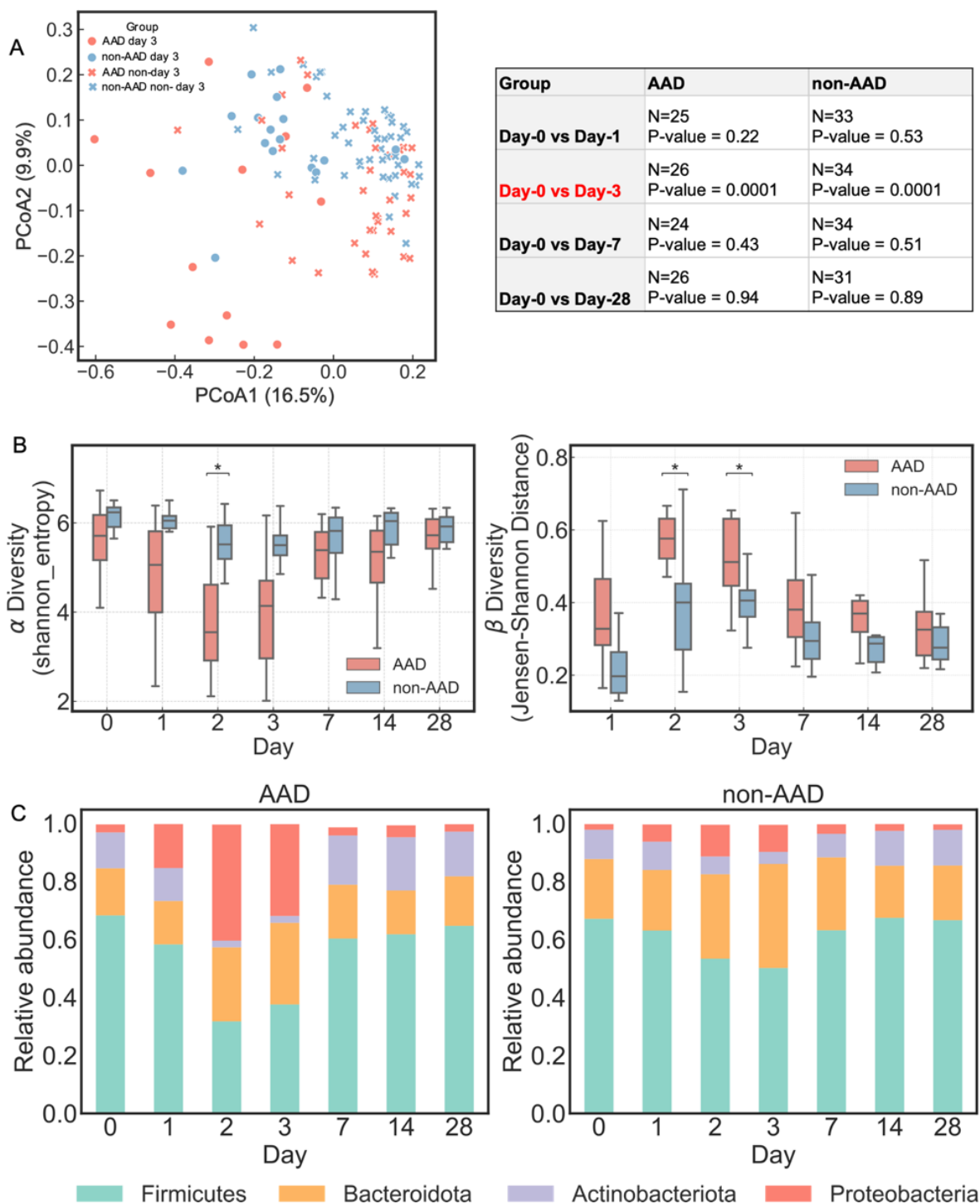


445

446 **Figure 1. AAD vs non-AAD classification across the duration of the study.** (A) Study
 447 design of 30 healthy adult volunteers administered amoxicillin-clavulanate with paired faecal
 448 sampling. (B) The segregation of AAD and non-AAD groups according to the maximum
 449 Bristol Stool Scale on either day 1, 2 or 3 of antibiotic treatment. The total number of
 450 episodes indicates the number of episodes across day 1 to 3.

