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

Helen Kettle[1,*], Petra Louis[2], and Harry J. Flint[2]

[1]Biomathematics and Statistics Scotland, James Clerk Maxwell Building, Peter Guthrie Tait Road, Edinburgh, EH9 3FD

[2]Gut Health Group, Rowett Institute, University of Aberdeen, Aberdeen, UK

[*]Corresponding author: Helen.Kettle@bioss.ac.uk

September 15, 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

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

Introduction

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

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

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

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

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

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

1. Total SCFA (TSCFA) concentration in the proximal, transverse and distal compartments should be around 123, 117 and 80 mM respectively according to sudden death human autopsies (Cummings et al., 1987)

2. Acetate:Propionate:Butyrate ratios are similar (around 3:1:1) in all regions of the colon and around 60:20:20 mM (Cummings et al., 1987)

3. Over 95% of SCFA are absorbed by the host (Topping and Clifton, 2001)

4. Approx. 90% of incoming water is absorbed by the host (Phillips and Giller, 1973)

5. pH in the proximal, transverse and distal compartments should be around 5.7, 6.2 and 6.6 respectively (Cummings et al. (1987), and telemetry data from Mikolajczyk et al. (2015), Bown et al. (1974))

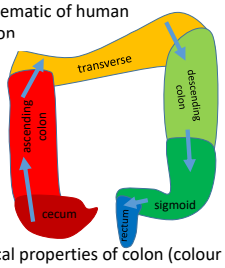
6. Normal daily fecal output in Britain is 100-200 g d^{-1} of which 25-50 g is solid matter (i.e. 50-175 g d^{-1} is water). Bacteria make up about 55% of the solid matter i.e. 14-28 g d^{-1} of microbes emitted (Stephen and Cummings, 1980).

7. TSCFA concentration decreases with transit time (Lewis and Heaton, 1997)

After model verification we examine the effects of including meals, bowel movements and fixed/varying pH into the model. We then use the model to look at how carbohydrate composition (based on the fractions of resistant starch (RS) 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.

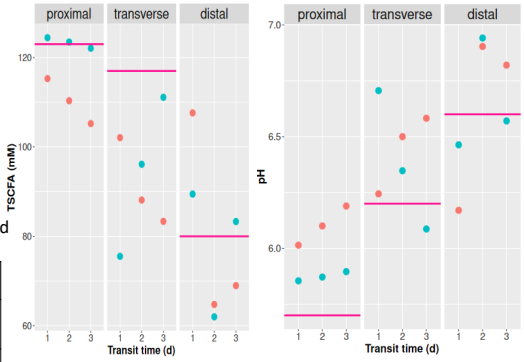
Although gut microbiota are highly complex and not fully understood, here we show that it is nonetheless possible to develop predictive models of key components of this ecological system. An important goal of our modelling is to aid and inform the interpretation of data obtained, mostly from faecal samples,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

A. Schematic of human colon



B. Physical properties of colon (colour matched with schematic A)

	Model region	Len (cm)	Vol (L)	pH	TSCFA (mM)
Cecum	Prox	21	0.41	5.7	123
Ascending					
Transverse	Trans	50	0.98	6.2	117
Descending	Dist	60	1.63	6.6	80
Sigmoid					
rectum					



C. Comparison of model results for time-averaged TSCFA and pH with observed properties shown in Table B (pink lines).

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.

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/microPopGutCode) and instructions on how to use the package are given in the supplementary file 'getStartedWithMicroPopGut.pdf'.

Results

Standard Model

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

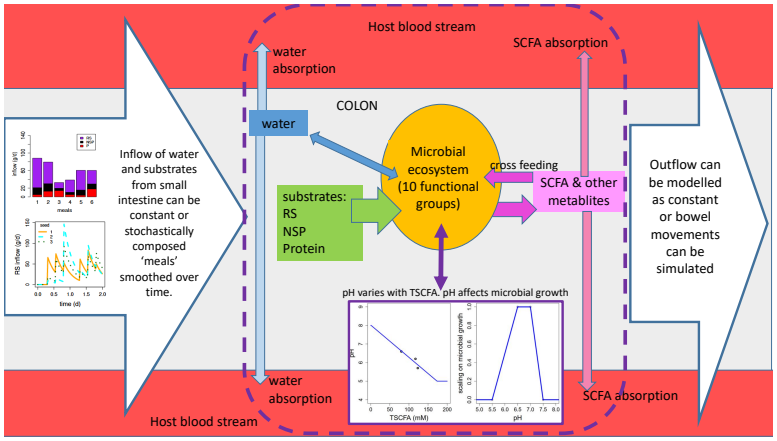


Figure 2: Model system with the microbial ecosystem comprising 10 microbial functional groups (Table 1) which consume substrates (RS, NSP and protein) and water. The microbes produce metabolites some of which are consumed by other MFGs ('cross-feeding'). SCFA and water are absorbed through the colon 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, transerve, 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	*Bacteroides* spp., *Prevotella* spp., *Akkermansia muciniphila* (Verrucomicrobia)
NoButyStarchDeg	NBSD	Non-butyrate-forming starch degraders	*Ruminococcaceae* related to *Ruminococcus bromii*. Also includes certain *Lachnospiraceae*
NoButyFibreDeg	NBFD	Non-butyrate-forming fibre degraders	*Ruminococcaceae* related to *Ruminococcus albus*, *Ruminococcus flavefaciens*. Also includes certain *Lachnospiraceae*
LactateProducers	LP	Lactate producers	*Actinobacteria*, especially *Bifidobacterium* spp., *Collinsella aerofaciens*
ButyrateProducers1	BP1	Butyrate Producers	*Lachnospiraceae* related to *Eubacterium rectale*, *Roseburia* spp.
ButyrateProducers2	BP2	Butyrate Producers	Certain *Ruminococcaceae*, in particular *Faecalibacterium prausnitzii*
PropionateProducers	PP	Propionate producers	*Veillonellaceae* e.g. *Veillonella* spp., *Megasphaera elsdenii*
ButyrateProducers3	BP3	Butyrate Producers	*Lachnospiraceae* related to *Eubacterium hallii*, *Anaerostipes* spp.
Acetogens	A	Acetate Producers	Certain *Lachnospiraceae*, e.g. *Blautia hydrogenotrophica*
Methanogens	M	Methanogenic archaea	*Methanobrevibacter smithii*

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

Table 3 gives summary results of the model simulations without bowel movements but with varying pH for each bowel region. Fig. 3 shows results from more simulations but for the distal colon only. Fig. 3a shows that although bowel movements make a difference to the total biomass and the TSCFA they do not have a large effect on the community composition or the SCFA ratios. Thus in the interests of model simplicity we decide to not include bowel movements 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

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

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

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

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

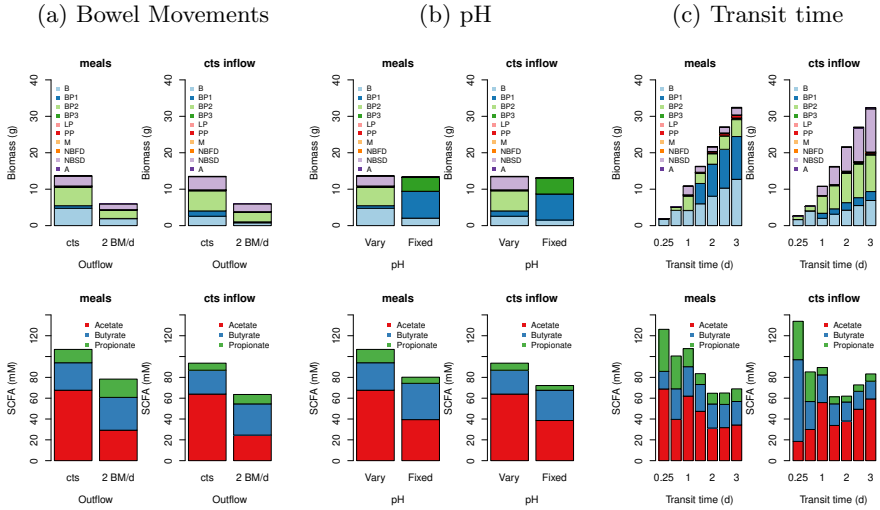


Figure 3: Summary results (averaged over days 7-28 and over random seeds) for the distal compartment for continuous inflow or fluctuating inflow (i.e. 'meals') for continuous outflow from colon or for 2 bowel movements per day ('2 BM/d'). The RS fraction is 0.78 (i.e. 78% of the dietary carbohydrate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

Model Experiments

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

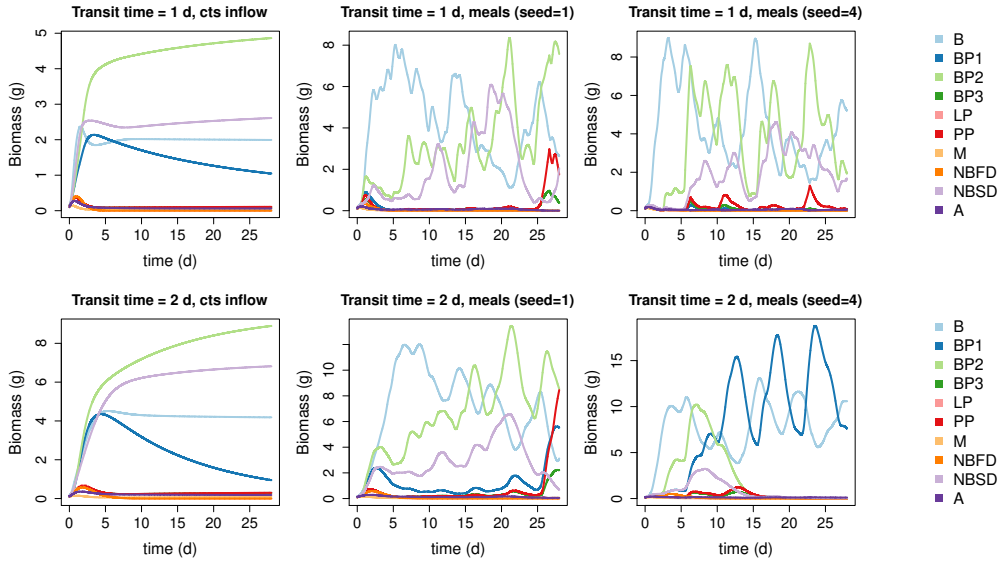


Figure 4: Simulation results for the distal compartment for continuous inflow (first plot on each row) or fluctuating inflow (i.e. 'meals') for transit times of 1 d (top row) and 2 d (bottom row) and for 2 random seeds. Modelled pH varies with TSCFA and the RS fraction is 0.78. There are no bowel movements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

Effects of total dietary carbohydrate

In this model experiment we investigate the effects of decreasing carbohydrate on the microbial community. Here we compare our results qualtitatively 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 g d^{-1} (our default value) and then increase inflowing non-digestable (ND) carbohydrate from 10 g d^{-1} to 60 g d^{-1} in 10 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 carbohydrate intake (Fig. 7). Analysis of 10 human studies involving 163 subjects has shown a highly significant increase in percentage butyrate with increasing total SCFA concentration in faecal samples (LaBouyer et al., 2022).

