

# Process-based modelling of microbial community dynamics in the human colon

Helen Kettle<sup>1,\*</sup>, Petra Louis<sup>2</sup>, and Harry J. Flint<sup>2</sup>

<sup>1</sup>Biomathematics and Statistics Scotland, James Clerk Maxwell Building, Peter Guthrie Tait Road, Edinburgh, EH9 3FD

<sup>2</sup>Gut Health Group, Rowett Institute, University of Aberdeen, Aberdeen, UK

\*Corresponding author: Helen.Kettle@bioss.ac.uk

September 6, 2022

## **Abstract**

The human colon contains a dynamic microbial community whose composition has important implications for human health. In this work we build a process-based model of the colonic microbial ecosystem and compare with general empirical observations and the results of in-vivo experiments. Our model comprises a complex microbial ecosystem along with absorption of short chain fatty acids (SCFA) and water by the host through the gut wall, variations in incoming dietary substrates (in the form of “meals” whose composition varies in time), bowel movements, feedback on microbial growth from changes in pH resulting from SCFA production, and multiple compartments to represent the proximal, transverse and distal colon. We verify our model against a number of observed criteria, e.g. total SCFA concentrations, SCFA ratios, mass of bowel movements, pH and water absorption over the transit time; and

15 then run simulations investigating the effect of colonic transit time, and  
16 the composition and amount of indigestible carbohydrate in the host diet,  
17 which we compare with in-vivo studies. The code is available as an R  
18 package (`microPopGut`) to aid future research.

## 19 **Author Summary**

20 Kettle wrote the model code and led the writing of the manuscript. Flint con-  
21 tributed to writing the manuscript. Flint and Louis contributed to all aspects  
22 of microbiology and all authors contributed critically to the drafts and gave final  
23 approval for publication.

## 24 **Introduction**

25 The human colon harbours a dense and diverse community of microbiota whose  
26 interactions with the host can have a profound effect on human health (e.g.  
27 Rios-Covian et al. (2016), Morrison and Preston (2016)). Due to the location of  
28 this community within its host, data collection and experimentation are prob-  
29 lematic. Information on this system mostly comes from volunteer experiments  
30 in which diet and stool samples are monitored or from laboratory experiments  
31 using the microbes found in stool samples. Another approach is to put current  
32 knowledge into a mathematical framework and run simulations of the system to  
33 test our understanding and identify knowledge gaps. To this end a number of  
34 mathematical models of this system have been developed - e.g. Cremer et al.  
35 (2016), Cremer et al. (2017), Munoz-Tamayo et al. (2010), Smith et al. (2021),  
36 Moorthy et al. (2015).

37 When developing a model, a number of assumptions about the system are  
38 made in order to reduce complexity/dimensionality so that the model is easier to  
39 parameterise, run and analyse. Some modellers choose to reduce the microbial  
40 complexity and focus on the physics of the gut (e.g. Cremer et al. (2016),  
41 Cremer et al. (2017)), some try to achieve a balance of both (e.g. Munoz-Tamayo

<sup>42</sup> et al. (2010)) and some choose to develop the microbial community (e.g. Smith  
<sup>43</sup> et al. (2021)). The model described here focuses on the microbial community  
<sup>44</sup> dynamics and on interactions with the host, with a fairly simple model of the  
<sup>45</sup> colon. We include the simulation of ‘meals’ (of random composition and size)  
<sup>46</sup> arriving at the colon and look at the effects of bowel movements, both of which,  
<sup>47</sup> as far as we are aware, have not been previously incorporated into such models.  
<sup>48</sup> Having developed a complex model of human gut microbiota in a fermentor  
<sup>49</sup> system (Kettle et al., 2015), and publicly available software (microPop - an  
<sup>50</sup> R package for modelling microbial communities (Kettle et al., 2018)) we now  
<sup>51</sup> incorporate this 10-group microbial ecosystem model (Table 1) into a model of  
<sup>52</sup> the human gut in order to simulate the effects of diet and host on the microbial  
<sup>53</sup> composition and subsequent short chain fatty acid (SCFA) production.

<sup>54</sup> Since approximately 95% of the SCFA produced by the microbes during  
<sup>55</sup> growth are absorbed by the host through the gut wall this represents a strong  
<sup>56</sup> interaction between the microbes and the host. Indeed the ratio of the 3 main  
<sup>57</sup> SCFAs (acetate, butyrate and propionate) is known to have a significant effect  
<sup>58</sup> on human health (Louis et al. (2014), Morrison and Preston (2016)). Thus,  
<sup>59</sup> we prioritise information on the values of these ratios in our model verification.  
<sup>60</sup> Similarly approximately 90% of the water flowing into the colon is absorbed.  
<sup>61</sup> Changes in the volume of water have a significant effect on the concentration of  
<sup>62</sup> the molecules in the colon which in turn affects pH which then affects microbial  
<sup>63</sup> growth, all of which are included in our model.

<sup>64</sup> Due to its shape within the body, the colon is commonly divided into 3  
<sup>65</sup> different regions - the proximal, transverse and distal sections running from  
<sup>66</sup> beginning to end (Fig. 1A and B). The availability of substrate, microbial  
<sup>67</sup> growth and hence pH vary along the colon, therefore, although our model is not  
<sup>68</sup> spatial we simulate these three regions explicitly, with flow from one to another.  
<sup>69</sup> Furthermore, as well as incorporating varying substrate inflow in the form of  
<sup>70</sup> meals we also add in the release of mucins along the length of the colon which  
<sup>71</sup> can be microbially broken down to release proteins and carbohydrates, allowing

72 for further microbial growth away from the beginning of the colon where the  
73 substrates enter. A graphical summary of the model is shown in Fig. 2, the  
74 microbial functional groups are shown in Table 1 and the model state variables  
75 are summarised in Table 2.

76 We use the following criteria to verify our model captures the main features  
77 established for the system:

- 78 1. Total SCFA (TSCFA) concentration in the proximal, transverse and distal  
79 compartments should be around 123, 117 and 80 mM respectively accord-  
80 ing to sudden death human autopsies (Cummings et al., 1987)
  - 81 2. Acetate:Propionate:Butyrate ratios are similar (around 3:1:1) in all regions  
82 of the colon and around 60:20:20 mM (Cummings et al., 1987)
  - 83 3. Over 95% of SCFA are absorbed by the host (Topping and Clifton, 2001)
  - 84 4. Approx. 90% of incoming water is absorbed by the host (Phillips and  
85 Giller, 1973)
  - 86 5. pH in the proximal, transverse and distal compartments should be around  
87 5.7, 6.2 and 6.6 respectively (Cummings et al. (1987), and telemetry data  
88 from Mikolajczyk et al. (2015), Bown et al. (1974))
  - 89 6. Normal daily fecal output in Britain is  $100\text{-}200 \text{ g d}^{-1}$  of which 25-50 g is  
90 solid matter (i.e.  $50\text{-}175 \text{ g d}^{-1}$  is water). Bacteria make up about 55%  
91 of the solid matter i.e.  $14\text{-}28 \text{ g d}^{-1}$  of microbes emitted (Stephen and  
92 Cummings, 1980).
  - 93 7. TSCFA concentration decreases with transit time (Lewis and Heaton,  
94 1997)
- 95 After model verification we examine the effects of including meals, bowel move-  
96 ments and fixed/varying pH into the model. We then use the model to look at  
97 how carbohydrate composition (based on the fractions of resistant starch (RS)

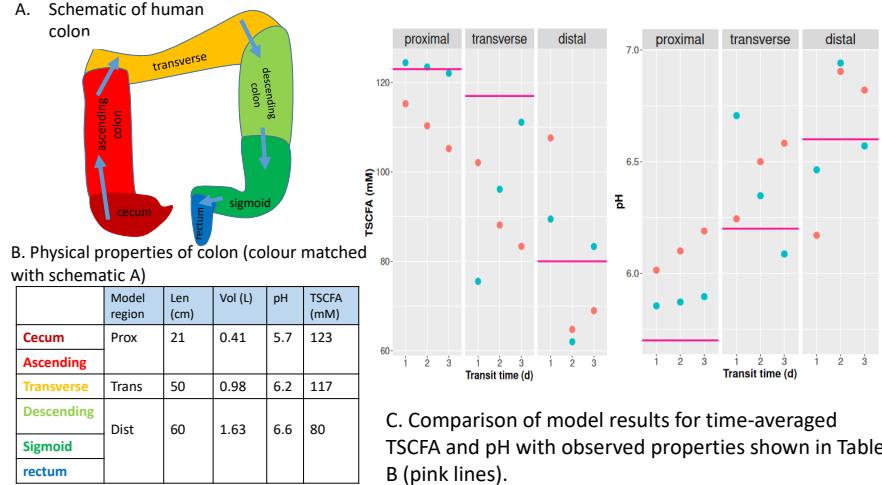


Figure 1: A. Colon schematic, B. Table of typical values for physical properties (length, volume, pH and TSCFA) and C. plots of summarised model simulations for average TSCFA and pH for comparison with typical values. Red dots show results from meals-inflow averaged over 4 random seeds; blue dots show results from continuous substrate inflow.

and non-starch polysaccharides (NSP)) and total carbohydrate affect the microbial community and SCFA composition. The simulations are then compared with in-vivo data from human volunteer experiments.

Although gut microbiota are highly complex and not fully understood, here we show that it is nonetheless possible to develop predictive models of key components of this ecological system. An important goal of our modelling is to aid and inform the interpretation of data obtained, mostly from faecal samples, in studies on diet and health in humans. Our results show promise and we believe this model represents a significant step forward in analysing this highly complex system. We refer to the model as “microPopGut” and to aid future research the code is available as an R-package on github (<https://github.com/HelenKettle/microPopGut>) and instructions on how to use the package are given in the supplementary file ‘gettingStartedWithMicroPopGut.pdf’.