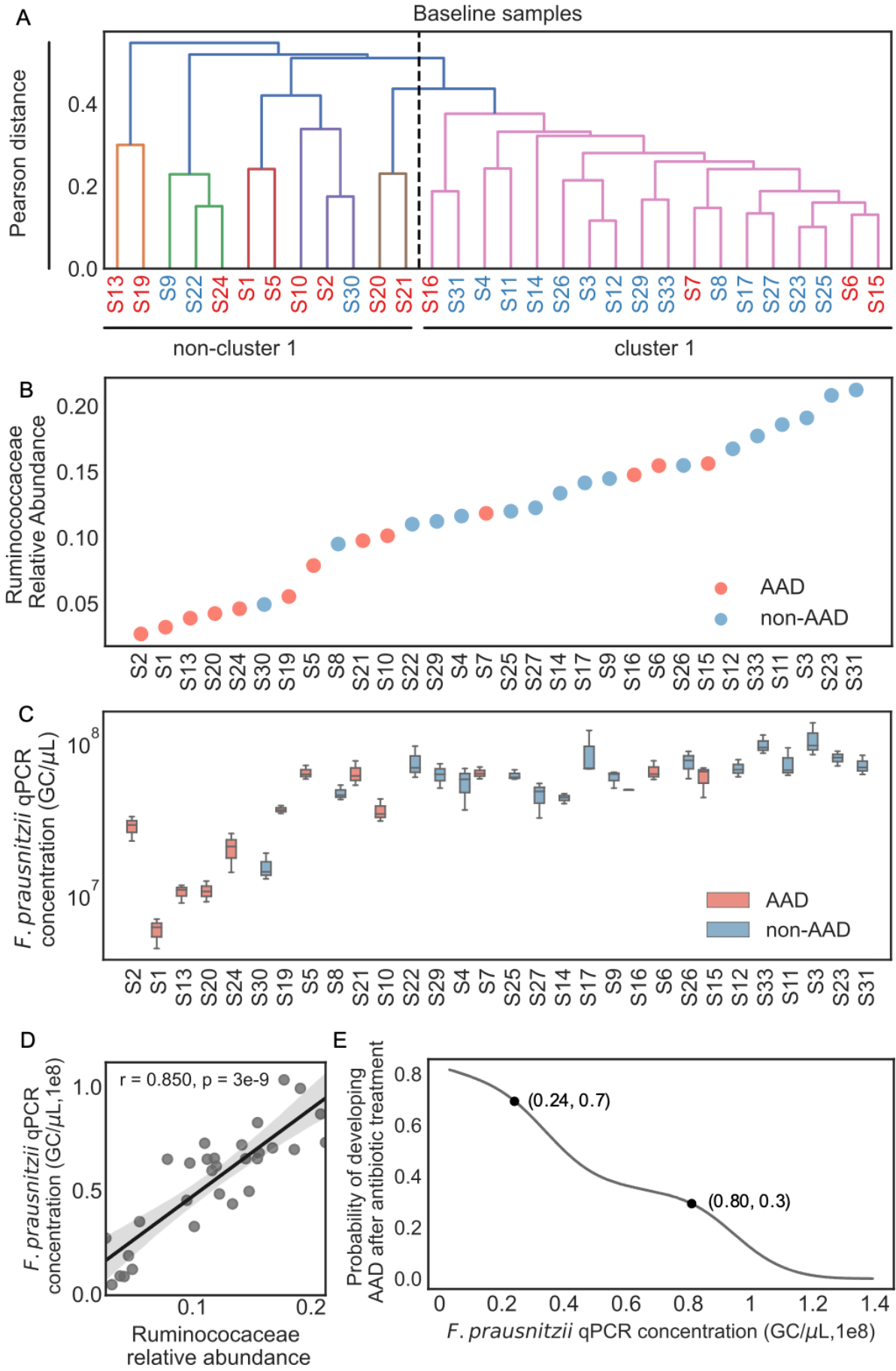


451
 452 **Figure 2. Ruminococcaceae differentiates the AAD and non-AAD groups across the**
 453 **duration of the study.** (A) Profound community changes were found in the AAD group
 454 compared to the non-AAD group at the family level. Solid lines represent the mean; color
 455 shadings represent 95% confidence intervals. (B) Dynamics of 5 most abundant families
 456 across the duration of the study between the AAD and non-AAD groups (Bonferroni-
 457 corrected, two-sided Mann-Whitney U test, $P \leq 0.05$, *; $P \leq 0.01$, **). Error bars represent 68%
 458 confidence intervals. (C) Pie charts of 3 most abundant genera in Ruminococcaceae across
 459 the duration of the study (D) Dynamics of 3 most abundant genera in Ruminococcaceae,
 460 *Faecalibacterium*, *Subdoligranulum* and *Ruminococcus* across the duration of the study
 461 between AAD and non-AAD groups (Bonferroni-corrected, two-sided Mann-Whitney U test,
 462 $P \leq 0.05$, *; $P \leq 0.01$, **).

