



Universiteit  
Leiden  
The Netherlands

**The effect of mucus on the infant microbiota**  
Wijtkamp, C.

**Citation**

Wijtkamp, C. *The effect of mucus on the infant microbiota.*

Version: Not Applicable (or Unknown)

License: [License to inclusion and publication of a Bachelor or Master thesis in the Leiden University Student Repository](#)

Downloaded from: <https://hdl.handle.net/1887/4171481>

**Note:** To cite this publication please use the final published version (if applicable).

LEIDEN UNIVERSITY

MATHEMATICAL INSTITUTE  
INSTITUTE BIOLOGY LEIDEN

---

**The effect of mucus on the infant  
microbiota**

---

*Author*  
C. WIJTKAMP

*Supervisor*  
Prof.dr. R.M.H. MERKS  
D.M. VERSLUIS

December 13, 2021

## Abstract

Microbiota development in the infant gut begins shortly after birth, and can influence the health of the infant later in life. The epithelium of the gastrointestinal tract is covered in a layer of mucus. Apart from being a defence barrier, mucus also serves as a potential nutrition source for the microbiota. We hypothesize that mucus increases the abundances of species that are able to feed on mucus related compounds, and that mucus thus plays a role in the diversity of the gut microbiota. To study the potential effect of mucus, we use a multiscale spatiotemporal dynamic model that considers metabolism, spatial bacterial population dynamics and resource dynamics and represents the colon of the infant. The model predicts that the species *Bacteroides vulgatus* benefits the most from mucus as an additional nutrition source. This project is an exploratory study that contributes to a better additional understanding of the formation of microbiota in the infant.

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Mucus background information</b>	<b>4</b>
<b>3</b>	<b>Model</b>	<b>7</b>
3.1	Previous models . . . . .	7
3.2	Model outline . . . . .	8
3.2.1	Species composition . . . . .	9
3.2.2	Mucin glycans . . . . .	9
3.2.3	Metabolites . . . . .	11
3.2.4	Flux Balance Analysis . . . . .	12
3.2.5	Output of the model . . . . .	14
<b>4</b>	<b>Results</b>	<b>15</b>
4.1	Model predicts <i>B. vulgatus</i> dominance with increasing MG81 generation . . . . .	15
4.2	<i>P. distasonis</i> grows less on lactose compared to other species . . . . .	17
4.3	Co-existence is possible between the two most abundant species <i>B. vulgatus</i> and <i>B. longum</i> . . . . .	18
4.4	<i>B. vulgatus</i> consumes more lactose when there is more addition of MG81 . . . . .	19
4.5	All possible intermediate reactions in the degradation of MG81 are performed . . . . .	22
4.6	The output metabolites consist mainly of acetate, hydrogen and tn antigen . . . . .	22
4.7	Spatial patterning of the species shows that lactose uptake is related to the volume of the species . . . . .	24
4.8	More cross-feeding at the distal part compared to the proximal part of the lattice . . . . .	29
4.9	Distinctive patterns of the abundances of the species result in a different composition of metabolites . . . . .	30
4.10	<i>B. vulgatus</i> benefits less with a change in mucin glycan . . . . .	31
4.11	<i>B. vulgatus</i> knock-out for MG81 degradation shows less advantage of MG81 compared to the normal <i>B. vulgatus</i> . . . . .	35
<b>5</b>	<b>Discussion</b>	<b>39</b>
<b>6</b>	<b>Supplementary</b>	<b>41</b>
6.1	Exchange metabolites related to MG81 . . . . .	41
6.2	Movie . . . . .	41
6.2.1	Video. Video of the simulation where <i>B. vulgatus</i> is the dominant species . . . . .	43
6.2.2	Video. Video of the simulation where <i>B. longum</i> is the dominant species . . . . .	43
6.3	Specification for cross-feeding conditions . . . . .	44

# 1 Introduction

The human gastrointestinal tract harbours hundreds of microbial species. These species together form individual microbiotas and these play important roles in human health and diseases. Therefore they are the subject of extensive research, establishing its involvement in human metabolism, nutrition, physiology, and immune function. A large part of these studies focuses on the microbiota of an adult host. However, some studies have already shown that the early development of the intestinal microbiota has a crucial impact on diseases and health later in life [1, 2]. Better understanding of the infant gut colonization is therefore needed.