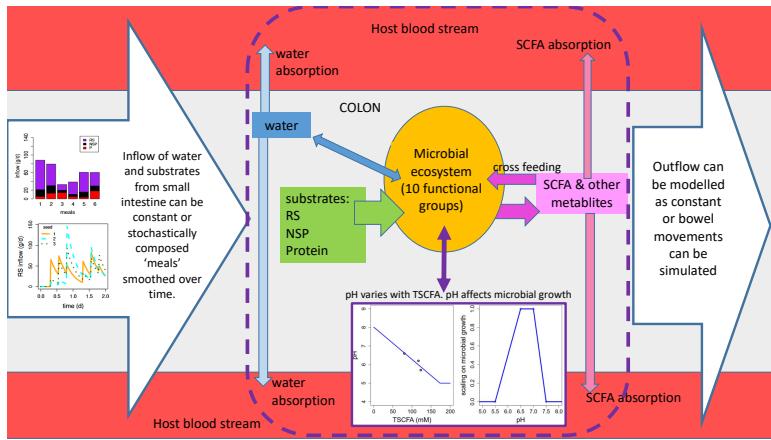


Figure 2: Model system with the microbial ecosystem comprising 10 microbial functional groups (Table 1) which consume substrates (RS, NSP and protein) and water. The microbes produce metabolites some of which are consumed by other MFGs ('cross-feeding'). SCFA and water are absorbed through the colon wall (at a different specific rates). The system shown within the dashed line is repeated in each of the modelled regions of the colon (proximal, transverse, and distal) with the contents of the previous region, flowing into the next. The first compartment (proximal) has inflow from the small intestine - this can be constant inflow or simulated meals whose composition varies randomly in time. The third model compartment (distal) has outflow to stool which can be constant or evacuation via bowel movements can be simulated. pH varies with the TSCFA concentration and affects the rate of microbial growth differently for each MFG.

Table 1: Microbial functional groups included in the model (and the R package microPop (Kettle et al., 2018)) and described by Kettle et al. (2015). Users should be aware that the parameter values given in the data frames in the software will almost certainly change with increasing knowledge of gut microbiota and in some cases are simply a “best guess”.

microPop Name	Abbr.	Description	Examples
Bacteroides	B	Acetate-propionate-succinate group	<i>Bacteroides</i> spp., <i>Prevotella</i> spp., <i>Akkermansia muciniphila</i> (Verrucomicrobia)
NoButyStarchDeg	NBSD	Non-butyrate-forming starch degraders	<i>Ruminococcaceae</i> related to <i>Ruminococcus bromii</i> . Also includes certain <i>Lachnospiraceae</i>
NoButyFibreDeg	NBFD	Non-butyrate-forming fibre degraders	<i>Ruminococcaceae</i> related to <i>Ruminococcus albus</i> , <i>Ruminococcus flavefaciens</i> . Also includes certain <i>Lachnospiraceae</i>
LactateProducers	LP	Lactate producers	<i>Actinobacteria</i> , especially <i>Bifidobacterium</i> spp., <i>Collinsella aerofaciens</i>
ButyrateProducers1	BP1	Butyrate Producers	<i>Lachnospiraceae</i> related to <i>Eubacterium rectale</i> , <i>Roseburia</i> spp.
ButyrateProducers2	BP2	Butyrate Producers	Certain <i>Ruminococcaceae</i> , in particular <i>Faecalibacterium prausnitzii</i>
PropionateProducers	PP	Propionate producers	<i>Veillonellaceae</i> e.g. <i>Veillonella</i> spp., <i>Megasphaera elsdenii</i>
ButyrateProducers3	BP3	Butyrate Producers	<i>Lachnospiraceae</i> related to <i>Eubacterium hallii</i> , <i>Anaerostipes</i> spp.
Acetogens	A	Acetate Producers	Certain <i>Lachnospiraceae</i> , e.g. <i>Blautia hydrogenotrophica</i>
Methanogens	M	Methanogenic archaea	<i>Methanobrevibacter smithii</i>

## 111 Results

### 112 Standard Model

113 The model settings which give the best fit to our criteria are shown in Table  
 114 4 (colon parameters and dietary inflow). The microbial group parameters are  
 115 listed in Supp. Info. (section 3). These define our default model. From this we  
 116 investigate the effects of different model configurations, e.g. with/without bowel  
 117 movements, meals and variable pH, for a range of transit times. Simulations  
 118 with meals have a random component therefore the model is run for a number of  
 119 different starting seed values. Due to the random fluctuations these simulations  
 120 will not reach steady state therefore the summary values are taken as the mean

Table 2: State variables included in the model. They are all in units of mass (g; with the exception of pH) and they are computed for each model compartment (e.g. prox., trans. and dist.). They are derived automatically from the substrates and metabolites specified for each microbial functional group (MFG) in the input file/dataframe to the R package microPop Kettle et al. (2018).

Name	Details
Microbial biomass	Computed for each of the 10 MFGs (Table 1)
water	from dietary intake or from microbial metabolism
Protein	from dietary intake or mucin
Resistant starch (RS)	from dietary intake or mucin
Non-starch polysaccharides (NSP)	from intake or mucin
pH	Computed from TSCFA (Eq. 5)
Acetate	metabolite
Butyrate	metabolite
Propionate	metabolite
Formate	metabolite
Carbon dioxide	metabolite
Methane	metabolite
Ethanol	metabolite
Lactate	metabolite
Succinate	metabolite
Hydrogen	metabolite

121 from day 7 (to remove the effect of the initial conditions) to the end of the  
 122 simulation (28 days) and are averaged over multiple seeds.

123 Table 3 gives summary results of the model simulations without bowel move-  
 124 ments but with varying pH for each bowel region. Fig. 3 shows results from  
 125 more simulations but for the distal colon only. Fig. 3a shows that although  
 126 bowel movements make a difference to the total biomass and the TSCFA they  
 127 do not have a large effect on the community composition or the SCFA ratios.  
 128 Thus in the interests of model simplicity we decide to not include bowel move-  
 129 ments in later simulations. However, varying pH with TSCFA can be seen to  
 130 make a large difference to the microbial community (Fig. 3b) and also improves  
 131 the SCFA ratios with respect to our verification criteria. The addition of meals  
 132 makes a significant difference which increases with increasing transit time (Fig.  
 133 3c). In Fig. 4 the time series output from the model shows how the meals-  
 134 inflow allows the community to experience large shifts over time (on a much  
 135 longer time scale than the variations in the input), as opposed to the fixed state

<sup>136</sup> approached using a constant substrate inflow.

<sup>137</sup> Fig. 1C shows the average pH and TSCFA for the proximal, transverse  
<sup>138</sup> and distal compartments. A decrease in TSCFA (and concomitant increase in  
<sup>139</sup> pH) with longer transit time is predicted in the proximal colon both for meal  
<sup>140</sup> inflow and continuous input and this is in broad agreement with experimental  
<sup>141</sup> findings (Lewis and Heaton 1997). In section 2 of the Supp. Info. we suggest  
<sup>142</sup> a mathematical explanation for this based on the supposition that the specific  
<sup>143</sup> rate of absorption of water through the gut wall is slower than that for SCFA.

<sup>144</sup> Regarding Table 3, for some criteria, e.g. pH, the continuous inflow setting  
<sup>145</sup> gives results closer to our verification values, but in other cases, e.g. A:B:P  
<sup>146</sup> in distal colon, simulating meals gives closer results. Note that we consider a  
<sup>147</sup> transit time of 1 day the most typical of the three transit times, and the one  
<sup>148</sup> that should be compared with our verification criteria, the others are included to  
<sup>149</sup> show the variation in results. Ideally TSCFA should be 123, 117 and 80 mM for  
<sup>150</sup> prox., trans., dist. but the best match we have to this is for a 3 d transit time and  
<sup>151</sup> continuous inflow. This is most likely due to the fact that our model has fixed  
<sup>152</sup> rates of specific absorption of SCFA and water throughout the colon. However,  
<sup>153</sup> our TSCFA values are within a reasonable range and display the general trend  
<sup>154</sup> of decreasing TSCFA from the proximal to distal colon. The microbe output,  
<sup>155</sup> i.e. the outflow of fecal microbes is steady at around  $20 \text{ g d}^{-1}$  in all cases which  
<sup>156</sup> fits well in the verification range ( $14\text{-}28 \text{ g d}^{-1}$ ). The water fraction is the ratio  
<sup>157</sup> of the rate of fecal water over the rate of water flowing into the colon, since 90%  
<sup>158</sup> of water is absorbed this should be 0.1. This is approximately correct for our 1  
<sup>159</sup> d simulations (0.14) but, as expected, when transit time increases this decreases  
<sup>160</sup> significantly. In summary, comparing these simulation results with our list of  
<sup>161</sup> model verification criteria shows that in general our model is fit for purpose,  
<sup>162</sup> and that the inclusion of meals-inflow and varying pH improve our simulations.