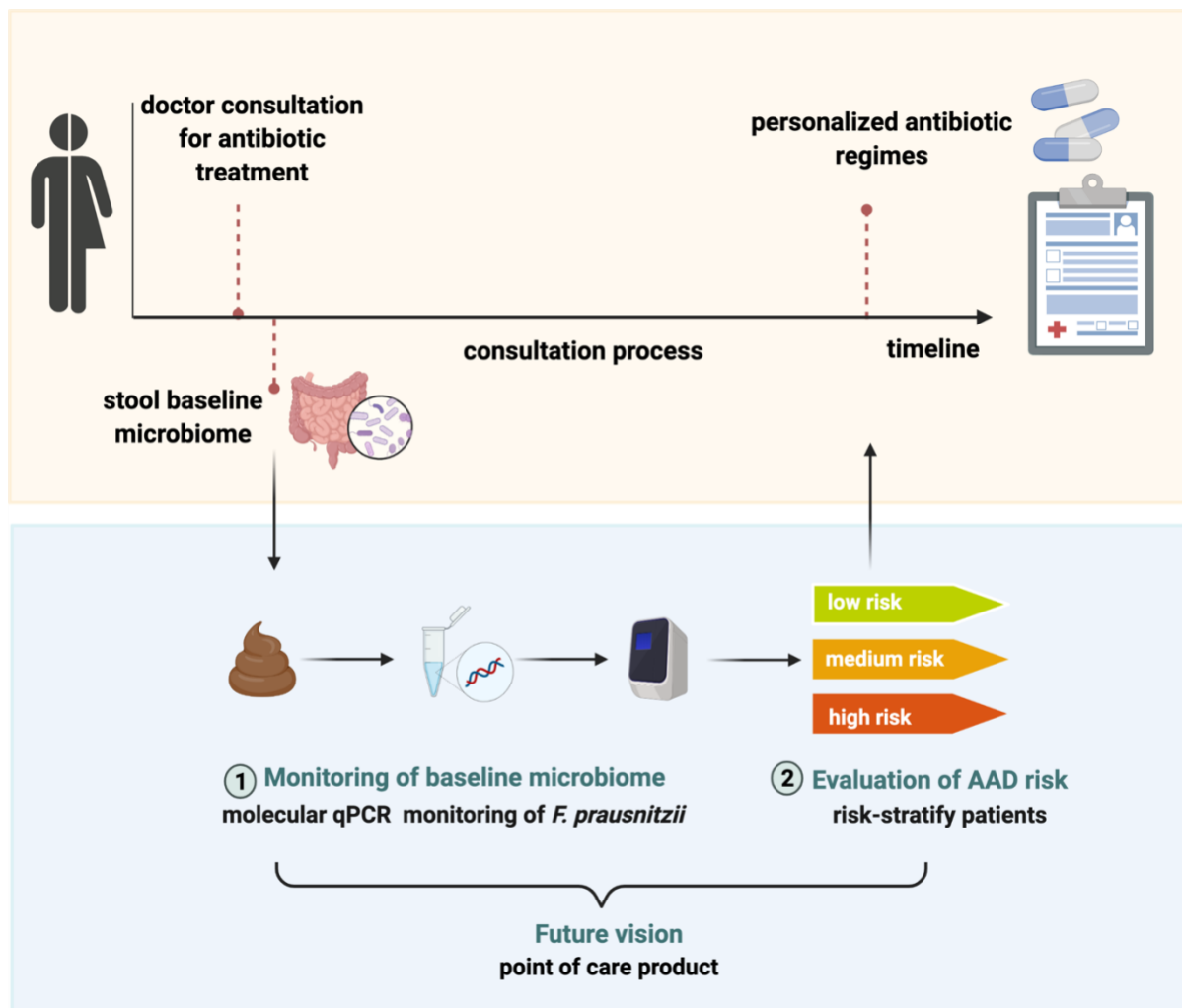


465 **Figure 3. Amoxicillin-clavulanate causes greater gut microbiome diversity loss and**
 466 **community disturbance in the AAD group compared to the non-AAD group.** (A)
 467 Principal coordinates analysis (PCoA) based on ASV-level Bray-Curtis dissimilarity. Display
 468 is based on sample scores on the primary axis (PCoA1, 16.5% variance explained) and
 469 secondary axis (PCoA2, 9.9% variance explained). To reduce the redundancy of sample
 470 points on the plot, we picked microbiomes on day 3 to represent the post-dosing period (days
 471 1-3). Days 0, 3, 7 and 28 were included as the data points with days 0, 7 and 28 represented
 472 simply as ‘non-day 3’. The greatest variation observed in the AAD group occurs on day 3.
 473 Individuals return to their baseline microbiomes from day 7. PERMANOVA results show

474 that there were no significant differences between day 0 and 7 in both AAD ($P = 0.43$, $N = 24$)
475 and non-AAD groups ($P = 0.51$, $N = 24$). (B) Within-sample species diversity (α diversity of
476 ASVs, Shannon entropy index) greatly decreased in the AAD group compared to the non-
477 AAD group on day 2. The similarity of each individual's gut microbiota to their baseline
478 communities (β diversity of ASVs, Jensen–Shannon distance) greatly decreased in the AAD
479 group compared to the non-AAD group cross days 2-3. Significant difference between the
480 AAD and non-AAD groups are labelled with asterisks (Bonferroni-corrected, two-sided
481 Mann–Whitney U test, $P \leq 0.05$, *; $P \leq 0.01$, **). (C) A sharp increase of Proteobacteria, and