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

Diet	PI ($g\ d^{-1}$)	CI ($g\ d^{-1}$)	SI ($g\ d^{-1}$)	NSPI ($g\ d^{-1}$)
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

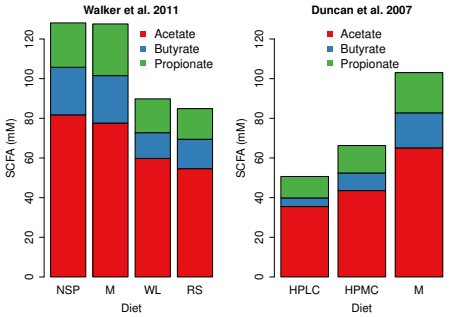


Figure 5: Table on left shows the dietary intake for two human studies (Duncan et al. (2007) and Walker et al. (2011). PI, CI, SI and NSPI refer to ingested dietary protein, carbohydrate, starch and NSP. Note, starch value for the high RS diet in the Walker et al. (2011) study included 26 g commercial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

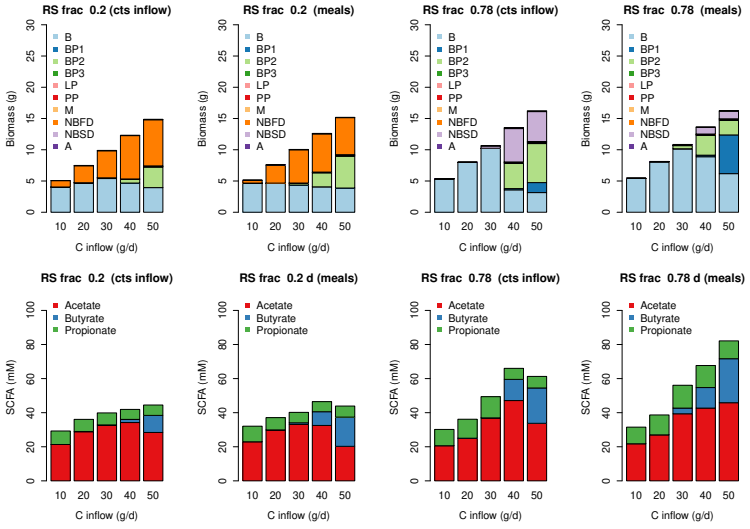


Figure 6: Simulated Biomass and SCFA results for increasing carbohydrate inflow. Simulations are run with continuous substrate inflow (cts) and with 'meals' for a transit time of 1.5 days. The results are the average over the last 3 weeks of a 28 day simulation and 'meals' is the average over 4 stochastically-generated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

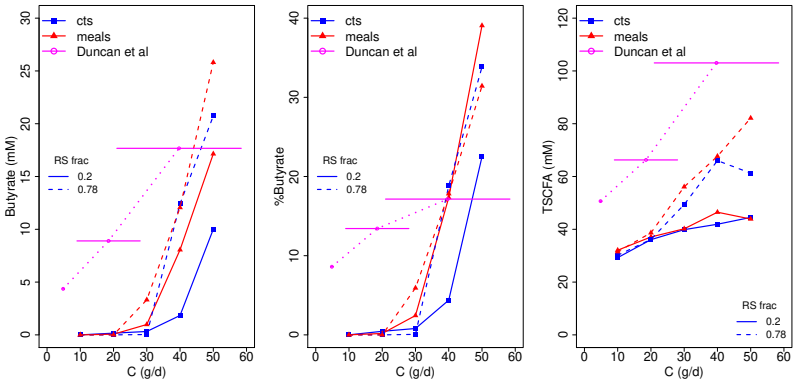


Figure 7: Plot of modelled butyrate, %butyrate and TSCFA against grams of carbohydrate entering the colon each day. Data from Duncan et al. (2007) is shown in magenta - due to uncertainties in converting ingested starch to RS entering the colon there are large error bars on the amount of C (g/d). Error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

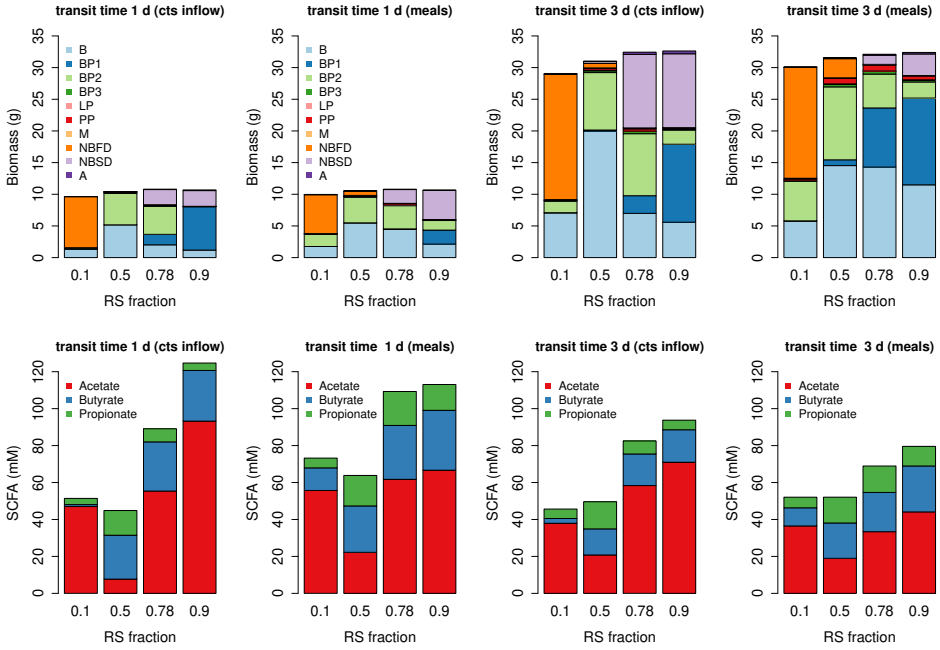


Figure 8: Biomass and SCFA results for changing the RS fraction of inflowing carbohydrate with continuous substrate inflow ('cts inflow') and with 'meals'. Protein and carbohydrate inflow are 10 and 50 g d^{-1} respectively. The results are the average over the last 3 weeks of a 28 day simulation and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

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

Assignments of microbial taxa to our ten functional groups were based initially on evidence from cultured isolates. These assignments have since been supported and greatly extended by analysis of genes diagnostic for different fermentation pathways within genomes and metagenomes (Reichardt et al., 2014) and by molecular detection of species enriched within the community by defined growth substrates in chemostat experiments (Duncan et al., 2016) and dietary intervention studies (Salonen et al., 2014). Nevertheless, these assignments inevitably remain provisional and incomplete and we do not claim that the model 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 for each group to represent inter-individual variation, groups can also be easily added or removed from the model through the input argument 'microbeNames'. Furthermore, it is also possible to include any number of strains (with varying parameter values) within each functional group in order to add more variation in outcome (see Kettle et al. (2015)) but we did not do this here in the interests of computational time. It should also be noted that the parameter values are highly uncertain in many cases and within each of our functional groups there will be large variability due to adaptation and evolution. Given this, we do not claim that the model response is necessarily representative of what may happen in an individual's gut, rather it can be used as an aid to gain insight

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

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

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

- Adding more functional groups or pathway switches in the existing functional groups. For example at present only the Bacteroides group can utilise protein but it is now known that some butyrate producers can also utilise amino acids (Louis and Flint, 2017)

- Our pH relation with TSCFA is very simplistic and could potentially be improved, although host secretions mean this is not necessarily straightforward.

- Currently we set the transit time for the colon and then this is split between the 3 model compartments based on their relative sizes. An interesting addition would be to alter transit time based on the composition of the various substrates entering the colon. For example, increasing residence time for high protein and/or low fibre diets. Due to variation in individual response this may need to include significant uncertainty ranges.

- Related to this is changing the absorption rate of water through the gut wall based on the diet, for example more water could remain in the gut on a high fibre diet.

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

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

Materials and methods

Software

To facilitate continued research and future model development by other researchers we provide all model code on github (https://github.com/HelenKettle/microPopGutCode).

The R package microPopGut is contained in the file microPopGut_1.0.tar.gz. This can be downloaded and installed in R using install.packages('microPopGut_1.0.tar.gz'). Furthermore instructions on how to use the package are given in the supplementary file 'gettingStartedWithMicroPopGut.pdf'.

Microbial Model

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., 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 summary, or refer to Kettle et al. (2015) for more detail). The model and its equations are described in detail by Kettle et al. (2015) and Kettle et al. (2018) so only a brief overview is given here The microbial groups are defined as data frames within the R package and these are shown in section 3 of the Supp. Info.. Although this application uses the microbial parameters (e.g. maximum growth rates, yields etc) that are in the package's intrinsic data frames, these can be easily changed by either modifying the dataframe in R or by providing a new dataframe - either as an input csv file or by creating one in R. One of the input arguments to the function *microPopGut()* is *microbeNames* which allows the user to also enter other microbial groups.

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

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

Inflow to colon

Incoming substrates and water

The main sources of nutrient for microbiota in the colon are complex dietary carbohydrates that are not absorbed higher up the digestive tract. We use a default value of 50 g d^{-1} of carbohydrate, C, in our model and we vary the proportion of this which is NSP or RS using the RS fraction (i.e. RS/(RS+NSP) where RS+NSP=C). Based on Cremer et al. (2017) and references therein, about 15 g of bio-available NSP and 30-40 g of RS enter the colon per day which gives us an RS fraction of 0.67-0.9 with average value of 0.78 which we use as our default value. According to Yao et al. (2016) less is known regarding dietary proteins, P, that escape digestion to reach the large intestine, although it is estimated that around 6 - 18 g P reaches the large intestine daily, the majority 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^{-1}
$\dot{C}_{diet}$	carbohydrate inflow rate	50 g d^{-1}
$\dot{W}_{diet}$	water inflow rate	1100 g d^{-1}
$\dot{M}$	mucin inflow rate	5 g d^{-1}
K_M	half saturation constant for Mucin breakdown	0.5 g l^{-1}
a_W	rate of water absorption by host	3 d^{-1}
a_Z	rate of SCFA absorption by host	9.6 d^{-1}

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

Meals

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

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

time-series is then passed through a one-compartment ordinary differential equation model representing the time spent in the stomach and small intestine (estimated to take 7 hours), i.e.