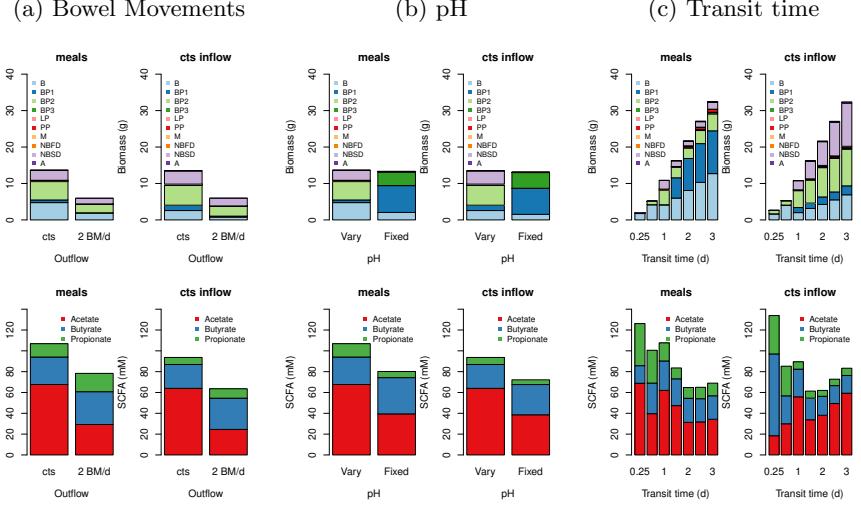


Figure 3: Summary results (averaged over days 7-28 and over random seeds) for the distal compartment for continuous inflow or fluctuating inflow (i.e. ‘meals’) for continuous outflow from colon or for 2 bowel movements per day (‘2 BM/d’). The RS fraction is 0.78 (i.e. 78% of the dietary carbohydrate is resistant starch and 22% is NSP) and the transit time is 0.93 d for a), 1.25 d for b) and at 0.25, 0.5, 1, 1.5, 2, 2.5 and 3 days for c). The top row shows the biomass of each group, the bottom row shows the SCFA.

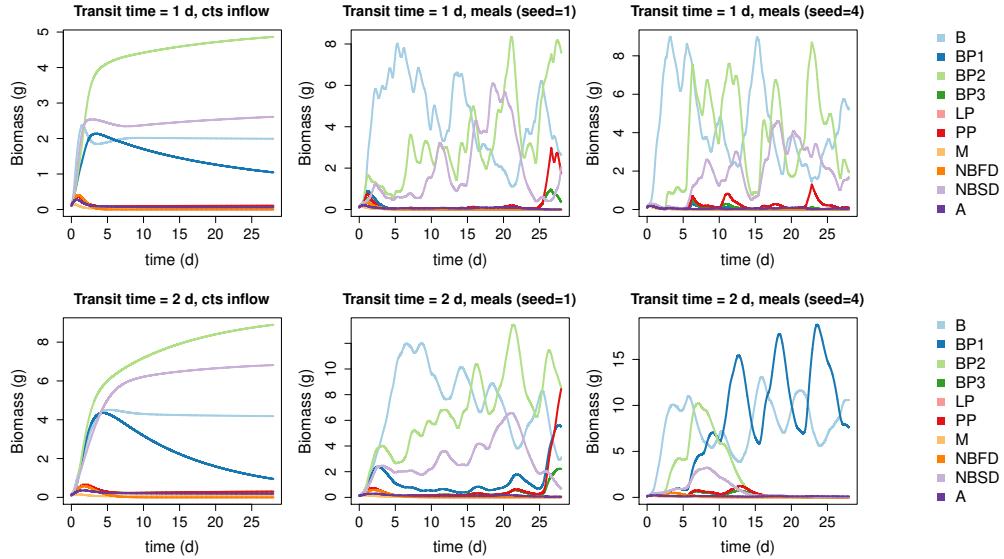


Figure 4: Simulation results for the distal compartment for continuous inflow (first plot on each row) or fluctuating inflow (i.e. ‘meals’) for transit times of 1 d (top row) and 2 d (bottom row) and for 2 random seeds. Modelled pH varies with TSCFA and the RS fraction is 0.78. There are no bowel movements (i.e. outflow is continuous). See Table 1 for microbial groups.

Table 3: Summary of model results (for comparison with our list of criteria) for 3 different transit times, with meals or continuous inflow and with pH varying with TSCFA. Microbe output is the mass of microbes leaving the colon per day and the water fraction is amount of water leaving the colon per day divided by the amount entering. All simulations were run for 28 days and the results shown are the average over days 7-28. The results for the simulations with meals are averaged over 4 random seeds. ‘A:B:P dist’ refers to the Acetate:Butyrate:Propionate ratio (mM) in the distal colon.

	Meals			Continuous inflow		
	1d	2d	3d	1d	2d	3d
TSCFA prox (mM)	115.3	110.3	105.2	124.4	123.5	122.1
TSCFA trans (mM)	102.1	88.1	83.4	75.5	96.1	111.1
TSCFA dist (mM)	107.6	64.8	69.0	89.5	62.0	83.3
A:B:P dist (mM)	62:28:17	31:23:10	34:23:12	56:27:7	38:18:6	59:17:7
pH prox	6.0	6.1	6.2	5.9	5.9	5.9
pH trans	6.2	6.5	6.6	6.7	6.4	6.1
pH dist	6.2	6.9	6.8	6.5	6.9	6.6
microbe output ( $\text{g d}^{-1}$ )	20.2	20.1	20.1	20.0	20.1	20.0
water fraction	0.14	0.04	0.02	0.14	0.04	0.02

## 163 Model Experiments

164 We now use our model to simulate two scenarios – firstly, the effects of decreasing  
 165 total carbohydrate intake and secondly, the effects of changing carbohydrate  
 166 composition (whilst keeping total intake fixed) on the microbial community and  
 167 associated SCFA production. Comparing our simulations with data from human  
 168 volunteer experiments is not straightforward since in order to run our model,  
 169 ingested food must be translated to **non-digestable** substrates reaching the colon.  
 170 This is problematic due to unknown water consumption and transit times and  
 171 uncertainties associated with the absorption rates of the ingested carbohydrate  
 172 and protein higher up the digestive tract. Thus we do not attempt to reproduce  
 173 human experiments **exactly** but rather we run simulations based on variations  
 174 to our standard model set up **which are qualitatively similar** and then compare  
 175 our results with the trends in the available data.

### 176 Effects of total dietary carbohydrate

177 In this model experiment we investigate the effects of decreasing carbohydrate  
 178 on the microbial community. Here we compare our results qualitatively with

the human dietary study of Duncan et al. (2007) which explored the impacts of carefully controlled decreases in carbohydrate intake upon weight loss and microbial fermentation products in obese subjects using 3 diets – a maintenance (M) diet, a high protein, moderate carbohydrate diet (HPMC) and a high protein, low carbohydrate diet (HPLC) (see Fig. 5 for details). This is of course, the composition for ingested food, which is not easily translated into substrate concentrations entering the colon. However, we can look at the general trends in SCFA and microbial composition with changing colonic carbohydrate intake rate. Thus, in these model experiments we keep protein inflow to the colon at  $10 \text{ g d}^{-1}$  (our default value) and then increase inflowing non-digestable (ND) carbohydrate from  $10 \text{ g d}^{-1}$  to  $60 \text{ g d}^{-1}$  in  $10 \text{ g d}^{-1}$  intervals. To include the effects of different ND-carbohydrate composition we run the model for an resistant starch (RS) fraction of either 0.2 or 0.78 (the default value), with non-starch polysaccharides (NSP) making up the remaining ND-carbohydrate in each case. Although subject to large uncertainties, we estimate the RS fractions for the Duncan et al. (2007) experiments of 0-0.6 (M diet), 0-0.68 (HPMC) and 0-0.12 (HPLC) (based on RS is 0–20% of ingested starch (Capuano et al., 2018) and bio-available NSP is 75% of ingested NSP (Slavin et al., 1981)). Due to the low fibre nature of many of these simulations we run the model with a slightly longer transit time of 1.5 d and for both continuous inflow and meals.

Fig. 6 shows the SCFA results from our model experiment and Fig. 5 shows the results from the in vivo experiment. It is clear, from both the model and in vivo results that the proportion of butyrate increases as the amount of ND-carbohydrate in the diet increases. Furthermore, both model and in vivo results show an increase in TSCFA with ND-carbohydrate intake rate. Since Duncan et al. (2007) also look at the relationship between butyrate concentration and grams of carbohydrate eaten per day, we plot butyrate against carbohydrate entering the colon (Fig. 7) to compare with their Fig. 1. In both cases, butyrate concentration increases with incoming carbohydrate. Furthermore, as seen in both the model and the data, the percentage of butyrate increases with carbo-

Diet	PI (g d <sup>-1</sup> )	CI (g d <sup>-1</sup> )	SI (g d <sup>-1</sup> )	NSPI (g d <sup>-1</sup> )
Duncan				
M	94	399	187	28
HPMC	127	164	95	12
HPLC	120	24	2.7	6.1
Walker				
M	103	427	230	28
High NSP	102	427	138	42
High RS	109	434	275	13
WL	144	201	110	22

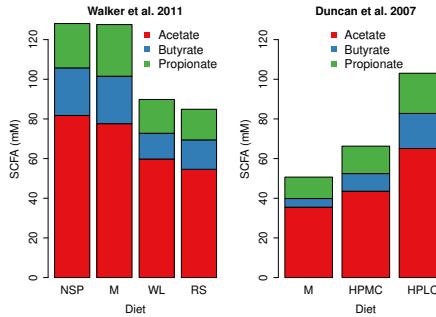


Figure 5: Table on left shows the dietary intake for two human studies (Duncan et al. (2007) and Walker et al. (2011)). PI, CI, SI and NSPI refer to ingested dietary protein, carbohydrate, starch and NSP. Note, starch value for the high RS diet in the Walker et al. (2011) study included 26 g commercial RS. Bar plots show SCFA data from these studies.

209 hydrate intake (Fig. 7). Analysis of 10 human studies involving 163 subjects  
210 has shown a highly significant increase in percentage butyrate with increasing  
211 total SCFA concentration in faecal samples (LaBouyer et al., 2022).

