Genetic determinants of gut microbiota composition and bile acid profiles in mice

Julia H. Kemis[1], Vanessa Linke[2], Kelsey L. Barrett[1], Frederick J. Boehm[4], Lindsay L. Traeger[1], Mark P. Keller[3], Mary E. Rabaglia[3], Kathryn L. Schueler[3], Donald S. Stapleton[3], Daniel M. Gatti[8], Gary A. Churchill[8], Daniel Amador-Noguez[1], Jason D. Russell[2], Brian S. Yandell[4], Karl W. Broman[5], Joshua J. Coon[2,6,7], Alan D. Attie[3] and Federico E. Rey*[1]

[1] Department of Bacteriology, University of Wisconsin – Madison, Madison, WI, USA

[2] Department of Chemistry, University of Wisconsin – Madison, Madison, WI, USA

[3] Department of Biochemistry, University of Wisconsin – Madison, Madison, WI, USA

[4] Department of Statistics, University of Wisconsin – Madison, Madison, WI, USA

[5] Department of Biostatistics and Medical Informatics, University of Wisconsin – Madison, Madison, WI, USA

[6] Morgridge Institute for Research, Madison, WI 53715

[7] Department of Biomolecular Chemistry, University of Wisconsin – Madison, Madison, WI, USA

[8] Jackson Laboratory, Bar Harbor, ME, USA

Key words: Gut microbiome, genetics, bile acids

***Corresponding author:**

Federico E. Rey

Email: ferey@wisc.edu

Abstract

The microbial communities that inhabit the distal gut of humans and other mammals exhibit large inter-individual variation. While host genetics is a known factor that influences gut microbiota composition, the mechanisms underlying this variation remain largely unknown. Bile acids (BAs) are hormones that are produced by the host and chemically modified by gut bacteria. BAs serve as environmental cues and nutrients to microbes, but they can also have antibacterial effects. We hypothesized that host genetic variation in BA metabolism and homeostasis influence gut microbiota composition. To address this, we used the Diversity Outbred (DO) stock, a population of genetically distinct mice derived from eight founder strains. We characterized the fecal microbiota composition and plasma and cecal BA profiles from 400 DO mice maintained on a high-fat high-sucrose diet for ~22 weeks. Using quantitative trait locus (QTL) analysis, we identified several genomic regions associated with variations in both bacterial and BA profiles. Notably, we found overlapping QTL for *Turicibacter sp.* and plasma cholic acid, which mapped to a locus containing the gene for the ileal bile acid transporter, *Slc10a2*. Mediation analysis and subsequent follow-up validation experiments suggest that differences in *Slc10a2* gene expression associated with the different strains influences levels of both traits and revealed novel interactions between *Turicibacter* and BAs. This work illustrates how systems genetics can be utilized to generate testable hypotheses and provide insight into host-microbe interactions.

Author summary

Inter-individual variation in the composition of the intestinal microbiota can in part be attributed to host genetics. However, the specific genes and genetic variants underlying differences in the microbiota remain largely unknown. To address this, we profiled the fecal microbiota composition

of 400 genetically distinct mice, for which genotypic data is available. We identified many loci of the mouse genome associated with changes in abundance of bacterial taxa. One of these loci is also associated with changes in the abundance of plasma bile acids—metabolites generated by the host that influence both microbiota composition and host physiology. Follow up validation experiments provide mechanistic insights linking host genetic differences, with changes in ileum gene expression, bile acid-bacteria interactions and bile acid homeostasis. Together, this work demonstrates how genetic approaches can be used to generate testable hypothesis to yield novel insight into how host genetics shape gut microbiota composition.

Introduction

The intestinal microbiota has profound effects on host physiology and health (1–3). The composition of the gut microbiota is governed by a combination of environmental factors, including diet, drugs, maternal seeding, cohabitation, and host genetics (4–7). Together, these factors cause substantial inter-individual variation in microbiota composition and modulate disease risk (8,9). Alterations in the composition of the microbiota are associated with a spectrum of cognitive, inflammatory and metabolic disorders (10–12) and a number of bacterial taxa have been causally linked with modulation of disease (13–15). A major challenge in the field is deciphering how host genetics and environmental factors interact to shape the composition of the gut microbiota. This knowledge is key for designing strategies aimed at modifying gut microbiota composition to improve health outcomes.

Several mouse and human studies have examined the role of host genetics in shaping the composition of the gut microbiota (16). Mouse studies comparing gut bacterial communities from inbred mouse strains (17,18) and strains harboring mutations in immune-related genes (19–22)

71 support this notion. Additionally, quantitative trait locus (QTL) analyses in mice have identified
72 genetic regions associated with the abundance of several bacterial taxa and community structure
73 (23–26). Twin studies and genome-wide association studies (GWAS) in humans have identified
74 heritable bacterial taxa and SNPs associated with specific gut microbes. While comparing these
75 studies is often difficult due to differences in environmental variables among populations, some
76 associations are consistently detected among geographically discrete populations, such as the
77 association between *Bifidobacterium* abundance and the lactase (*LCT*) gene locus (27–29),
78 indicating the abundance of specific taxa is influenced by host genetic variation.

79 Gut microbes and the host communicate through the production and modification of metabolites,
80 many of which impact host physiology (30–34). Bile Acids (BAs) are host-derived and microbial-
81 modified metabolites that regulate both the gut microbiome and host metabolism (35–37). BAs are
82 synthesized in the liver from cholesterol, stored in the gallbladder and are secreted in the proximal
83 small intestine where they facilitate absorption of fat-soluble vitamins and lipids. Once in the
84 intestine, BAs can be metabolized by gut bacteria through different reactions, including
85 deconjugation, dehydroxylation, epimerization, and dehydrogenation, to produce secondary BAs
86 with differential effects on the host (33,35). In addition to their direct effects on the host, BAs
87 shape the gut microbiota composition through antimicrobial activities (38,39). The detergent
88 properties of BAs cause plasma membrane damage. The bactericidal activity of a BA molecule
89 corresponds to its hydrophobicity (40). Additionally, the microbiota modulates primary BA
90 synthesis through regulation of the nuclear factor FXR (41). Thus, we hypothesized that host
91 genetic variation associated with changes in BA homeostasis mediates alterations in gut microbiota
92 composition.

93 To investigate how genetic variation affects gut microbiota and BA profiles, we used the
94 Diversity Outbred (DO) mouse population, which is a heterogenous population derived from eight
95 founder strains: C57BL6/J (B6), A/J (A/J), 1291/SvImJ (129), NOD/ShiLtJ (NOD), NZO/HiLtJ
96 (NZO), CAST/EiJ (CAST), PWK/PhJ (PWK), and WSB/EiJ (WSB) (42,43). These eight strains
97 capture a large breadth of the genetic diversity found in inbred mouse strains. Additionally, the
98 founder strains harbor distinct gut microbial communities and exhibit disparate metabolic
99 responses to diet-induced metabolic disease (18,44,45). The DO population is maintained by an
100 outbreeding strategy aimed at maximizing the heterozygosity of the outbred stock. The genetic
101 diversity and large number of generations of outbreeding make it an ideal resource for high-
102 resolution genetic mapping of microbial and metabolic traits (43).
103 We characterized the intestinal microbiota composition and plasma and cecal BA profiles in ~400
104 genetically distinct DO mice fed a high-fat/high-sucrose diet for ~22 weeks and performed
105 quantitative trait loci (QTL) analysis to identify host genetic loci associated with these traits.
106 Specifically, we focused our analysis on potentially pleiotropic loci, which we defined as a single
107 genetic locus that associates with both bacterial and BA traits. Our analysis revealed several
108 instances of bacterial and metabolite traits attributed to the same DO founder haplotypes mapping
109 to the same position of the mouse genome, including a locus associated with plasma BA levels and
110 the disease-modulating organism *Akkermansia muciniphila.* Additionally, we identified the ileal
111 BA transporter *Slc10a2* as a candidate gene that regulates both the abundance of *Turicibacter sp.*
112 and plasma levels of cholic acid.
113
114

Results and discussion

Phenotypic variation among Diversity Outbred (DO) mice fed high-fat and high-sucrose diet

We investigated the impact of genetic variation on gut microbiota composition and bile acid (BA) profiles using a cohort of ~400 DO mice maintained on a high-fat high-sucrose diet (45% kcal from fat and 34% from sucrose) for ~22 weeks (range 21-25 weeks), starting at weaning. Additionally, we incorporated in our analyses previously published clinical weight traits collected from the same mice (46) (Fig 1A). All animals were individually housed throughout the duration of the study to minimize microbial exchange.

We performed LC-MS analyses of plasma and cecal contents to assess abundance of 27 BAs. There was substantial variation in the plasma and cecal BA profiles across the 400 mice (Fig 1C and 1D; S1 Table). Additionally, we examined gut microbiota composition using 16S rRNA gene amplicon sequencing of DNA extracted from fecal samples collected at the end of the experiment. Within the cohort, there were 907 unique Exact Sequence Variants (ESVs), (100% operational taxonomic units defined with dada2 (47)), which were agglomerated into 151 lower taxonomic rankings (genus, family, order, class, phyla) (S1 Table). The microbial traits represented each of the major phyla found in the intestine and the relative abundance of these phyla was highly variable among the DO mice (Fig 1B). For instance, the abundance of taxa classified to the Bacteroidetes phylum ranged from 1.17 – 89.28%.

For subsequent analysis, we identified a core measurable microbiota (CMM), which we defined as taxon found in at least 20% of the mice (24). This was done to remove the effects of excessive variation in the data due to bacterial taxa that were low abundance and/or sparsely distributed. In total, the CMM was comprised of 86 ESVs and 42 agglomerated taxa (S2 Table). The CMM traits

represent a small fraction of the total microbes detected, but account for 94.5% of the rarefied sequence reads, and therefore constitute a significant portion of the identifiable microbiota.

Since mice were received in waves of 100, we examined whether animals in each wave were more similar to each other than mice in other waves. The fecal microbiota composition significantly clustered by wave ($p < 0.001$, PERMANOVA) and sex ($p < 0.001$, PERMANOVA) (S1 Fig). PCA analysis of plasma and cecal bile acids showed a significant effect of sex, but not wave, on both plasma ($p < 0.0001$, Kruskal Wallis) and cecal BA profiles ($p < 0.05$, Kruskal Wallis) (S2 Fig).

There is substantial evidence implicating gut microbiota and BAs in metabolic disease development (36,37). To identify potential relationships among these traits, we performed correlation analysis which yielded many significant associations after FDR correction (FDR < 0.05) (S3 Table, discussed in S1 Data).

Abundance of gut bacterial taxa and bile acids are associated with host genetics

To identify associations between regions of the mouse genome and the clinical and molecular traits discussed above, we performed QTL analysis using the R/qtl2 package (48). We used sex, days on the diet, and experimental cohort (wave) as covariates. We identified 459 QTL for bacterial (306), bile acid (131), and body weight (22) traits (Fig 2, S4 Table) with a LOD score > 5.5.

Of the microbial QTL, we found 190 QTL for 76 distinct bacterial ESVs from four phyla that met a cut-off LOD > 5.5. ESVs with the strongest QTL (LOD > 8) are classified to the Clostridiales order and map on chr 12 at ~33 Mbp, the Lachnospiraceae family on chr 2 at 164 Mbp, and the S24-7 family on chr 2 at ~115 Mbp. We also identified 116 QTL for microbial taxa

160 collapsed by taxonomic assignment (i.e., genus to phylum). The genera *Lactococcus* and
161 *Akkermansia* were also associated with host genetic variation, which is consistent with previous
162 studies (23,24,49,50).

163 Similarly, BA QTL mapped to multiple loci spanning the mouse genome and most BA
164 traits mapped to multiple positions. BA synthesis and metabolism are regulated by multiple host
165 signaling pathways: there are >17 known host enzymes involved in the production of BAs (36),
166 transporters, which play a critical role in maintaining the enterohepatic circulation and BA
167 homeostasis, and receptors that respond to BA in a variety of host tissues (51–53). Therefore, it is
168 not surprising that our results indicate that BA levels are polygenic and shaped by multiple host
169 factors.