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

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

Mucin

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

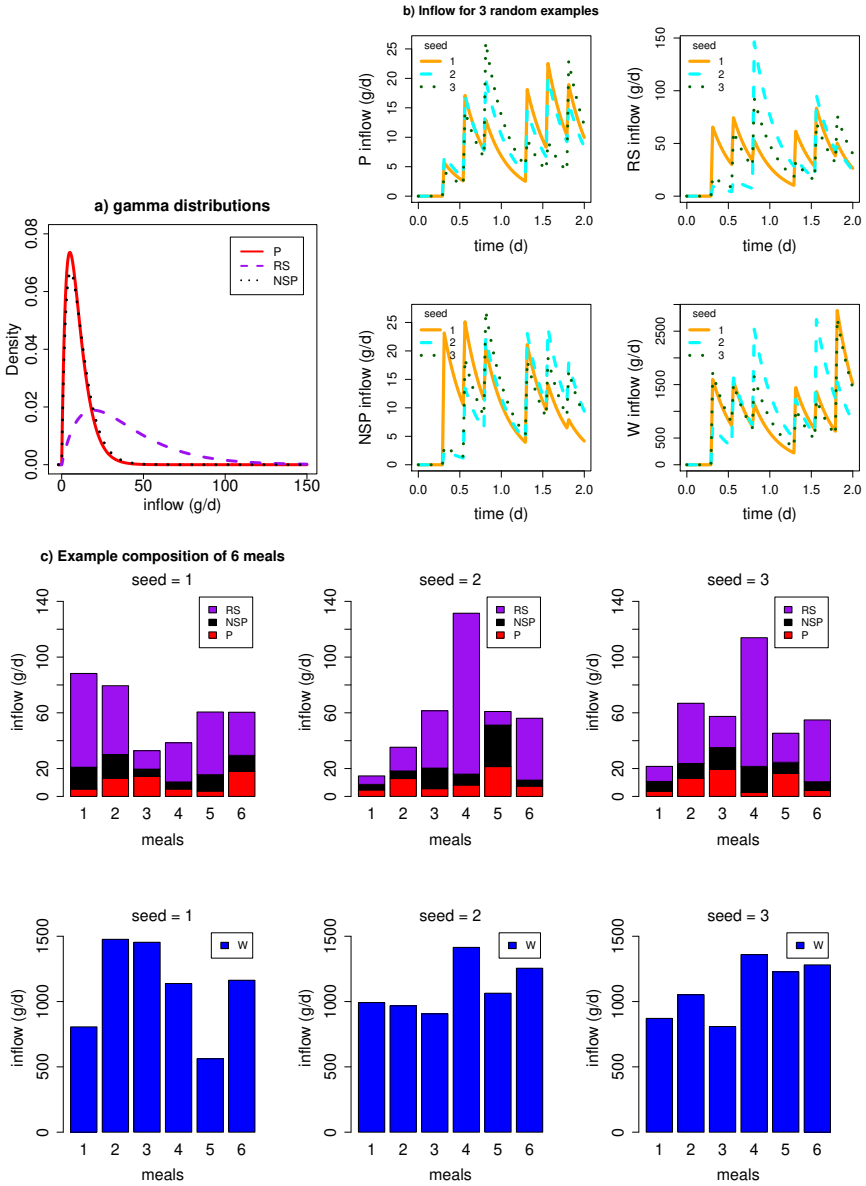


Figure 9: a) Gamma distribution from which random values are drawn to generate the composition of each meal (note water is not shown due to the large difference in magnitude between water and dietary substrates). b) The substrate inflow time series to the proximal colon after passing through the small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

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

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

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

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

Absorption by host

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

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

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

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

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

pH

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

$$\text{pH} = 8.02 - 0.0174 \times \text{TSCFA}. \tag{5}$$

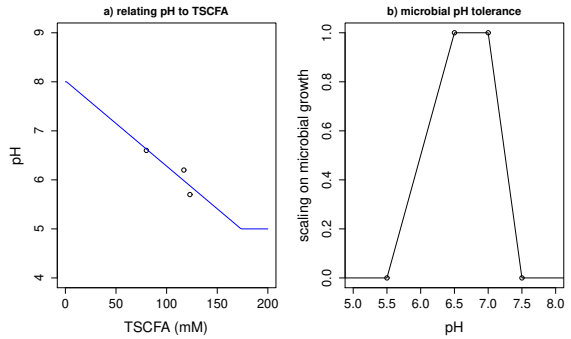


Figure 10: a) Relating pH to TSCFA using Eq. 5 and data from (Cummings et al., 1987). b) Example of microbial tolerance to pH. A pH tolerance function of this form is specified individually for each microbial group in our model.

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

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

Fecal outflow

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

volume fraction, thus

$$T_i = \frac{v_i}{v_{colon}} T_t \tag{6}$$

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

If we introduce bowel movements then, assuming the distal colon is approximately emptied for each bowel movement, the total transit time is given by

$$T_t = \Sigma_{i=1}^{2} T_i + \frac{1}{N_{BM}} \tag{7}$$

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

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

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

Author Contributions

Kettle wrote the model code and led the writing of the manuscript. Flint contributed to writing the manuscript. Flint and Louis contributed to all aspects of microbiolgy and all authors contributed critically to the drafts and gave final

approval for publication.

Acknowledgments

We thank the Scottish Goverment's Rural and Environment Science and Analytical Services Division (RESAS) for funding this research.

Data Accessibility

All model code is on github (https://github.com/HelenKettle/microPopGutCode). The R package microPopGut is contained in the file microPopGut_1.0.tar.gz. This can be downloaded and installed in R using install.packages('microPopGut_1.0.tar.gz'). Furthermore instructions on how to use the package are given in the supplementary file 'getStartedWithMicroPopGut.pdf'. The model output for the simulations described in this manuscript are included on figshare in the file ModelRuns.tar.gz (https://doi.org/10.6084/m9.figshare.21094558.v1). The plotting code is provided in the github repository.

Funding Statement

The Scottish Goverment's Rural and Environment Science and Analytical Services Division (RESAS) funded this research.

References

Bown, R., Gibson, J., Sladen, G., Hicks, B., and Dawson, A. (1974). Effects of lactulose and other laxatives on ileal and colonic ph as measured by a radiotelemetry device. *Gut*, 15(12):999–1004.

Capuano, E., Oliviero, T., Fogliano, V., and Pellegrini, N. (2018). Role of the food matrix and digestion on calculation of the actual energy content of food. *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 dietarysulphate: absorption and excretion in humans. *Gut*,
615 32:766–773.

Hamer, H. M., Jonkers, D., Venema, K., Vanhoutvin, S., Troost, F., and Brummer, R.-J. (2008). The role of butyrate on colonic function. *Alimentary pharmacology & therapeutics*, 27(2):104–119.

Kettle, H., Holtrop, G., Louis, P., and Flint, H. J. (2018). micropop: Modelling microbial populations and communities in r. *Methods in Ecology and Evolution*, 9(2):399–409.

Kettle, H., Louis, P., Holtrop, G., Duncan, S. H., and Flint, H. J. (2015). Modelling the emergent dynamics and major metabolites of the human colonic microbiota. *Enviromental Microbiology*, 17(5):1615–1630.

LaBouyer, M., Holtrop, G., Horgan, G., Gratz, S. W., Belenguer, A., Smith, N., Walker, A. W., Duncan, S. H., Johnstone, A. M., Louis, P., et al. (2022). Higher total faecal short-chain fatty acid concentrations correlate with increasing proportions of butyrate and decreasing proportions of branched-chain fatty acids across multiple human studies. *Gut Microbiome*, 3(e2):1–14.

Lewis, S. and Heaton, K. (1997). Increasing butyrate concentration in the distal colon by accelerating intestinal transit. *Gut*, 41:245–251.

Louis, P. and Flint, H. J. (2017). Formation of propionate and butyrate by the human colonic microbiota. *Environmental Microbiology*, 19(1):29–41.

Louis, P., Hold, G. L., and Flint, H. J. (2014). The gut microbiota, bacterial metabolites and colorectal cancer. *Nature reviews microbiology*, 12(10):661–672.

Mikolajczyk, A. E., Watson, S., Surma, B. L., and Rubin, D. T. (2015). Assessment of tandem measurements of ph and total gut transit time in healthy volunteers. *Clinical and Translational Gastroenterology*, 6(7).

Moorthy, A. S., S.P.J., B., M., K., and H.J., E. (2015). Continuous model of carbohydrate digestion and transport processes in the colon. *PLoS ONE*, 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.

Slavin, J. L., Brauer, P. M., and Marlett, J. A. (1981). Neutral Detergent Fiber, Hemicellulose and Cellulose Digestibility in Human Subjects. *The Journal of Nutrition*, 111(2):287–297.

Smith, N., Shorten, P., and Altermann, E. (2021). Examination of hydrogen cross-feeders using a colonic microbiota model. *BMC Bioinformatics*, 22(3).

Stephen, A. and Cummings, J. (1980). The microbial contribution to human fecal mass. *J. Medical Microbiology*, 13(1):45–56.

Sung, J., Kim, S., Cabatbat, J., Jang, S., Jin, Y., Jung, G. Y., Chia, N., and Kim, P. (2017). Global metabolic interaction network of the human gut microbiota for context-specific community-scale analysis. *Nature communications*.

Topping, D. and Clifton, P. (2001). Short-chain fatty acids and human colonic function: roles of resistant starch and nonstarch polysaccharides. *Physiological Review*, 81(3):1031–1064.

Walker, A. W., Ince, J., Duncan, S. H., Webster, L. M., Holtrop, G., Ze, X., Brown, D., Stares, M. D., Scott, P., Bergerat, A., Louis, P., McIntosh, F., Johnstone, A. M., Lobley, G. E., Parkhill, J., and Flint, H. J. (2011). Dominant and diet-responsive groups of bacteria within the human colonic microbiota. *The ISME Journal*, 5:220–230.

Wang, S. P., Rubio, L. A., Duncan, S. H., Donachie, G. E., Holtrop, G., Lo, G., Farquharson, F. M., Wagner, J., Parkhill, J., Louis, P., Walker, A. W., and Flint, H. J. (2020). Pivotal roles for ph, lactate, and lactate-utilizing bacteria in the stability of a human colonic microbial ecosystem. *mSystems*, 5(5):e00645–20.

Wu, G., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y., Keilbaugh, S., Bewtra, M., Knights, D., Walters, W., Knight, R., and Sinha, R. (2011). Linking long-term dietary patterns with gut microbial enterotypes. *Science*, 334(6052):105–108.

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

microPopGut: R package for simulating microbial populations in the human colon

Helen Kettle

Overview of microPopGut

MicroPopGut is an R package which simulates/predicts the growth of interacting microbial populations in the human colon based on the solution of a system of ordinary differential equations (ODEs). It models the colon as three compartments (proximal, transverse and distal) with a microbial community based on 10 microbial functional groups (MFGs) in each one. Protein and carbohydrates (resistant starch and non-starch polysaccharide (NSP)) enter the proximal compartment. This inflow can be modelled as a constant rate or with fluctuations to simulate meals (which have passed through the stomach and small intestine). The outflow from the distal compartment can be modelled as a constant flow or bowel movements can be simulated.

To simulate microbial growth we use our previous R package, microPop, which is a generic package for modelling microbial communities. The 10 microbial functional groups used in microPopGut are instrinsic data frames in microPop (see Table 1 in the paper Kettle et al. 2018, in the link below for details of the MFGs). The main difference to the usual use of microPop used here is that we include water as a state variable. Water is injested by the host and absorbed through the colon walls and its volume is not constant. When calculating concentrations we use the current volume of water in each compartment.

the microPop package

For background info, the microPop package is described in the paper:

Kettle H, G Holtrop, P Louis, HJ Flint. 2018. microPop: Modelling microbial populations and communities in R. Methods in Ecology and Evolution, 9(2), p399-409. doi: 10.1111/2041-210X.12873

https://besjournals.onlinelibrary.wiley.com/doi/full/10.1111/2041-210X.12873

The user specifies the system via a number of input files (csv files that become dataframes) and the function `microPopModel()` will construct and solve the necessary equations (ordinary differential equations) and provide an output containing the solution (e.g. the concentrations of microbes, substrates and metabolites at the required time points) as well as all the settings/parameters involved in the simulation, and plots of the microbes and resources over time.