212 In terms of microbial composition, Fig. 6 shows the results from our simulations  
213 are reasonably consistent across inflow type (meals or continuous), with B  
214 dominating at low carbohydrate intake. When the RS fraction is low (i.e. when  
215 ND-carbohydrate is made up of 80% NSP) then NBFD increase with increased  
216 C intake. Whereas when C is mostly RS then NBSD and BP1 increase with C.  
217 In both cases BP2 increase with increasing C intake.

### 218 Effects of carbohydrate composition

219 Here we use the model to simulate the effects of changing carbohydrate compo-  
220 sition on the microbial community composition by changing the ratio of RS to  
221 NSP whilst keeping the same amount of total incoming carbohydrate. Fig. 8  
222 show a summary of the model results. Although there are differences between  
223 the continuous inflow/meals, and also for the different transit times (1 d and 3  
224 d), the modelled trends are generally similar, showing a significant shift in com-  
225 munity as the fraction of RS increases, an increase in TSCFA and changes in  
226 the SCFA ratios. We compare our results with a human dietary study (Walker

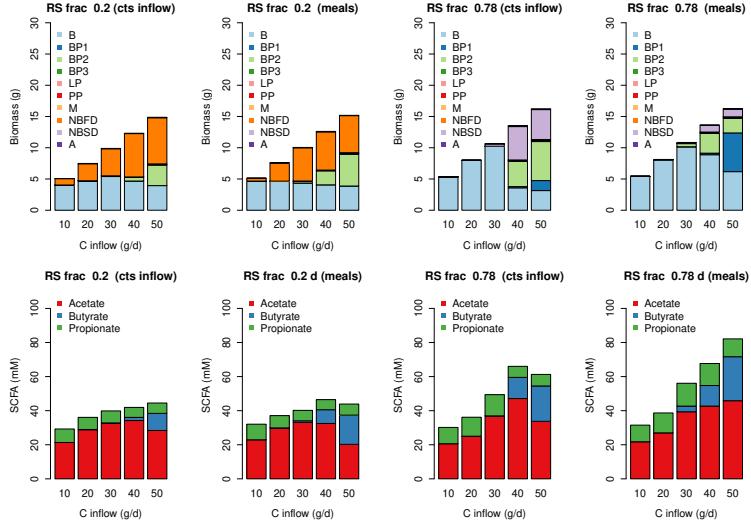


Figure 6: Simulated Biomass and SCFA results for increasing carbohydrate inflow. Simulations are run with continuous substrate inflow (cts) and with ‘meals’ for a transit time of 1.5 days. The results are the average over the last 3 weeks of a 28 day simulation and ‘meals’ is the average over 4 stochastically-generated simulations.

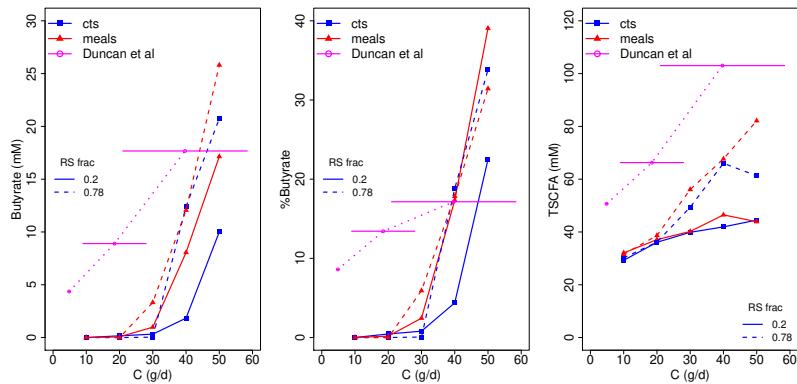


Figure 7: Plot of modelled butyrate, %butyrate and TSCFA against grams of carbohydrate entering the colon each day. Data from Duncan et al. (2007) is shown in magenta - due to uncertainties in converting ingested starch to RS entering the colon there are large error bars on the amount of C (g/d). Error bars show C estimated by the sum of 75% of ingested NSP plus 0-20% of ingested starch.

<sup>227</sup> et al. (2011), Salonen et al. (2014) and references therein) examining the im-  
<sup>228</sup> pact of switching the major type of ND-carbohydrate from wheat bran (NSP)  
<sup>229</sup> to resistant starch. Volunteers were provided successively with a maintenance  
<sup>230</sup> diet, diets high in RS or NSPs and a reduced carbohydrate weight loss (WL)  
<sup>231</sup> diet, over 10 weeks (Fig. 5).

<sup>232</sup> There are large discrepancies between the SCFA predicted by our model (Fig.  
<sup>233</sup> 8) and the measured SCFA data (Fig. 5). Our model predicts an increase in  
<sup>234</sup> TSCFA as proportion of RS increases whereas total fecal SCFA were significantly  
<sup>235</sup> lower for the RS and WL diets compared to the other two diets (in which NSP  
<sup>236</sup> is higher). One possible explanation is that fermentation of RS occurs in more  
<sup>237</sup> proximal regions of the colon compared with NSP fibre fermentation, such that  
<sup>238</sup> there is greater absorption of the SCFA products. A second possibility, also  
<sup>239</sup> likely, is that transit times were longer for the RS diet than for the NSP diet,  
<sup>240</sup> which we predict would result in decreased SCFA concentrations. In our model  
<sup>241</sup> the effect of the RS fraction on TSCFA is greater than the effect of transit time  
<sup>242</sup> so we do not see this in Fig. 8.

<sup>243</sup> The human study also included detailed compositional analysis of the fecal  
<sup>244</sup> microbiota (Walker et al. (2011), Salonen et al. (2014)) that revealed specific  
<sup>245</sup> responses mainly by different groups of Firmicutes bacteria to the RS and NSP  
<sup>246</sup> diets. This information was particularly important for the phylogenetic assign-  
<sup>247</sup> ments to the functional groups used here and previously in the model of Kettle  
<sup>248</sup> et al. (2015). Our modelling predicts striking shifts in the microbial commu-  
<sup>249</sup> nity, especially involving the NBSD, NBFD and butyrate-producing groups,  
<sup>250</sup> with changing proportions of RS and NSP fibre (Fig. 8). We should also note  
<sup>251</sup> that in the volunteer experiments many bacterial species were not significantly  
<sup>252</sup> altered by the RS-NSP switch *in vivo* (Walker et al., 2011) indicating that many  
<sup>253</sup> may be generalists, able to switch quickly between energy sources.

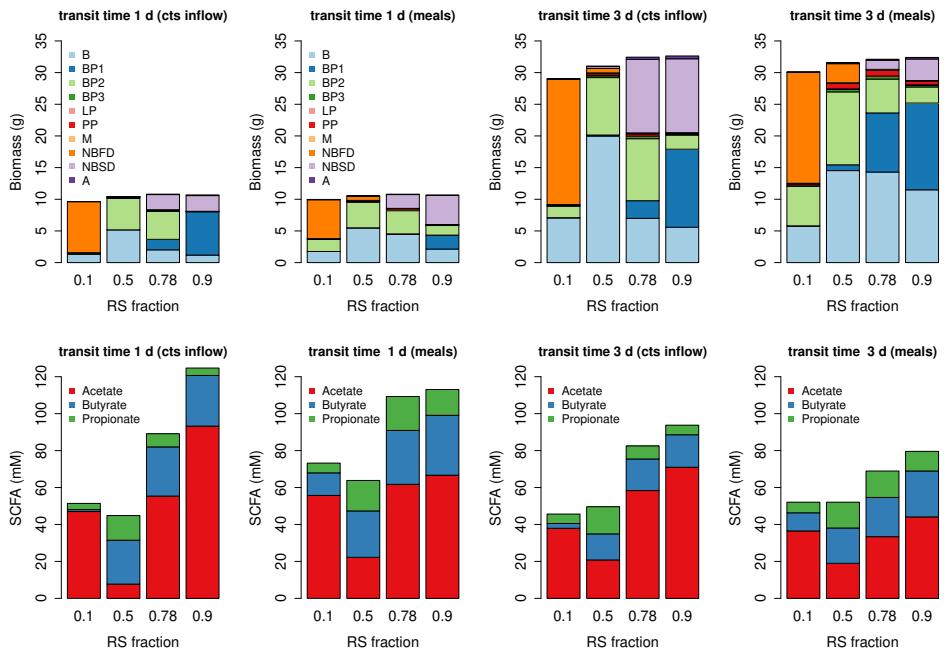


Figure 8: Biomass and SCFA results for changing the RS fraction of inflowing carbohydrate with continuous substrate inflow ('cts inflow') and with 'meals'. Protein and carbohydrate inflow are 10 and 50 g d<sup>-1</sup> respectively. The results are the average over the last 3 weeks of a 28 day simulation and 'meals' is the average over 4 stochastically-generated simulations.

254 **Discussion**

255 The development of a complex model of the microbial community in the human  
256 colon, whose simulations compare well with data, represents a significant step  
257 forward. Previous models have been based on simpler microbial models (e.g.  
258 Cremer et al. (2017), Munoz-Tamayo et al. (2010), Moorthy et al. (2015)), or  
259 have not shown such a good agreement with data (e.g. Smith et al. (2021)). Our  
260 previous complex model community consisted of 10 functional groups, but the  
261 model was designed only to simulate continuous culture conditions in a chemo-  
262 stat (Kettle et al., 2015). Translating this 10-group model into an in vivo setting  
263 has required introducing multiple gut compartments, and the absorption of wa-  
264 ter and SCFA, followed by comparison with generally observed characteristics  
265 of the system. We were then able to use this model to examine the predicted  
266 impact of changes in the amount and type of non-digestible carbohydrate (fibre)  
267 present in human diets upon concentrations of fermentation products (SCFA)  
268 in different gut compartments and in stool. At the same time, we predict the  
269 likely impact of dietary changes and variations in gut transit upon microbiota  
270 composition and fermentation products. The model must be regarded as work  
271 in progress particularly with respect to microbiota composition. Predictions can  
272 however become improved and refined as more information becomes available  
273 in time.