Colonization in the infant gut occurs shortly after birth. Previous studies have shown that there is a wide variation in initial gut microbiota among individuals and that the composition of the microbiota differs from that of adults [1, 3]. It is already known that certain factors influence the colonization of the infant, these include: gestation time, mode of delivery, type of feeding, and antibiotic treatment. Various research has shown that these factors have an effect on the abundance of different species in the infant's intestine. For example, infants delivered vaginally possess a microbiota with a high proportion of Bifidobacteria compared to caesarean delivered infants. Over the first few years of life a mature gut microbiota composition is developed [4].

The gastrointestinal epithelium is covered with a mucus layer. Apart from being a physical barrier protecting the underlying epithelium from pathogens and shear induced stress the mucus also serves as a potential source of nutrition for microbiota [5]. Thus, there is an interplay between the mucus layer and the gut microbiota that also has a crucial role on the homeostasis. Studies in sterile animals have shown that there is an interdependence link between the mucus layer and gut microbiota, showing that the mucus layer is a carbon source for the microbiota, while gut microorganisms influence the composition of the mucus. This indicates that mucus potentially can have a significant effect on the infants microbiota [6].

The microbiota composition of the infant influences the health of its host. *Bifidobacterium* are associated with beneficial health effects such as protecting the host against pathogenic bacteria, contributing to the priming of the mucosal immune system and reduced probability to be overweight later in life [7, 8]. The symbiotic relationship between the mucus and the bacteria are important to further investigate for its effect on the abundances of species and consequently the host health.

The aim of this thesis is to construct a simple model of the infant's gut and thereby investigating the effect of mucus on the microbiota. We hypothesize that mucus increases the abundances of species that are able to feed on mucus related compounds, and that mucus thus plays a role in the diversity of the gut microbiota. We will use flux balance analysis (FBA) to investigate the effect of mucus *in silico*. FBA calculates the optimal flow of metabolites through a metabolic network, thereby making it possible to predict the growth rate of an organism or the rate of production of a metabolite [9]. To simulate the colon, we use a multiscale spatiotemporal dynamic model. The colon is represented by a regular square lattice with two long and two short sides, where lattice sites can contain metabolites and bacterial populations. A mucus layer is placed at the top and bottom long sides. The species are represented by their own genome-scale metabolic models (GEM) and these GEMS together with local metabolite concentration are used to obtain a FBA solution to calculate growth of the local populations. The purpose of this model is to provide insight in the colonization of the bacteria and their behaviour spatially and temporally. Furthermore, we are able to investigate the mucus related metabolic activity of the species.

## 2 Mucus background information

The human gut consists on macroscopic scale of three layers: mucosa, submucosa, and serosa. We will focus on the most inner layer, mucosa, since this layer is in direct contact with the microbiota. The mucosa consists of two layers, the inner layer is called the epithelial layer and the outer layer the lamina propria. The epithelial layer is coated in a layer of mucus and contains different cell types such as surface absorptive cells, goblet cells, enteric endocrine cells, stem/progenitor cells, and undifferentiated crypt cells. The lamina propria contains capillaris, enteric neurons, and immune cells, as well as a thin layer of smooth muscle (the lamina muscularis mucosae) [10]. Our model represents a simplification of the layer of mucus and therefore we will mainly discuss this part of the intestine wall. Furthermore, we can use the literature about mucus in general since the production of mucus is initiated before birth, and a complete mucus layer has already developed several days after birth [11]. There is not much known about possible differences between adult and infant mucus. However, some research suggest that the ratio of neutral to acidic mucins differs between infant and adult mucus [12]. Generally, relatively more acidic mucins are present between birth and the weaning period and the ratio then decreases after weaning. This is also seen in newborn piglets where the mucins were found highly sulphated and sialylated. The presence of acidic mucins in early life stages may be important for the newborns defense barrier since the immune system is not fully functional yet. Acidic mucins may protect passage of (pathogenic) bacteria from the external environment into the internal environment by being less degradable by bacterial glycosidases and host proteases [13].