Also see the webpage:

https://www.bioss.ac.uk/people/helen/microPop/microPop.html

Getting started with microPopGut

Install the package and add the library:

```
install.packages('microPopGut')
```

```
#> Loading required package: usethis
#> i Loading microPopGut
#> Loading required package: microPop
#> Loading required package: deSolve
#> Loading required package: visNetwork
```

```
library(microPopGut)
```

Basic model run

Note that simulating a complex microbial community of 10 MFGs in 3 compartments is fairly slow to run, so the code evaluated in this tutorial only simulates very short time periods.

In this first example we only use two microbial groups and have constant inflow and outflow to and from the colon and we run the simulation for 2 hours.

```
sim.time.h=2 #time to simulate in hours

m.out=microPopGut(
    numDays=sim.time.h/24, #number of days to simulate
    time.step=1/24/60, #time step at which you want output
    transitTime=1.25,  #time taken to move through colon (days)
    microbeNames=c('Bacteroides','ButyrateProducers1'), #vector of MFG names
                  #(these are data frames provided in the microPop package)
    microbeNames.short=c('Bacteroides'='B','ButyrateProducers1'='BP1') #abbreviated names used for plot

)
#> [1] "simulating growth in compartment 1 - please wait!"
#> [1] "using microPopGut"
#> [1] "Set up completed, ODE solver called..."
#> [1] "simulating growth in compartment 2 - please wait!"
#> [1] "using microPopGut"
#> [1] "Set up completed, ODE solver called..."
#> [1] "simulating growth in compartment 3 - please wait!"
#> [1] "using microPopGut"
#> [1] "Set up completed, ODE solver called..."
```

Here we can see several messages telling us which compartment the processor is currently simulating - this is helpful if the model is running over several hours.

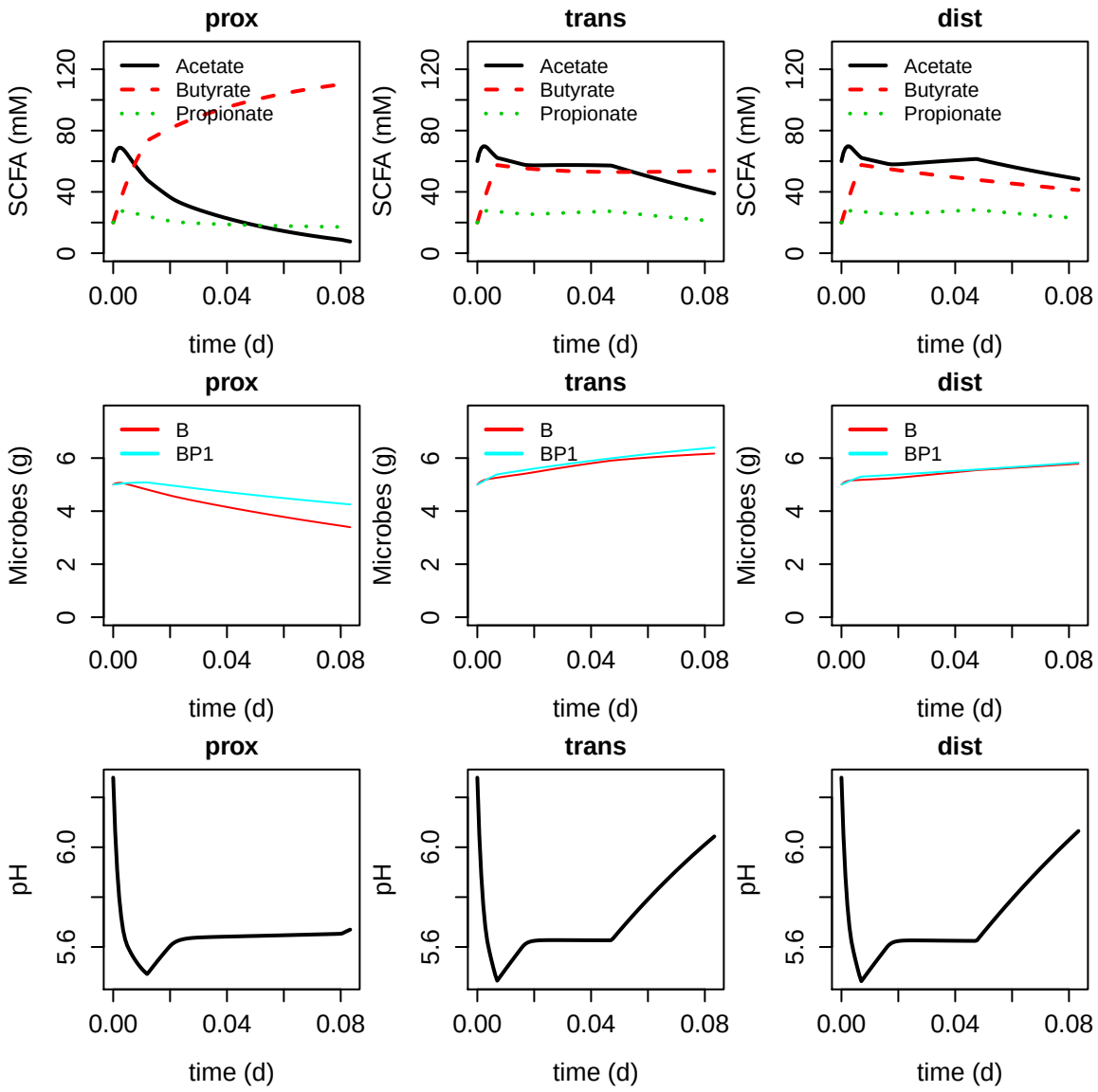
To look at the results of the model there are two inbuilt functions. The first is **verification()** (see below) which gives a summary of the concentrations of the short chain fatty acids (SCFA) in each compartment, which are released as the microbes grow (averaged over the time period **start.av** to **fin.av**), as well as the fraction of the incoming water that reaches the end of the colon (should be about 0.1, i.e. 10%) and the microbial outflow from the colon (should be around 16 g/d). Note, TSCFA is the total SCFA.

```
time=m.out$solution[[1]][,'time']
verification(m.out,start.av=0.8*max(time),fin.av=max(time))
#>                   prox trans  dist
#> TSCFA (mM)      136.1 118.5 117.7
#> Acetate (mM)     10.1  42.8  51.1
#> Butyrate (mM)   108.7  53.4  42.7
```

```
#> Propionate (mM)  17.3  22.3  23.9
#> [1] "Fraction of the incoming water that leaves colon is 0.12"
#> [1] "fecal microbe output rate is 17.04 g/d"
```

Next we use **plotMPG()** to see a summary plot of the SCFA concentration, the mass of each microbial group and the pH in each model compartment.

```
plotMPG(m.out)
```



Changing host diet

A standard western diet is the default setting in the **microPopGut()** but this can be changed via the input arguments, **init** (i.e. the starting mass in each of the 3 model compartments) and **inflow** which is the mass inflow each day of carbohydrate, protein and water. Here you can also specify the division of carbohydrate into resistant starch and NSP using **RS.frac**. In this example we change from 10 g/d of protein to zero and add an extra 10 g/d to the carbohydrate default of 50 g/d. Results are shown in Fig. 1.

```
m.out=microPopGut(
    numDays=1,
    time.step=1/24/60,
    transitTime=1.25,
    microbeNames=c('Bacteroides','ButyrateProducers1'),
    microbeNames.short=c('Bacteroides'='B','ButyrateProducers1'='BP1'),

    #initial mass in each compartment
    init=list(
        C=2, #carbohydrate (g)
        P=0, #protein (g)
        B=10, #biomass (g)
        Acetate=0.3606, #g
        Propionate=0.1482, #g
        Butyrate=0.1762, #g
        W=100), #water (g)

    #inflow from diet
    inflow=list(
        C=60, #carbohydrate (g/d)
        P=0, #protein (g/d)
        W=1100, #water (g/d)
        RS.frac=0.78) #fraction of C that is resistant starch (rest is NSP)

)

time=m.out$solution[[1]][,'time']
verification(m.out,start.av=0.8*max(time),fin.av=max(time))
dev.new()
plotMPG(m.out)
```

Meal composition/fluctuations

The default setting in **microPopGut()** is constant inflow but there is also the option to include a more realistic inflow which aims to represent intermittent host eating. This is specified in the **meals** input list.

The **gamma.mag** option in the **meals** list controls the magnitude of the fluctuations in meal composition and size by controlling the spread of the gamma distribution that the values are drawn from for each substrate.

```
m.out=microPopGut(
    numDays=1,
    time.step=1/60/24,
    transitTime=1.25,
    microbeNames='Bacteroides',
    microbeNames.short=c('Bacteroides'='B'),
    meals=list(
        seed=1,
        fluc.inflow=TRUE,
        fluc.subst.comp=TRUE,
        plotInflow=TRUE,
        saveInflowFig=TRUE,
        breakfast.start=7, #time (24 h clock) to start breakfast
```

