

# 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 14, 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 **Introduction**

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

32 When developing a model, a number of assumptions about the system are  
33 made in order to reduce complexity/dimensionality so that the model is easier to  
34 parameterise, run and analyse. Some modellers choose to reduce the microbial  
35 complexity and focus on the physics of the gut (e.g. Cremer et al. (2016),  
36 Cremer et al. (2017)), some try to achieve a balance of both (e.g. Munoz-Tamayo  
37 et al. (2010)) and some choose to develop the microbial community (e.g. Smith  
38 et al. (2021)). The model described here focuses on the microbial community  
39 dynamics and on interactions with the host, with a fairly simple model of the  
40 colon. We include the simulation of ‘meals’ (of random composition and size)  
41 arriving at the colon and look at the effects of bowel movements, both of which,  
42 as far as we are aware, have not been previously incorporated into such models.

43 Having developed a complex model of human gut microbiota in a fermentor  
44 system (Kettle et al., 2015), and publicly available software (microPop - an  
45 R package for modelling microbial communities (Kettle et al., 2018)) we now  
46 incorporate this 10-group microbial ecosystem model (Table 1) into a model of  
47 the human gut in order to simulate the effects of diet and host on the microbial  
48 composition and subsequent short chain fatty acid (SCFA) production.

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

59 Due to its shape within the body, the colon is commonly divided into 3  
60 different regions - the proximal, transverse and distal sections running from  
61 beginning to end (Fig. 1A and B). The availability of substrate, microbial  
62 growth and hence pH vary along the colon, therefore, although our model is not  
63 spatial we simulate these three regions explicitly, with flow from one to another.  
64 Furthermore, as well as incorporating varying substrate inflow in the form of  
65 meals we also add in the release of mucins along the length of the colon which  
66 can be microbially broken down to release proteins and carbohydrates, allowing  
67 for further microbial growth away from the beginning of the colon where the  
68 substrates enter. A graphical summary of the model is shown in Fig. 2, the  
69 microbial functional groups are shown in Table 1 and the model state variables  
70 are summarised in Table 2.

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

- 73        1. Total SCFA (TSCFA) concentration in the proximal, transverse and distal  
74        compartments should be around 123, 117 and 80 mM respectively accord-  
75        ing to sudden death human autopsies (Cummings et al., 1987)
- 76        2. Acetate:Propionate:Butyrate ratios are similar (around 3:1:1) in all regions  
77        of the colon and around 60:20:20 mM (Cummings et al., 1987)
- 78        3. Over 95% of SCFA are absorbed by the host (Topping and Clifton, 2001)
- 79        4. Approx. 90% of incoming water is absorbed by the host (Phillips and  
80        Giller, 1973)
- 81        5. pH in the proximal, transverse and distal compartments should be around  
82        5.7, 6.2 and 6.6 respectively (Cummings et al. (1987), and telemetry data  
83        from Mikolajczyk et al. (2015), Bown et al. (1974))
- 84        6. Normal daily fecal output in Britain is  $100\text{-}200 \text{ g d}^{-1}$  of which 25-50 g is  
85        solid matter (i.e.  $50\text{-}175 \text{ g d}^{-1}$  is water). Bacteria make up about 55%  
86        of the solid matter i.e.  $14\text{-}28 \text{ g d}^{-1}$  of microbes emitted (Stephen and  
87        Cummings, 1980).
- 88        7. TSCFA concentration decreases with transit time (Lewis and Heaton,  
89        1997)

90        After model verification we examine the effects of including meals, bowel move-  
91        ments and fixed/varying pH into the model. We then use the model to look at  
92        how carbohydrate composition (based on the fractions of resistant starch (RS)  
93        and non-starch polysaccharides (NSP)) and total carbohydrate affect the mi-  
94        crobial community and SCFA composition. The simulations are then compared  
95        with in-vivo data from human volunteer experiments.

96        Although gut microbiota are highly complex and not fully understood, here  
97        we show that it is nonetheless possible to develop predictive models of key com-  
98        ponents of this ecological system. An important goal of our modelling is to aid  
99        and inform the interpretation of data obtained, mostly from faecal samples, in

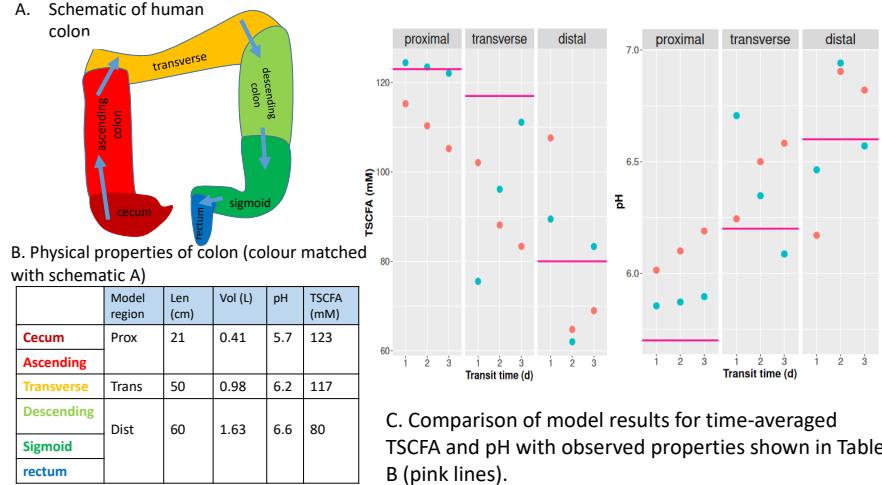


Figure 1: A. Colon schematic, B. Table of typical values for physical properties (length, volume, pH and TSCFA) and C. plots of summarised model simulations for average TSCFA and pH for comparison with typical values. Red dots show results from meals-inflow averaged over 4 random seeds; blue dots show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

100 studies on diet and health in humans. Our results show promise and we believe  
 101 this model represents a significant step forward in analysing this highly complex  
 102 system. We refer to the model as “microPopGut” and to aid future research the  
 103 code is available as an R-package on github (<https://github.com/HelenKettle/microPopGutCode>)  
 104 and instructions on how to use the package are given in the supplementary file  
 105 ‘getStartedWithMicroPopGut.pdf’.