Mucus is a biological hydrogel that coats the epithelial surface of the intestine and all other organs that are in contact with the external environment such as the eyes and lungs. Mucus serves as a lubricant to protect epithelia against shear induced damage from mechanical forces such as digestion and acts as a selective physicochemical barrier by excluding foreign or harmful molecules while permitting the passage of desirable agents such as nutrients. mucus consists of a loose outer layer and a dense epithelium-attached layer. The outer layer is circa 715  $\mu\text{m}$  and is inhabited by the microbiota [14]. It is composed primarily of water (95%), mucins, lipids, salts, and proteins involved in defense such as immunoglobulins [15]. Mucins are the most important components of mucus and are a solution of glycoproteins. Mucins are generated in and secreted from goblet cells. There exist 21 different types of mucin and these can be divided into two distinct groups: secreted mucins and tethered, cell surface-associated (also called transmembrane) mucins [16]. Only a small subset of the 21 different mucin types are expressed in the intestine: the gel-forming mucin MUC2 and several tethered, cell surface-associated mucins [17]. Usually, secreted mucins form a protective layer over epithelial lined surfaces, while the membrane anchored mucins function as cell surface receptors and respond to various external stimuli for different cell responses like cell growth, differentiation, cell proliferation and apoptosis [18]. Microbiota can live in or next to the gel forming layer and mucus is therefore a potential glycan source. For this thesis MUC2 is the most interesting type since it is the most abundant one in the gel forming layer of the colon and because of that assumable the most consumed type. Research in mice showed that mucins have more extended structures and higher fucosylation in the presence of microbiota compared to mice that were grown germ free [19]. This is because the levels of enzymes that play a role in the O-glycosylation are higher in normal mice than in germ free mice. These extended structures are also related to health. Fucosylated mucins are involved in cellular processes such as immune responses and sialyted mucins aids in the defense against pathogens [20,21]. Another important finding in the same study was that in particular the MUC2 O-glycans are probably very important for the composition of the microbiota. This idea is supported by an experiment where zebrafish and mice were cross-colonized with the other species' microbiota. Then, the

original zebrafish- and mice-characteristic microbiota was selected out from the donor bacteria. However, there are other mechanisms as well that influence the selection of microbiota [19].

The protein backbone of the mucin molecule is composed of a variable number of tandem repeats rich in proline, threonine, and/or serine (PTS domains), as well as cysteine-rich regions at the amino terminus, at the carboxy terminus, and interspersed between the PTS domains [22]. A frequent modification to the mucin molecule is mucin-type O-glycosylation at the PTS domain (Fig. 1). Glycosylation is the addition of sugar chains to proteins and with mucin-type O-linked is meant that the glycosylation takes place to serine or threonine. The first step is addition of a (N-acetyl-D-galactosamine) GalNAc sugar to the hydroxyl group of serine or threonine residues on the PTS domains of mucin. Then extension of the sugar chain occurs in a step-wise manner, yielding several higher order glycan structures. The serine or threonine with the glycan structure attached is called the mucin glycan and these can become complex polysaccharides. Apart from the GalNAc sugar the mucin glycan is built of L-fucose, D-galactose, N-acetyl-D-galactosamine, N-acetyl-D-glucosamine, and N-acetylneuraminic acid which is also known as sialic acid. In our model, it is difficult to add complete mucin molecules and therefore we add these mucin glycans. In the next chapter we will go into further detail how the modelling is performed.

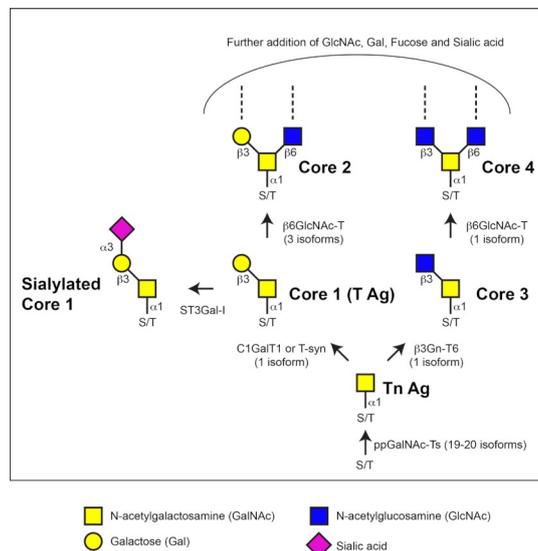


Figure 1: Schematic representation of the mucin-type O-Glycosylation [22].