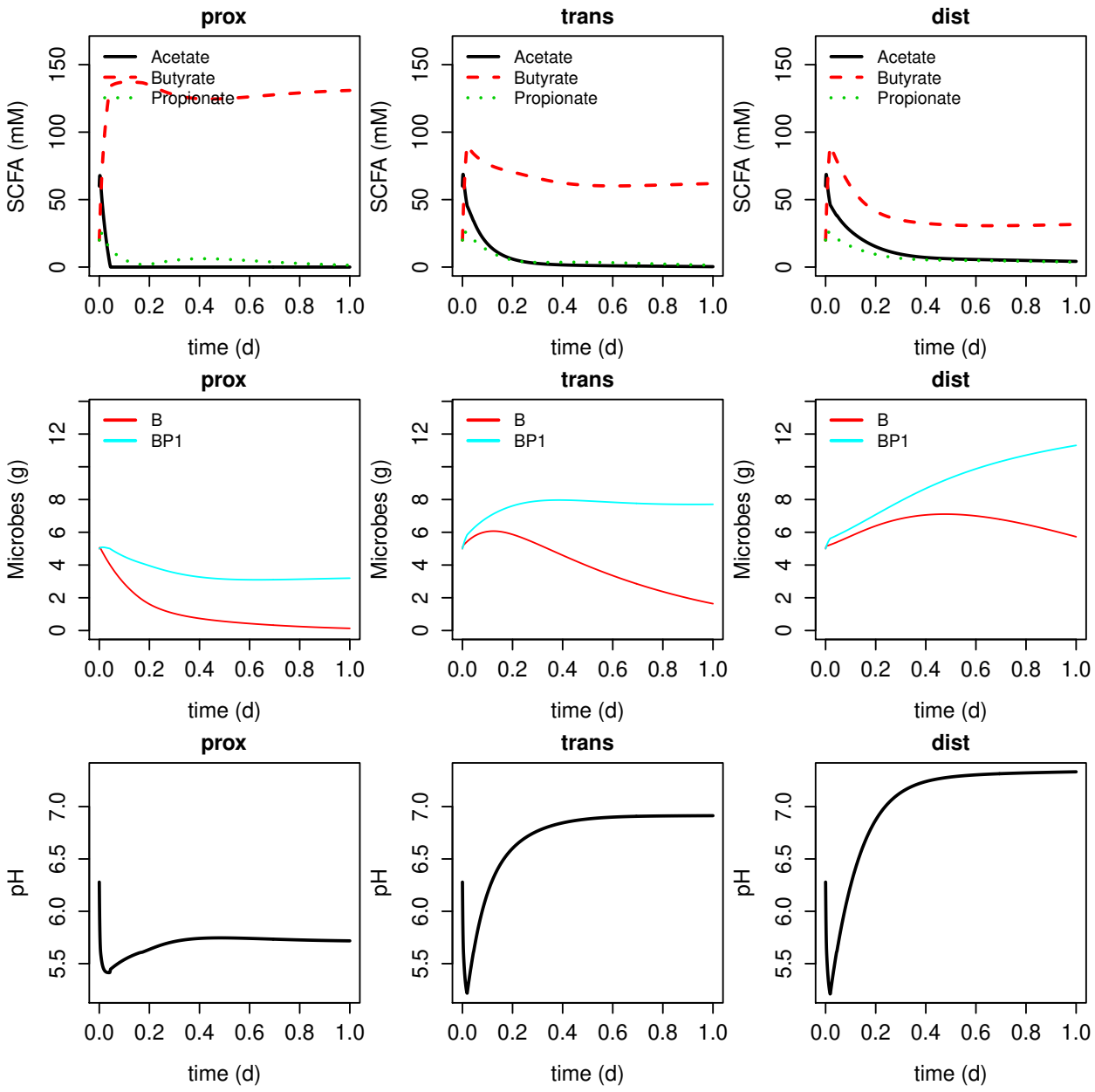


Figure 1: Summary results for a diet with no protein and 2 MFGs

```
        lunch.start=13,     #time (24 h clock) to start lunch
        dinner.start=19,    #time (24 h clock) to start dinner
        meal.duration.h=0.5, #length of time eating one meal in hours
        time.to.reach.colon.h=7, #time take to pass through stomach and small intestine in hours
        gamma.mag=1 #scaling factor to control the fluctuations in meals
    )
)
```

Figs. 2-4 shows meals over 7 days for gamma.mag equal to 0.1, 1 and 1.9 respectively for 3 arbitrary substrates A, B and C, and water. Note that after stochastically generating the composition the output is scaled so that the mean substrate values are maintained over the time period (this information is printed to screen when you run the model). Due to smoothing delays as the meals pass through the small intestine and stomach the final inflow to the colon may have slightly smaller means (this is also printed to screen).

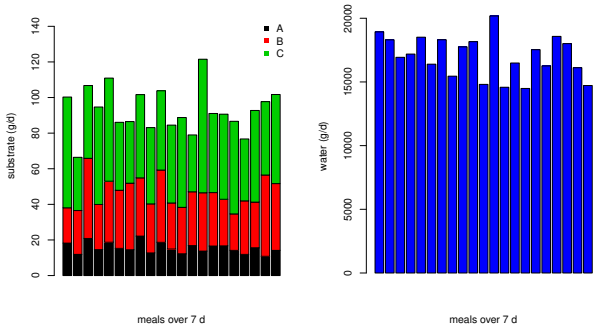


Figure 2: Meal composition for gamma.mag=0.1

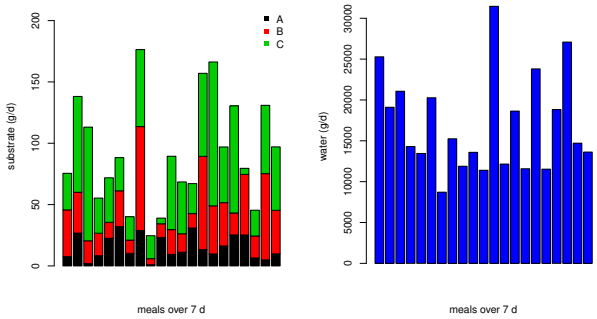


Figure 3: Meal composition for gamma.mag=1

The "meals" are then passed through an ODE model representing the stomach and small intestine - the amount of smoothing caused by this is controlled by the time.to.reach.colon.h item in the meals list. The fig. below shows the input to the colon for gamma.mag=1.9 and time.to.reach.colon.h=7 hours.

Full gut model (10 MFGs with meals for 7 days)

In this example we show the settings for simulating the full model. In this example we also change the pH tolerance and the maximum growth rate on resistant starch of the LactateProducers group. The code below

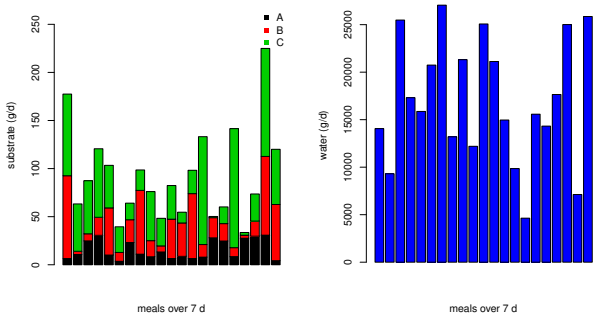


Figure 4: Meal composition for gamma.mag=1.9

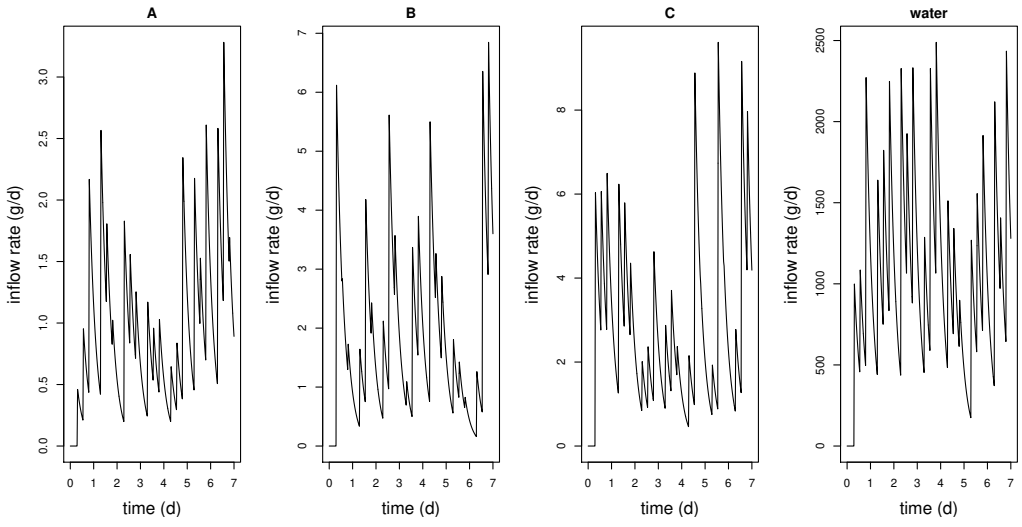


Figure 5: Substrate inflows to colon (after passing through stomach and small intestine model) for gamma.mag=1.9

takes over an hour to run so if you want to try it out it is better to change **num.days** from 5, to 1 or less, or to try fewer groups e.g. microbeNames=microbeNames[1:2].

```
microbeNames = c('Bacteroides','ButyrateProducers1',
                 'ButyrateProducers2','ButyrateProducers3',
                 'LactateProducers','PropionateProducers',
                 'Methanogens','NoButyFibreDeg',
                 'NoButyStarchDeg','Acetogens')

microbeNames.short = c('Bacteroides'='B','ButyrateProducers1'='BP1',
                   'ButyrateProducers2'='BP2','ButyrateProducers3'='BP3',
                   'LactateProducers'='LP','PropionateProducers'='PP',
                   'Methanogens'='M','NoButyFibreDeg'='NBFD',
                   'NoButyStarchDeg'='NBSD','Acetogens'='A')

#Change pH tolerance & Gmax(RS) for lactate producers
#read LPc from batch file
LactateProducers['pHcorners',2:5]=c(4.5,5.25,7.2,7.95)
LactateProducers['maxGrowthRate','RS']=7

m.out=microPopGut(
    numDays=5,
    time.step=1/24/60,
    transitTime=1.25,
    microbeNames=microbeNames,
    microbeNames.short=microbeNames.short,
    meals=list(
        seed=1,
        fluc.inflow=TRUE,
        fluc.subst.comp=TRUE,
        plotInflow=TRUE,
        saveInflowFig=TRUE,
        breakfast.start=7,
        lunch.start=13,
        dinner.start=19,
        meal.duration.h=0.5,
        time.to.reach.colon.h=7,
        gamma.mag=1
    )
)
```

See Fig. 6 for the results of **plotMPG(m.out)**

Bowel Movements

By default the model is set up for zero bowel movements per day (BMpd) i.e there is constant outlow. To turn on bowel movements, set **BMpd** in the **bowel.movements** input to either 1, 2 or 3 (see code below). The **start.BM.time** list gives the start times for 1, 2 or 3 bowel movements per day and the bowel movement duration is set using **BM.duration.h**. In our model we assume that bowel movements only affect the distal part of the colon.

```
  microbeNames = c('Bacteroides','ButyrateProducers1','ButyrateProducers2','ButyrateProducers3','Lactat

microbeNames.short = c('Bacteroides'='B','ButyrateProducers1'='BP1','ButyrateProducers2'='BP2','Butyrat
```