## 106 Results

### 107 Standard Model

108 The model settings which give the best fit to our criteria are shown in Table  
 109 4 (colon parameters and dietary inflow). The microbial group parameters are  
 110 listed in Supp. Info. (section 3). These define our default model. From this we  
 111 investigate the effects of different model configurations, e.g. with/without bowel  
 112 movements, meals and variable pH, for a range of transit times. Simulations  
 113 with meals have a random component therefore the model is run for a number of

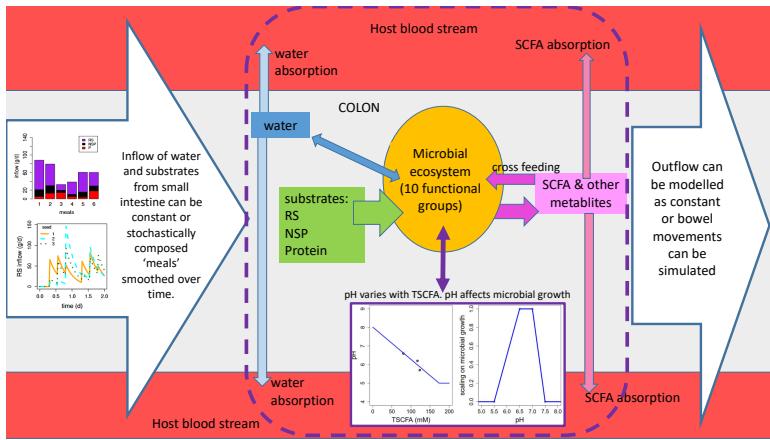


Figure 2: Model system with the microbial ecosystem comprising 10 microbial functional groups (Table 1) which consume substrates (RS, NSP and protein) and water. The microbes produce metabolites some of which are consumed by other MFGs ('cross-feeding'). SCFA and water are absorbed through the colon 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>

<sup>114</sup> different starting seed values. Due to the random fluctuations these simulations  
<sup>115</sup> will not reach steady state therefore the summary values are taken as the mean  
<sup>116</sup> from day 7 (to remove the effect of the initial conditions) to the end of the  
<sup>117</sup> simulation (28 days) and are averaged over multiple seeds.

<sup>118</sup> Table 3 gives summary results of the model simulations without bowel move-  
<sup>119</sup> ments but with varying pH for each bowel region. Fig. 3 shows results from  
<sup>120</sup> more simulations but for the distal colon only. Fig. 3a shows that although  
<sup>121</sup> bowel movements make a difference to the total biomass and the TSCFA they  
<sup>122</sup> do not have a large effect on the community composition or the SCFA ratios.  
<sup>123</sup> Thus in the interests of model simplicity we decide to not include bowel move-  
<sup>124</sup> ments in later simulations. However, varying pH with TSCFA can be seen to

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

125 make a large difference to the microbial community (Fig. 3b) and also improves  
 126 the SCFA ratios with respect to our verification criteria. The addition of meals  
 127 makes a significant difference which increases with increasing transit time (Fig.  
 128 3c). In Fig. 4 the time series output from the model shows how the meals-  
 129 inflow allows the community to experience large shifts over time (on a much  
 130 longer time scale than the variations in the input), as opposed to the fixed state  
 131 approached using a constant substrate inflow.

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

139 Regarding Table 3, for some criteria, e.g. pH, the continuous inflow setting

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

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
transit time						
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

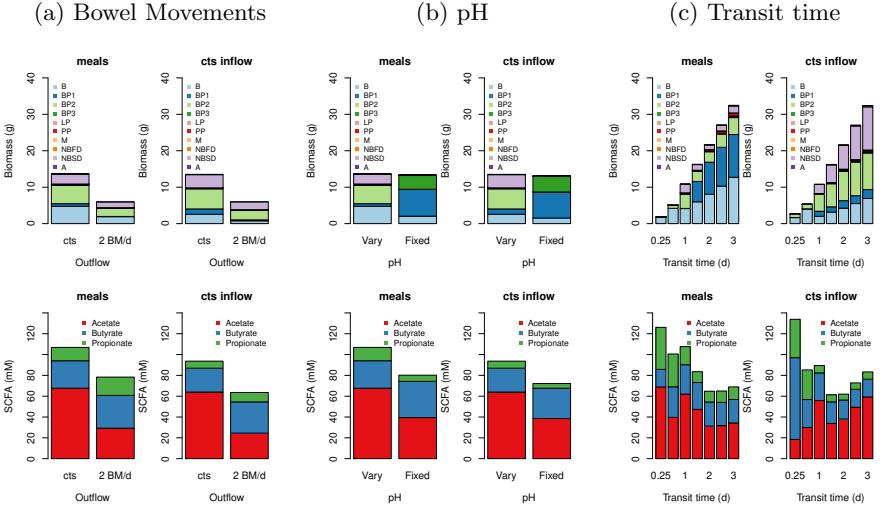


Figure 3: Summary results (averaged over days 7-28 and over random seeds) for the distal compartment for continuous inflow or fluctuating inflow (i.e. ‘meals’) for continuous outflow from colon or for 2 bowel movements per day (‘2 BM/d’). The RS fraction is 0.78 (i.e. 78% of the dietary carbohydrate 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.

## 158 Model Experiments

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

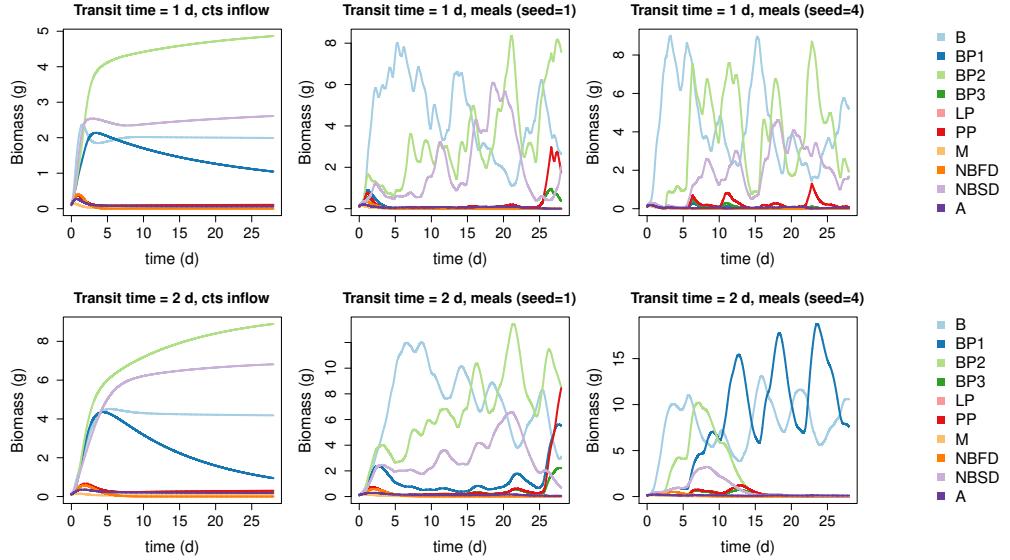


Figure 4: Simulation results for the distal compartment for continuous inflow (first plot on each row) or fluctuating inflow (i.e. ‘meals’) for transit times of 1 d (top row) and 2 d (bottom row) and for 2 random seeds. Modelled pH varies with TSCFA and the RS fraction is 0.78. There are no bowel movements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

#### **171 Effects of total dietary carbohydrate**

172 In this model experiment we investigate the effects of decreasing carbohydrate  
 173 on the microbial community. Here we compare our results qualitatively with  
 174 the human dietary study of Duncan et al. (2007) which explored the impacts of  
 175 carefully controlled decreases in carbohydrate intake upon weight loss and mi-  
 176 crobial fermentation products in obese subjects using 3 diets – a maintenance  
 177 (M) diet, a high protein, moderate carbohydrate diet (HPMC) and a high pro-  
 178 tein, low carbohydrate diet (HPLC) (see Fig. 5 for details). This is of course,  
 179 the composition for ingested food, which is not easily translated into substrate  
 180 concentrations entering the colon. However, we can look at the general trends  
 181 in SCFA and microbial composition with changing colonic carbohydrate intake  
 182 rate. Thus, in these model experiments we keep protein inflow to the colon at  
 183  $10 \text{ g d}^{-1}$  (our default value) and then increase inflowing non-digestable (ND)  
 184 carbohydrate from  $10 \text{ g d}^{-1}$  to  $60 \text{ g d}^{-1}$  in  $10 \text{ g d}^{-1}$  intervals. To include the

185 effects of different ND-carbohydrate composition we run the model for an resis-  
186 tant starch (RS) fraction of either 0.2 or 0.78 (the default value), with non-starch  
187 polysaccharides (NSP) making up the remaining ND-carbohydrate in each case.  
188 Although subject to large uncertainties, we estimate the RS fractions for the  
189 Duncan et al. (2007) experiments of 0-0.6 (M diet), 0-0.68 (HPMC) and 0-0.12  
190 (HPLC) (based on RS is 0–20% of ingested starch (Capuano et al., 2018) and  
191 bio-available NSP is 75% of ingested NSP (Slavin et al., 1981)). Due to the  
192 low fibre nature of many of these simulations we run the model with a slightly  
193 longer transit time of 1.5 d and for both continuous inflow and meals.