As mentioned before mucins are a potential glycan source for the gut microbiota. Consumption of complex sugars requires a great variety of enzymes that perform the breakdown reactions. It is common that multiple species are necessary to fully degrade complex sugars, which is done through cross-feeding. Cross-feeding is the phenomenon when the production of a molecule such as a by-product from metabolism is used by the producing organism and/or other microbes in the environment. Not all species in the colon are able to completely digest mucin since it requires specific enzymes. As discussed previously, mucin glycans are complex and diverse structures and to degrade these compounds mucin degrading species rely on the cooperative action of a number of enzymes, consisting of proteases, sulfatases, and glycoside hydrolases (GH). Figure 2 illustrates schematically where different enzymes break linkages from the mucin glycan molecule. Every building block of the structure can be detached by the enzymes [6]. After

degradation, by-products are further degraded by the same or other species thereby enabling cross-feeding [11]. Figure 3 shows a schematic example of multiple cross-feeding actions, where every arrow represents an example of cross-feeding. Species 1 ( $S_1$ ) is able to consume metabolite 1 ( $M_1$ ) and produces secondary metabolites  $M_2$  and  $M_3$ , which in turn are consumed by  $S_2$  and  $S_3$ . These species produce other tertiary metabolites that benefit other (or the same) species. Competition among species can occur when species feed on the same metabolite.

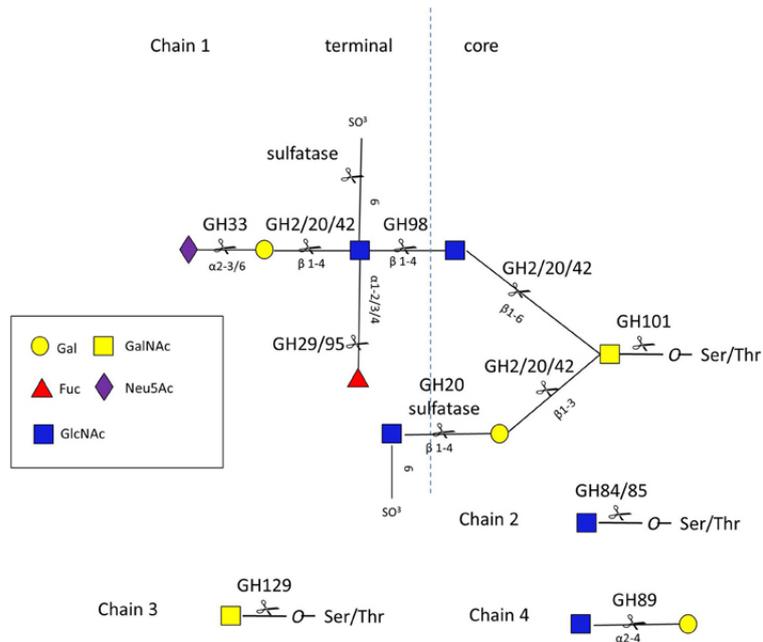


Figure 2: Schematic representation of degradation of mucin glycans. Chain 1 up to chain 4 represent different mucin glycan structures, and sites of action by the enzymes belonging to sulfatases and GH are indicated with the scissors [6].

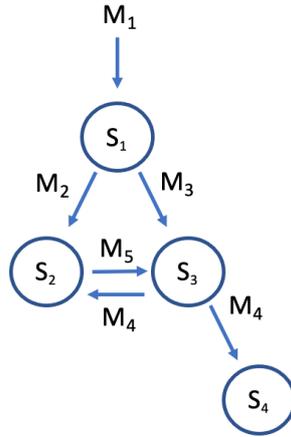


Figure 3: Schematic representation of multiple cross-feeding actions with species  $S_i$  with  $i \in \{1\dots4\}$  and metabolites  $M_i$  with  $i \in \{1\dots5\}$ .