274 Assignments of microbial taxa to our ten functional groups were based ini-  
275 tially on evidence from cultured isolates. These assignments have since been  
276 supported and greatly extended by analysis of genes diagnostic for different fer-  
277 mentation pathways within genomes and metagenomes (Reichardt et al., 2014)  
278 and by molecular detection of species enriched within the community by defined  
279 growth substrates in chemostat experiments (Duncan et al., 2016) and dietary  
280 intervention studies (Salonen et al., 2014). Nevertheless, these assignments in-  
281 evitably remain provisional and incomplete and we do not claim that the model  
282 predictions can be made precise at a phylogenetic level. More emphasis is placed

in our model on the prediction of metabolic outputs based on microbial transformations and interactions. While there is relatively little phylogenetic overlap for example between producers of propionate and butyrate (Reichardt et al. (2014), Louis and Flint (2017)) there are many cases where individual species are known to use multiple alternative substrates as energy sources, which complicates assignments. For this reason, more weight was given to fermentation pathways than to substrate preferences in defining the functional groups. However, it would also be possible to define completely different groupings that relate to other outputs (e.g. bile acid metabolism, or vitamin/micronutrient supply) in order to address specific questions. Furthermore, it may well be worthwhile to increase the number of functional groups in the future. The large B group for example currently includes members of the Bacteroidetes phylum, but its characteristics are mainly based on well-studied members of the *Bacteroides* genus. We know that *Prevotella* is another highly abundant genus of *Bacteroidetes* in the human colon, but the two genera tend not to co-occur at high levels in the same individuals (Wu et al. (2011), Chung et al. (2020)). Less is known about human colonic *Prevotella*, for which there are relatively few cultured representatives, making it premature to create a separate grouping, but this would clearly be desirable in the future as their prevalence is reported to affect health and responses to dietary intervention. In future it should become possible to define the relative abundance of functional groups (MFGs) and their relationship to phylogeny directly from genomic and metagenome analysis, by examining genes diagnostic for particular pathways and functions (e.g. Reichardt et al, 2014).

The parameter values for the microbial groups used in our model are from the intrinsic data frames in the microPop package (the only changes are to LactateProducers). Although the work presented here did not attempt to fit particular parameters to data, as we focussed on expanding the scope of the model (i.e. changing the environment from fermentor to colon), these values are easy to alter, e.g. Wang et al. (2020) changed many of these parameters to achieve a better model fit to their data. As well as adjusting the parameters

313 for each group to represent inter-individual variation, groups can also be easily  
314 added or removed from the model through the input argument ‘microbeNames’.

315 Furthermore, it is also possible to include any number of strains (with varying  
316 parameter values) within each functional group in order to add more variation  
317 in outcome (see Kettle et al. (2015)) but we did not do this here in the interests  
318 of computational time. It should also be noted that the parameter values are  
319 highly uncertain in many cases and within each of our functional groups there  
320 will be large variability due to adaptation and evolution. Given this, we do  
321 not claim that the model response is necessarily representative of what may  
322 happen in an individual’s gut, rather it can be used as an aid to gain insight  
323 into the relative importance of the different processes we are currently aware of  
324 and potentially to highlight, those we are not.

325 In addition to this, it must be noted that the default diet chosen here with  
326 10 g of protein and 50 g of carbohydrate fibre reaching the colon each day could  
327 be revised for any given population. However, converting from quantities of  
328 ingested food to substrate inflow to the colon is highly uncertain with large  
329 variations between studies, as well as technical issues with measuring this accu-  
330 rately. With more time, it would be interesting to investigate a larger range of  
331 typical diets but this was beyond the scope of the current work.

332 In summary, although performing reasonably well, the model has the poten-  
333 tial to be considerably improved simply by altering the parameter values and  
334 existing settings, however, more fundamental changes such as those listed below  
335 could also be investigated in future work:

- 336 • Adding more functional groups or pathway switches in the existing func-  
337 tional groups. For example at present only the Bacteroides group can  
338 utilise protein but it is now known that some butyrate producers can also  
339 utilise amino acids (Louis and Flint, 2017)
- 340 • Our pH relation with TSCFA is very simplistic and could potentially be  
341 improved, although host secretions mean this is not necessarily straight-

342 forward.

- 343 • Currently we set the transit time for the colon and then this is split be-  
344 tween the 3 model compartments based on their relative sizes. An interest-  
345 ing addition would be to alter transit time based on the composition of the  
346 various substrates entering the colon. For example, increasing residence  
347 time for high protein and/or low fibre diets. Due to variation in individual  
348 response this may need to include significant uncertainty ranges.
- 349 • Related to this is changing the absorption rate of water through the gut  
350 wall based on the diet, for example more water could remain in the gut  
351 on a high fibre diet.
- 352 • A longer term goal would be to model the processes in the gastrointestinal  
353 tract preceding the colon in order to simulate how substrates entering the  
354 colon relate to dietary intake. This would allow more accurate prediction  
355 of microbial metabolite production based on diet.

356 To conclude, our model helps to explain some important, but poorly under-  
357 stood, relationships that have been reported in human studies, including the  
358 increase in butyrate proportion with increasing total faecal SCFA (LaBouyer  
359 et al., 2022). This phenomenon has important implications in view of the  
360 claimed benefits of butyrate supply for colorectal cancer prevention and the  
361 health of the colonic mucosa (Louis et al. (2014), Hamer et al. (2008)). The  
362 model also predicts increasing total faecal SCFA with greater fibre intake and  
363 more rapid gut transit. Gut transit is also shown to have potentially important  
364 consequences for microbiota composition and gut metabolism. In addition, the  
365 model confirms that the amount and type of non-digestible carbohydrate in the  
366 diet has the potential to cause major changes in microbiota composition. The  
367 nature of such changes is, however, predicted to be influenced by patterns of  
368 meal feeding and by any effects of dietary components (e.g. dietary fibre) upon  
369 gut transit. Human studies suggest that they will also depend on the initial  
370 microbiota composition. There is potential to use the model to explore how the

371 presence of particular functional groups (such as lactate-utilizers (Wang et al  
372 2020)) within an individual's microbiota can influence their gut metabolism and  
373 response to dietary intervention. This may indeed be one of the most intriguing  
374 and fruitful applications of such modelling approaches in the future.

## 375 Materials and methods

### 376 Software

377 To facilitate continued research and future model development by other re-  
378 searchers we provide all model code on github (<https://github.com/HelenKettle/microPopGut>).  
379 The R package microPopGut is contained in the file microPopGut\_1.0.tar.gz.  
380 This can be downloaded and installed in R using `install.packages('microPopGut_1.0.tar.gz')`.  
381 Furthermore instructions on how to use the package are given in the supplemen-  
382 tary file '`gettingStartedWithMicroPopGut.pdf`'.

### 383 Microbial Model

384 The microbial functional group model is based on the model described by Ket-  
385 tle et al. (2015) and implemented using the R package microPop (Kettle et al.,  
386 2018). The microbial groups include producers of the three major SCFA de-  
387 tected in fecal samples (acetate, butyrate and propionate) together with uti-  
388 lizers of acetate, lactate, succinate, formate and hydrogen (see Table 1 for a  
389 summary, or refer to Kettle et al. (2015) for more detail). The model and its  
390 equations are described in detail by Kettle et al. (2015) and Kettle et al. (2018)  
391 so only a brief overview is given here. The microbial groups are defined as data  
392 frames within the R package and these are shown in section 3 of the Supp. Info..  
393 Although this application uses the microbial parameters (e.g. maximum growth  
394 rates, yields etc) that are in the package's intrinsic data frames, these can be  
395 easily changed by either modifying the dataframe in R or by providing a new  
396 dataframe - either as an input csv file or by creating one in R. One of the input  
397 arguments to the function `microPopGut()` is `microbeNames` which allows the

398 user to also enter other microbial groups.

399 The growth substrates available in the large intestine are divided into four  
400 categories: protein (P), non starch polysaccharides (NSP), resistant starch (RS)  
401 and sugars (and oligosaccharides and sugar alcohols); for simplicity, all carbo-  
402 hydrate units are regarded as being hexoses. NSP comprise major components  
403 of dietary fibre including the structural polysaccharides of the plant cell wall  
404 (cellulose, xylan, pectin), whereas RS refers to the fraction of dietary starch  
405 that resists digestion in the small intestine. We consider 10 major metabolites  
406 that arise from substrate fermentation: acetate, propionate, butyrate, lactate,  
407 succinate, formate, hydrogen, carbon dioxide, methane and ethanol. Six of these  
408 metabolites (acetate, lactate, succinate, formate, hydrogen and carbon dioxide)  
409 are also considered as substrates, because they are known to be consumed by  
410 some groups (cross-feeding). It is well known that pH affects growth rate there-  
411 fore each group is assigned a preferred range of pH within which it can reach its  
412 maximum growth rate, but outside of which, its growth is reduced or zero. We  
413 model the rate of bacterial growth using Monod kinetics and assume that from  
414 1 g of resource, Y g of biomass is produced. We assume that resource that is  
415 taken up by microbes, but not used to produce biomass, is converted to metabo-  
416 lites. If not all of the resource is converted to biomass or to the metabolites  
417 represented in our model, it is discarded. This applies, for example, to many  
418 diverse fermentation products of proteins (e.g. phenols, amines) that are not  
419 among the 10 major products covered by the model. Although the model was  
420 initially developed to be run with multiple strains within each functional group,  
421 in the current work we do not do this due to the high CPU time associated with  
422 multiple compartments.