170 We observed multiple instances of related BA species associating to the same genetic
171 locus. These overlapping QTL may indicate the presence of a pleiotropic locus. Interestingly,
172 several of these loci associate with levels of related BA species in different stages of microbial
173 modification. For example, cecal taurocholic acid (TCA) and plasma CA QTL overlap on chr 7 at
174 122 Mbp. Likewise, four BA QTL that are all derivatives of the secondary BA DCA, including
175 plasma TDCA and cecal DCA, isodeoxycholic acid (IDCA), and HDCA overlap on chr 12
176 between ~99 – 104 Mbp. For the cecal BA, the WSB founder haplotype was associated with higher
177 levels of these three BA, while the NOD founder haplotype was associated with lower levels. The
178 opposite pattern was observed for plasma TDCA, where the NOD and WSB haplotype were
179 associated with higher and lower levels, respectively (S3A-S3D Fig).

180 We also identified overlapping QTLs on chr 11 at ~71 Mbp for cecal levels of the
181 secondary BAs lithocholic acid (LCA) and isolithocholic acid (ILCA), the isomer of LCA
182 produced by bacterial 3α-hydroxylation (S3E Fig). Higher levels of these cecal BAs are associated

183 with the 129 founder haplotype and lower levels are associated with the A/J founder haplotype
184 (S3F-S3G Fig). We identified the positional candidate gene *Slc13a5* (S3H Fig), which is a sodium-
185 dependent transporter that mediates cellular uptake of citrate, an important precursor in the
186 biosynthesis of fatty acids and cholesterol (54). Recent evidence indicates that *Slc13a5* influences
187 host metabolism and energy homeostasis (55–57). *Slc13a5* is a transcriptional target of pregnane
188 X receptor (PXR) (58), which also regulates the expression of genes involved in the biosynthesis,
189 transport, and metabolism of BAs (59).
190
191 **Co-mapping analyses identifies novel interactions between bacterial taxa and bile acid**
192 **homeostasis**
193 We searched for regions of the chromosome that were associated with both BA and
194 bacterial abundance, as this may provide evidence of interactions between the traits (60). We
195 identified 17 instances of overlapping microbial and BA QTL on 12 chromosomes. This QTL
196 overlap indicates there might be QTL with pleiotropic effects on BAs and the microbiota, suggest
197 that genetic variation influencing host BA profiles has an effect on compositional features of the
198 gut microbiota, or genetic-driven variation in microbiota composition alters BAs. Examples of
199 notable instances of overlapping bacterial and BA QTL are discussed in the Supporting
200 Information (S1 Data).
201 We focused our co-mapping analysis on chr 8 at ~ 5.5 Mbp, where *Turicibacter sp.* QTL
202 and plasma cholic acid (CA) QTL overlap (Fig 3A and 3B). These traits were particularly
203 interesting because both have been shown to be influenced by host genetics by previous studies.
204 *Turicibacter* has been identified as highly heritable in both mouse and human genetic studies
205 (24,27,45,49), whereas multiple reports have found differences in CA levels as a function of host

206 genotype (18,61). Furthermore, CA levels are influenced by both host genetics and microbial
207 metabolism since it is synthesized by host liver enzymes from cholesterol and subsequently
208 modified by gut microbes in the intestine. Notably, these co-mapping traits also share the same
209 allele effects pattern, where the A/J and WSB haplotypes have strong positive and negative
210 associations, respectively (Fig 3C and 3D).

211 To assess whether the trait patterns observed in the DO founder strains correspond to the
212 observed allelic effects in the QTL mapping, we performed a separate characterization of the fecal
213 microbiota composition and plasma bile acids in age-matched A/J and WSB animals fed the
214 HF/HS diet. The founder strain allele patterns inferred from the QTL mapping closely resembled
215 the observed levels of *Turicibacter sp.* (Fig 3E) and plasma CA in the founder strains (Fig 3F),
216 where A/J animals had significantly higher levels of *Turicibacter sp.* and CA than WSB animals.
217 However, *Turicibacter* levels in the founder strains do not complete mirror the estimated allele
218 effects. This may be due to other genetic factors that also influence *Turicibacter* levels, as this taxa
219 may be influenced by multiple host genes and levels of *Turicibacter* have previously been
220 associated on chr 7 (24), 9 and 11 (49). Furthermore, *Turicibacter* and plasma CA were positively
221 correlated in the DO mice ($r = 0.43$, $p = 3.53e^{-10}$). This finding is consistent with a previous study
222 that found positive correlations between *Turicibacter* and unconjugated cecal BAs (62). Taken
223 together, the overlap between the *Turicibacter sp.* QTL and plasma CA QTL, along with the
224 similar allele effects pattern, which reflect the values observed in the founder strains, provide
225 strong evidence suggesting that these traits are related and they are responding to the common
226 genetic driver.

227

228 **_Slc10a2_ is a candidate gene for _Turicibacter sp._ and plasma cholic acid**

We searched in the QTL confidence interval for candidate genes via high-resolution association mapping on chr 8 and identified SNPs associated with both traits. Among these we identified SNPs upstream of the gene *Slc10a2*, which encodes for the apical sodium-bile transporter (Fig 3G). Slc10a2 is responsible for ~95% of BA reabsorption in the distal ileum and plays a key role in BA homeostasis (63). In humans, mutations in this gene are responsible for primary BA malabsorption, resulting in interruption of enterohepatic circulation of BAs and decreased plasma cholesterol levels (64). Likewise, *Slc10a2*$^{-/-}$ mice have a reduced total BA pool size, increased fecal BA concentrations and reduced total plasma cholesterol in comparison to wild-type mice (63). Additionally, a comparison between germ-free and conventionally-raised mice found that expression of *Slc10a2* is downregulated in presence of the gut microbiota, suggesting microbes may influence the expression of the transporter (41).

Our analysis identified SNPs associated with levels of *Turicibacter sp.* and plasma CA at the QTL peak (Fig 3G). The SNPs with the strongest associations were attributed to the WSB and A/J haplotypes and fell on intergenic regions near *Slc10a2*. There is growing evidence that non-coding intergenic SNPs are often located in or closely linked to regulatory regions, suggesting that they may influence host regulatory elements and alter gene expression (65,66). To assess if candidate gene expression patterns in the DO founders corresponds to the estimated allelic effects in the QTL mapping, we quantified *Slc10a2* expression in distal ileum samples from A/J and WSB mice by quantitative reverse transcriptase PCR (qRT-PCR). A/J mice exhibited significantly higher expression of *Slc10a2* compared to WSB mice (Fig 3H), which is consistent with estimated allele patterns for the overlapping *Turicibacter* and plasma CA QTLs on chr 8 (Fig 3A and 3B). Remarkably, several studies have noted concomitant changes in microbiota composition and *Slc10a2* mRNA levels (67–69).

252

253 **A common genetic driver controls *Turicibacter sp.* and plasma cholic acid**

254 We mapped QTL for *Turicibacter sp.* and for plasma CA levels to a common locus on chr
255 8 at 5-7 Mbp. Since the LOD profiles and allelic effects are highly similar, the QTL may be due
256 to a single shared locus (pleiotropy) or multiple closely linked loci. We examined this question
257 using a likelihood ratio testing of the null hypothesis of pleiotropy versus the alternative of two
258 independent genetic regulators of these traits (70). Analysis of 1000 bootstrap samples resulted in
259 a p-value of 0.531, which is consistent with the presence of a single pleiotropic locus that affects
260 both traits.

261 We next sought to understand the causal relationships between the microbe and the BA.
262 We asked whether the relationship between the microbe and BA was causal, reactive or
263 independent. To establish the directionality of the relationship, we applied mediation analysis
264 where we conditioned one trait on the other (71). When we conditioned *Turicibacter sp.* on plasma
265 CA (QTL → BA → Microbe), we observed a LOD drop of 3.2 (Fig 4A and 4B). Likewise, when
266 we conditioned the plasma cholic acid on the microbe (QTL → Microbe → BA) there was a LOD
267 drop of 3.32 (Fig 4C and 4D). The partial mediation seen in both models suggests that the
268 relationship between the microbe and the BA could be bidirectional, where they exert an effect on
269 one another.

270 From this analysis, we can hypothesize this relationship can be explained by a pleiotropic
271 model, where a single locus influences a microbial and a BA trait, and the microbial trait is also
272 reactive to changes in the BA trait. It is important to note that statistical inference only partially
273 explains the relationship between the traits and there may be other hidden variables that may
274 further explain the relationship. The complex relationship depicted by the causal inference testing

275 is consistent with the interplay between gut microbes and BAs in the intestine and their known
276 ability to influence the other.
277
278 **Bile acids inhibit *Turicibacter sanguinis* growth at physiologically relevant concentrations**
279 Due to the strong correlative relationship between the QTL, we tested whether there was a
280 direct interaction between bile acids and *Turicibacter*. *Turicibacter* inhabits the small intestine
281 where BAs are secreted upon consumption of a meal (73,74). We screened the human isolate
282 *Turicibacter sanguinis* for deconjugation and transformation activity *in vitro* by HPLC/MS-MS.
283 We found that *T. sanguinis* deconjugated ~96-100% of taurocholic acid and
284 glycochenodeoxycholic acid (Fig 5A) within 24 hours. It also transformed ~6 and 8 % of CA and
285 CDCA to 7-dHCA and 7-ketolithocholic acid (7-KLCA), respectively (Fig 5B and 5C). The
286 percent transformed did not increase after 24 hours (data not shown). Both of these transformations
287 require the action of the bacterial 7α-hydroxysteroid dehydrogenase.
288 Based on these results, we asked if conjugated and unconjugated bile acids differentially
289 modulate *T. sanguinis* growth. BA concentrations range from ~1-10 mM along the small intestine
290 (75) to ~0.2-1 mM in the cecum (76). Therefore, we grew *T. sanguinis* in the presence of either
291 conjugated or unconjugated bile acids at physiologically relevant concentrations ranging from 0.1
292 – 5 mM. *T. sanguinis* growth decreased with increasing concentrations of BAs and growth was
293 completely inhibited at 1 mM for unconjugated BAs and 5 mM for conjugated BAs (Fig 5D and
294 5E). Growth rate was significantly slower in the presence of 1 mM conjugated and 0.5mM
295 unconjugated bile acids (Fig 5F). These results suggest that levels of BAs may affect abundance
296 of *Turicibacter* in the gut.

297 To compare *T. sanguinis* sensitivity to conjugated bile acids relative to other small intestine
298 colonizers, we grew four taxa (*Bacteroides thetaiotaomicron*, *Clostridium asparagaiforme*,
299 *Lactobacillus reuteri* and *Escherichia coli* MS200-1) known to colonize this region of the intestine
300 with or without 1 mM conjugated bile acids. Members of these genera are known to have bile salt
301 hydrolase (BSH) activity to deconjugate bile acids (35). Unlike *T. sanguinis*, the addition of high
302 levels of conjugated bile acids had little to no effect on the growth of these four gut microbes (S4
303 Fig). Consistent with these findings, *Turicibacter* abundance was negatively correlated with cecal
304 TCA levels in the DO mice ($r = -0.262$, $p = 0.0035$).

305 Taken together, these data indicate that *T. sanguinis* is sensitive to higher concentrations
306 of BA compared to other small intestine colonizers. These reciprocal effects between the BA and
307 the bacterium provide biological evidence for the correlative relationship shown by the causal
308 model testing. In summary, using a genetic approach, we identified and provide validation of a
309 relationship between a genetic locus containing the BA transporter *Slc10a2*, and levels of
310 *Turicibacter* and plasma cholic acid. Based on our findings, we hypothesize that the identified
311 locus regulates expression of *Slc10a2*, altering active BA reabsorption in the ileum, leading to
312 increased intestinal BA concentrations and alterations in the intestinal BA environment.
313 Consequently, the resulting environmental change provides an unfavorable habitat for
314 *Turicibacter*. In turn, lower levels of *Turicibacter* BA deconjugation activity leads to a decrease
315 in circulating free plasma cholic acid levels.