194 Fig. 6 shows the SCFA results from our model experiment and Fig. 5 shows  
195 the results from the in vivo experiment. It is clear, from both the model and  
196 in vivo results that the proportion of butyrate increases as the amount of ND-  
197 carbohydrate in the diet increases. Furthermore, both model and in vivo results  
198 show an increase in TSCFA with ND-carbohydrate intake rate. Since Duncan  
199 et al. (2007) also look at the relationship between butyrate concentration and  
200 grams of carbohydrate eaten per day, we plot butyrate against carbohydrate en-  
201 tering the colon (Fig. 7) to compare with their Fig. 1. In both cases, butyrate  
202 concentration increases with incoming carbohydrate. Furthermore, as seen in  
203 both the model and the data, the percentage of butyrate increases with carbo-  
204 hydrate intake (Fig. 7). Analysis of 10 human studies involving 163 subjects  
205 has shown a highly significant increase in percentage butyrate with increasing  
206 total SCFA concentration in faecal samples (LaBouyer et al., 2022).

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

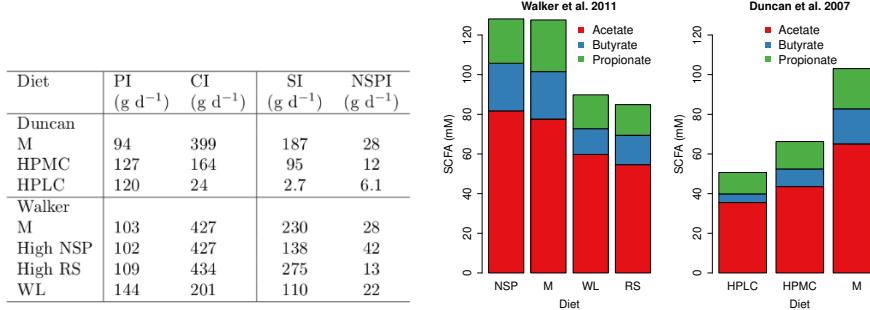


Figure 5: Table on left shows the dietary intake for two human studies (Duncan et al. (2007) and Walker et al. (2011)). PI, CI, SI and NSPI refer to ingested dietary protein, carbohydrate, starch and NSP. Note, starch value for the high RS diet in the Walker et al. (2011) study included 26 g commercial RS. Bar plots show SCFA data from these studies. The bars in the plots have been ordered to show increasing RS fraction (estimated by SI/CI) for the Walker study (for comparison with Fig. 8) and increasing carbohydrate for the Duncan study (for comparison with Fig. 6).

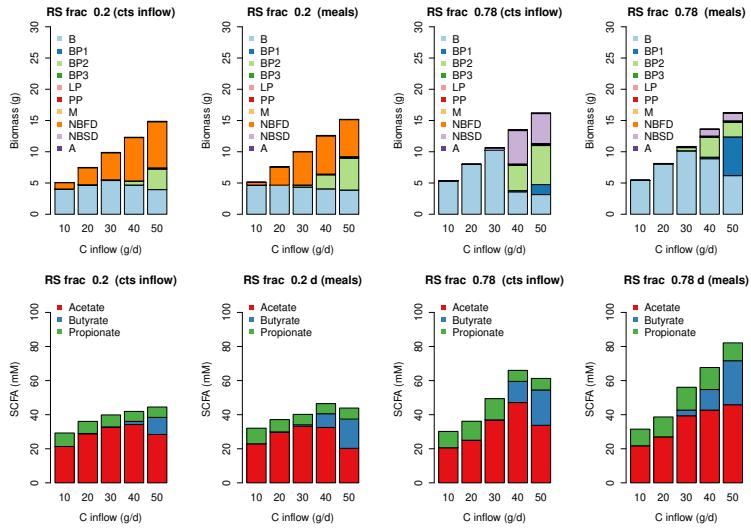


Figure 6: Simulated Biomass and SCFA results for increasing carbohydrate inflow. Simulations are run with continuous substrate inflow (cts) and with ‘meals’ for a transit time of 1.5 days. The results are the average over the last 3 weeks of a 28 day simulation and ‘meals’ is the average over 4 stochastically-generated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

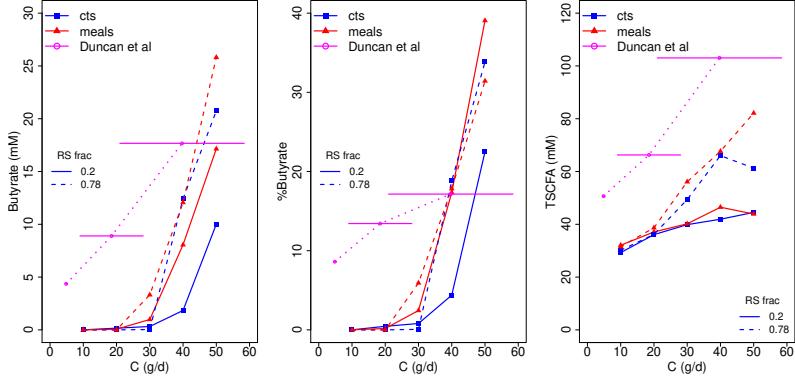


Figure 7: Plot of modelled butyrate, %butyrate and TSCFA against grams of carbohydrate entering the colon each day. Data from Duncan et al. (2007) is shown in magenta - due to uncertainties in converting ingested starch to RS entering the colon there are large error bars on the amount of C (g/d). Error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

### Effects of carbohydrate composition

Here we use the model to simulate the effects of changing carbohydrate composition on the microbial community composition by changing the ratio of RS to NSP whilst keeping the same amount of total incoming carbohydrate. Fig. 8 show a summary of the model results. Although there are differences between the continuous inflow/meals, and also for the different transit times (1 d and 3 d), the modelled trends are generally similar, showing a significant shift in community as the fraction of RS increases, an increase in TSCFA and changes in the SCFA ratios. We compare our results with a human dietary study (Walker et al. (2011), Salonen et al. (2014) and references therein) examining the impact of switching the major type of ND-carbohydrate from wheat bran (NSP) to resistant starch. Volunteers were provided successively with a maintenance diet, diets high in RS or NSPs and a reduced carbohydrate weight loss (WL) diet, over 10 weeks (Fig. 5).