<sup>423</sup> **Inflow to colon**

<sup>424</sup> **Incoming substrates and water**

<sup>425</sup> The main sources of nutrient for microbiota in the colon are complex dietary  
<sup>426</sup> carbohydrates that are not absorbed higher up the digestive tract. We use  
<sup>427</sup> a default value of  $50 \text{ g d}^{-1}$  of carbohydrate, C, in our model and we vary the  
<sup>428</sup> proportion of this which is NSP or RS using the RS fraction (i.e.  $\text{RS}/(\text{RS+NSP})$ )  
<sup>429</sup> where  $\text{RS+NSP}=C$ ). Based on Cremer et al. (2017) and references therein,  
<sup>430</sup> about 15 g of bio-available NSP and 30-40 g of RS enter the colon per day  
<sup>431</sup> which gives us an RS fraction of 0.67-0.9 with average value of 0.78 which we  
<sup>432</sup> use as our default value. According to Yao et al. (2016) less is known regarding  
<sup>433</sup> dietary proteins, P, that escape digestion to reach the large intestine, although it  
<sup>434</sup> is estimated that around 6 - 18 g P reaches the large intestine daily, the majority  
<sup>435</sup> from the diet and a small proportion from endogenous origins. Given this, here  
<sup>436</sup> we assume that  $10 \text{ g d}^{-1}$  of undigested P reaches the colon from dietary intake  
<sup>437</sup> along with a small amount from mucin degradation (approx.  $1 \text{ g d}^{-1}$ ). Phillips  
<sup>438</sup> and Giller (1973) state that water enters at approximately  $1.5 \text{ l d}^{-1}$  and about  
<sup>439</sup> 90% of this is absorbed by the colon. Stephen and Cummings (1980) states that  
<sup>440</sup> normal fecal daily output in Britain is  $100\text{-}200 \text{ g d}^{-1}$  of which  $25\text{-}50 \text{ g d}^{-1}$  is  
<sup>441</sup> solid matter and the rest ( $50\text{-}175 \text{ g d}^{-1}$ ) is water. Thus if 90% is absorbed then  
<sup>442</sup> this indicates water inflow in the range  $0.5\text{-}1.75 \text{ l d}^{-1}$ . The midpoint of this  
<sup>443</sup> range is  $110 \text{ g d}^{-1}$  of water outflow which, if 90% is absorbed, implies that the  
<sup>444</sup> inflow of water is approximately  $1100 \text{ g d}^{-1}$ . This will clearly vary depending  
<sup>445</sup> on the host's oral water intake but we use  $1100 \text{ g d}^{-1}$  as our default value. The  
<sup>446</sup> default inflow values are summarised in Table 4.

<sup>447</sup> **Meals**

<sup>448</sup> The normal human diet does not consist of continuous fixed inflow of substrate;  
<sup>449</sup> for a more realistic substrate inflow to the colon we simulate eating 3 meals a  
<sup>450</sup> day with randomly varying composition. We then approximate the passage of

Table 4: Summary of default values used in the model. Parameter values for the microbial groups are given in the Supp. Info. (section 3)

Symbol	Description	Default Value
$T_t$	transit time through colon	1.25 d
$\dot{P}_{diet}$	protein inflow rate	10 g d <sup>-1</sup>
$\dot{C}_{diet}$	carbohydrate inflow rate	50 g d <sup>-1</sup>
$\dot{W}_{diet}$	water inflow rate	1100 g d <sup>-1</sup>
$\dot{M}$	mucin inflow rate	5 g d <sup>-1</sup>
$K_M$	half saturation constant for Mucin breakdown	0.5 g l <sup>-1</sup>
$a_w$	rate of water absorption by host	3 d <sup>-1</sup>
$a_z$	rate of SCFA absorption by host	9.6 d <sup>-1</sup>

451 these meals through the stomach and small intestine to obtain a smoothed time  
 452 series for substrate entering the colon. Note that we are not simulating all the  
 453 food ingested by the host (most of which will not reach the colon) but rather  
 454 simply trying to produce a more realistic time series for the substrates that we  
 455 know reach the colon.

456 We specify three meals per day each with a duration of 30 minutes. This  
 457 time-series is then passed through a one-compartment ordinary differential equa-  
 458 tion model representing the time spent in the stomach and small intestine (es-  
 459 timated to take 7 hours), i.e.

$$\frac{ds(t)}{dt} = \dot{s}(t)_{in} - vs(t) \quad (1)$$

460 where  $v=3.4$  d<sup>-1</sup> (inverse of 7 h transit time in days);  $\dot{s}(t)_{in}$  is time series  
 461 representing 3 meals a day (g d<sup>-1</sup>) and  $t$  is time in days. The inflow to the  
 462 colon (i.e. the outflow from small intestine) is given by  $vs(t)$ . The composition  
 463 (in terms of P, NSP, RS and water (W)) of these meals varies randomly around  
 464 the mean of each component (Table 4) for each meal. To generate such random  
 465 fluctuations we draw samples for each meal from a gamma distribution (since  
 466 this is always above zero) defined by a scale parameter ( $\gamma_s$ ) and the daily average  
 467 inflow of the substrate (g d<sup>-1</sup>). We assume the magnitude of the substrate  
 468 fluctuations are proportional to the mean value. Preliminary simulations showed  
 469 that  $\gamma_s$  equal to half the mean value of each substrate gave a good variation for

<sup>470</sup> P, RS and NSP, and for water variation we assumed  $\gamma_s$  was one tenth of the  
<sup>471</sup> incoming daily flow. The distributions and flow patterns are shown in Fig. 9.

## <sup>472</sup> Mucin

<sup>473</sup> There is a further input of protein and carbohydrate from the host via the  
<sup>474</sup> breakdown of host-released mucin by many strains in the B group (Ravcheev  
<sup>475</sup> and Thiele, 2017) and in our NBFD group (Crost et al., 2013). It is estimated  
<sup>476</sup> that  $2.7\text{-}7.3 \text{ g d}^{-1}$  of mucin, denoted  $\dot{M}$ , is secreted into the colon (Florin  
<sup>477</sup> et al., 1991), therefore we take the midpoint value  $5 \text{ g/d}$ . We assume our mucin  
<sup>478</sup> degraders break down  $1 \text{ g}$  of mucin into  $0.05 \text{ g}$  sulphate,  $0.2 \text{ g}$  P and  $0.75 \text{ g}$  C,  
<sup>479</sup> based on Sung et al. (2017), but consider their yield on mucin to be negligible  
<sup>480</sup> compared with growth on other substrates. We split C equally between NSP  
<sup>481</sup> and RS - this arbitrary choice did not affect model results since C from mucin  
<sup>482</sup> ( $3.75 \text{ g d}^{-1}$  maximum) is much less than dietary C ( $50 \text{ g d}^{-1}$ ), but this should  
<sup>483</sup> be revised if considering very different dietary drivers. Since the compartments  
<sup>484</sup> of the colon are not equal-sized we assume that the rate of mucin entering the  
<sup>485</sup> colon is divided through the model compartments proportional to their relative  
<sup>486</sup> volumes. We assume this enters the colon at a fixed, continuous rate and mucin-  
<sup>487</sup> derived P and C are a function of the mass of mucin degraders,  $D_M$  (B and  
<sup>488</sup> NBFD), such that,

$$\dot{C}(t) = 0.75 \frac{D_M(t)}{D_M(t) + K_M W_v(t)} \dot{M} \quad (2)$$

$$\dot{P}(t) = 0.2 \frac{D_M(t)}{D_M(t) + K_M W_v(t)} \dot{M} \quad (3)$$

<sup>489</sup> where  $C$ ,  $P$ ,  $D_M$  are in mass units and the over dot indicates a rate (e.g.  $\text{g}$   
<sup>490</sup>  $\text{d}^{-1}$ ),  $t$  is time and  $W_v$  is the volume of water in the model compartment.  $K_M$   
<sup>491</sup> ( $\text{g l}^{-1}$ ) is chosen such that if  $D_M \ll K_M W_v$  then there is minimal breakdown  
<sup>492</sup> of mucin and if  $D_M \gg K_M W_v$  there is maximal breakdown. The smaller the  
<sup>493</sup> value of  $K_M$  the more breakdown there will be at low concentrations of mucin  
<sup>494</sup> degraders. We set  $K_M=0.5 \text{ g l}^{-1}$  based on preliminary model simulations.