316

317 ## Conclusion

318 In this study, we performed the first known genetic mapping integration of gut microbiome
319 and BA profiles. Using DO mice, we identified multiple QTL for gut microbes and bile acids

spanning the host genome. These included loci that associated with individual microbial and BA traits, as well as loci with potential pleiotropic effects, where a single genetic region influenced both the abundance of a gut microbe and levels of a BA. While several studies suggest that host genetic variation has a minor impact on microbiota composition, there are overlapping findings among different studies in both human and mouse populations that indicate that specific bacterial taxa are influenced by host genetics. Our results in the DO population corroborate several of these key findings (discussed in S1 Data). *Turicibacter sp*. is among the microbes consistently associated with host genetics. This work plus data from previous reports suggest that alterations in the BA pool driven by *Slc10a2* genetic variation and concomitant changes in expression/activity elicit an impact on gut microbiota community structure and influence the ability of *Turicibacter* to colonize and persist in the intestine. Although this microbe deconjugates primary BAs, we found that it is also sensitive to elevated concentrations of both conjugated and unconjugated BAs. Future experiments are needed to examine how a decrease in *Slc10a2* expression changes intestinal BA profiles and the consequences on *Turicibacter* colonization. Additionally, this work identified multiple host-microbe-metabolite interactions that need to be validated with additional molecular studies. More broadly, our work demonstrates the power of genetics to identify novel interactions between microbial and metabolite traits and provides new testable hypotheses to further dissect factors that shape gut microbiota composition.

Materials and methods

Animals and sample collection. Animal care and study protocols were approved by the University of Wisconsin-Madison Animal Care and Use Committee. DO mice were obtained from the Jackson Laboratories (Bar Harbor, ME, USA) at ~4 weeks of age and maintained in the

Department of Biochemistry vivarium at the University of Wisconsin-Madison. Mice were housed on a 12-hour light:dark cycle under temperature- and humidity-controlled conditions. Five waves of 100 DO mice each from generations, 17, 18, 19, 21, and 23 were obtained at intervals of 3-6 months. Each wave was composed of equal numbers of male and female mice. All mice were fed a high-fat high-sucrose diet (TD.08811, Envigo Teklad, 44.6% kcal fat, 34% carbohydrate, and 17.3% protein) *ad libitum* upon arrival to the facility. Mice were kept in the same vivarium room and were individually housed to monitor food intake and prevent coprophagy between animals. DO mice were sacrificed at 22-25 weeks of age.

The eight DO founder strains (C57BL/6J, A/J, 129S1/SvImJ, NOD/ShiLtJ, NZO/HILtJ, PWK/PhJ, WSB/EiJ and CAST/EiJ) were obtained from the Jackson Laboratories. Mice were bred at the University of Wisconsin-Madison Biochemistry Department. Mice were housed by strain and sex (2-5 mice/cage), with the exception of CAST that required individual housing. Inbred founder mice were housed under the same environmental conditions as the DO animals. Like the DO mice, the eight founder strains were maintained on the HF/HS diet and were sacrificed at 22 weeks of age, except for NZO males that were sacrificed at 14 weeks, due to high mortality attributable to severe disease.

For both DO and founder mice, fecal samples for 16S rRNA sequencing were collected immediately before sacrifice after a 4 hour fast. Cecal contents, plasma, and additional tissues were harvested promptly after sacrifice and all samples were immediately flash frozen in liquid nitrogen and stored at -80°C until further processing.

DNA extraction. DNA was isolated from feces using a bead-beating protocol {Turnbaugh:2009ei}. Mouse feces (~1 pellet per animal) were re-suspended in a solution

367 containing 500µl of extraction buffer [200mM Tris (pH 8.0), 200mM NaCl, 20mM EDTA], 210µl
368 of 20% SDS, 500µl phenol:chloroform:isoamyl alcohol (pH 7.9, 25:24:1) and 500µl of 0.1-mm
369 diameter zirconia/silica beads. Cells were mechanically disrupted using a bead beater (BioSpec
370 Products, Barlesville, OK; maximum setting for 3 min at room temperature), followed by
371 extraction with phenol:chloroform:isoamyl alcohol and precipitation with isopropanol.
372 Contaminants were removed using QIAquick 96-well PCR Purification Kit (Qiagen, Germantown,
373 MD, USA). Isolated DNA was eluted in 5 mM Tris/HCL (pH 8.5) and was stored at -80°C until
374 further use.
375
376 **16S rRNA Sequencing.** PCR was performed using universal primers flanking the variable 4 (V4)
377 region of the bacterial 16S rRNA gene (102). Genomic DNA samples were amplified in duplicate.
378 Each reaction contained 10-30 ng genomic DNA, 10 µM each primer, 12.5 µl 2x HiFi HotStart
379 ReadyMix (KAPA Biosystems, Wilmington, MA, USA), and water to a final reaction volume of
380 25 µl. PCR was carried out under the following conditions: initial denaturation for 3 min at 95°C,
381 followed by 25 cycles of denaturation for 30 s at 95°C, annealing for 30 s at 55°C and elongation
382 for 30 s at 72°C, and a final elongation step for 5 min at 72°C. PCR products were purified with
383 the QIAquick 96-well PCR Purification Kit (Qiagen, Germantown, MD, USA) and quantified
384 using Qubit dsDNA HS Assay kit (Invitrogen, Oregon, USA). Samples were equimolar pooled
385 and sequenced by the University of Wisconsin – Madison Biotechnology Center with the MiSeq
386 2x250 v2 kit (Illumina, San Diego, CA, USA) using custom sequencing primers.
387
388 **16S analysis.** Demultiplexed paired end fastq files generated by CASAVA (Illumina) and a
389 mapping file were used as input files. Sequences were processed, quality filtered and analyzed

with QIIME2 (version 2018.4) (https://qiime2.org), a plugin-based microbiome analysis platform (103). DADA2 (47) was used to denoise sequencing reads with the q2-dada2 plugin for quality filtering and identification of *de novo* exact sequence variants (ESVs) (i.e. 100% exact sequence match). This resulted in 20,831,573 total sequences with an average of 52,078 sequences per sample for the DO mice, and 2,128,796 total sequences with an average of 34,335.4 sequences per sample for the eight DO founder strains. Sequence variants were aligned with mafft (104) with the q2-alignment plugin. The q2-phylogeny plugin was used for phylogenetic reconstruction via FastTree (105). Taxonomic classification was assigned using classify-sklearn (106) against the Greengenes 13_8 99% reference sequences (107). Alpha- and beta-diversity (weighted and unweighted UniFrac (108) analyses were performed using q2-diversity plugin at a rarefaction depth of 10000 sequences per sample. For the DO mice, one sample (DO071) was removed from subsequent analysis because it did not reach this sequencing depth. For analysis of the eight DO founder strains, one sample (NOD5) was removed because it did not reach this sequencing depth. Subsequent processing and analysis were performed in R (v.3.5.1), and data generated in QIIME2 was imported into R using Phyloseq (109). Sequencing data was normalized by cumulative sum scaling (CSS) using MetagenomeSeq (110). Summaries of the taxonomic distributions were generated by collapsing normalized ESV counts into higher taxonomic levels (genus to phylum) by phylogeny. We defined a core measurable microbiota (CMM) (24) to include only microbial traits present in 20% of individuals in the QTL mapping. In total, 86 ESVs and 42 collapsed microbial taxonomies comprised the CMM.

Sample preparation for plasma bile acid analysis. 40 µL of DO plasma collected at sacrifice (30 µL used for founder strains) were aliquoted into a tube with 10 µL SPLASH Lipidomix internal

standard mixture (Avanti Polar Lipids, Inc.). Protein was precipitated by addition of 215 µL MeOH. After the mixture was vortexed for 10 s, 750 µL methyl tert-butyl ether (MTBE) were added as extraction solvent and the mixture was vortexed for 10 s and mixed on an orbital shaker for 6 min. Phase separation was induced by adding 187.5 µL of water followed by 20 s of vortexing. All steps were performed at 4 °C on ice. Finally, the mixture was centrifuged for 4 min at 14,000 x g at 4 °C and stored at -80 °C. For targeted bile acids analysis, samples were thawed on ice. 400 µL of ethanol were added to further precipitate protein, as well as 15 µL of isotope-labeled internal standard mix (12.5 µM d4-TαMCA, 10 µM d4-CDCA). The samples were vortexed for 20 s and centrifuged for 4 min at 14,000 g at 4 °C after which the supernatant (ca. 1000 µL) was taken out and dried down. Dried supernatants were resuspended in 60 µL mobile phase (50 %B), vortexed for 20 s, centrifuged for 4 min at 14,000 g and then 50 µL were transferred to vials with glass inserts for MS analysis.

Sample preparation for cecal bile acid analysis. 30 ± 7.5 mg cecal contents along with 10 µL SPLASH Lipidomix internal standard mixture were aliquoted into a tube with a metal bead and 270 µL MeOH were added for protein precipitation. To each tube, 900 µL MTBE and 225 µL of water were added as extraction solvents. All steps were performed at 4 °C on ice. The mixture was homogenized by bead beating for 8 min at 25 Hz. Finally, the mixture was centrifuged for 4-8 min at 11,000 x g at 4 °C. Subsequent processing for the DO mice and eight DO founder strains differed due to other analyses performed on the samples that are not presented in this paper. For DO samples, 100 µL of the aqueous and 720 µL of organic layer were combined and stored at -80 °C. For analysis, these were thawed on ice and 400 µL of ethanol were added to further precipitate protein, as well as 15 µL of isotope-labeled internal standard mix (12.5 µM d4-TαMCA, 10 µM

436 d4-CDCA). The samples were vortexed for 20 s and centrifuged for 4 min at 14,000 g at 4 °C after
437 which the supernatant (ca. 1000 µL) was taken out and dried down. Dried supernatants were
438 resuspended in 100 µL mobile phase (50 %B), vortexed for 20 s, centrifuged for 8 min at 14,000
439 g and then 50 µL were transferred to vials with glass inserts for MS analysis. For the eight DO
440 founder strains, the mixture was dried down including all solid parts and stored dried at -80 °C.
441 For targeted bile acid analysis, these dried down samples were then thawed on ice and reconstituted
442 in 270 µL of methanol, 900 µL of MTBE, and 225 µL of water. 400 µL of ethanol were added to
443 further precipitate protein, as well as 15 µL of isotope-labeled internal standard mix (12.5 µM d4-
444 TαMCA, 10 µM d4-CDCA). The mixture was bead beat for 8 min at 25 Hz and centrifuged at
445 14,000 g for 8 minutes after which the supernatant (ca. 1500 µL) was taken out and dried down.
446 Dried supernatants were resuspended in 100 µL mobile phase (50 %B), vortexed for 20 s,
447 centrifuged for 4 min at 14,000 g and then 90 µL were transferred to vials with glass inserts for
448 MS analysis.
449
450 **Measurement and analysis of mouse bile acids.** LC-MS analysis was performed in randomized
451 order using an Acquity CSH C18 column held at 50 °C (100 mm × 2.1 mm × 1.7 µm particle size;
452 Waters) connected to an Ultimate 3000 Binary Pump (400 µL/min flow rate; Thermo Scientific).
453 Mobile phase A consisted of 10 mM ammonium acetate containing 1 mL/L ammonium hydroxide.
454 Mobile phase B consisted of MeOH with the same additives (111). Mobile phase B was initially
455 held at 50% for 1.5 min and then increased to 70% over 13.5 min. Mobile phase B was further
456 increased to 99% over 0.5 min and held for 2.5 min. The column was re-equilibrated for 5.5 min
457 before the next injection. Twenty microliters of plasma sample or ten microliters of cecum sample
458 were injected by an Ultimate 3000 autosampler (Thermo Scientific). The LC system was coupled

459 to a TSQ Quantiva Triple Quadrupole mass spectrometer (Thermo Scientific) by a heated ESI
460 source kept at 325°C (Thermo Scientific). The inlet capillary was kept at 350 °C, sheath gas was
461 set to 15 units, auxiliary gas to 10 units, and the negative spray voltage was set to 2,500 V. For
462 targeted analysis the MS was operated in negative single reaction monitoring (SRM) mode
463 acquiring scheduled, targeted scans to quantify selected bile acid transitions, with two transitions
464 for each species' precursor and 3 min retention time windows. Collision energies were optimized
465 for each species and ranging from 20-55 V. Due to insufficient fragmentation for unconjugated
466 bile acids, the precursor was monitored as one transition with a CE of 20 V. MS acquisition
467 parameters were 0.7 FWHM resolution for Q1 and Q3, 1 s cycle time, 1.5 mTorr CID gas and 3 s
468 Chrom filter. In total, 27 bile acids, including 14 unconjugated, 9 tauro- and 4 glycine-conjugated
469 species, were measured. The resulting bile acid data were processed using Skyline 3.6.0.10493
470 (University of Washington). For each species, one transition was picked for quantitation, while the
471 other was used for retention time confirmation. Normalization of the quantitative data was
472 performed to the internal standard d4-CDCA as indicated in Equation 1.