### 3 Model

This section gives background information about previous models of metabolic analysis and describes the current model we use in this thesis. The current model is mainly built by David Versluis and we ourselves only made small extensions to the model [23].

#### 3.1 Previous models

Modelling of the gut microbiota has improved in recent years. At first, much attention was given to characterizing the composition of the microbiome and on comparative analyses. However, this neglects the important interactions between species. Therefore, new computational modelling of the human microbiome was needed to analyse the interaction between species [24]. Improved knowledge about metabolic networks of species, for example GEMs, and improved computational methods have led to enhanced modelling. A widely used constraint based modelling approach called flux balance analysis (FBA) is a frequently used method among metabolic models and calculates the flow of metabolites through a metabolic network. In short, all the reactions that can take place are converted to a matrix. Then, based on certain constraints and an objective function a optimal solution that gives the flux through all the reactions is found using convex optimization. More details about this method will be provided in the section “Flux Balance Analysis”. An example of individual based metabolic modelling is the BacArena method, where they also studied the dynamics of gut microbes interacting within the epithelial mucus layer [25]. They developed an R package where each organism is represented individually on a two-dimensional grid to model a spatial environment. Furthermore, temporal dynamics are modeled by including time steps in which the state of each individual and the environment is updated. FBA on the underlying genome-scale metabolic model of the particular species is used to model the metabolic activity of each individual. To study the effect of mucus, they added a mucus glycan gradient with higher concentration to the wall and lower concentration at the lumen and they found that mucus glycans are important for niche formations which shape the overall community structure. The mucus layer was mostly dominated by *Bacteroides thetaiotataomicron*. Experimental data

also shows a similar niche separation where mucus degrading microbiota are close to the gut epithelial layer and other microbiota in the lumen [26].

Our model is based on the multiscale spatiotemporal dynamic flux balance analysis model of the human gut introduced by Van Hoek and Merks [27]. Instead of looking at different bacterial species they constructed a hypothetical supra-organism. This organism was based on the GEM of *Lactobacillus plantarum* and was extended with four additional metabolic pathways. Furthermore, they allowed for evolution, with mutations occurring at random. As a consequence, pathways could become more or less efficient or could even disappear. They used flux balance analysis with molecular crowding to account for the space used by the enzymes catalyzing used reactions:  $\sum a_i f_i \leq V_{prot}$ . Where  $a_i$  represents the space occupied by the enzyme catalyzing  $R_i$ ,  $f_i$  is the flux through  $R_i$ , and  $V_{prot}$  represents the volume fraction of macromolecules devoted to metabolic enzymes. Our model is based on this model but major differences are that we model individual species, do not take evolution into account and we use different constraints. These changes were made to enhance the simulation of the infant gut. An important finding of the study is that a stratified structure of metabolic niches formed in the tube with primary source consumers in front, followed by strata inhabited by secondary and tertiary consumers. To quantify the amount of cross-feeding, they defined a cross-feeding factor,  $C(i)$ , with  $i$  identifying the species. They found that the cross-feeding factor in the front part is close to 0, meaning almost no cross-feeding, and increases to the end part of the lattice, where it is almost 1, meaning complete cross-feeding.