There are large discrepancies between the SCFA predicted by our model (Fig. 8) and the measured SCFA data (Fig. 5). Our model predicts an increase in TSCFA as proportion of RS increases whereas total fecal SCFA were significantly

lower for the RS and WL diets compared to the other two diets (in which NSP is higher). One possible explanation is that fermentation of RS occurs in more proximal regions of the colon compared with NSP fibre fermentation, such that there is greater absorption of the SCFA products. A second possibility, also likely, is that transit times were longer for the RS diet than for the NSP diet, which we predict would result in decreased SCFA concentrations. In our model the effect of the RS fraction on TSCFA is greater than the effect of transit time so we do not see this in Fig. 8.

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

## Discussion

The development of a complex model of the microbial community in the human colon, whose simulations compare well with data, represents a significant step forward. Previous models have been based on simpler microbial models (e.g. Cremer et al. (2017), Munoz-Tamayo et al. (2010), Moorthy et al. (2015)), or have not shown such a good agreement with data (e.g. Smith et al. (2021)). Our previous complex model community consisted of 10 functional groups, but the model was designed only to simulate continuous culture conditions in a chemostat (Kettle et al., 2015). Translating this 10-group model into an *in vivo* setting

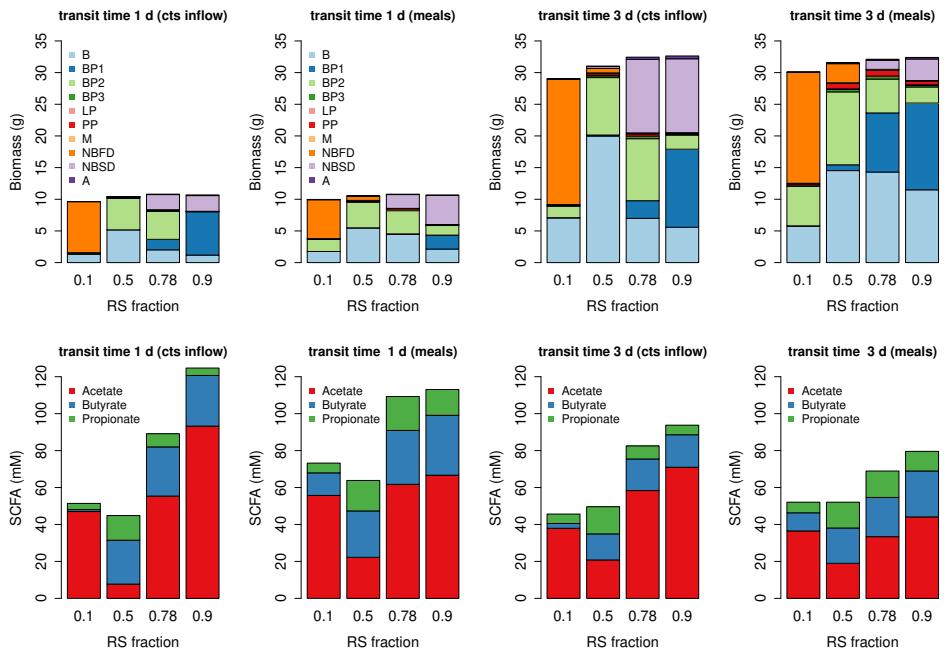


Figure 8: Biomass and SCFA results for changing the RS fraction of inflowing carbohydrate with continuous substrate inflow ('cts inflow') and with 'meals'. Protein and carbohydrate inflow are 10 and 50 g d<sup>-1</sup> respectively. The results are the average over the last 3 weeks of a 28 day simulation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

258 has required introducing multiple gut compartments, and the absorption of wa-  
259 ter and SCFA, followed by comparison with generally observed characteristics  
260 of the system. We were then able to use this model to examine the predicted  
261 impact of changes in the amount and type of non-digestible carbohydrate (fibre)  
262 present in human diets upon concentrations of fermentation products (SCFA)  
263 in different gut compartments and in stool. At the same time, we predict the  
264 likely impact of dietary changes and variations in gut transit upon microbiota  
265 composition and fermentation products. The model must be regarded as work  
266 in progress particularly with respect to microbiota composition. Predictions can  
267 however become improved and refined as more information becomes available  
268 in time.

269 Assignments of microbial taxa to our ten functional groups were based ini-  
270 tially on evidence from cultured isolates. These assignments have since been  
271 supported and greatly extended by analysis of genes diagnostic for different fer-  
272 mentation pathways within genomes and metagenomes (Reichardt et al., 2014)  
273 and by molecular detection of species enriched within the community by defined  
274 growth substrates in chemostat experiments (Duncan et al., 2016) and dietary  
275 intervention studies (Salonen et al., 2014). Nevertheless, these assignments in-  
276 evitably remain provisional and incomplete and we do not claim that the model  
277 predictions can be made precise at a phylogenetic level. More emphasis is placed  
278 in our model on the prediction of metabolic outputs based on microbial transfor-  
279 mations and interactions. While there is relatively little phylogenetic overlap for  
280 example between producers of propionate and butyrate (Reichardt et al. (2014),  
281 Louis and Flint (2017)) there are many cases where individual species are known  
282 to use multiple alternative substrates as energy sources, which complicates as-  
283 signments. For this reason, more weight was given to fermentation pathways  
284 than to substrate preferences in defining the functional groups. However, it  
285 would also be possible to define completely different groupings that relate to  
286 other outputs (e.g. bile acid metabolism, or vitamin/micronutrient supply) in  
287 order to address specific questions. Furthermore, it may well be worthwhile to

288 increase the number of functional groups in the future. The large B group for  
289 example currently includes members of the Bacteroidetes phylum, but its char-  
290 acteristics are mainly based on well-studied members of the *Bacteroides* genus.  
291 We know that *Prevotella* is another highly abundant genus of *Bacteroidetes* in  
292 the human colon, but the two genera tend not to co-occur at high levels in the  
293 same individuals (Wu et al. (2011), Chung et al. (2020)). Less is known about  
294 human colonic *Prevotella*, for which there are relatively few cultured representa-  
295 tives, making it premature to create a separate grouping, but this would clearly  
296 be desirable in the future as their prevalence is reported to affect health and  
297 responses to dietary intervention. In future it should become possible to define  
298 the relative abundance of functional groups (MFGs) and their relationship to  
299 phylogeny directly from genomic and metagenome analysis, by examining genes  
300 diagnostic for particular pathways and functions (e.g. Reichardt et al, 2014).

301 The parameter values for the microbial groups used in our model are from  
302 the intrinsic data frames in the microPop package (the only changes are to  
303 LactateProducers). Although the work presented here did not attempt to fit  
304 particular parameters to data, as we focussed on expanding the scope of the  
305 model (i.e. changing the environment from fermentor to colon), these values  
306 are easy to alter, e.g. Wang et al. (2020) changed many of these parameters to  
307 achieve a better model fit to their data. As well as adjusting the parameters  
308 for each group to represent inter-individual variation, groups can also be easily  
309 added or removed from the model through the input argument ‘microbeNames’.  
310 Furthermore, it is also possible to include any number of strains (with varying  
311 parameter values) within each functional group in order to add more variation  
312 in outcome (see Kettle et al. (2015)) but we did not do this here in the interests  
313 of computational time. It should also be noted that the parameter values are  
314 highly uncertain in many cases and within each of our functional groups there  
315 will be large variability due to adaptation and evolution. Given this, we do  
316 not claim that the model response is necessarily representative of what may  
317 happen in an individual’s gut, rather it can be used as an aid to gain insight

318 into the relative importance of the different processes we are currently aware of  
319 and potentially to highlight, those we are not.