482 a decrease of Firmicutes and Actinobacteria were observed in the AAD group on days 2 and
483 3.



485 **Figure 4. Predicting the risk of AAD based on the relative abundance of**
486 **Ruminococcaceae at baseline.** Red and blue labels denote individuals with AAD and non-
487 AAD, respectively (A) Inter-individual microbial community variation of amplicon sequence
488 variants (ASVs) at baseline. Hierarchical clustering of the baseline faecal bacterial
489 composition of 30 healthy individuals (average linkage, with correlation matrix). The colour
490 threshold for signifying clusters was set to a pearson distance of '0.4'. Majority (14/17) of
491 non-AAD individuals were grouped into Cluster 1. Majority of AAD individuals (9/13) were
492 excluded from Cluster 1 to form several individual branches. (B) Distribution of
493 Ruminococcaceae relative abundance at baseline among the AAD and non-AAD groups.
494 Each number refers to the individual at baseline. (C) qPCR concentration for the species *F.*
495 *prausnitzii* was normalized to 16S rRNA gene copy (Data represent median±IQR range, n=3).
496 Each number refers to the individual at baseline. (D) Correlations between Ruminococcaceae
497 relative abundance and *F. prausnitzii* median absolute quantification (Spearman's $\rho = 0.850$,
498 $p = 3.0e-9$). Line depicts the best linear fit and blue shading the 95% confidence interval of
499 the linear fit. (E) Calculated predictive precision of developing AAD using *F. prausnitzii*
500 absolute abundance quantified by qPCR assay.