### 3.2 Model outline

The infant colon is represented by a regular square lattice consisting of 225 by 8 lattice sites, where each lattice site represents 4 mm<sup>2</sup> of space. This leads to a simulation of the infant colon of 450 by 16mm. Each lattice site can contain metabolites and a single bacterial population. Every time step the populations use metabolites to produce ATP and also excrete products into the grid. Figure 4 gives a schematic overview of the different compartments in a lattice site. Metabolites are present in the medium and with an exchange reaction, if contained in the GEM of the species, they can be transported to the extracellular compartment. Furthermore, the metabolite can move to the intracellular compartment through a transport reaction provided that the species can perform that reaction. The intracellular compartment represents the intracellular compartment of all the bacteria that form the population. Populations can move to random adjacent lattice sites and the metabolites diffuse to all adjacent sites and are subjected to advection distally. This choice has been made since in the actual gut the flow of the contents (and thus metabolites) is also distally. We implemented diffusion with the Euler method, using a 9-point stencil. The populations grow by producing ATP and when they are large enough they divide into empty, adjacent lattice sites. During initialization, populations are added randomly to the grid and during the simulation populations are also randomly removed. The latter condition ensures that populations need to reproduce themselves in order to survive. Stochasticity is introduced in the model due to the random placement, movement, and removal of the populations. The first 21 days of the development of the infant colon is modelled, where one time step simulates 3 minutes. As initial nutrition source we use lactose and mucin glycans. Lactose is added every 3 hours proximally and then moves distally like all other metabolites. Mucin glycans are added along the two walls and every time step a certain amount is added if the glycans are consumed. Figure 5 shows a schematic representation of this process.

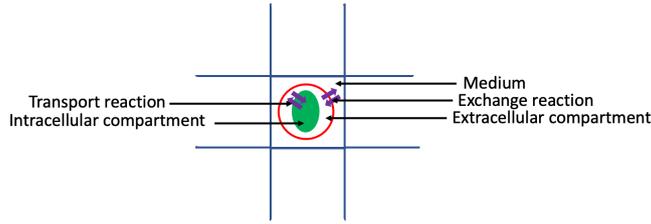


Figure 4: Schematic representation of a lattice site.

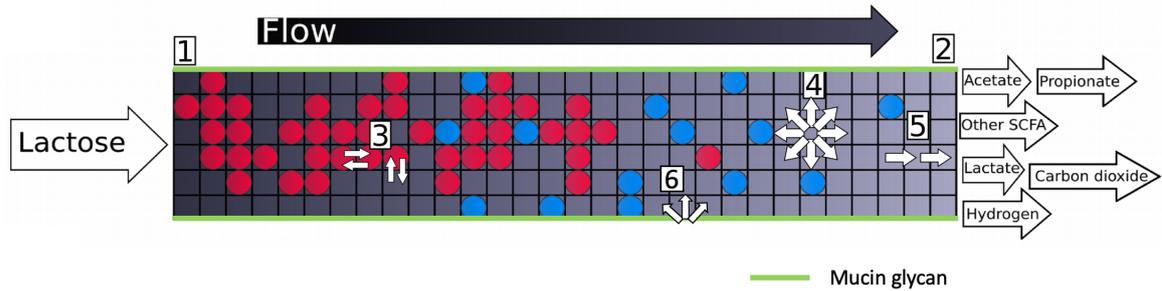


Figure 5: Schematic representation of modelling bacterial metabolism in the colon. The red and blue dots represent two different bacterial species and diffuse through the swapping of the content of lattice sites (3). Lactose is added proximally (1) and then moves together with other formed metabolites distally (5). Mucin glycans, in green, are continuously added along the two walls and diffuse a little bit to adjacent sites (6). The metabolites diffuse between adjacent lattice sites (4) and leave the system distally representing the feces (2) (adapted from [23]).

### 3.2.1 Species composition

The used species were selected based on a study performed on 98 Swedish infants [28]. The 19 most prevalent species on genus level were chosen. Then, the most prevalent species are determined by another data set of the study or when not available determined by hand based on the species in the Virtual Metabolic Human (VMH) database [29]. Table 1 shows which species we use in our model and also whether these possess enzymes that can degrade mucin glycans.

The metabolic network of these species is based on a genome-scale metabolic model (GEM), which were created in the AGORA project. GEMs are genome-scale metabolic network reconstructions that contain all the metabolic reactions in an organism and the genes that encode each enzyme. If a species contained reactions that could degrade any mucin glycan then we indicated mucin glycan degradation in table 1 with yes.