320 In addition to this, it must be noted that the default diet chosen here with  
321 10 g of protein and 50 g of carbohydrate fibre reaching the colon each day could  
322 be revised for any given population. However, converting from quantities of  
323 ingested food to substrate inflow to the colon is highly uncertain with large  
324 variations between studies, as well as technical issues with measuring this accu-  
325 rately. With more time, it would be interesting to investigate a larger range of  
326 typical diets but this was beyond the scope of the current work.

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

- 331 • Adding more functional groups or pathway switches in the existing func-  
332 tional groups. For example at present only the Bacteroides group can  
333 utilise protein but it is now known that some butyrate producers can also  
334 utilise amino acids (Louis and Flint, 2017)
- 335 • Our pH relation with TSCFA is very simplistic and could potentially be  
336 improved, although host secretions mean this is not necessarily straight-  
337 forward.
- 338 • Currently we set the transit time for the colon and then this is split be-  
339 tween the 3 model compartments based on their relative sizes. An interest-  
340 ing addition would be to alter transit time based on the composition of the  
341 various substrates entering the colon. For example, increasing residence  
342 time for high protein and/or low fibre diets. Due to variation in individual  
343 response this may need to include significant uncertainty ranges.
- 344 • Related to this is changing the absorption rate of water through the gut  
345 wall based on the diet, for example more water could remain in the gut  
346 on a high fibre diet.

- 347     ● A longer term goal would be to model the processes in the gastrointestinal  
348       tract preceding the colon in order to simulate how substrates entering the  
349       colon relate to dietary intake. This would allow more accurate prediction  
350       of microbial metabolite production based on diet.

351       To conclude, our model helps to explain some important, but poorly under-  
352       stood, relationships that have been reported in human studies, including the  
353       increase in butyrate proportion with increasing total faecal SCFA (LaBouyer  
354       et al., 2022). This phenomenon has important implications in view of the  
355       claimed benefits of butyrate supply for colorectal cancer prevention and the  
356       health of the colonic mucosa (Louis et al. (2014), Hamer et al. (2008)). The  
357       model also predicts increasing total faecal SCFA with greater fibre intake and  
358       more rapid gut transit. Gut transit is also shown to have potentially important  
359       consequences for microbiota composition and gut metabolism. In addition, the  
360       model confirms that the amount and type of non-digestible carbohydrate in the  
361       diet has the potential to cause major changes in microbiota composition. The  
362       nature of such changes is, however, predicted to be influenced by patterns of  
363       meal feeding and by any effects of dietary components (e.g. dietary fibre) upon  
364       gut transit. Human studies suggest that they will also depend on the initial  
365       microbiota composition. There is potential to use the model to explore how the  
366       presence of particular functional groups (such as lactate-utilizers (Wang et al  
367       2020)) within an individual's microbiota can influence their gut metabolism and  
368       response to dietary intervention. This may indeed be one of the most intriguing  
369       and fruitful applications of such modelling approaches in the future.

## 370       **Materials and methods**

### 371       **Software**

372       To facilitate continued research and future model development by other re-  
373       searchers we provide all model code on github (<https://github.com/HelenKettle/microPopGutCode>).

<sup>374</sup> The R package microPopGut is contained in the file microPopGut\_1.0.tar.gz.  
<sup>375</sup> This can be downloaded and installed in R using `install.packages('microPopGut_1.0.tar.gz')`.  
<sup>376</sup> Furthermore instructions on how to use the package are given in the supplementary  
<sup>377</sup> file ‘gettingStartedWithMicroPopGut.pdf’.

### <sup>378</sup> Microbial Model

<sup>379</sup> The microbial functional group model is based on the model described by Kettle et al. (2015) and implemented using the R package microPop (Kettle et al.,  
<sup>380</sup> 2018). The microbial groups include producers of the three major SCFA detected in fecal samples (acetate, butyrate and propionate) together with utilizers of acetate, lactate, succinate, formate and hydrogen (see Table 1 for a  
<sup>383</sup> summary, or refer to Kettle et al. (2015) for more detail). The model and its  
<sup>384</sup> equations are described in detail by Kettle et al. (2015) and Kettle et al. (2018)  
<sup>385</sup> so only a brief overview is given here. The microbial groups are defined as data  
<sup>386</sup> frames within the R package and these are shown in section 3 of the Supp. Info..  
<sup>387</sup> Although this application uses the microbial parameters (e.g. maximum growth  
<sup>388</sup> rates, yields etc) that are in the package’s intrinsic data frames, these can be  
<sup>389</sup> easily changed by either modifying the dataframe in R or by providing a new  
<sup>390</sup> dataframe - either as an input csv file or by creating one in R. One of the input  
<sup>391</sup> arguments to the function `microPopGut()` is `microbeNames` which allows the  
<sup>392</sup> user to also enter other microbial groups.

<sup>394</sup> The growth substrates available in the large intestine are divided into four  
<sup>395</sup> categories: protein (P), non starch polysaccharides (NSP), resistant starch (RS)  
<sup>396</sup> and sugars (and oligosaccharides and sugar alcohols); for simplicity, all carbo-  
<sup>397</sup> hydrate units are regarded as being hexoses. NSP comprise major components  
<sup>398</sup> of dietary fibre including the structural polysaccharides of the plant cell wall  
<sup>399</sup> (cellulose, xylan, pectin), whereas RS refers to the fraction of dietary starch  
<sup>400</sup> that resists digestion in the small intestine. We consider 10 major metabolites  
<sup>401</sup> that arise from substrate fermentation: acetate, propionate, butyrate, lactate,  
<sup>402</sup> succinate, formate, hydrogen, carbon dioxide, methane and ethanol. Six of these

403 metabolites (acetate, lactate, succinate, formate, hydrogen and carbon dioxide)  
404 are also considered as substrates, because they are known to be consumed by  
405 some groups (cross-feeding). It is well known that pH affects growth rate there-  
406 fore each group is assigned a preferred range of pH within which it can reach its  
407 maximum growth rate, but outside of which, its growth is reduced or zero. We  
408 model the rate of bacterial growth using Monod kinetics and assume that from  
409 1 g of resource, Y g of biomass is produced. We assume that resource that is  
410 taken up by microbes, but not used to produce biomass, is converted to metabo-  
411 lites. If not all of the resource is converted to biomass or to the metabolites  
412 represented in our model, it is discarded. This applies, for example, to many  
413 diverse fermentation products of proteins (e.g. phenols, amines) that are not  
414 among the 10 major products covered by the model. Although the model was  
415 initially developed to be run with multiple strains within each functional group,  
416 in the current work we do not do this due to the high CPU time associated with  
417 multiple compartments.