501

502 **Figure 5. Risk-stratifying patients at risk of AAD using stool *F. prausnitzii* qPCR .** We
503 propose a workflow to evaluate AAD risk at baseline prior to antibiotic administration. This
504 would enable clinicians to risk-stratify patients, and identify those in whom use of
505 amoxicillin-clavulanate should be avoided due to higher risk of AAD. Although
506 quantification of *F. prausnitzii* is currently via qPCR, this could potentially be translated into
507 rapid point-of-care diagnostics in future. This framework could also serve as basis for
508 identifying potential drivers in AAD caused by other classes of antibiotics. (This figure was
509 created using BioRender.com.)

510

511 REFERENCES

- 512 1. Allen SJ, et al. Lactobacilli and bifidobacteria in the prevention of antibiotic-associated
513 diarrhoea and *Clostridium difficile* diarrhoea in older inpatients (PLACIDE): a
514 randomised, double-blind, placebo-controlled, multicentre trial. *The Lancet*
515 2013;382(9900):1249–1257.
- 516 2. McFarland L V. Diarrhoea associated with antibiotic use. *British Medical Journal*
517 2007;335(7610).
- 518 3. Larcombe S, Hutton ML, Lyras D. Involvement of Bacteria Other Than *Clostridium*
519 *difficile* in Antibiotic-Associated Diarrhoea. *Trends in Microbiology* 2016;24(6):463–
520 476.
- 521 4. Beaugerie L, Petit JC. Antibiotic-associated diarrhoea. *Best Practice and Research:*
522 *Clinical Gastroenterology* 2004;18(2):337–352.
- 523 5. Mylonakis E, Ryan ET, Calderwood SB. *Clostridium difficile*-associated diarrhea: A
524 review. *Archives of Internal Medicine* 2001;161(4):525–533.
- 525 6. Larcombe S, et al. Diverse bacterial species contribute to antibiotic-associated diarrhoea
526 and gastrointestinal damage. *Journal of Infection* 2018;77(5):417–426.
- 527 7. Zhang W, et al. *Bacteroides fragilis* protects against antibiotic-associated Diarrhea in
528 rats by modulating intestinal defenses. *Frontiers in Immunology* 2018;9(MAY).
- 529 8. Clausen MR, et al. Colonic fermentation to short-chain fatty acids is decreased in
530 antibiotic-associated diarrhea. *Gastroenterology* 1991;101(6):1497–1504.

- 531 9. Young VB, Schmidt TM. Antibiotic-Associated Diarrhea Accompanied by Large-Scale
532 Alterations in the Composition of the Fecal Microbiota. *Journal of Clinical*
533 *Microbiology* 2004;42(3):1203–1206.
- 534 10. WHO. 2019 WHO AWaRe Classification Database of Antibiotics for evaluation and
535 monitoring of use2019; doi:(WHO/EMP/IAU/2019.11).
- 536 11. Bartlett JG. Antibiotic-associated diarrhea. *Clinical Infectious Disease, Second Edition*
537 2015;18(2):349–351.
- 538 12. Kuehn J, et al. Reported rates of diarrhea following oral penicillin therapy in pediatric
539 clinical trials. *Journal of Pediatric Pharmacology and Therapeutics* 2015;20(2).
- 540 13. Kubota H, et al. Development of TaqMan-Based Quantitative PCR for Sensitive and
541 Selective Detection of Toxigenic *Clostridium difficile* in Human Stools. *PLoS ONE*
542 2014;9(10):e111684.
- 543 14. Carroll IM, et al. Alterations in composition and diversity of the intestinal microbiota in
544 patients with diarrhea-predominant irritable bowel syndrome. *Neurogastroenterology*
545 *and Motility* 2012;24(6).
- 546 15. De Weirdt R, Van De Wiele T. Micromanagement in the gut: Microenvironmental
547 factors govern colon mucosal biofilm structure and functionality. *npj Biofilms and*
548 *Microbiomes* 2015;1(October):1–6.
- 549 16. Joossens M, et al. Dysbiosis of the faecal microbiota in patients with Crohn’s disease
550 and their unaffected relatives. *Gut* 2011;60(5).
- 551 17. Morgan XC, et al. Dysfunction of the intestinal microbiome in inflammatory bowel
552 disease and treatment.. *Genome biology* 2012;13(9).