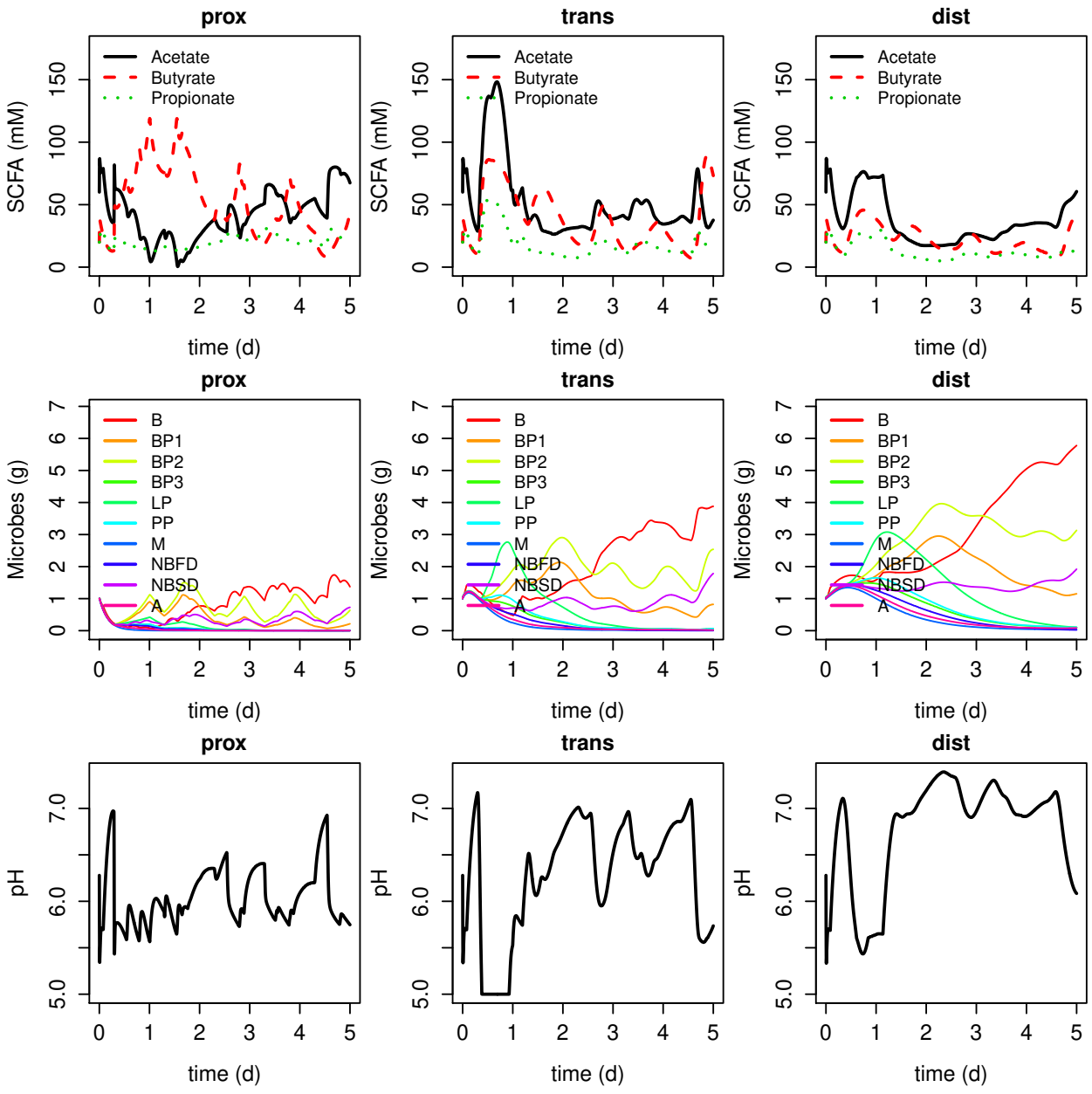


Figure 6: Summary results for 10 MFGs over 5 days

```
sim.time.h=48 #time to simulate in hours

m.out=microPopGut(
    numDays=sim.time.h/24,
    time.step=1/24/60,
    transitTime=1.25,
    microbeNames=microbeNames[1:2],
    microbeNames.short=microbeNames.short[1:2],
    bowel.movements = list(
        BM.duration.h = 15/60,
        frac.distal.emptied = 0.9,
        BMpd = 3,
        start.BM.time = list(7, c(7, 19), c(7, 15, 21)))
)

time=m.out$solution[[1]][,'time']
verification(m.out,start.av=0.8*max(time),fin.av=max(time))
dev.new()
plotMPG(m.out)
```

Multiple strains per group

By default there is only one strain in each MFG but using the functionality of microPop (on which microPopGut is based) we can add multiple strains in each group. In versions of microPop from 1.6 onwards we can have different numbers of strains in each group. The parameters for each strain in a group are drawn randomly from a given range for that particular group. The particulars of this are controlled by the **strainOptions** input argument.

```
m.out=microPopGut(
    numDays=2,
    time.step=1/24/60,
    transitTime=1.25,
    microbeNames=c('Bacteroides','ButyrateProducers1'),
    microbeNames.short=('Bacteroides'='B','ButyrateProducers1'='BP1'),
    numStrains=c('Bacteroides'=3,'ButyrateProducers1'=2),
    strainOptions = list(
        randomParams = c("halfSat", "yield", "maxGrowthRate",
        "pHtrait"),
        seed = 3,
        distribution = "uniform",
        percentTraitRange = 10,
        maxPHshift = 0.1,
        applyTradeOffs = FALSE,
        tradeOffParams = NULL,
        paramsSpecified = FALSE,
        paramDataName = NULL)
)

time=m.out$solution[[1]][,'time']
```

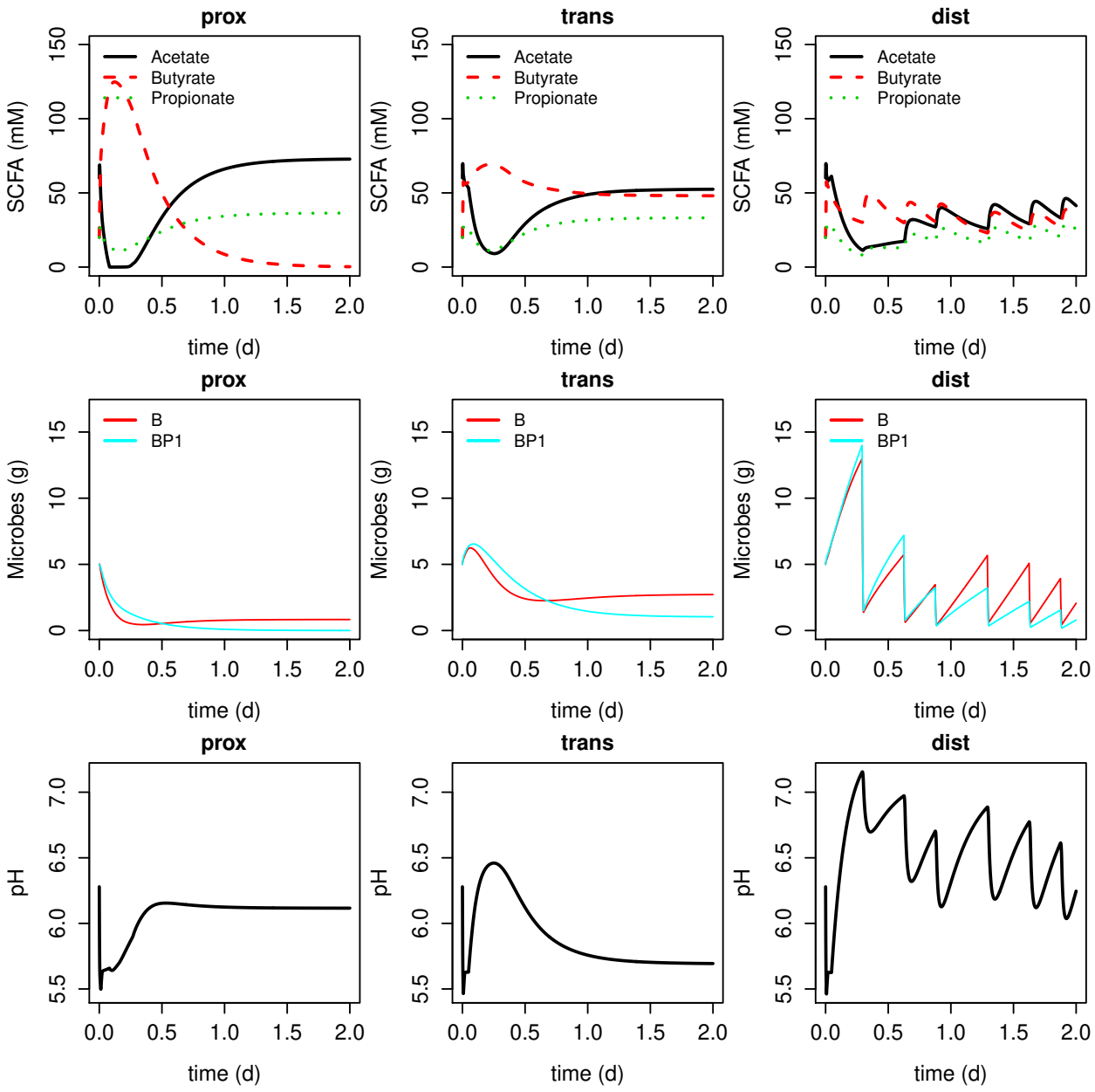


Figure 7: Summary results for 2 MFGs over 2 days with 3 bowel movements per day

```
verification(m.out,start.av=0.8*max(time),fin.av=max(time))
plotMPG(m.out)
```

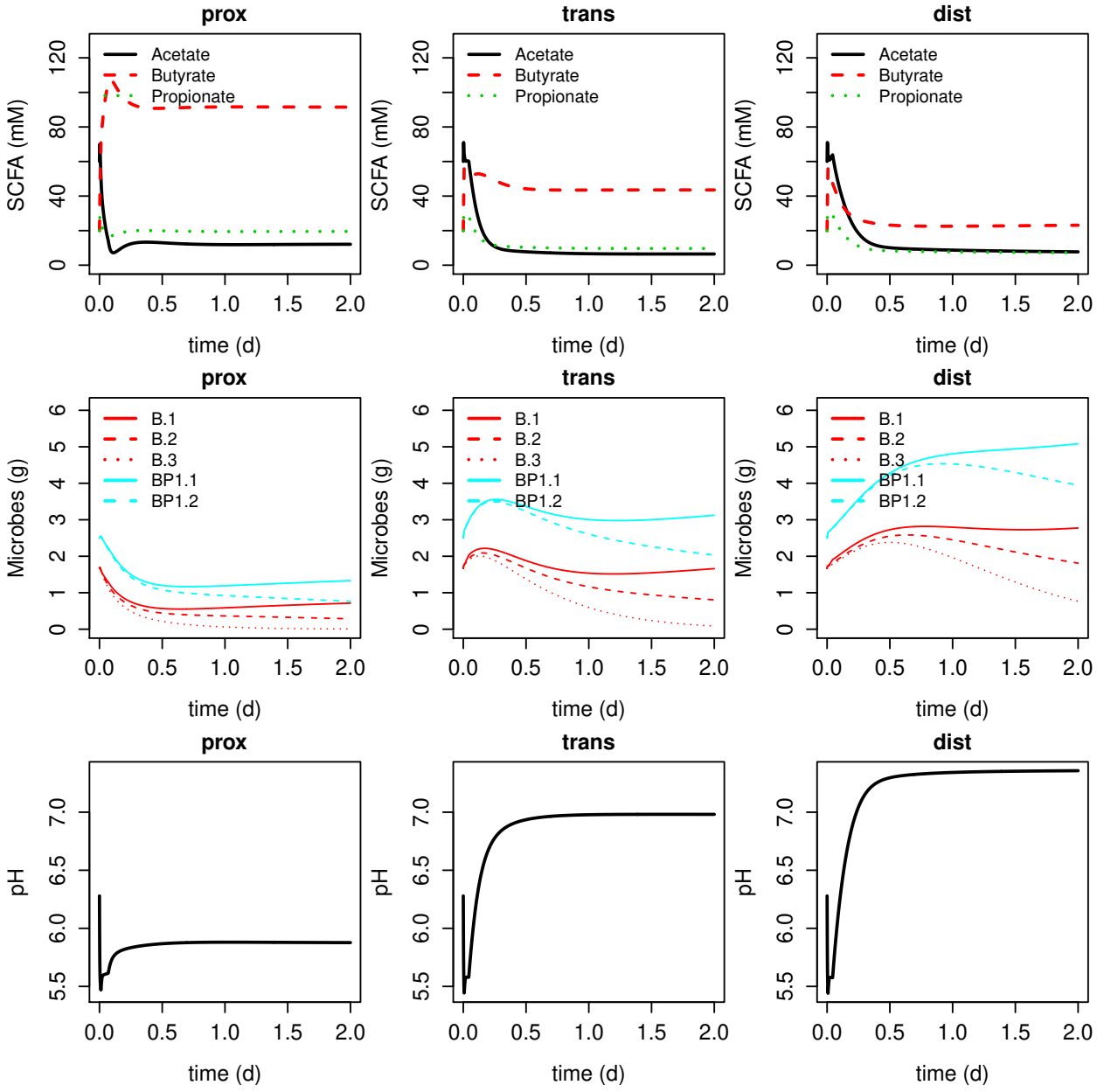


Figure 8: Summary results for 2 MFGs: Bacteroides with 3 strains and ButyrateProducers1 with 2 strains

Troubleshooting

Failure of ODE solver

If you get warnings about time steps then this mean the ODE solver is failing - generally because the problem is too stiff i.e. there are rapid changes in time. Ways to deal with this are to make your time step

(microPopGut input argument) smaller, to make gamma.mag smaller (nearer to zero) which will decrease the size of the meal fluctuations, or to alter the tolerances in the ODE solver (see ode.options list of microPopGut input arguments). If you are running the system with only one group, adding more groups often makes the system more stable as this slows down the growth of each individual group due to resource competition.