#### 418 **Inflow to colon**

##### 419 **Incoming substrates and water**

420 The main sources of nutrient for microbiota in the colon are complex dietary  
421 carbohydrates that are not absorbed higher up the digestive tract. We use  
422 a default value of  $50 \text{ g d}^{-1}$  of carbohydrate, C, in our model and we vary the  
423 proportion of this which is NSP or RS using the RS fraction (i.e.  $\text{RS}/(\text{RS+NSP})$ )  
424 where  $\text{RS+NSP}=C$ ). Based on Cremer et al. (2017) and references therein,  
425 about 15 g of bio-available NSP and 30-40 g of RS enter the colon per day  
426 which gives us an RS fraction of 0.67-0.9 with average value of 0.78 which we  
427 use as our default value. According to Yao et al. (2016) less is known regarding  
428 dietary proteins, P, that escape digestion to reach the large intestine, although it  
429 is estimated that around 6 - 18 g P reaches the large intestine daily, the majority  
430 from the diet and a small proportion from endogenous origins. Given this, here

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>

431 we assume that 10 g d<sup>-1</sup> of undigested P reaches the colon from dietary intake  
 432 along with a small amount from mucin degradation (approx. 1 g d<sup>-1</sup>). Phillips  
 433 and Giller (1973) state that water enters at approximately 1.5 l d<sup>-1</sup> and about  
 434 90% of this is absorbed by the colon. Stephen and Cummings (1980) states that  
 435 normal fecal daily output in Britain is 100-200 g d<sup>-1</sup> of which 25-50 g d<sup>-1</sup> is  
 436 solid matter and the rest (50-175 g d<sup>-1</sup>) is water. Thus if 90% is absorbed then  
 437 this indicates water inflow in the range 0.5 - 1.75 l d<sup>-1</sup>. The midpoint of this  
 438 range is 110 g d<sup>-1</sup> of water outflow which, if 90% is absorbed, implies that the  
 439 inflow of water is approximately 1100 g d<sup>-1</sup>. This will clearly vary depending  
 440 on the host's oral water intake but we use 1100 g d<sup>-1</sup> as our default value. The  
 441 default inflow values are summarised in Table 4.

#### 442 Meals

443 The normal human diet does not consist of continuous fixed inflow of substrate;  
 444 for a more realistic substrate inflow to the colon we simulate eating 3 meals a  
 445 day with randomly varying composition. We then approximate the passage of  
 446 these meals through the stomach and small intestine to obtain a smoothed time  
 447 series for substrate entering the colon. Note that we are not simulating all the  
 448 food ingested by the host (most of which will not reach the colon) but rather  
 449 simply trying to produce a more realistic time series for the substrates that we  
 450 know reach the colon.

451 We specify three meals per day each with a duration of 30 minutes. This

452 time-series is then passed through a one-compartment ordinary differential equa-  
 453 tion model representing the time spent in the stomach and small intestine (es-  
 454 timated to take 7 hours), i.e.

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

455 where  $v=3.4 \text{ d}^{-1}$  (inverse of 7 h transit time in days);  $\dot{s}(t)_{in}$  is time series  
 456 representing 3 meals a day ( $\text{g d}^{-1}$ ) and  $t$  is time in days. The inflow to the  
 457 colon (i.e. the outflow from small intestine) is given by  $vs(t)$ . The composition  
 458 (in terms of P, NSP, RS and water (W)) of these meals varies randomly around  
 459 the mean of each component (Table 4) for each meal. To generate such random  
 460 fluctuations we draw samples for each meal from a gamma distribution (since  
 461 this is always above zero) defined by a scale parameter ( $\gamma_s$ ) and the daily average  
 462 inflow of the substrate ( $\text{g d}^{-1}$ ). We assume the magnitude of the substrate  
 463 fluctuations are proportional to the mean value. Preliminary simulations showed  
 464 that  $\gamma_s$  equal to half the mean value of each substrate gave a good variation for  
 465 P, RS and NSP, and for water variation we assumed  $\gamma_s$  was one tenth of the  
 466 incoming daily flow. The distributions and flow patterns are shown in Fig. 9.

#### 467 Mucin

468 There is a further input of protein and carbohydrate from the host via the  
 469 breakdown of host-released mucin by many strains in the B group (Ravcheev  
 470 and Thiele, 2017) and in our NBFD group (Crost et al., 2013). It is estimated  
 471 that  $2.7\text{-}7.3 \text{ g d}^{-1}$  of mucin, denoted  $\dot{M}$ , is secreted into the colon (Florin  
 472 et al., 1991), therefore we take the midpoint value 5 g/d. We assume our mucin  
 473 degraders break down 1 g of mucin into 0.05 g sulphate, 0.2 g P and 0.75 g C,  
 474 based on Sung et al. (2017), but consider their yield on mucin to be negligible  
 475 compared with growth on other substrates. We split C equally between NSP  
 476 and RS - this arbitrary choice did not affect model results since C from mucin  
 477 ( $3.75 \text{ g d}^{-1}$  maximum) is much less than dietary C ( $50 \text{ g d}^{-1}$ ), but this should