473 ***Equation 1***: (Peak Area / d4-CDCA Peak Area) · Average of d4-CDCA Peak Area

474

475 **Genotyping.** Genotyping was performed on tail biopsies as previously described (42) using the
476 Mouse Universal Genotyping Array (GigaMUGA) [143,259 markers] (112) at Neogen (Lincoln,
477 NE). Genotypes were converted to founder strain-haplotype reconstructions using a hidden
478 Markov model (HMM) implemented in the R/qtl2 package (48). We interpolated the GigaMUGA
479 markers onto an evenly spaced grid with 0.02-cM spacing and added markers to fill in regions with
480 sparse physical representation, which resulted in 69,005 pseudomarkers.

481

QTL mapping. We performed QTL mapping using the R package R/qtl2 (48). QTL mapping was done through a regression of the phenotype on the founder haplotype probabilities estimated with an HMM designed for multi-parental populations. Genome scans were performed for each phenotype with sex, cohort (wave), and days on diet included as additive covariates. Genetic similarity between mice was accounted for using a kinship matrix based on the leave-one-chromosome-out (LOCO) methods (113). For microbial QTL mapping, normalized gut microbiota abundance data transformed to normal quantiles. For bile acid QTL mapping, normalized plasma and cecal bile acid levels were log2 transformed. The mapping statistic reported is log of the odds ratio (LOD). The significance thresholds were determined by performing 1000 permutations of genome-wide scans by shuffling phenotypic data in relation to individual genotypes. Significant QTL were determined at a genome-wide *P*-value of < 0.05 and the QTL support interval was defined using the 95% Bayesian confidence interval.

Mediation/Pleiotropy analysis. To assess whether two co-mapping traits were caused by a pleiotropic locus, we used a likelihood ratio test implemented with the open source R package R/qtl2pleio (70). Here, we compared the alternative hypothesis of two distinct loci with the null hypothesis of pleiotropy for two traits that map to the same genetic region. Parametric bootstrapping was used to determine statistical significance. Mediation analysis was applied to identify whether a microbe or bile acid were likely to be a causal mediator of the QTL as presented in Li et al. (114). This analysis was adapted from a general approach previously described to differentiate target from mediator variables (115). The effect of a mediator on a target was evaluated by performing an allele scan or SNP scan using the target adjusted by mediator. Only individuals with both values for both traits were considered for mediation analysis. Traits with a

505 LOD drop >2 after controlling for the mediator were considered for further causality testing. To
506 statistically assess causality between microbial and bile acid trait sets (causal, reactive,
507 independent, undecided), a causal model selection test (72) was applied using the R packages
508 R/intermediate and R/qtl2. Causal model selection tests were evaluated on both alleles and SNPs
509 in peak region.
510
511 **RNA extraction.** Total RNA was extracted from flash-frozen distal ileum tissues by TRIzol
512 extraction and further cleaned using the RNeasy Mini Kit (Qiagen, Germantown, MD, USA). DNA
513 was removed by on-column DNase digestion (Qiagen). Purified RNA was quantified using a
514 Nanodrop 2000 spectrophotometer.
515
516 **Quantitative Real-Time PCR.** SuperScript II Reverse Transcriptase with oligo(dT) primer (all
517 from Invitrogen, Carlsbad, CA, USA) was used to synthesize 20 µl cDNA templates from 1 µg
518 purified RNA. cDNA was diluted 2X before use and qRT-PCR reactions were prepared in a 10 µl
519 volume using SsoAdvanced Universal SYBR Green Supermix (Bio-Rad, Hercules, CA, USA) and
520 400 nM specific primers targeting the gene of interest (SLC10A2-F [5'-
521 TGGGTTTCTTCCTGGCTAGACT-3']; SLC10A2-R [5'- TGTTCTGCATTCCAGTTTCCAA-
522 3'] (116)). All reactions were performed in triplicate. Reactions were run on a CFX96 Real-Time
523 PCR System (Bio-Rad, Hercules, CA, USA). The $2^{-\Delta\Delta Ct}$ method (117) was used to calculate
524 relative changes in gene expression and all results were normalized to GAPDH.
525
526 **Bacterial culturing.** Bacterial strains were obtained from DSMZ and ATCC. All strains were
527 cultured at 37°C under anaerobic conditions using an anaerobic chamber (Coy Laboratory

Products) with a gas mix of 5% hydrogen, 20% carbon dioxide and 75% nitrogen. Strains were grown in rich medium (S5 Table) that was filter sterilized and stored in the anaerobic chamber at least 24 hours prior to use. *L. reuteri* was grown in medium supplemented with 20 mM glucose. For all *in vitro* assays, cultures used for inoculation were grown overnight at 37°C in 10 mL 14b medium in anaerobic Hungate tubes. Stock solutions of conjugated bile acids (TCA, GCDCA) and unconjugated bile acids (CA, CDCA, DCA) were prepared to a final concentration of 100 mM and used for all *in vitro* assays. All bile acids used were soluble in methanol.

Microbial bile acid metabolism screen. Stock solutions of conjugated and unconjugated bile acids (100 mM) were added to 3 ml 14b medium to obtain a final concentration of 100 μM total bile acid. Tubes were inoculated with a *T. sanguinis* cultured overnight, then incubated in the anaerobic chamber at 37°C for 48 hours. At the 24- and 48-hour timepoints, 1 mL of each culture was removed and the supernatant was collected after brief centrifugation. Each culture supernant was diluted 10x in initial running solvent (30:70 MeOH:10 mM ammonium acetate). Samples were spun at max speed for 3 minutes to remove suspended particles prior to loading on the uHPLC. Samples were analyzed using a uHPLC coupled with a high-resolution mass spectrometer.

Microbial bile acid screen uHPLC-MS/MS parameters. 10 μL aliquots of diluted supernatant samples were analyzed using a uHPLC-MS/MS system consisting of a Vanquish uHPLC coupled by electrospray ionization (ESI) (negative mode) to a hybrid quadrupole-high-resolution mass spectrometer (Q Exactive Orbitrap; Thermo Scientific). Liquid chromatography separation was achieved on an Acquity UPLC BEH C_{18} column (2.1-by 100-mm column, 1.7-μm particle size) heated to 50˚C. Solvent A was 10 mM Ammonium acetate, pH 6; solvent B was 100% methanol.

551 The total run time was 31.5 minutes with the following gradient: 0 min, 30% B; 0.5 min, 30% B;
552 24 min, 100% B; 29 min, 100% B; 29 min, 30% B; 31.5 min, 30% B. Bile acid peaks were
553 identified using the Metabolomics Analysis and Visualization Engine (MAVEN) (118).
554
555 **Growth curves**. Bacterial growth rate was measured in medium 14b supplemented with either 100
556 µM, 300 µM, 1 mM bile acids or methanol control. Medium was dispensed inside an anerobic
557 chamber into Hungate tubes. Tubes containing 10 mL of medium were inoculated with 30 µL of
558 an overnight culture and incubated at 37°C for 24 hours. *T. sanguinis* was grown with shaking to
559 disrupt the formation of flocculent colonies. Growth was monitored as the increase in absorbance
560 at 600 nm in a Spectronic 20D+ spectrophotometer (Thermo Scientific, Waltham, MA, USA).
561 Growth rate was determined as $\mu = \ln(X/X_o)/T$, where X is the OD_{600} value during the linear portion
562 of growth and T is time in hours. Values given are the mean µ values from two independent cultures
563 done in triplicate.
564
565 **Statistical analysis.** All statistical analyses were performed in R (v.3.5.1) (119). Unless otherwise
566 indicated in the figure legends, differences between groups were evaluated using unpaired two-
567 tailed Welch's t-test. For multiple comparisons, Krustkal-Wallis test was used if ANOVA
568 conditions were not met, followed by Mann-Whitney/Wilcoxon rank-sum for multiple
569 comparisons and adjusted for multiple testing using the Benjamini-Hochberg FDR procedure. The
570 correlation between the abundance of microbial taxa was performed using Spearman's correlation
571 in the "Hmisc" (v.4.1-1) R package (120). The p-values were adjusted using the Benjamini and
572 Hochberg method, and correlation coefficients were visualized using the "pheatmap" (v.1.0.10)
573 (121). Multiple groups were compared by Kruskal-Wallis test and adjusted for multiple testing

574 using the Benjamini-Hochberg FDR procedure. Significance was determined as p-value < 0.05.
575 To assess magnitude of variability of the CMMs, summary statistics were calculated on each CMM
576 (taxa and ESVs). Non-parametric-based PERMANOVA statistical test (122) with 999 Monte
577 Carlo permutations was used to compare microbiota compositions among groups using the Vegan
578 R package (123).
579

580 ## Acknowledgements

581 The authors thank the University of Wisconsin Biotechnology Center DNA Sequencing
582 Facility for providing sequencing and support services, and the University of Wisconsin Center
583 for High Throughput Computing (CHTC) in the Department of Computer Sciences for providing
584 computational resources, support, and assistance. We also thank Paul Dawson for his feedback.
585

586 ## References

587 1. Clemente JC, Ursell LK, Parfrey LW, Knight R. The impact of the gut microbiota on human
588 health: an integrative view. Cell. 2012 Mar 16;148(6):1258–70.
589 2. Le Chatelier E, Nielsen T, Qin J, Prifti E, Hildebrand F, Falony G, et al. Richness of human
590 gut microbiome correlates with metabolic markers. Nature. 2013 Aug 29;500(7464):541–
591 6.
592 3. Sommer F, Bäckhed F. The gut microbiota--masters of host development and physiology.
593 Nat Rev Microbiol. 2013 Apr 1;11(4):227–38.
594 4. Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R. Diversity, stability and
595 resilience of the human gut microbiota. Nature. 2012 Sep 13;489(7415):220–30.
596 5. Zhernakova A, Kurilshikov A, Bonder MJ, Tigchelaar EF, Schirmer M, Vatanen T, et al.

Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. Science. 2016 Apr 29;352(6285):565–9.

6. Rothschild D, Weissbrod O, Barkan E, Kurilshikov A, Korem T, Zeevi D, et al. Environment dominates over host genetics in shaping human gut microbiota. Nature. 2018 Mar 8;555(7695):210–5.
7. Dill-McFarland K, Tang Z-Z, Kemis J, Kerby R, Chen G, Palloni A, et al. Social relationships, social isolation, and the human gut microbiota. bioRxiv. 2018 Sep 27;428938.
8. Ussar S, Fujisaka S, Kahn CR. Interactions between host genetics and gut microbiome in diabetes and metabolic syndrome. Mol Metab. 2016 Sep 1;5(9):795–803.
9. Hall AB, Tolonen AC, Xavier RJ. Human genetic variation and the gut microbiome in disease. Nat Rev Genet. 2017 Nov 1;18(11):690–9.
10. Karlsson F, Tremaroli V, Nielsen J, Backhed F. Assessing the Human Gut Microbiota in Metabolic Diseases. Diabetes. 2013 Sep 24;62(10):3341–9.
11. Petersen C, Round JL. Defining dysbiosis and its influence on host immunity and disease. Cell Microbiol. 2014 Jul;16(7):1024–33.
12. Petra AI, Panagiotidou S, Hatziagelaki E, Stewart JM, Conti P, Theoharides TC. Gut-Microbiota-Brain Axis and Its Effect on Neuropsychiatric Disorders With Suspected Immune Dysregulation. Clin Ther. 2015 May 1;37(5):984–95.
13. Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R, et al. Human genetics shape the gut microbiome. Cell. 2014 Nov 6;159(4):789–99.
14. Plovier H, Everard A, Druart C, Depommier C, Van Hul M, Geurts L, et al. A purified membrane protein from Akkermansia muciniphila or the pasteurized bacterium improves metabolism in obese and diabetic mice. Nat Med. 2017 Jan 1;23(1):107–13.