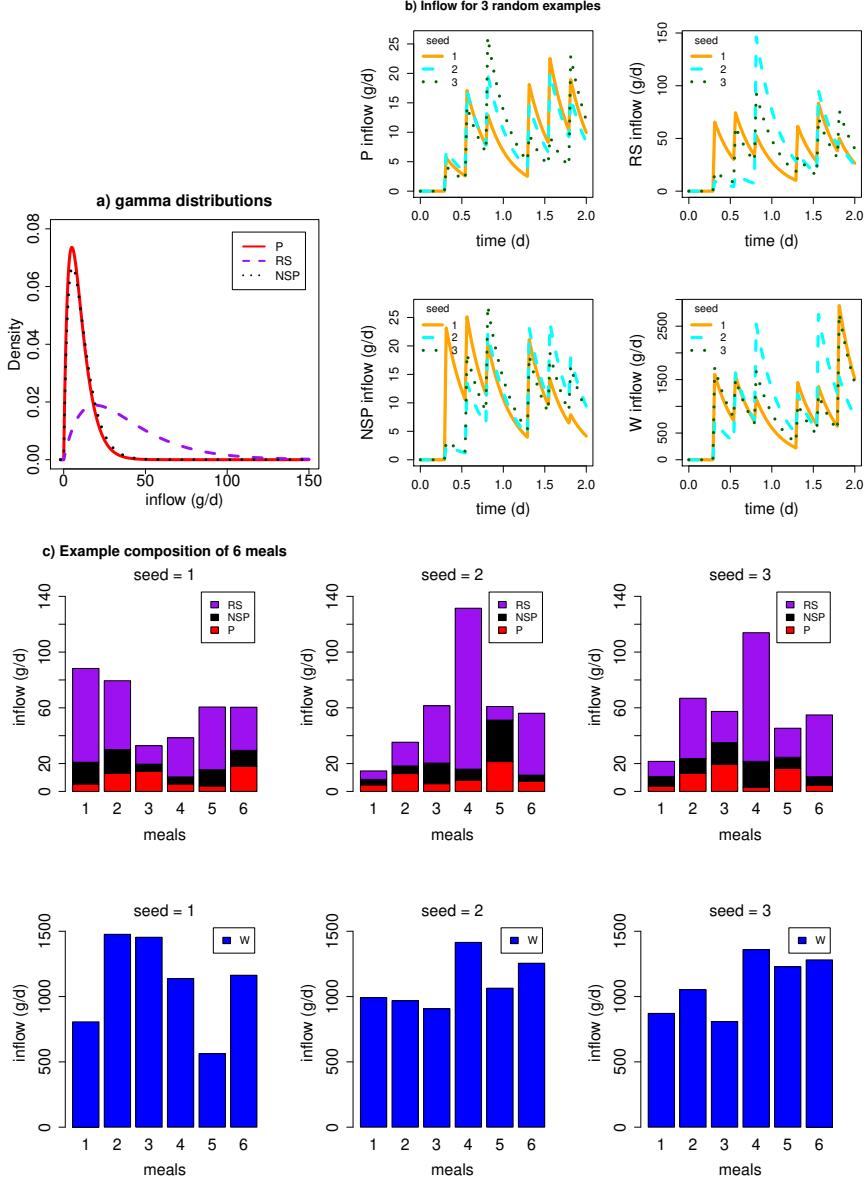


Figure 9: a) Gamma distribution from which random values are drawn to generate the composition of each meal (note water is not shown due to the large difference in magnitude between water and dietary substrates). b) The substrate inflow time series to the proximal colon after passing through the small intestine. Examples shown for 3 stochastic simulations starting with different seeds. c) Barplots showing the composition of 6 meals over 2 days for 3 different stochastic simulations.

495 **Absorption by host**

496 SCFA and water are both absorbed by the host through the gut wall; over 95%  
497 of SCFA (Topping and Clifton, 2001) and approximately 90% of incoming water  
498 is absorbed (Phillips and Giller, 1973). Experiments by Ruppin et al. (1980)  
499 found that the absorption rates of SCFA to be approximately  $0.4 \text{ h}^{-1}$  (i.e.  $9.6$   
500  $\text{d}^{-1}$ ) with little difference in rates between the different SCFA (Ruppin et al.  
501 (1980), Topping and Clifton (2001)).

502 We can estimate mathematically the specific water absorption rate required  
503 to give 90% absorption of inflowing water for a given number of compartments  
504 in the colon ( $N$ ) and a given transit time,  $T_t$ , using

$$a_W = \frac{16.95 - 9.72N + 1.77N^2}{T_t} \quad (4)$$

505 (see Supp. Info. Section 1.3 for the derivation). As a rough estimation, for  
506 a 3 compartment model with a transit time 1-1.5 days, gives  $a_W \approx 3 \text{ d}^{-1}$   
507 (Supp. Info. Fig. S1a). Given this will not be significantly affected by the  
508 microbial model (microbial uptake/production of water is small) this is a robust  
509 estimation.

510 To estimate the value of the specific absorption rate of SCFA,  $a_Z$ , we used  
511 a simple model (see Supp. Info. sections 1.1 and 1.4). Estimating the value  
512 of the specific absorption rate of SCFA based on the values of SCFA given in  
513 the verification criteria and given our estimate for  $a_W$  we found that it was  
514 necessary for the specific absorption rate to change along the colon (see Supp.  
515 Info. section 1.4). The best estimates were given by  $a_Z$  values of 25.2, 4.2 and  
516  $9.2 \text{ d}^{-1}$  in the proximal, transverse and distal colon respectively. However, in  
517 the interests of a robust model (i.e. the fewer parameter values, the better)  
518 we made the decision to use one value for  $a_Z$ . Since the experimental value of  
519  $9.6 \text{ d}^{-1}$  compares well with our estimate in the distal colon we set  $a_Z=9.6 \text{ d}^{-1}$   
520 throughout. It should be noted though that our model results could potentially  
521 be improved by varying  $a_Z$  between model compartments.

522 **pH**

523 Calculating pH in our model is not straightforward due to a lack of necessary  
524 state variables as well as pH buffering via secretions from the host. However,  
525 observations tell us the pH in the colon goes from 5.7 in the proximal, 6.2 in  
526 the transverse and 6.6 in the descending colon and TSCFA in these regions is  
527 around 123 mM, 117 mM and 80 mM respectively (Cummings et al., 1987).  
528 Therefore an approximate approach is to simply make pH a function of TSCFA.  
529 Fitting a line through the above points gives us the following relationship

$$\text{pH} = 8.02 - 0.0174 \times \text{TSCFA}. \quad (5)$$

530 which we further limit by setting the minimum and maximum pH values at 5  
531 and 8 respectively i.e. if the TSCFA values give predicted pH outside of this  
532 range (Fig. 10).

533 The impact of pH on microbial growth is modelled via a pH limitation func-  
534 tion whereby there is a range over which there is no limit on growth but outside  
535 of this range the growth rate decreases linearly to reach zero at the specified  
536 outer limits. Thus there are 4 parameters used to describe the pH tolerance – 2  
537 for the inner range where there is no limit on growth and 2 for the outer range  
538 outside which there is no growth – an example is shown in Fig. 10. The pH  
539 tolerance range for each microbial group is specified under the entry ‘pHcorners’  
540 in the data frame for each group and shown in Supp. Info. section 3.

541 **Fecal outflow**

542 Fecal outflow ( $\text{g d}^{-1}$ ) at time,  $t$ , is given by  $m_d(t)V_d$  where  $m_d(t)$  is the mass  
543 in the distal colon (i.e. microbes, unconsumed substrate, microbial metabolites  
544 and water) and  $V_d$  is the specific wash out rate from the colon (the inverse of  
545 the time spent in the distal colon). For continuous outflow (as is used in most  
546 gut models) we compute the specific wash out rate from each compartment by  
547 assuming the fraction of time spent in compartment,  $i$ , is proportional to its

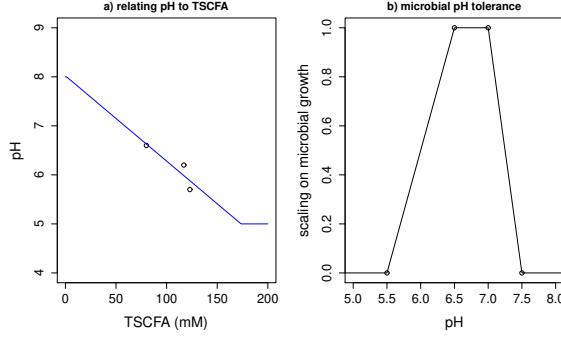


Figure 10: a) Relating pH to TSCFA using Eq. 5 and data from (Cummings et al., 1987). b) Example of microbial tolerance to pH. A pH tolerance function of this form is specified individually for each microbial group in our model.

548 volume fraction, thus

$$T_i = \frac{v_i}{v_{colon}} T_t \quad (6)$$

549 where  $v_i$  is the volume of compartment  $i$  and  $v_{colon}$  is the total volume of the  
550 colon. The specific wash out rate is then  $V_i = 1/T_i$ .

551 If we introduce bowel movements then, assuming the distal colon is approx-  
552 imately emptied for each bowel movement, the total transit time is given by

$$T_t = \sum_{i=1}^2 T_i + \frac{1}{N_{BM}} \quad (7)$$

553 where  $N_{BM}$  is the number of bowel movements per day. For example, using vol-  
554 ume measurements (Table 1B) and assuming a total transit time of 1 day would  
555 mean about 45% of the transit time is spent in the proximal and transverse  
556 colon and about 55% of the day spent in the distal, which would be similar to 2  
557 bowel movements per day. In model experiments where we vary the number of  
558 bowel movements per day we also change the time spent in the rest of the colon  
559 since we assume increased bowel movements are indicative of a general increase  
560 in passage rate. We estimate the wash out rate from the colon during a bowel  
561 movement,  $V_{BM}$ , by

$$V_{BM} = -\frac{\ln(f_d)}{\Delta t_{BM}} \quad (8)$$

562 where  $f_d$  is the fraction of mass left in the distal colon after the bowel movement  
563 and  $\Delta t_{BM}$  is the time taken for the bowel movement (d). For example, if a bowel  
564 movement takes 10 minutes to remove 90% of the contents of the distal colon  
565 then  $V_{BM}$  is  $332 \text{ d}^{-1}$ . This is not affected by the number of bowel movements  
566 per day.

## 567 Acknowledgments

568 We thank the Scottish Goverment's Rural and Environment Science and Ana-  
569 lytical Services Division (RESAS) for funding this research.