- 553 18. Sokol H, et al. Faecalibacterium prausnitzii is an anti-inflammatory commensal
554 bacterium identified by gut microbiota analysis of Crohn disease patients. *Proceedings*
555 *of the National Academy of Sciences of the United States of America*
556 2008;105(43):16731–16736.
- 557 19. Bajaj JS, et al. Linkage of gut microbiome with cognition in hepatic encephalopathy.
558 *American Journal of Physiology - Gastrointestinal and Liver Physiology* 2012;302(1).
- 559 20. Antharam VC, et al. Intestinal dysbiosis and depletion of butyrogenic bacteria in
560 *Clostridium difficile* infection and nosocomial diarrhea. *Journal of Clinical*
561 *Microbiology* 2013;51(9).
- 562 21. Wong JMW, et al. Colonic health: Fermentation and short chain fatty acids. *Journal of*
563 *Clinical Gastroenterology* 2006;40(3):235–243.
- 564 22. Harig JM, et al. Treatment of Diversion Colitis with Short-Chain-Fatty Acid Irrigation.
565 *New England Journal of Medicine* 1989;320(1).
- 566 23. Scarpellini E, et al. Efficacy of butyrate in the treatment of diarrhoea-predominant
567 irritable bowel syndrome. *Digestive and Liver Disease Supplements* 2007;1(1).
- 568 24. Scheppach W, et al. Effect of butyrate enemas on the colonic mucosa in distal ulcerative
569 colitis. *Gastroenterology* 1992;103(1).
- 570 25. Mekonnen SA, et al. Molecular mechanisms of probiotic prevention of antibiotic-
571 associated diarrhea. *Current Opinion in Biotechnology* 2020;61:226–234.
- 572 26. Hempel S, et al. Probiotics for the prevention and treatment of antibiotic-associated
573 diarrhea: A systematic review and meta-analysis. *JAMA - Journal of the American*
574 *Medical Association* 2012;307(18):1959–1969.

- 575 27. Johnston BC, Goldenberg JZ, Parkin PC. Probiotics and the prevention of antibiotic-
576 associated diarrhea in infants and children. *JAMA - Journal of the American Medical*
577 *Association* 2016;316(14).
- 578 28. Perceval C, et al. Prophylactic use of probiotics for gastrointestinal disorders in children.
579 *The Lancet Child and Adolescent Health* 2019;3(9):655–662.
- 580 29. Kołodziej M, Szajewska H. *Lactobacillus reuteri* DSM 17938 in the prevention of
581 antibiotic-associated diarrhoea in children: a randomized clinical trial. *Clinical*
582 *Microbiology and Infection* 2019;25(6).
- 583 30. Caporaso JG, et al. Global patterns of 16S rRNA diversity at a depth of millions of
584 sequences per sample. *Proceedings of the National Academy of Sciences of the United*
585 *States of America* 2011;108(SUPPL. 1).
- 586 31. T M, SL S. FLASH: fast length adjustment of short reads to improve genome
587 assemblies. *Bioinformatics (Oxford, England)* 2011;27(21):2957–2963.
- 588 32. Caporaso JG, et al. QIIME allows analysis of high-throughput community sequencing
589 data. *Nature Methods* 2010;7(5):335–336.
- 590 33. RC E, et al. UCHIME improves sensitivity and speed of chimera detection.
591 *Bioinformatics (Oxford, England)* 2011;27(16):2194–2200.
- 592 34. Bolyen E, et al. Reproducible, interactive, scalable and extensible microbiome data
593 science using QIIME 2. *Nature Biotechnology* 2019;37(8).
- 594 35. Callahan BJ, et al. DADA2: High-resolution sample inference from Illumina amplicon
595 data. *Nature Methods* 2016;13(7).

596 36. Callahan BJ, McMurdie PJ, Holmes SP. Exact sequence variants should replace
597 operational taxonomic units in marker-gene data analysis. *ISME Journal* 2017;11(12).

598 37. Bokulich NA, et al. Optimizing taxonomic classification of marker-gene amplicon
599 sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome* 2018;6(1).

600 38. Van Hul M, et al. From correlation to causality: the case of *Subdoligranulum*. *Gut*
601 *Microbes* 2020;12(1):1–13.

602