15. Kasahara K, Krautkramer KA, Org E, Romano KA, Kerby RL, Vivas EI, et al. Interactions between Roseburia intestinalis and diet modulate atherogenesis in a murine model. Nat Microbiol. 2018 Nov 5;1.
16. Kurilshikov A, Wijmenga C, Fu J, Zhernakova A. Host Genetics and Gut Microbiome: Challenges and Perspectives. Trends Immunol. 2017 Sep 1;38(9):633–47.
17. Parks BW, Nam E, Org E, Kostem E, Norheim F, Hui ST, et al. Genetic control of obesity and gut microbiota composition in response to high-fat, high-sucrose diet in mice. Cell Metab. 2013 Jan 8;17(1):141–52.
18. Kreznar JH, Keller MP, Traeger LL, Rabaglia ME, Schueler KL, Stapleton DS, et al. Host Genotype and Gut Microbiome Modulate Insulin Secretion and Diet-Induced Metabolic Phenotypes. Cell Rep. 2017 Feb 14;18(7):1739–50.
19. Vijay-Kumar M, Aitken JD, Carvalho FA, Cullender TC, Mwangi S, Srinivasan S, et al. Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5. Science. 2010 Apr 9;328(5975):228–31.
20. Henao-Mejia J, Elinav E, Jin C, Hao L, Mehal WZ, Strowig T, et al. Inflammasome-mediated dysbiosis regulates progression of NAFLD and obesity. Nature. 2012 Feb 1;482(7384):179–85.
21. Rehman A, Sina C, Gavrilova O, Hasler R, Ott S, Baines JF, et al. Nod2 is essential for temporal development of intestinal microbial communities. Gut. 2011 Oct 1;60(10):1354–62.
22. Lamas B, Richard ML, Leducq V, Pham H-P, Michel M-L, Da Costa G, et al. CARD9 impacts colitis by altering gut microbiota metabolism of tryptophan into aryl hydrocarbon receptor ligands. Nat Med. 2016 Jun 9;22(6):598–605.

643 23. Leamy LJ, Kelly SA, Nietfeldt J, Legge RM, Ma F, Hua K, et al. Host genetics and diet,
644 but not immunoglobulin A expression, converge to shape compositional features of the gut
645 microbiome in an advanced intercross population of mice. Genome Biol. 2014 Jan
646 1;15(12):552.
647 24. Benson AK, Kelly SA, Legge R, Ma F, Low SJ, Kim J, et al. Individuality in gut microbiota
648 composition is a complex polygenic trait shaped by multiple environmental and host genetic
649 factors. Proc Natl Acad Sci U S A. 2010 Nov 2;107(44):18933–8.
650 25. McKnite AM, Perez-Munoz ME, Lu L, Williams EG, Brewer S, Andreux PA, et al. Murine
651 gut microbiota is defined by host genetics and modulates variation of metabolic traits. White
652 BA, editor. PLoS One. 2012 Jan 1;7(6):e39191.
653 26. Belheouane M, Gupta Y, Künzel S, Ibrahim S, Baines JF. Improved detection of gene-
654 microbe interactions in the mouse skin microbiota using high-resolution QTL mapping of
655 16S rRNA transcripts. Microbiome. 2017 Jun 6;5(1):59.
656 27. Goodrich JK, Davenport ER, Beaumont M, Jackson MA, Knight R, Ober C, et al. Genetic
657 Determinants of the Gut Microbiome in UK Twins. Cell Host Microbe. 2016 May
658 11;19(5):731–43.
659 28. Blekhman R, Goodrich JK, Huang K, Sun Q, Bukowski R, Bell JT, et al. Host genetic
660 variation impacts microbiome composition across human body sites. Genome Biol. 2015
661 Sep 15;16(1):191.
662 29. Bonder MJ, Kurilshikov A, Tigchelaar EF, Mujagic Z, Imhann F, Vila AV, et al. The effect
663 of host genetics on the gut microbiome. Nat Genet. 2016 Nov 1;48(11):1407–12.
664 30. Wang Z, Klipfell E, Bennett BJ, Koeth R, Levison BS, Dugar B, et al. Gut flora metabolism
665 of phosphatidylcholine promotes cardiovascular disease. Nature. 2011 Apr

7;472(7341):57–63.

31. Herrema H, IJzerman RG, Nieuwdorp M. Emerging role of intestinal microbiota and microbial metabolites in metabolic control. Diabetologia. 2017 Apr 1;60(4):613–7.
32. Krautkramer KA, Kreznar JH, Romano KA, Vivas EI, Barrett-Wilt GA, Rabaglia ME, et al. Diet-Microbiota Interactions Mediate Global Epigenetic Programming in Multiple Host Tissues. Mol Cell. 2016 Dec 1;64(5):982–92.
33. Ridlon JM, Harris SC, Bhowmik S, Kang D-J, Hylemon PB. Consequences of bile salt biotransformations by intestinal bacteria. Gut Microbes. 2016 Jan 1;7(1):22–39.
34. Romano KA, Martinez-Del Campo A, Kasahara K, Chittim CL, Vivas EI, Amador-Noguez D, et al. Metabolic, Epigenetic, and Transgenerational Effects of Gut Bacterial Choline Consumption. Cell Host Microbe. 2017 Sep 13;22(3):279–290.e7.
35. Ridlon JM, Kang D-J, Hylemon PB. Bile salt biotransformations by human intestinal bacteria. J Lipid Res. 2006 Feb 1;47(2):241–59.
36. Wahlström A, Sayin SI, Marschall H-U, Bäckhed F. Intestinal Crosstalk between Bile Acids and Microbiota and Its Impact on Host Metabolism. Cell Metab. 2016 Jun 15;24(1):41–50.
37. Kuipers F, Bloks VW, Groen AK. Beyond intestinal soap--bile acids in metabolic control. Nat Rev Endocrinol. 2014 Aug 1;10(8):488–98.
38. Islam KBMS, Fukiya S, Hagio M, Fujii N, Ishizuka S, Ooka T, et al. Bile acid is a host factor that regulates the composition of the cecal microbiota in rats. Gastroenterology. 2011 Nov 1;141(5):1773–81.
39. Zheng X, Huang F, Zhao A, Lei S, Zhang Y, Xie G, et al. Bile acid is a significant host factor shaping the gut microbiome of diet-induced obese mice. BMC Biol. 2017 Dec 14;15(1):120.

40. Begley M, Gahan CGM, Hill C. The interaction between bacteria and bile. FEMS Microbiol Rev. 2005 Sep 1;29(4):625–51.

41. Sayin SI, Wahlström A, Felin J, Jäntti S, Marschall H-U, Bamberg K, et al. Gut microbiota regulates bile acid metabolism by reducing the levels of tauro-beta-muricholic acid, a naturally occurring FXR antagonist. Cell Metab. 2013 Feb 5;17(2):225–35.

42. Svenson KL, Gatti DM, Valdar W, Welsh CE, Cheng R, Chesler EJ, et al. High-resolution genetic mapping using the Mouse Diversity outbred population. Genetics. 2012 Feb 1;190(2):437–47.

43. Churchill GA, Gatti DM, Munger SC, Svenson KL. The Diversity Outbred mouse population. Mamm Genome. 2012 Oct 1;23(9–10):713–8.

44. Kovacs A, Ben-Jacob N, Tayem H, Halperin E, Iraqi FA, Gophna U. Genotype is a stronger determinant than sex of the mouse gut microbiota. Microb Ecol. 2011 Feb 1;61(2):423–8.

45. O'Connor A, Quizon PM, Albright JE, Lin FT, Bennett BJ. Responsiveness of cardiometabolic-related microbiota to diet is influenced by host genetics. Mamm Genome. 2014 Dec 1;25(11–12):583–99.

46. Keller MP, Gatti DM, Schueler KL, Rabaglia ME, Stapleton DS, Simecek P, et al. Genetic Drivers of Pancreatic Islet Function. Genetics. 2018 May 1;209(1):335–56.

47. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016 Jul 1;13(7):581–3.

48. Broman KW, Gatti DM, Simecek P, Furlotte NA, Prins P, Sen Ś, et al. R/qtl2: Software for Mapping Quantitative Trait Loci with High-Dimensional Data and Multi-parent Populations. Genetics. 2018 Dec 27;genetics.301595.2018.

712 49. Org E, Parks BW, Joo JWJ, Emert B, Schwartzman W, Kang EY, et al. Genetic and
713 environmental control of host-gut microbiota interactions. Genome Res. 2015 Oct
714 1;25(10):1558–69.
715 50. Davenport ER, Cusanovich DA, Michelini K, Barreiro LB, Ober C, Gilad Y. Genome-Wide
716 Association Studies of the Human Gut Microbiota. White BA, editor. PLoS One. 2015 Jan
717 1;10(11):e0140301.
718 51. de Aguiar Vallim TQ, Tarling EJ, Edwards PA. Pleiotropic roles of bile acids in
719 metabolism. Cell Metab. 2013 May 7;17(5):657–69.
720 52. Russell DW. The enzymes, regulation, and genetics of bile acid synthesis. Annu Rev
721 Biochem. 2003 Jan 1;72(1):137–74.
722 53. Martinot E, Sèdes L, Baptissart M, Lobaccaro J-M, Caira F, Beaudoin C, et al. Bile acids
723 and their receptors. Mol Aspects Med. 2017 Aug 1;56:2–9.
724 54. Inoue K, Zhuang L, Maddox DM, Smith SB, Ganapathy V. Structure, function, and
725 expression pattern of a novel sodium-coupled citrate transporter (NaCT) cloned from
726 mammalian brain. J Biol Chem. 2002 Oct 18;277(42):39469–76.
727 55. Pesta DH, Perry RJ, Guebre-Egziabher F, Zhang D, Jurczak M, Fischer-Rosinsky A, et al.
728 Prevention of diet-induced hepatic steatosis and hepatic insulin resistance by second
729 generation antisense oligonucleotides targeted to the longevity gene mIndy (Slc13a5).
730 Aging (Albany NY). 2015 Dec 1;7(12):1086–93.
731 56. Birkenfeld AL, Lee H-Y, Guebre-Egziabher F, Alves TC, Jurczak MJ, Jornayvaz FR, et al.
732 Deletion of the mammalian INDY homolog mimics aspects of dietary restriction and
733 protects against adiposity and insulin resistance in mice. Cell Metab. 2011 Aug
734 3;14(2):184–95.

735 57. von Loeffelholz C, Lieske S, Neuschäfer-Rube F, Willmes DM, Raschzok N, Sauer IM, et
736 al. The human longevity gene homolog INDY and interleukin-6 interact in hepatic lipid
737 metabolism. Hepatology. 2017 Aug 1;66(2):616–30.
738 58. Li L, Li H, Garzel B, Yang H, Sueyoshi T, Li Q, et al. SLC13A5 is a novel transcriptional
739 target of the pregnane X receptor and sensitizes drug-induced steatosis in human liver. Mol
740 Pharmacol. 2015 Apr 1;87(4):674–82.
741 59. Staudinger JL, Goodwin B, Jones SA, Hawkins-Brown D, MacKenzie KI, LaTour A, et al.
742 The nuclear receptor PXR is a lithocholic acid sensor that protects against liver toxicity.
743 Proc Natl Acad Sci U S A. 2001 Mar 13;98(6):3369–74.
744 60. Civelek M, Lusis AJ. Systems genetics approaches to understand complex traits. Nat Rev
745 Genet. 2014 Jan 1;15(1):34–48.
746 61. Sehayek E, Hagey LR, Fung Y-Y, Duncan EM, Yu HJ, Eggertsen G, et al. Two loci on
747 chromosome 9 control bile acid composition: evidence that a strong candidate gene,
748 Cyp8b1, is not the culprit. J Lipid Res. 2006 Sep 1;47(9):2020–7.
749 62. Theriot CM, Bowman AA, Young VB. Antibiotic-Induced Alterations of the Gut
750 Microbiota Alter Secondary Bile Acid Production and Allow for Clostridium difficile Spore
751 Germination and Outgrowth in the Large Intestine. Ellermeier CD, editor. mSphere. 2016
752 Jan 1;1(1):e00045-15.
753 63. Dawson PA, Haywood J, Craddock AL, Wilson M, Tietjen M, Kluckman K, et al. Targeted
754 deletion of the ileal bile acid transporter eliminates enterohepatic cycling of bile acids in
755 mice. J Biol Chem. 2003 Sep 5;278(36):33920–7.
756 64. Oelkers P, Kirby LC, Heubi JE, Dawson PA. Primary bile acid malabsorption caused by
757 mutations in the ileal sodium-dependent bile acid transporter gene (SLC10A2). J Clin

758 Invest. 1997 Apr 15;99(8):1880–7.

759 65. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic
760 localization of common disease-associated variation in regulatory DNA. Science. 2012 Sep
761 7;337(6099):1190–5.

762 66. Chen J, Tian W. Explaining the disease phenotype of intergenic SNP through predicted long
763 range regulation. Nucleic Acids Res. 2016 Oct 14;44(18):8641–54.