## 570 References

- 571 Bown, R., Gibson, J., Sladen, G., Hicks, B., and Dawson, A. (1974). Effects  
572 of lactulose and other laxatives on ileal and colonic ph as measured by a  
573 radiotelemetry device. *Gut*, 15(12):999–1004.
- 574 Capuano, E., Oliviero, T., Fogliano, V., and Pellegrini, N. (2018). Role of the  
575 food matrix and digestion on calculation of the actual energy content of food.  
576 *Nutrition Reviews*, 76(4):274–289.
- 577 Chung, W. S., Walker, A. W., Bosscher, D., Garcia-Campayo, V., Wagner, J.,  
578 Parkhill, J., Duncan, S. H., and Flint, H. J. (2020). Relative abundance of  
579 the *Prevotella* genus within the human gut microbiota of elderly volunteers  
580 determines the inter-individual responses to dietary supplementation with  
581 wheat bran arabinoylan-oligosaccharides. *BMC Microbiology*, 20(1):1–4.
- 582 Cremer, J., Arnoldini, M., and Hwa, T. (2017). Effect of water flow and chemical  
583 composition in the human colon. *PNAS*, 114(25):6438–6443.
- 584 Cremer, J., Segota, I., y Yang, C., Arnoldini, M., Sauls, J., Zhang, Z., Gutierrez,  
585 E., Groisman, A., and Hwa, T. (2016). Effect of flow and peristaltic mixing  
586 on bacterial growth in a gut-like channel. *PNAS*, 113:11414–11419.

- 587 Crost, E., Tailford, L., Gall, G. L., Fons, M., Henrissat, B., and Juge, N. (2013).  
588 Utilisation of mucin glycans by the human gut symbiont *ruminococcus gnavus*  
589 is strain-dependent. *PLoS ONE*, 8(10):e76341.
- 590 Cummings, J. H., Pomare, E., Branch, W., Naylor, C., and Macfarlane, G.  
591 (1987). Short chain fatty acids in human large intestine, portal, hepatica and  
592 venous blood. *Gut*, 28:1221–1227.
- 593 Duncan, S. H., Belenguer, A., Holtrop, G., Johnstone, A., Flint, H. J., and  
594 Lobley, G. E. (2007). Reduced dietary intake of carbohydrates by obese sub-  
595 jects results in decreased concentrations of butyrate and butyrate-producing  
596 bacteria in feces. *Applied and Environmental Microbiology*, 73(4):1073–1078.
- 597 Duncan, S. H., Russell, W. R., Quartieri, A., Rossi, M., Parkhill, J., Walker,  
598 A. W., and Flint, H. J. (2016). Wheat bran promotes enrichment within the  
599 human colonic microbiota of butyrateproducing bacteria that release ferulic  
600 acid. *Environmental Microbiology*, 18(7):2214–2225.
- 601 Florin, T., Neale, G., Gibson, G., Christl, S., and Cummings, J. (1991).  
602 Metabolism of dietary sulphate: absorption and excretion in humans. *Gut*,  
603 32:766–773.
- 604 Hamer, H. M., Jonkers, D., Venema, K., Vanhoutvin, S., Troost, F., and Brum-  
605 mer, R.-J. (2008). The role of butyrate on colonic function. *Alimentary  
606 pharmacology & therapeutics*, 27(2):104–119.
- 607 Kettle, H., Holtrop, G., Louis, P., and Flint, H. J. (2018). micropop: Mod-  
608 elling microbial populations and communities in r. *Methods in Ecology and  
609 Evolution*, 9(2):399–409.
- 610 Kettle, H., Louis, P., Holtrop, G., Duncan, S. H., and Flint, H. J. (2015).  
611 Modelling the emergent dynamics and major metabolites of the human colonic  
612 microbiota. *Enviromental Microbiology*, 17(5):1615–1630.

- 613 LaBouyer, M., Holtrop, G., Horgan, G., Gratz, S. W., Belenguer, A., Smith,  
614 N., Walker, A. W., Duncan, S. H., Johnstone, A. M., Louis, P., et al. (2022).  
615 Higher total faecal short-chain fatty acid concentrations correlate with in-  
616 creasing proportions of butyrate and decreasing proportions of branched-chain  
617 fatty acids across multiple human studies. *Gut Microbiome*, 3(e2):1–14.
- 618 Lewis, S. and Heaton, K. (1997). Increasing butyrate concentration in the distal  
619 colon by accelerating intestinal transit. *Gut*, 41:245–251.
- 620 Louis, P. and Flint, H. J. (2017). Formation of propionate and butyrate by the  
621 human colonic microbiota. *Environmental Microbiology*, 19(1):29–41.
- 622 Louis, P., Hold, G. L., and Flint, H. J. (2014). The gut microbiota, bacterial  
623 metabolites and colorectal cancer. *Nature reviews microbiology*, 12(10):661–  
624 672.
- 625 Mikolajczyk, A. E., Watson, S., Surma, B. L., and Rubin, D. T. (2015). As-  
626 sessment of tandem measurements of ph and total gut transit time in healthy  
627 volunteers. *Clinical and Translational Gastroenterology*, 6(7).
- 628 Moorthy, A. S., S.P.J., B., M., K., and H.J., E. (2015). Continuous model of  
629 carbohydrate digestion and transport processes in the colon. *PLoS ONE*,  
630 10(12):e0145309.
- 631 Morrison, D. J. and Preston, T. (2016). Formation of short chain fatty acids by  
632 the gut microbiota and their impact on human metabolism. *Gut microbes*,  
633 7(3):189–200.
- 634 Munoz-Tamayo, R., Laroche, B., Walter, E., Dore, J., and Leclerc, M. (2010).  
635 Mathematical modelling of carbohydrate degradation by human colonic mi-  
636 crobiota. *Journal of Theoretical Biology*, 266(1):189 – 201.
- 637 Phillips, S. and Giller, J. (1973). The contribution of the colon to electrolyte  
638 and water conservation in man. *J Lab Clin Med*, 81(5):733–746.

- 639 Ravcheev, D. A. and Thiele, I. (2017). Comparative genomic analysis of the  
640 human gut microbiome reveals a broad distribution of metabolic pathways for  
641 the degradation of host-synthetized mucin glycans and utilization of mucin-  
642 derived monosaccharides. *Frontiers in Genetics*, 8:111.
- 643 Reichardt, N., Duncan, S. H., Young, P., Belenguer, A., Leitch, C. M., Scott,  
644 K. P., Flint, H. J., and Louis, P. (2014). Phylogenetic distribution of three  
645 pathways for propionate production within the human gut microbiota. *The  
646 ISME Journal*, 8(6):1323–1335.
- 647 Rios-Covian, D., Ruas-Madiedo, P., Margolles, A., Gueimonde, M., de los  
648 Reyes-Gavilan, C. G., and Salazar, N. (2016). Intestinal short chain fatty  
649 acids and their link with diet and human health. *Front. Microbiol.*, 17.
- 650 Ruppin, H., Bar-Maeir, S., Soergel, K., Wood, C., and Schmitt, M. (1980). Ab-  
651 sorption of short chain fatty acids by the colon. *Gastroenterology*, 78(6):1500–  
652 1507.
- 653 Salonen, A., Lahti, L., Saloja, J., Holtrop, G., Korpela, K., Duncan, S. H.,  
654 Date, P., Farquharson, F., Johnstone, A. M., Lobley, G. E., Louis, P., Flint,  
655 H. J., and de Vos, W. M. (2014). Impact of diet and individual variation on  
656 intestinal microbiota composition and fermentation products in obese men.  
657 *The ISME Journal*, pages 1–13.
- 658 Slavin, J. L., Brauer, P. M., and Marlett, J. A. (1981). Neutral Detergent Fiber,  
659 Hemicellulose and Cellulose Digestibility in Human Subjects. *The Journal of  
660 Nutrition*, 111(2):287–297.
- 661 Smith, N., Shorten, P., and Altermann, E. (2021). Examination of hydrogen  
662 cross-feeders using a colonic microbiota model. *BMC Bioinformatics*, 22(3).
- 663 Stephen, A. and Cummings, J. (1980). The microbial contribution to human  
664 fecal mass. *J. Medical Microbiology*, 13(1):45–56.

- 665 Sung, J., Kim, S., Cabatbat, J., Jang, S., Jin, Y., Jung, G. Y., Chia, N., and  
666 Kim, P. (2017). Global metabolic interaction network of the human gut micro-  
667 biota for context-specific community-scale analysis. *Nature communications*.
- 668 Topping, D. and Clifton, P. (2001). Short-chain fatty acids and human colonic  
669 function: roles of resistant starch and nonstarch polysaccharides. *Physiological  
670 Review*, 81(3):1031–1064.
- 671 Walker, A. W., Ince, J., Duncan, S. H., Webster, L. M., Holtrop, G., Ze, X.,  
672 Brown, D., Stares, M. D., Scott, P., Bergerat, A., Louis, P., McIntosh, F.,  
673 Johnstone, A. M., Lobley, G. E., Parkhill, J., and Flint, H. J. (2011). Domi-  
674 nant and diet-responsive groups of bacteria within the human colonic micro-  
675 biota. *The ISME Journal*, 5:220–230.
- 676 Wang, S. P., Rubio, L. A., Duncan, S. H., Donachie, G. E., Holtrop, G., Lo,  
677 G., Farquharson, F. M., Wagner, J., Parkhill, J., Louis, P., Walker, A. W.,  
678 and Flint, H. J. (2020). Pivotal roles for ph, lactate, and lactate-utilizing  
679 bacteria in the stability of a human colonic microbial ecosystem. *mSystems*,  
680 5(5):e00645–20.
- 681 Wu, G., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y., Keilbaugh, S., Bewtra,  
682 M., Knights, D., Walters, W., Knight, R., and Sinha, R. (2011). Linking long-  
683 term dietary patterns with gut microbial enterotypes. *Science*, 334(6052):105–  
684 108.
- 685 Yao, C. K., Muir, J. G., and Gibson, P. R. (2016). Review article: insights  
686 into colonic protein fermentation, its modulation and potential health impli-  
687 cations. *Alimentary Pharmacology and Therapeutics*, 43(2):181–196.