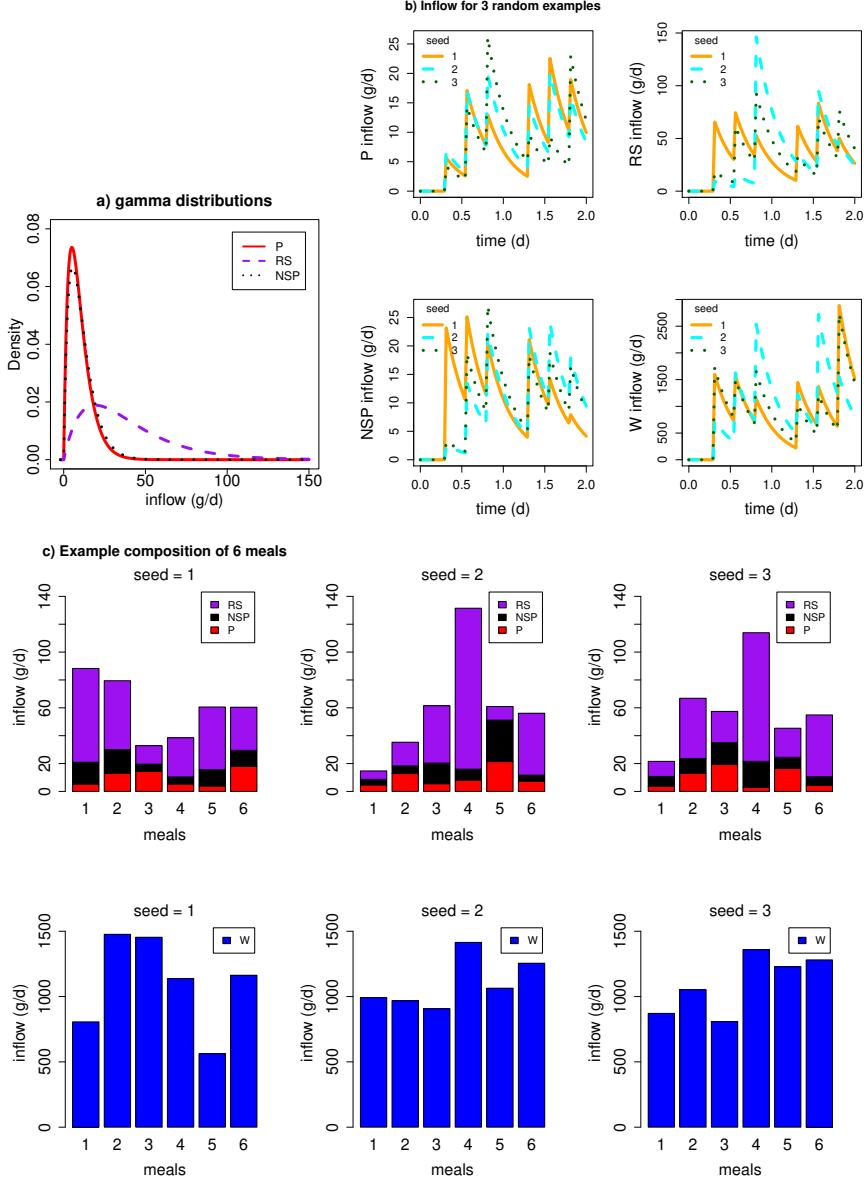


Figure 9: a) Gamma distribution from which random values are drawn to generate the composition of each meal (note water is not shown due to the large difference in magnitude between water and dietary substrates). b) The substrate inflow time series to the proximal colon after passing through the small 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.

478 be revised if considering very different dietary drivers. Since the compartments  
 479 of the colon are not equal-sized we assume that the rate of mucin entering the  
 480 colon is divided through the model compartments proportional to their relative  
 481 volumes. We assume this enters the colon at a fixed, continuous rate and mucin-  
 482 derived P and C are a function of the mass of mucin degraders,  $D_M$  (B and  
 483 NBFD), such that,

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

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

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

## 490 Absorption by host

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

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

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

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

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

## 517 pH

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

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

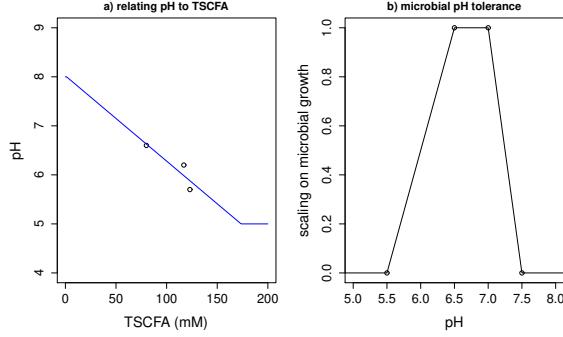


Figure 10: a) Relating pH to TSCFA using Eq. 5 and data from (Cummings et al., 1987). b) Example of microbial tolerance to pH. A pH tolerance function of this form is specified individually for each microbial group in our model.

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

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

### 536 Fecal outflow

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

543 volume fraction, thus

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

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

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

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

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

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

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

## 562 Author Contributions

563 Kettle wrote the model code and led the writing of the manuscript. Flint con-  
564 tributed to writing the manuscript. Flint and Louis contributed to all aspects  
565 of microbiology and all authors contributed critically to the drafts and gave final

<sup>566</sup> approval for publication.

## <sup>567</sup> Acknowledgments

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

## <sup>570</sup> Data Accessibility

<sup>571</sup> All model code is on github (<https://github.com/HelenKettle/microPopGutCode>).  
<sup>572</sup> The R package microPopGut is contained in the file microPopGut\_1.0.tar.gz.  
<sup>573</sup> This can be downloaded and installed in R using install.packages('microPopGut\_1.0.tar.gz').  
<sup>574</sup> Furthermore instructions on how to use the package are given in the supplemen-  
<sup>575</sup> tary file 'getStartedWithMicroPopGut.pdf'. The model output for the simula-  
<sup>576</sup> tions described in this manuscript are included on figshare in the file Mod-  
<sup>577</sup> elRuns.tar.gz (<https://doi.org/10.6084/m9.figshare.21094558.v1>). The plotting  
<sup>578</sup> code is provided in the github repository.

## <sup>579</sup> Funding Statement

<sup>580</sup> The Scottish Government's Rural and Environment Science and Analytical Ser-  
<sup>581</sup> vices Division (RESAS) funded this research.

## <sup>582</sup> References

- <sup>583</sup> Bown, R., Gibson, J., Sladen, G., Hicks, B., and Dawson, A. (1974). Effects  
<sup>584</sup> of lactulose and other laxatives on ileal and colonic pH as measured by a  
<sup>585</sup> radiotelemetry device. *Gut*, 15(12):999–1004.
- <sup>586</sup> Capuano, E., Oliviero, T., Fogliano, V., and Pellegrini, N. (2018). Role of the  
<sup>587</sup> food matrix and digestion on calculation of the actual energy content of food.  
<sup>588</sup> *Nutrition Reviews*, 76(4):274–289.

- 589 Chung, W. S., Walker, A. W., Bosscher, D., Garcia-Campayo, V., Wagner, J.,  
590 Parkhill, J., Duncan, S. H., and Flint, H. J. (2020). Relative abundance of  
591 the *Prevotella* genus within the human gut microbiota of elderly volunteers  
592 determines the inter-individual responses to dietary supplementation with  
593 wheat bran arabinoxylan-oligosaccharides. *BMC Microbiology*, 20(1):1–4.
- 594 Cremer, J., Arnoldini, M., and Hwa, T. (2017). Effect of water flow and chemical  
595 composition in the human colon. *PNAS*, 114(25):6438–6443.
- 596 Cremer, J., Segota, I., y Yang, C., Arnoldini, M., Sauls, J., Zhang, Z., Gutierrez,  
597 E., Groisman, A., and Hwa, T. (2016). Effect of flow and peristaltic mixing  
598 on bacterial growth in a gut-like channel. *PNAS*, 113:11414–11419.
- 599 Crost, E., Tailford, L., Gall, G. L., Fons, M., Henrissat, B., and Juge, N. (2013).  
600 Utilisation of mucin glycans by the human gut symbiont *ruminococcus gnavus*  
601 is strain-dependent. *PLoS ONE*, 8(10):e76341.
- 602 Cummings, J. H., Pomare, E., Branch, W., Naylor, C., and Macfarlane, G.  
603 (1987). Short chain fatty acids in human large intestine, portal, hepatica and  
604 venous blood. *Gut*, 28:1221–1227.
- 605 Duncan, S. H., Belenguer, A., Holtrop, G., Johnstone, A., Flint, H. J., and  
606 Lobley, G. E. (2007). Reduced dietary intake of carbohydrates by obese sub-  
607 jects results in decreased concentrations of butyrate and butyrate-producing  
608 bacteria in feces. *Applied and Environmental Microbiology*, 73(4):1073–1078.
- 609 Duncan, S. H., Russell, W. R., Quartieri, A., Rossi, M., Parkhill, J., Walker,  
610 A. W., and Flint, H. J. (2016). Wheat bran promotes enrichment within the  
611 human colonic microbiota of butyrateproducing bacteria that release ferulic  
612 acid. *Environmental Microbiology*, 18(7):2214–2225.
- 613 Florin, T., Neale, G., Gibson, G., Christl, S., and Cummings, J. (1991).  
614 Metabolism of dietary sulphate: absorption and excretion in humans. *Gut*,  
615 32:766–773.