764 67. Janssen AWF, Dijk W, Boekhorst J, Kuipers F, Groen AK, Lukovac S, et al. ANGPTL4
765 promotes bile acid absorption during taurocholic acid supplementation via a mechanism
766 dependent on the gut microbiota. Biochim Biophys Acta. 2017 Oct 1;1862(10 Pt A):1056–
767 67.

768 68. Out C, Patankar J V, Doktorova M, Boesjes M, Bos T, de Boer S, et al. Gut microbiota
769 inhibit Asbt-dependent intestinal bile acid reabsorption via Gata4. J Hepatol. 2015 Sep
770 1;63(3):697–704.

771 69. Miyata M, Yamakawa H, Hamatsu M, Kuribayashi H, Takamatsu Y, Yamazoe Y.
772 Enterobacteria modulate intestinal bile acid transport and homeostasis through apical
773 sodium-dependent bile acid transporter (SLC10A2) expression. J Pharmacol Exp Ther.
774 2011 Jan 1;336(1):188–96.

775 70. Boehm F. qtl2pleio: Hypothesis test of close linkage vs pleiotropy in multiparental
776 populations. 2018.

777 71. MacKinnon DP, Fairchild AJ, Fritz MS. Mediation analysis. Annu Rev Psychol. 2007 Jan
778 1;58(1):593–614.

779 72. Neto EC, Broman AT, Keller MP, Attie AD, Zhang B, Zhu J, et al. Modeling causality for
780 pairs of phenotypes in system genetics. Genetics. 2013 Mar 1;193(3):1003–13.

781 73. Onishi JC, Campbell S, Moreau M, Patel F, Brooks AI, Zhou YX, et al. Bacterial
782 communities in the small intestine respond differently to those in the caecum and colon in
783 mice fed low- and high-fat diets. Microbiology. 2017 Aug 1;163(8):1189–97.
784 74. Li D, Chen H, Mao B, Yang Q, Zhao J, Gu Z, et al. Microbial Biogeography and Core
785 Microbiota of the Rat Digestive Tract. Sci Rep. 2017 Apr 4;8(1):45840.
786 75. Northfield TC, McColl I. Postprandial concentrations of free and conjugated bile acids
787 down the length of the normal human small intestine. Gut. 1973 Jul 1;14(7):513–8.
788 76. Hamilton JP, Xie G, Raufman J-P, Hogan S, Griffin TL, Packard CA, et al. Human cecal
789 bile acids: concentration and spectrum. Am J Physiol Gastrointest Liver Physiol. 2007 Jul
790 1;293(1):G256-63.
791 77. Falony G, Joossens M, Vieira-Silva S, Wang J, Darzi Y, Faust K, et al. Population-level
792 analysis of gut microbiome variation. Science. 2016 Apr 29;352(6285):560–4.
793 78. Wang J, Thingholm LB, Skiecevičienė J, Rausch P, Kummen M, Hov JR, et al. Genome-
794 wide association analysis identifies variation in vitamin D receptor and other host factors
795 influencing the gut microbiota. Nat Genet. 2016 Nov 1;48(11):1396–406.
796 79. Edenharder R, Pfützner A, Hammann R. Characterization of NAD-dependent 3 alpha- and
797 3 beta-hydroxysteroid dehydrogenase and of NADP-dependent 7 beta-hydroxysteroid
798 dehydrogenase from Peptostreptococcus productus. Biochim Biophys Acta. 1989 Aug
799 8;1004(2):230–8.
800 80. Chen C, Sargent C, Quilter C, Yang Z, Ren J, Affara N, et al. Cloning, mapping and
801 molecular characterization of porcine progesterone receptor membrane component 2
802 (PGRMC2) gene. Genet Mol Biol. 2010 Jul 1;33(3):471–4.
803 81. Gerdes D, Wehling M, Leube B, Falkenstein E. Cloning and tissue expression of two

804 putative steroid membrane receptors. Biol Chem. 1998 Jul 1;379(7):907–11.

805 82. Wendler A, Wehling M. PGRMC2, a yet uncharacterized protein with potential as tumor
806 suppressor, migration inhibitor, and regulator of cytochrome P450 enzyme activity.
807 Steroids. 2013 Jun 1;78(6):555–8.

808 83. Hughes AL, Powell DW, Bard M, Eckstein J, Barbuch R, Link AJ, et al. Dap1/PGRMC1
809 binds and regulates cytochrome P450 enzymes. Cell Metab. 2007 Feb 1;5(2):143–9.

810 84. Pierre JF, Martinez KB, Ye H, Nadimpalli A, Morton TC, Yang J, et al. Activation of bile
811 acid signaling improves metabolic phenotypes in high-fat diet-induced obese mice. Am J
812 Physiol Gastrointest Liver Physiol. 2016 Aug 1;311(2):G286-304.

813 85. Van den Bossche L, Hindryckx P, Devisscher L, Devriese S, Van Welden S, Holvoet T, et
814 al. Ursodeoxycholic Acid and Its Taurine- or Glycine-Conjugated Species Reduce
815 Colitogenic Dysbiosis and Equally Suppress Experimental Colitis in Mice. Elkins CA,
816 editor. Appl Environ Microbiol. 2017 Apr 1;83(7):e02766-16.

817 86. van der Ark KCH, Nugroho ADW, Berton-Carabin C, Wang C, Belzer C, de Vos WM, et
818 al. Encapsulation of the therapeutic microbe Akkermansia muciniphila in a double emulsion
819 enhances survival in simulated gastric conditions. Food Res Int. 2017 Dec 1;102:372–9.

820 87. Derrien M, Vaughan EE, Plugge CM, de Vos WM. Akkermansia muciniphila gen. nov., sp.
821 nov., a human intestinal mucin-degrading bacterium. Int J Syst Evol Microbiol. 2004 Sep
822 1;54(Pt 5):1469–76.

823 88. Belzer C, de Vos WM. Microbes inside--from diversity to function: the case of
824 Akkermansia. ISME J. 2012 Aug 1;6(8):1449–58.

825 89. Shekels LL, Lyftogt CT, Ho SB. Bile acid-induced alterations of mucin production in
826 differentiated human colon cancer cell lines. Int J Biochem Cell Biol. 1996 Feb

1;28(2):193–201.

90. Klinkspoor JH, Mok KS, Van Klinken BJW, Tytgat GNJ, Lee SP, Groen AK. Mucin secretion by the human colon cell line LS174T is regulated by bile salts. Glycobiology. 1999 Jan 1;9(1):13–9.

91. Everard A, Belzer C, Geurts L, Ouwerkerk JP, Druart C, Bindels LB, et al. Cross-talk between Akkermansia muciniphila and intestinal epithelium controls diet-induced obesity. Proc Natl Acad Sci U S A. 2013 May 28;110(22):9066–71.

92. Cani PD, de Vos WM. Next-Generation Beneficial Microbes: The Case of Akkermansia muciniphila. Front Microbiol. 2017 Jan 1;8:1765.

93. Derrien M, Van Baarlen P, Hooiveld G, Norin E, Müller M, de Vos WM. Modulation of Mucosal Immune Response, Tolerance, and Proliferation in Mice Colonized by the Mucin-Degrader Akkermansia muciniphila. Front Microbiol. 2011 Jan 1;2:166.

94. Arakawa R, Bagashev A, Song L, Maurer K, Sullivan KE. Characterization of LRRFIP1. Biochem Cell Biol. 2010 Dec 1;88(6):899–906.

95. Shi L, Song L, Fitzgerald M, Maurer K, Bagashev A, Sullivan KE. Noncoding RNAs and LRRFIP1 regulate TNF expression. J Immunol. 2014 Apr 1;192(7):3057–67.

96. Maravillas-Montero JL, Burkhardt AM, Hevezi PA, Carnevale CD, Smit MJ, Zlotnik A. Cutting edge: GPR35/CXCR8 is the receptor of the mucosal chemokine CXCL17. J Immunol. 2015 Jan 1;194(1):29–33.

97. Su Z, Cox A, Shen Y, Stylianou IM, Paigen B. Farp2 and Stk25 are candidate genes for the HDL cholesterol locus on mouse chromosome 1. Arterioscler Thromb Vasc Biol. 2009 Jan 1;29(1):107–13.

98. Purcell-Huynh DA, Weinreb A, Castellani LW, Mehrabian M, Doolittle MH, Lusis AJ.

Genetic factors in lipoprotein metabolism. Analysis of a genetic cross between inbred mouse strains NZB/BINJ and SM/J using a complete linkage map approach. J Clin Invest. 1995 Oct 1;96(4):1845–58.

99. Ishimori N, Li R, Kelmenson PM, Korstanje R, Walsh KA, Churchill GA, et al. Quantitative trait loci analysis for plasma HDL-cholesterol concentrations and atherosclerosis susceptibility between inbred mouse strains C57BL/6J and 129S1/SvImJ. Arterioscler Thromb Vasc Biol. 2004 Jan 1;24(1):161–6.

100. Su Z, Ishimori N, Chen Y, Leiter EH, Churchill GA, Paigen B, et al. Four additional mouse crosses improve the lipid QTL landscape and identify Lipg as a QTL gene. J Lipid Res. 2009 Oct 1;50(10):2083–94.

101. Fu J, Bonder MJ, Cenit MCMC, Tigchelaar EF, Maatman A, Dekens JAMM, et al. The Gut Microbiome Contributes to a Substantial Proportion of the Variation in Blood Lipids. Circ Res. 2015 Oct 9;117(9):817–24.

102. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Appl Environ Microbiol. 2013 Sep 1;79(17):5112–20.

103. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 2010 May 1;7(5):335–6.

104. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013 Apr 1;30(4):772–80.

105. Price MN, Dehal PS, Arkin AP. FastTree 2--approximately maximum-likelihood trees for

large alignments. Poon AFY, editor. PLoS One. 2010 Mar 10;5(3):e9490.

106. Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, et al. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome. 2018 May 17;6(1):90.

107. McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, et al. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME J. 2012 Mar 1;6(3):610–8.

108. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. Appl Environ Microbiol. 2005 Dec 1;71(12):8228–35.

109. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. Watson M, editor. PLoS One. 2013 Jan 1;8(4):e61217.

110. Paulson JN, Stine OC, Bravo HC, Pop M. Differential abundance analysis for microbial marker-gene surveys. Nat Methods. 2013 Dec 1;10(12):1200–2.

111. Scherer M, Gnewuch C, Schmitz G, Liebisch G. Rapid quantification of bile acids and their conjugates in serum by liquid chromatography-tandem mass spectrometry. J Chromatogr B. 2009;877:3920–5.

112. Morgan AP, Fu C-P, Kao C-Y, Welsh CE, Didion JP, Yadgary L, et al. The Mouse Universal Genotyping Array: From Substrains to Subspecies. G3 (Bethesda). 2015 Dec 18;6(2):263–79.

113. Yang J, Zaitlen NA, Goddard ME, Visscher PM, Price AL. Advantages and pitfalls in the application of mixed-model association methods. Nat Genet. 2014 Feb;46(2):100–6.

114. Li Y, Tesson BM, Churchill GA, Jansen RC. Critical reasoning on causal inference in genome-wide linkage and association studies. Trends Genet. 2010 Dec 1;26(12):493–8.

115. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. J Pers Soc Psychol. 1986 Dec 1;51(6):1173–82.

116. Rao A, Kosters A, Mells JE, Zhang W, Setchell KDR, Amanso AM, et al. Inhibition of ileal bile acid uptake protects against nonalcoholic fatty liver disease in high-fat diet-fed mice. Sci Transl Med. 2016 Sep 21;8(357):357ra122-357ra122.

117. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 2001 Dec 1;25(4):402–8.

118. Clasquin MF, Melamud E, Rabinowitz JD. LC-MS data processing with MAVEN: a metabolomic analysis and visualization engine. Curr Protoc Bioinforma. 2012 Mar 1;Chapter 14(1):Unit14.11-14.11.23.

119. R Core Team. R: A Language and environment for statistical computing. R Foundation for Statistical Computing;

120. Harrell Jr FE, others with contributions from CD and many. Hmisc: Harrell Miscellaneous. 2018.

121. Kolde R. pheatmap: Pretty Heatmaps. 2018.

122. McArdle BH, Anderson MJ. Fitting Multivariate Models to Community Data: A Comment on Distance-Based Redundancy Analysis. Ecology. 2001 Jan 1;82(1):290–7.

123. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. vegan: Community Ecology Package. 2018.