### 3.2.2 Mucin glycans

Modelling complete mucin molecules is complex, since these do not occur in the genome-scale metabolic models we use for the species [29]. The mucin glycans on the other hand do appear in the models. To add a complete mucin molecule, we would have to built it ourselves. The structure however is complex and also not completely known. For example, around 200 different mucin glycans exist each having their own structure. It is unknown which mucin glycans are

Species	Mucin glycan degradation
<i>Bacteroides vulgatus</i>	yes
<i>Bifidobacterium longum</i>	yes
<i>Escherichia coli</i>	yes
<i>Lactobacillus gasseri</i>	yes
<i>Parabacteroides distasonis</i>	yes
<i>Streptococcus oralis</i>	no
<i>Staphylococcus epidermidis</i>	no
<i>Clostridium butyricum</i>	no
<i>Gemella morbillorum</i>	no
<i>Rothia mucilaginosa</i>	no
<i>Collinsella aerofaciens</i>	no
<i>Enterococcus faecalis</i>	no
<i>Ruminococcus gnavus</i>	no
<i>Propionibacterium acnes</i>	no
<i>Haemophilus parainfluenzae</i>	no
<i>Eubacterium hallii</i>	no
<i>Veillonella dispar</i>	no
<i>Eggerthella sp.</i>	no

Table 1: The used species in the model and whether they can or cannot degrade mucin glycans.

exactly present on different mucin types. The inconvenience of modelling mucin glycans instead of mucin molecules is that we cannot add MUC2 (or make a distinction between different mucin types at all) to the system and thereby investigate the effects of this molecule. Instead we add very tiny pieces of the molecule to the system. We justify this by two reasons: the other parts of the mucin molecule are not consumable by the microbiota and these mucin glycans make up a big part of the mucin molecule (around 80%).

The next challenge is to decide which mucin glycan(s) should be added to the system in order to represent MUC2 to some degree. A study performed on 25 humans showed that most of the mucin glycans were based on the core 3 structure and also contained sialic acid [30]. Initially, we will look at mucin glycan number 81 (MG81) (Fig. 6). This glycan is chosen since it contains a core 3 structure and is sialylated. Additionally, sialylated mucins are associated with advantageous effects. Mother’s milk is a source of sialylated oligosaccharides and glycans and these are linked to beneficial effects for the immune system and development of the infant. Sialic acid-dependent pathogens recognise and can bind to sialylated molecules, this then prevents the adhesion of the pathogen to the epithelial cells of infants. Furthermore, sialylated oligosaccharides modulate the immunological system of infants at a cellular level, and are essential for brain development and probably for infant growth [21]. In section 4.10 we will consider a structure that is fucosylated which also has beneficial effects. Fucosylated host glycans play an important role in intestinal microbiota formation and are also a nutrition source [20].

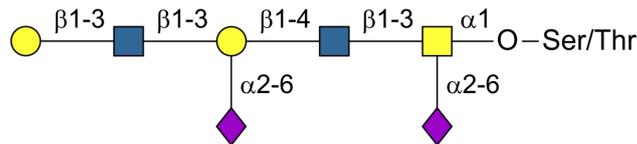


Figure 6: Schematic representation of MG81. This glycan consist of a core 3 structure and is furthermore highly sialylated (purple rhombus).

The mucus is *in vivo* layered along the whole wall in a 3D setting. Since we are working in a 2D grid, we have chosen to align the mucin glycans along both long walls. We have chosen that mucin glycans diffuse slightly to simulate the effect of shedding in the colon. This diffusion rate is twenty times smaller compared to the diffusion rate of the other metabolites in the system. Furthermore, there is no advection of the mucin glycans when adhered to the both long walls. The maximum generation rate per time step (MGR) is a maximum value for each lattice site along both walls to generate mucin glycans. Every timestep the algorithm checks if the mucin glycan concentration in those lattice sites is below a chosen value (CV) and if so it generates mucin glycans (G) with the difference between a chosen value and the actual value (AV), unless this difference is bigger than the maximum generation rate. In this case, the generation is equal to the maximum generation rate. Thus, we generate mucin glycans per time step with the following formula:  $G = \min(MGR, (CV - AV))$ . The generation rate is difficult to determine and therefore we performed a sensitivity analysis. To estimate appropriate rates, we used the effect on the abundances of the species.