Supp. Info. for "Process-based modelling of microbial community dynamics in the human colon"

Helen Kettle, Petra Louis and Harry J. FLint

July 1, 2022

1 Mathematical Model

1.1 Simple Model

In this section we present a very simple model with one microbial group and one colon compartment that we then use to derive bounds on parameters (e.g. absorption of SCFA and water) and to look at the bulk properties of the system, e.g. the relationship between transit time and SCFA concentration.

This simple model consists of, bacteria (X), substrate (S), SCFA mass (Z) and water (W) all with units of mass. We set f_s as the fraction of the waste products of X that are SCFA, and Y is the amount of microbial growth for 1 g of S and a_Z and a_W are the absorption rates of Z and W. The rates of change are given by

$$\frac{dX(t)}{dt} = G(t)X(t) - X(t)V \tag{1}$$

$$\frac{dS(t)}{dt} = \dot{S_{in}} - \frac{G(t)X(t)}{Y} - S(t)V \tag{2}$$

$$\frac{dZ(t)}{dt} = f_s\left(\frac{1}{Y} - 1\right)G(t)X(t) - (V + a_Z)Z(t) \tag{3}$$

$$\frac{dW(t)}{dt} = \dot{W_{in}} - (V + a_W)W(t) \tag{4}$$

where microbial growth, G, is given by

$$G(t) = G^m \frac{S(t)}{S(t) + K} \tag{5}$$

where K is the half-saturation constant and G^m is the maximum growth rate of X on S. Transit time is incorporated via the washout rate, V such that $V = 1/T_t$.

Steady state analysis (i.e. when the system is not changing with time) of the one group model can be used to give us some bounds or checks on the bulk properties of the system. The steady state solution (at time, t_s), assuming $X > 0$, is given by

$$X(t_s) = (\dot{S_{in}}/V - S(t_s))Y \tag{6}$$

$$S(t_s) = \frac{VK}{G^{\max} - V} \quad (7)$$

$$Z(t_s) \approx VX(t_s)\frac{1-Y}{Y(a_Z + V)} \quad (8)$$

$$W(t_s) = \frac{\dot{W_{in}}}{a_W + V} \quad (9)$$

where $\dot{X_{in}}$ is the inflow rate of X.

1.2 Microbial yield and substrate inflow

Assuming the microbes consume all available substrate, then the steady state mass of microbes can be approximated by

$$X(t_s) \approx \frac{\dot{S_{in}}Y}{V} \quad (10)$$

where $\dot{S}_{in}$ is the dietary inflow of all substrates (i.e. dietary P, C and mucin); V is the wash out rate from the system and Y, the microbial yield. Assuming the output of microbes (given by $X_{t_s}V$) is 14-28 g d^{-1} (Stephen and Cummings, 1980) (with midpoint of 21) and the substrate inflow is about 65 g d^{-1}; Eq. 10 suggests that Y is 21/65 i.e. about 0.3 which matches very well with the yield values for our functional groups which have yield values around 0.28 or 0.33 (see other Supp. Info. file).

1.3 Specific water absorption, a_W

Extending Eq. 9 to N compartments with downstream flow from 1 to N, and assuming the specific absorption rate is the same in all, then at steady state the water in each compartment is given by,

$$W_1 = \frac{\dot{W_{in}}}{a_W + V_1}, \quad (11)$$

$$W_2 = \frac{W_1 V_1}{a_W + V_2}, ... \quad (12)$$

$$..., W_N = \frac{W_{N-1}V_{N-1}}{a_W + V_N} \quad (13)$$

Successively substituting for the unknowns gives

$$W_N = \frac{\dot{W_{in}}V_1 V2...V_{N-1}}{(a_W + V_1)(a_W + V_2)...(a_W + V_N)} \quad (14)$$

If 90% of water is absorbed over the transit time then in the last compartment, N, $W_N V_N = 0.1\dot{W_{in}}$. Substituting this into Eq. 14 gives

$$\prod_{k=1}^{N}(a_W + V_k) = 10\prod_{j=1}^{N} V_j. \quad (15)$$

This can be solved numerically where V_j is computed by dividing the colon into N compartments which each take fraction, f_j^T, of the total transit time to pass

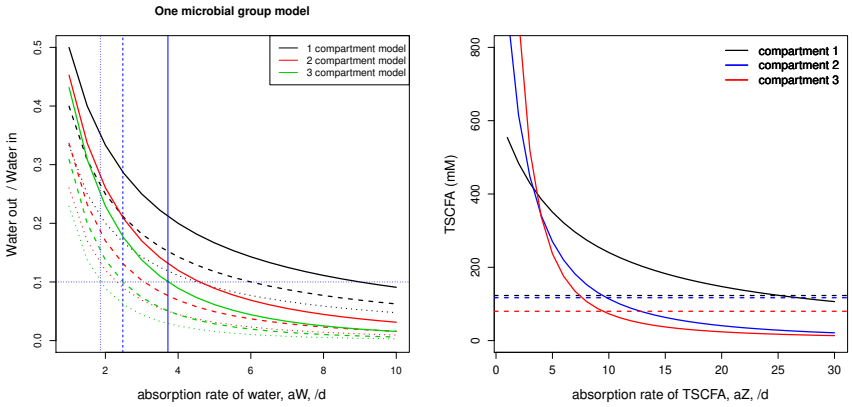


Figure S1: a) Achieving 90% water absorption for different transit times (1, 1.5 and 2 days represented by solid, dashed and dotted line respectively) and different number of compartments (1, 2, and 3 by colour, as shown in legend). The dotted horizontal line shows the required value for 90% incoming water absorption in the colon and the vertical blue lines show the a_W value which gives the correct total absorption for the 3 different transit times (which are as before, 1, 1.5 and 2 days represented by solid, dashed and dotted line respectively). b) Investigating SCFA absorption. TSCFA (mM) for different a_Z for transit time of 1.25 days, for the one group, three compartment model. a_Z is from 1,2,...30 /d with values constant throughout the colon. The dashed horizontal lines show the expected TSCFA in each compartment. The simulation is for 5 days and the results are the mean over the last day. Constant inflow and outflow (no meals or bowel movements) with a_W changing with transit time according to Eq. 16.

through. Using fractional times based on compartment volume (Fig. 1 in main manuscript) and assuming that a_W is the same in each compartment we find that for a one compartment model, $a_W = \frac{9}{T_t}$; for a two compartment model, $a_W = \frac{4.59}{T_t}$; and for a three compartment model, $a_W = \frac{3.72}{T_t}$ (see Fig. S1). This can be expressed exactly by

$$a_W = \frac{16.95 - 9.72N + 1.77N^2}{T_t} \tag{16}$$

where N is the number of compartments in the model. Note that this does not mean that specific water absorption changes with transit time, rather that to fulfill the 90% absorption criteria we can set a_W based on N and T_t. Once a typical transit time is chosen, the value of a_W can be fixed. As a rough estimation, $a_W \approx 3$ /d for a 3 compartment model with a transit time between one to one and a half days (Fig. S1). Given this will not be significantly affected by the microbial model (microbial uptake/production of water is small) this result will apply to all of the models in this work.

1.4 Specific SCFA absorption, a_Z

Using our one group microbial group model but adapted for 3 compartments, and our estimation for a_W based on transit time and the number of compartments (Eq. 16), we run the model for a transit times of 1.25 d with continuous inflow and outflow, over a range of a_Z from 1-30 d^{-1}. We compute TSCFA from our model by converting Z from g to mM using

$$Z_{mM} = 10^6 \frac{Z_g}{W_g m_Z} \tag{17}$$

where m_Z is computed by assuming TSCFA is in the ratio 3:1:1 (Ac:Bu:Pr) to give a weighted mean molar mass of TSCFA, m_Z of 68.4 g mol^{-1}. Fig. S1, shows the TSCFA in each model compartment versus a_Z. The horizontal dashed lines show the TSCFA value matching the model criteria, indicating the best estimates were a_Z equal to 25.2, 4.2 and 9.2 d^{-1} in the proximal, transverse and distal colon respectively. However, this was determined using a_Z constant through the colon so if a_Z varies between compartments this will change the results. In the interests of a robust model (i.e. the fewer parameter values, the better) we made the decision to use one value for a_Z . Given the experimental value of 9.6 d^{-1} compares well with our best estimate for the distal colon (9.2 d^{-1}) we decided to set aZ =9.6 d^{-1} throughout. It should be noted however that decreasing a_Z along the colon has been implemented in other models e.g. Labarthe et al. (2019).