Figure legends

Figure 1. Phenotypic variation among Diversity Outbred (DO) mice fed high-fat and high-sucrose diet. (A) Body weight at 6, 10, 14, and 21-25 (sacrifice) weeks in DO mice fed high-fat and high-sucrose diet (n = 500) (Adapted from Keller et al. (46)) (B) Distributions of the normalized relative abundance of bacterial phyla identified in DO fecal microbiota (n = 399). (C) Abundance (peak area) of primary bile acids detected in plasma and (D) cecal contents (n = 384).

Figure 2. Genetic architecture of quantitative trait loci (QTL) for microbial exact sequence variants (ESVs) and taxa abundance, and plasma and cecal bile acids in 400 Diversity Outbred (DO) mice. The outer layer shows the chromosome location where major tick marks correspond to 25 Mbp. Logarithm of the odds (LOD) range is shown for each track. Each dot represents a QTL on each chromosome of the mouse genome for a given trait. Grey dots denote QTLs with LOD < 5.5. Candidate genes discussed in text are denoted.

Figure 3. Co-mapping of *Turicibacter sp*. and plasma cholic acid (CA) QTL on chromosome 8. Association of (A) fecal abundance of *Turicibacter sp.* and (B) plasma CA levels on chromosome (chr) 8. The x-axis indicates the position in Mbp along chr 8. The y-axis for the top panel and the y-axis in the bottom panel is the LOD score. A/J and WSB founder alleles are associated with higher and lower levels of *Turicibacter* and plasma CA levels, respectively. The estimated founder strain abundance of (C) *Turicibacter* and (D) levels of plasma CA in the DO population reflects measured values observed in founder strains for (E) the abundance of *Turicibacter sp.* and (F) plasma cholic acid levels (n = 8 mice/genotype, 4 male and 4 female). (E) Spearman rank correlation between *Turicibacter sp*. and plasma CA in DO mice (n=192). (F) Spearman rank correlation between Turicibacter genera and plasma cholic acid in DO founder strains (n = 19).

941 (G) SNPs (top panel) and protein coding genes (bottom panel) under the QTL interval. Magenta
942 dots correspond to SNPs with the strongest association where the LOD drop < 1.5 from the top
943 SNP. (H) Relative expression of *Slc10a2* measured in the distal ileum by qRT-PCR in A/J and
944 WSB parental strains (n = 6, 3 male and 3 female). Data are presented as mean ± SEM; Welch's t
945 test; * $p < 0.05$. Correlation p-values adjusted for multiple tests using Benjamini and Hochberg
946 correction. ND – not detected.
947
948 **Figure 4.** Mediation analysis and causal inference testing suggest causal relationship between
949 *Turicibacter sp.* abundance and plasma cholic acid (CA) levels. (A) Hypothetical causal model
950 that proposes that cholic acid (CA) mediates the changes in *Turicibacter sp.* abundance. (B)
951 Change in LOD score of plasma CA when adjusting for *Turicibacter sp.* abundance. The x-axis
952 indicates the position in Mbp along chr 8. (C) Hypothetical causal model that proposes that
953 *Turicibacter sp.* mediates changes in abundance of plasma CA levels. (D) Change in LOD score
954 of *Turicibacter sp.* when controlling for plasma CA levels.
955
956 **Figure 5.** *Turicibacter sanguinis* and bile acid interactions. (A) Percent of conjugated bile acids
957 detected after 24-hour incubation with or without the presence of *T. sanguinis*. (B)
958 Transformation of cholic acid (CA) to 7-dehydrocholic acid (7-dHCA), and (C)
959 chenodeoxycholic acid (CDCA) to 7-ketolithocholic acid (7-KLCA) by *T. sanguinis* after 24
960 hours. Growth of *T. sanguinis* in the presence of 0.1 mM, 0.5 mM, 1 mM and 5 mM (D)
961 conjugated (equimolar pool of taurocholic acid (TCA) and glycochenodeoxycholic acid
962 (GCDCA)), and (E) unconjugated (equimolar pool of cholic CA, CDCA, and deoxycholic acid
963 (DCA)) bile acids over 24 hours. (F) Growth rate (μ) of *T. sanguinis* in medium supplemented

964 with varying concentrations of conjugated and unconjugated bile acids. Data shown are from one
965 experiment with three technical replicates. Data are presented as mean ± SEM; one-way
966 ANOVA followed by Tukey's multiple comparisons test; ** $p < 0.01$, *** $p < 0.001$, **** $p <$
967 0.0001.
968

969 ## Supporting information legends

970 **S1 Figure. Principal coordinate analysis (PCoA) of unweighted UniFrac distances for fecal**
971 **samples.** PCoA shows significant clustering by (A) sex ($F = 5.572$, $p = 0.001$) and (B) wave ($F =$
972 16.954, $p = 0.001$). Clustering by treatment evaluated by PERMANOVA.
973

974 **S2 Figure. Plasma and cecal bile acids group by sex, but not wave.** PCAs of plasma bile acid
975 profiles colored by (A) sex ($p < 0.0001$) and (B) wave ($p = 0.594$), and PCAs of cecal bile acid
976 profiles colored by (C) sex ($p = 0.011$) and (D) wave ($p = 0.207$). Kruskal Wallis one-way test
977 followed by Wilcoxon pair-wise multiple comparisons with Benjamini and Hochberg correction.
978

979 **S3 Figure. Related bile acid species map associate to same locus.** (A) Haplotype effects and
980 LOD scores of plasma taurodeoxycholic acid (TDCA), (B) cecal deoxycholic acid (DCA), (C)
981 cecal isodeoxycholic acid (IDCA) and (D) cecal hyodeoxycholic acid (HDCA). For each plot, the
982 x-axis is the physical position in Mbp along chr 12. The y-axis for the top panel is the effect
983 coefficient depicting the estimated contributions of each founder allele, and the y-axis in the
984 bottom panel is the LOD score. (E) Cecal levels of isolithocholic acid (ILCA) and lithocholic acid
985 (LCA) associate to same locus on chr 11. (F) Estimated founder allele effects for cecal ILCA and

(G) LCA. (H) Genes under cecal LCA and ILCA QTL interval. Dashed lines denote QTL confidence interval.

S4 Figure. Gut associated bacteria have differential growth responses to conjugated bile acids. Growth rate in the presence of 1 mM conjugated bile acids or methanol control for (A) *Bacteroides thetaiotaomicron*, (B) *Clostridium asparagiforme*, (C) *Escherichia coli* MS200-1, and (D) *Lactobacillus reuteri.* Data shown are from duplicate experiments with three technical replicates. Data are presented as mean ± SEM; Welch's *t* test; no significant differences were observed between growth conditions for any of the tested organisms.

S5 Figure. Peptostreptococcaceae and plasma bile acids co-map on chromosome (chr) 3. Haplotype effects and LOD scores of (A) Peptostreptococcaceae family, (B) plasma cholic acid (CA), (C) plasma chenodeoxycholic acid (CDCA), (D) plasma muricholic acid (MCA), (E) plasma ursodeoxycholic acid (UDCA), and (F) plasma 7-dehydrocholic acid (7-dHCA). For each plot, the x-axis is the physical position in Mbp along chr 3. The y-axis for the top panel is the effect coefficient depicting the estimated contributions of each founder allele, and the y-axis in the bottom panel is the LOD score. All overlapping QTL have positive association with the NOD allele. (G) Protein coding genes under QTL interval.

S6 Figure. Exact sequence variant of *Akkermansia muciniphila* and plasma bile acid QTL overlap on chromosome (chr) 1. Haplotype effects and LOD scores of (A) *A. muciniphila* (B) plasma cholic acid (CA), (C) plasma muricholic acid (MCA), and (D) plasma 7-dehydrocholic acid (7-dHCA). For each plot, the x-axis is the physical position in Mbp along chr 1. The y-axis

for the top panel is the effect coefficient depicting the estimated contributions of each founder allele, and the y-axis in the bottom panel is the LOD score. (E) Protein coding genes under 10 Mbp QTL interval. Spearman correlations in the DO mice between *A. muiniphila* and (F) plasma CA, (G) plasma MCA, and (H) plasma 7-dHCA levels. Correlation p-values adjusted for multiple tests using Benjamini and Hochberg correction. Higher levels of these microbial and bile acid traits were associated with the NZO haplotype and lower levels were associated with the 129 haplotype. (E) Protein coding genes under 10 Mbp QTL interval. Dashed lines denote QTL confidence interval. Spearman correlations in the DO mice between *A. muiniphila* and (F) plasma CA, (G) plasma MCA, and (H) plasma 7-dHCA levels. Correlation p-values adjusted for multiple tests using Benjamini and Hochberg correction.

S1 Table. Measures of variability of microbial exact sequence variants (ESVs) or taxon (phylum, class, order, family, genus) in DO mice. Data presented as normalized read counts; n = 399; SD, standard deviation.

S2 Table. Measures of variability of cecal and plasma bile acids in DO mice. Bile acid levels are presented as log2(peak area); n = 384; SD, standard deviation.

S3 Table. Correlations among microbial taxa, bile acid and weight traits. Spearman's rank correlation. Only microbial exact sequence variants, genera and family included in figure. Correlations shown passed FDR < 0.01 cut-off and correlation coefficient either < -0.35 or > 0.35. Correlating bile acids from same tissue removed from table for brevity.

1032 **S4 Table. QTL peaks for gut microbiota, plasma and cecal bile acid, and weight traits in the**

1033 **Diversity Outbred mice.** Only QTL with LOD > 5.5 shown. "Pos" is peak position is Mbp. "ci_lo"

1034 and "ci_hi" correspond to the positions for the 95% bayesian confidence interval.