- 616 Hamer, H. M., Jonkers, D., Venema, K., Vanhoutvin, S., Troost, F., and Brum-  
617 mer, R.-J. (2008). The role of butyrate on colonic function. *Alimentary*  
618 *pharmacology & therapeutics*, 27(2):104–119.
- 619 Kettle, H., Holtrop, G., Louis, P., and Flint, H. J. (2018). micropop: Mod-  
620 elling microbial populations and communities in r. *Methods in Ecology and*  
621 *Evolution*, 9(2):399–409.
- 622 Kettle, H., Louis, P., Holtrop, G., Duncan, S. H., and Flint, H. J. (2015).  
623 Modelling the emergent dynamics and major metabolites of the human colonic  
624 microbiota. *Environmental Microbiology*, 17(5):1615–1630.
- 625 LaBouyer, M., Holtrop, G., Horgan, G., Gratz, S. W., Belenguer, A., Smith,  
626 N., Walker, A. W., Duncan, S. H., Johnstone, A. M., Louis, P., et al. (2022).  
627 Higher total faecal short-chain fatty acid concentrations correlate with in-  
628 creasing proportions of butyrate and decreasing proportions of branched-chain  
629 fatty acids across multiple human studies. *Gut Microbiome*, 3(e2):1–14.
- 630 Lewis, S. and Heaton, K. (1997). Increasing butyrate concentration in the distal  
631 colon by accelerating intestinal transit. *Gut*, 41:245–251.
- 632 Louis, P. and Flint, H. J. (2017). Formation of propionate and butyrate by the  
633 human colonic microbiota. *Environmental Microbiology*, 19(1):29–41.
- 634 Louis, P., Hold, G. L., and Flint, H. J. (2014). The gut microbiota, bacterial  
635 metabolites and colorectal cancer. *Nature reviews microbiology*, 12(10):661–  
636 672.
- 637 Mikolajczyk, A. E., Watson, S., Surma, B. L., and Rubin, D. T. (2015). As-  
638 sessment of tandem measurements of ph and total gut transit time in healthy  
639 volunteers. *Clinical and Translational Gastroenterology*, 6(7).
- 640 Moorthy, A. S., S.P.J., B., M., K., and H.J., E. (2015). Continuous model of  
641 carbohydrate digestion and transport processes in the colon. *PLoS ONE*,  
642 10(12):e.0145309.

- 643 Morrison, D. J. and Preston, T. (2016). Formation of short chain fatty acids by  
644 the gut microbiota and their impact on human metabolism. *Gut microbes*,  
645 7(3):189–200.
- 646 Munoz-Tamayo, R., Laroche, B., Walter, E., Dore, J., and Leclerc, M. (2010).  
647 Mathematical modelling of carbohydrate degradation by human colonic mi-  
648 crobiota. *Journal of Theoretical Biology*, 266(1):189 – 201.
- 649 Phillips, S. and Giller, J. (1973). The contribution of the colon to electrolyte  
650 and water conservation in man. *J Lab Clin Med*, 81(5):733–746.
- 651 Ravcheev, D. A. and Thiele, I. (2017). Comparative genomic analysis of the  
652 human gut microbiome reveals a broad distribution of metabolic pathways for  
653 the degradation of host-synthetized mucin glycans and utilization of mucin-  
654 derived monosaccharides. *Frontiers in Genetics*, 8:111.
- 655 Reichardt, N., Duncan, S. H., Young, P., Belenguer, A., Leitch, C. M., Scott,  
656 K. P., Flint, H. J., and Louis, P. (2014). Phylogenetic distribution of three  
657 pathways for propionate production within the human gut microbiota. *The*  
658 *ISME Journal*, 8(6):1323–1335.
- 659 Rios-Covian, D., Ruas-Madiedo, P., Margolles, A., Gueimonde, M., de los  
660 Reyes-Gavilan, C. G., and Salazar, N. (2016). Intestinal short chain fatty  
661 acids and their link with diet and human health. *Front. Microbiol.*, 17.
- 662 Ruppin, H., Bar-Maeir, S., Soergel, K., Wood, C., and Schmitt, M. (1980). Ab-  
663 sorption of short chain fatty acids by the colon. *Gastroenterology*, 78(6):1500–  
664 1507.
- 665 Salonen, A., Lahti, L., Saloja, J., Holtrop, G., Korpela, K., Duncan, S. H.,  
666 Date, P., Farquharson, F., Johnstone, A. M., Lobley, G. E., Louis, P., Flint,  
667 H. J., and de Vos, W. M. (2014). Impact of diet and individual variation on  
668 intestinal microbiota composition and fermentation products in obese men.  
669 *The ISME Journal*, pages 1–13.

- 670 Slavin, J. L., Brauer, P. M., and Marlett, J. A. (1981). Neutral Detergent Fiber,  
671 Hemicellulose and Cellulose Digestibility in Human Subjects. *The Journal of*  
672 *Nutrition*, 111(2):287–297.
- 673 Smith, N., Shorten, P., and Altermann, E. (2021). Examination of hydrogen  
674 cross-feeders using a colonic microbiota model. *BMC Bioinformatics*, 22(3).
- 675 Stephen, A. and Cummings, J. (1980). The microbial contribution to human  
676 fecal mass. *J. Medical Microbiology*, 13(1):45–56.
- 677 Sung, J., Kim, S., Cabatbat, J., Jang, S., Jin, Y., Jung, G. Y., Chia, N., and  
678 Kim, P. (2017). Global metabolic interaction network of the human gut micro-  
679 biota for context-specific community-scale analysis. *Nature communications*.
- 680 Topping, D. and Clifton, P. (2001). Short-chain fatty acids and human colonic  
681 function: roles of resistant starch and nonstarch polysaccharides. *Physiological*  
682 *Review*, 81(3):1031–1064.
- 683 Walker, A. W., Ince, J., Duncan, S. H., Webster, L. M., Holtrop, G., Ze, X.,  
684 Brown, D., Stares, M. D., Scott, P., Bergerat, A., Louis, P., McIntosh, F.,  
685 Johnstone, A. M., Lobley, G. E., Parkhill, J., and Flint, H. J. (2011). Domi-  
686 nant and diet-responsive groups of bacteria within the human colonic micro-  
687 biota. *The ISME Journal*, 5:220–230.
- 688 Wang, S. P., Rubio, L. A., Duncan, S. H., Donachie, G. E., Holtrop, G., Lo,  
689 G., Farquharson, F. M., Wagner, J., Parkhill, J., Louis, P., Walker, A. W.,  
690 and Flint, H. J. (2020). Pivotal roles for ph, lactate, and lactate-utilizing  
691 bacteria in the stability of a human colonic microbial ecosystem. *mSystems*,  
692 5(5):e00645–20.
- 693 Wu, G., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y., Keilbaugh, S., Bewtra,  
694 M., Knights, D., Walters, W., Knight, R., and Sinha, R. (2011). Linking long-  
695 term dietary patterns with gut microbial enterotypes. *Science*, 334(6052):105–  
696 108.

697 Yao, C. K., Muir, J. G., and Gibson, P. R. (2016). Review article: insights  
698 into colonic protein fermentation, its modulation and potential health impli-  
699 cations. *Alimentary Pharmacology and Therapeutics*, 43(2):181–196.