2 Effect of Transit Time

Experimental evidence (e.g. (Lewis and Heaton, 1997)) shows that TSCFA (mM) decreases as transit time increases. We can explain why this is, mathematically, using a very simple one group model with monod growth, which we can solve analytically at steady state. To compute TSCFA in mM we need to use the fraction of P that is SCFA and then divide by the mean molor mass (m_m)and multiply by 1000 to find mmol. We then need to divide by W in litres, thus,

$$\text{TSCFA} = 10^6 \frac{P f_s}{W m_m} \tag{18}$$

Substituting for P and W, ignoring scaling constants and assuming remaining substrate at steady state is negligible, shows that TSCFA is linearly related to the expression

$$\frac{\dot{S_{in}}}{\dot{W_{in}}} \frac{a_W + V}{a_P + V} \tag{19}$$

To see the effect of simply changing the transit time through the colon on TSCFA we assume $\dot{S_{in}}$ and $\dot{W_{in}}$ are fixed and replace V by $1/Tt$ to get

$$TSCFA \propto \frac{a_W T_t + 1}{a_P T_t + 1} \tag{20}$$

Since we have a_W=3 and a_P=9.6, the denominator will increase much faster than the numerator as T_t increases thus, theoretically, TSCFA will decrease as transit time increases as SCFA are absorbed faster than water. Using realistic

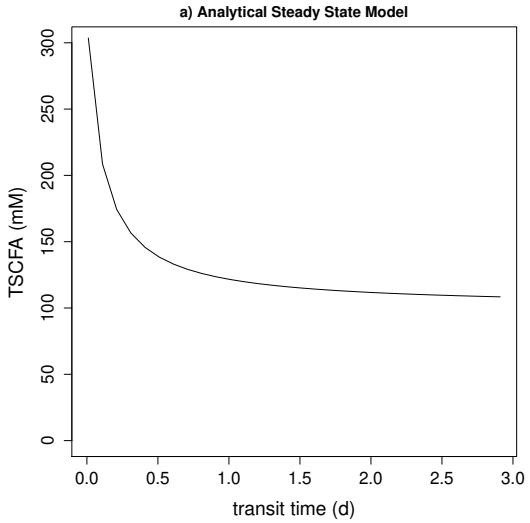


Figure S2: TSCFA as a function of transit time obtained from the solution of Eqs. 2-4. We convert from product mass, P, to moles using the average molar mass (weighted according to A:B:P = 3:1:1) of 68.5 g/mol and compute f_s as an average of the microPop microbial group stoichiometries to be approximately 0.5. We set inflowing substrate at 65 g/d (dietary substrate plus mucin) and inflowing water at 1100 g/d, with parameter values Y=0.3, $G^{\max}$=20 /d and K=0.001. [transitTimeModel.R]

parameter values in the above model (Eq. 2-4) allows us to plot TSCFA against transit time – see Fig S2 which compares very well with the experimental data shown in Fig. 1 by Lewis and Heaton (1997).

3 Microbial group parameter values

The parameters describing the different microbial groups are the same as the intrinsic functional groups given in the microPop R package (version 1.6), with one exception. We increased the maximum growth rate of Lactate Producers on RS from 6 d^{-1} to 7 d^{-1} and their pH tolerance were cordinates changed to tolerate lower pH (first two pH coordinates now 4.5 and 5.25, rather than 4.95 and 5.7) to ensure a better chance of their survival in the model. This section shows the data frames used for each microbial group in microPopGut. The following list explains the different entries in these data frames.

- 'Rtype' refers to the substrate type on the pathway:
  - 'X': not involved in pathway
  - 'S': substitutable substrate (this can be interchanged with other substitutable substrates)
  - 'Se': essential substrate (the microbes can not grow without this)
  - 'Sb': boosting substrate (if this is present the microbe can grow faster)

  - 'Sw': water
  - 'P': metabolic product

- 'halfsat' is the half saturation constant for monod growth
- 'yield' is the microbial mass produced from one gram of substrate
- 'maxGrowthRate' is the specific maximum growth rate of the microbes
- 'stoichiom' refers to the number of moles of each molecules involved in growth
- 'keyResource' is the substrate whose uptake rate is used to compute the uptake of the other substrates on the pathway according to the stoichiometry
- 'numPathways' defines how many metabolic pathways the microbial group has. When there is more than one pathway, numbered parameter names for the subsequent pathways are used.

For more details please refer to Kettle et al. (2015) and Kettle et al. (2018) or use the help function within the microPopGut package.

Table 1: Bacteroides

	units	Protein	NSP	RS	Acetate	Propionate	Succinate	H2	CO2	other
Rtype	none	X	S	S	P	P	P	P	P	X
halfSat	g/l		0.001	0.001						
yield	g/g		0.286	0.333						
maxGrowthRate	/d		12	24						
stoichiom	mol		2	2	2	1	1	2	1	
keyResource	none									
numPathways	none	2								
Rtype.2	none	S	X	X	P	P	P	P	P	P
halfSat.2	g/l	0.001								
yield.2	g/g	0.2								
maxGrowthRate.2	/d	24								
stoichiom.2	mol	6			2	1	1	2	1	7
keyResource.2	none									
pHcorners	pH	5.6	6.35	7.85	8.6					

Table 2: NoButyStarchDeg

	units	NSP	RS	Acetate	H2	CO2	H2O
Rtype	none	S	S	P	P	P	Sw
halfSat	g/l	0.001	0.001				
yield	g/g	0.286	0.333				
maxGrowthRate	/d	3.6	14.4				
stoichiom	mol	1	1	2	4	2	2
keyResource	none						
numPathways	none	1					
pHcorners	pH	5.35	6.1	7.6	8.35		

Table 3: NoButyFibreDeg

	units	NSP	RS	Acetate	Succinate	H2
Rtype	none	S	S	P	P	P
halfSat	g/l	0.001	0.001			
yield	g/g	0.286	0.333			
maxGrowthRate	/d	16.8	3.6			
stoichiom	mol	1	1	1	1	1
keyResource	none					
numPathways	none	1				
pHcorners	pH	5	5.75	7.25	8	

Table 4: LactateProducers

	units	NSP	RS	Sugars	Acetate	Lactate	Formate	Ethanol	H2O
Rtype	none	S	S	S	P	P	P	P	Sw
halfSat	g/l	0.001	0.001	0.001					
yield	g/g	0.286	0.333	0.333					
maxGrowthRate	/d	7.2	7	24					
stoichiom	mol	6	6	6	10	4	2	1	1
keyResource	none								
numPathways	none	1							
pHcorners	pH	4.5	5.25	7.2	7.95				

Table 5: ButyrateProducers1

	units	NSP	RS	Sugars	Acetate	Butyrate	H2	CO2	H2O
Rtype	none	S	S	S	Sb	P	P	P	P
halfSat	g/l	0.001	0.001	0.001	0.001				
yield	g/g	0.286	0.333	0.333					
maxGrowthRate	/d	8.4	8.4	24					
stoichiom	mol	2	2	2	2	3	2	4	2
keyResource	none	Hex							
numPathways	none	1							
nonBoostFrac	none	0.75							
pHcorners	pH	4.95	5.7	7.2	7.95				

Table 6: ButyrateProducers2

	units	NSP	RS	Sugars	Acetate	Butyrate	Lactate	Formate	CO2	H2O
Rtype	none	S	S	S	Sb	P	P	P	P	P
halfSat	g/l	0.001	0.001	0.001	0.001					
yield	g/g	0.286	0.333	0.333						
maxGrowthRate	/d	14.4	7.2	24						
stoichiom	mol	6	6	6	4	7	2	6	4	4
nonBoostFrac	none	0.1								
keyResource	none	Hex								
numPathways	none	1								
pHcorners	pH	4.85	5.6	7.1	7.85					

Table 7: PropionateProducers

	units	NSP	RS	Sugars	Acetate	Propionate	CO2	Lactate	H2O
Rtype	none	S	S	S	P	P	P	X	P
halfSat	g/l	0.001	0.001	0.001					
yield	g/g	0.286	0.333	0.333					
maxGrowthRate	/d	7.2	7.2	24					
stoichiom	moles	3	3	3	2	4	2		2
keyResource	none								
numPathways	none	2							
Rtype.2	none	X	X	X	P	P	P	Se	P
halfSat.2	g/l							0.001	
yield.2	g/g							0.111	
maxGrowthRate.2	/d							4.8	
stoichiom.2	moles				1	2	1	3	1
keyResource.2	none	Lactate							
pHcorners	pH	4.75	5.5	7	7.75				

Table 8: ButyrateProducers3

	units	NSP	RS	Sugars	Acetate	Butyrate	Formate	H2	CO2	Lactate	H2O
Rtype	none	S	S	S	P	P	P	P	P	X	Sw
halfSat	g/l	0.001	0.001	0.001							
yield	g/g	0.286	0.333	0.333							
maxGrowthRate	/d	7.2	7.2	24							
stoichiom	mol	10	10	10	2	9	12	10	8		2
keyResource	none										
numPathways	none	2									
Rtype.2	none	X	X	X	Se	P	X	P	P	Se	P
halfSat.2	g/l				0.001					0.001	
yield.2	g/g									0.111	
maxGrowthRate.2	/d									4.8	
stoichiom.2	mol				2	3		2	4	4	2
keyResource.2	none	Lactate									
pHcorners	pH	4.85	5.6	7.1	7.85						

Table 9: Acetogens

	units	NSP	RS	Sugars	Acetate	H2	CO2	Formate	H2O
Rtype	none	S	S	S	P	X	X	X	X
halfSat	g/l	0.001	0.001	0.001					
yield	g/g	0.286	0.333	0.333					
maxGrowthRate	/d	7.2	7.2	24					
stoichiom	moles	1	1	1	3				
keyResource	none								
numPathways	none	3							
Rtype.2	none	X	X	X	P	Se	Se	X	P
halfSat.2	g/l					0.001	0.001		
yield.2	g/g						0.03		
maxGrowthRate.2	/d						2.4		
stoichiom.2	moles				1	4	2		2
keyResource.2	none	CO2							
Rtype.3	none	S	S	S	P	P	P	Se	X
halfSat.3	g/l	0.001	0.001	0.001				0.001	
yield.3	g/g	0.286	0.333	0.333					
maxGrowthRate.3	/d	7.2	7.2	24					
stoichiom.3	moles	1	1	1	3	2	2	2	
keyResource.3	none	Hex							
pHcorners	pH	5.25	6	7.5	8.25				

Table 10: Methanogens

	units	H2	CO2	CH4	H2O	Formate
Rtype	none	Se	Se	P	P	X
halfSat	g/l	0.001	0.001			
yield	g/g		0.03			
maxGrowthRate	/d		2.4			
stoichiom	mol	4	1	1	2	
keyResource	none	CO2				
numPathways	none	2				
Rtype.2	none	X	P	P	P	Se
halfSat.2	g/l					0.001
yield.2	g/g					0.00724
maxGrowthRate.2	/d					2.4
stoichiom.2	mol		3	1	2	4
keyResource.2	none	Formate				
pHcorners	pH	5.25	6	7.5	8.25	

References

H Kettle, G Holtrop, P Louis, and Harry J. Flint. micropop: Modelling microbial populations and communities in r. *Methods in Ecology and Evolution*, 9 (2):399–409, 2018. doi: 10.1111/2041-210X.12873.

Helen Kettle, Petra Louis, G Holtrop, Sylvia H. Duncan, and Harry J. Flint. Modelling the emergent dynamics and major metabolites of the human colonic microbiota. *Enviromental Microbiology*, 17(5):1615–1630, 2015. doi: 10.1111/1462-2920.12599.

Simon Labarthe, Bastien Polizzi, Thuy Phan, Thierry Goudon, Magali Ribot, and Beatrice Laroche. A mathematical model to investigate the key drivers of the biogeography of the colon microbiota. *Journal of Theoretical Biology*, 462:552 – 581, 2019. ISSN 0022-5193. doi: https://doi.org/10.1016/j.jtbi.2018.12.009. URL `http://www.sciencedirect.com/science/article/pii/S002251931830599X`.

S.J. Lewis and K.W. Heaton. Increasing butyrate concentration in the distal colon by accelerating intestinal transit. *Gut*, 41:245–251, 1997.

AM Stephen and JH Cummings. The microbial contribution to human fecal mass. *J. Medical Microbiology*, 13(1):45–56, 1980. doi: https://doi.org/10.1099/00222615-13-1-45.