1035

1036 **S5 Table. Media used for bacterial culture. Medium 14(b) recipe.**

1037

Figure 1

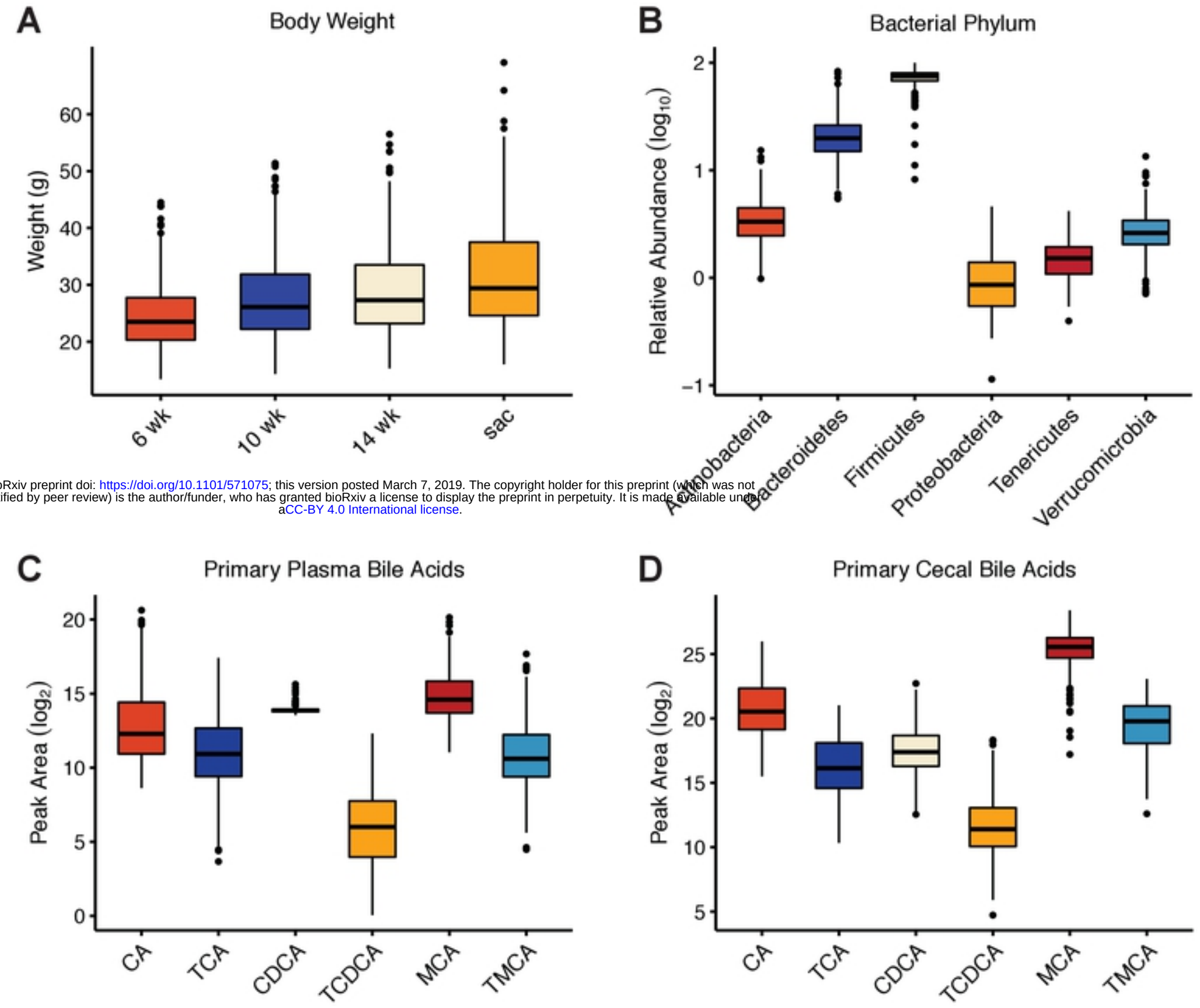

Figure 2

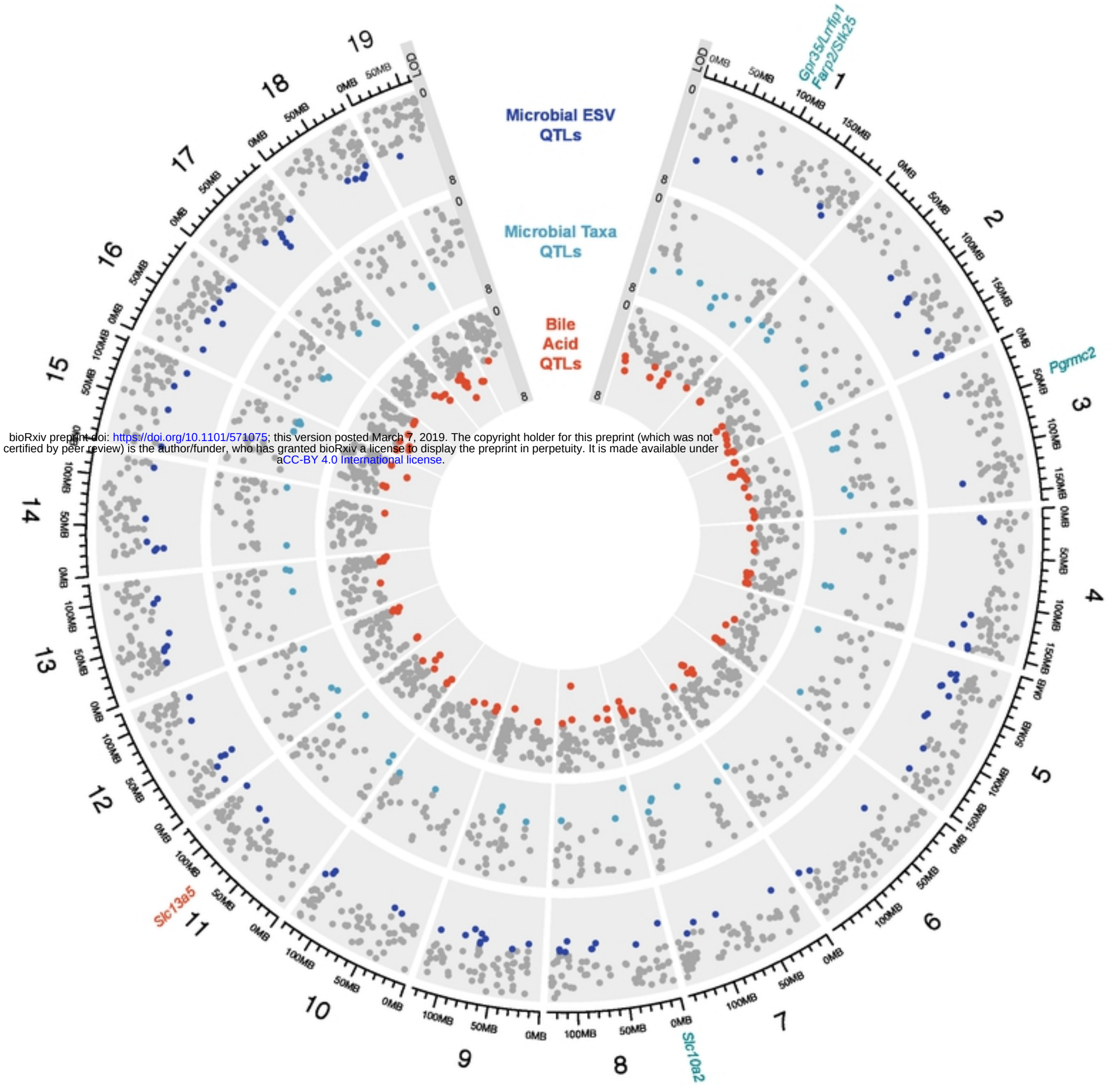

Figure 3

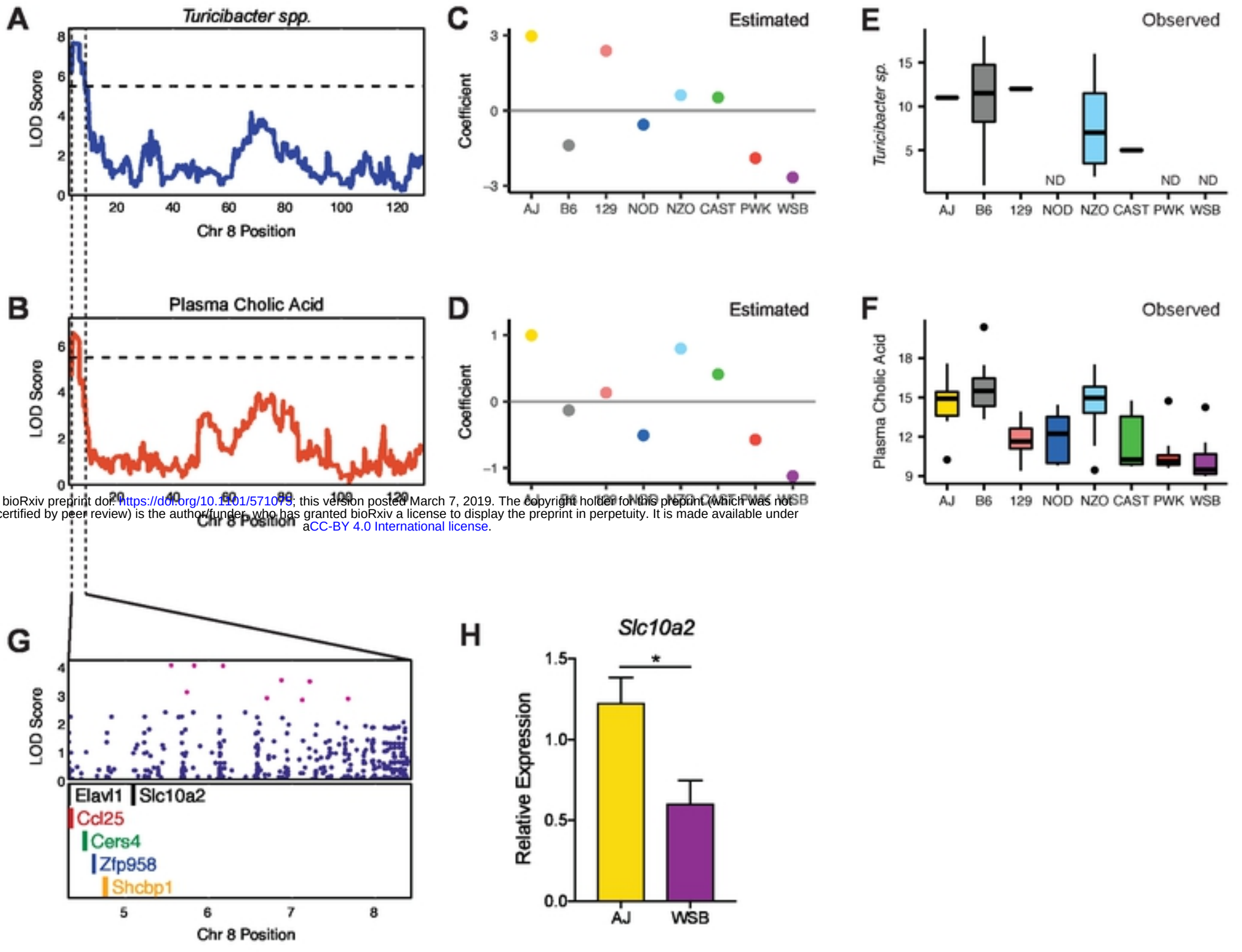

Figure 4

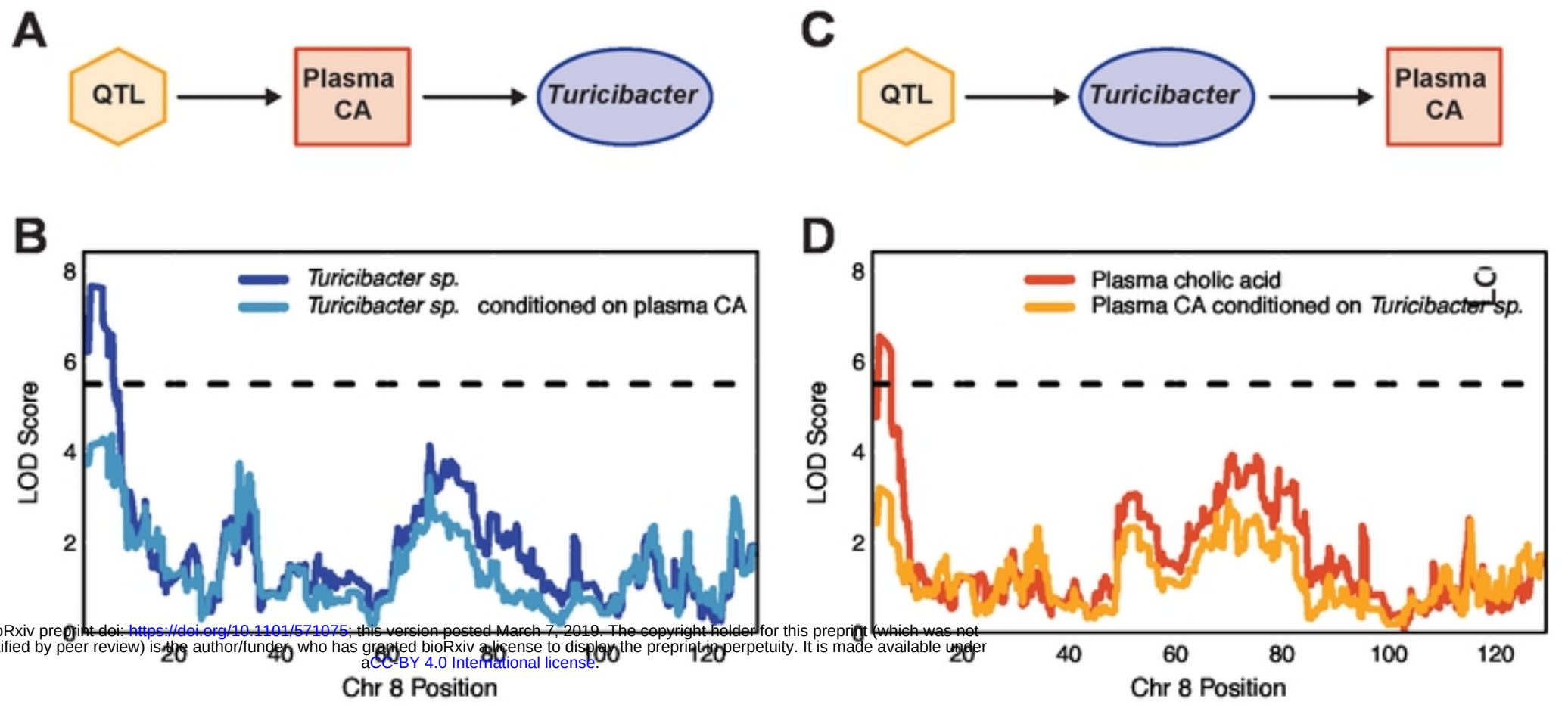

Figure 5