### 3.2.3 Metabolites

Each GEM contains all the reactions a particular species is able to carry out. Based on this information, we have made a pathway network that shows all the reactions that can take place in the simulation in the degradation from MG81 (Fig. 7). Each arrow represents a reaction and only the formed mucin glycan type is shown (and no by-products). For example, if we start with MG81, then the upper reaction leads to mucin glycan number 152 (MG152). The species *B. vulgatus* and *P. distasonis* are able to perform all reactions depicted apart from the reaction indicated with \*, which cannot be performed by *P. distasonis*. Furthermore, we indicated the reactions that can be performed by other species with different symbols: a triangle for *B. longum*, a square for *E. coli*, and a circle for *R. gnavus*. Moreover, every metabolite is also associated with a so called released form where the serine-threonine group is removed. For simplicity reasons, these reactions are not indicated here. Not all metabolites are released in the medium (white nodes), meaning that these metabolites do not appear in the simulation. This can have two reasons, first the reaction simply does not take place, and secondly the formed metabolite is immediately used in another reaction in the same FBA solution. When we look at for example the reactions that take place to generate MG2 and MG12, then we know that the second reason is for sure applicable to at least one of these two mucin glycans, because otherwise metabolite Tn antigen and core 3 could never be released in the medium.

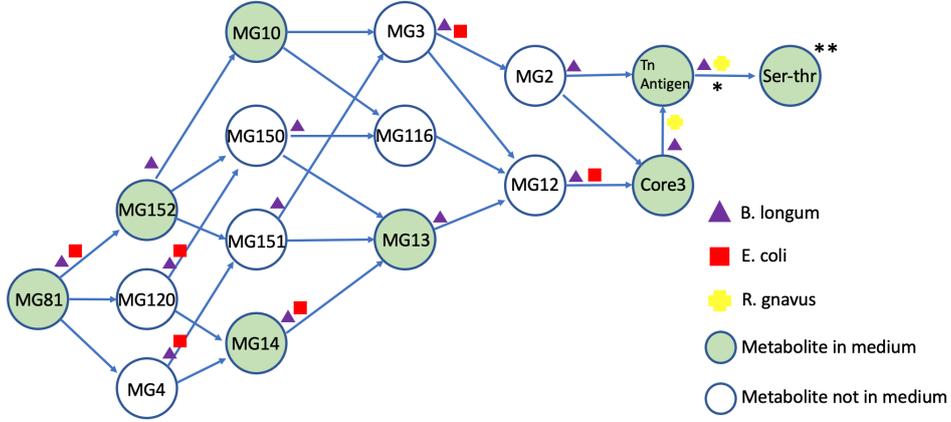


Figure 7: Schematic representation of the possible metabolic pathways for MG81 based on the GEMs. The arrows represent reactions and only the formed mucin glycan type is shown. The nodes in green represent metabolites that are released in the medium, and the nodes in white represent metabolites that are not released in the medium. The species *B. vulgatus* and *P. distasonis* are able to perform all reactions apart from the reaction indicated with \*, which cannot be performed by *P. distasonis*. The reactions that can be performed by the other species are indicated with a triangle for *B. longum*, a square for *E. coli*, and a circle for *R. gnavus*. \*\* The metabolite Ser-thr can also be formed at any other moment in the pathway.

### 3.2.4 Flux Balance Analysis

Flux balance analysis (FBA) is a mathematical approach for analyzing the flow of metabolites through a metabolic network. This is a constraint-based approach and is a key method for *in silico* study of microbial metabolism and in particular in modelling the gut microbial metabolism. In our model, each time step FBA is performed to calculate the metabolic activity. The first step is the conversion of each GEM to a stoichiometric matrix  $S$ . Reversible reactions are split into two separate irreversible reactions. Furthermore, the reactions classified as exchange, sink, or demand in the metabolic reconstruction are denoted as exchange reaction in the matrix. FBA assumes that all reactions are in steady state:

$$\frac{d\vec{x}}{dt} = S \cdot \vec{f} = 0. \quad (1)$$

Where  $\vec{x}$  is a vector of all metabolites and  $\vec{f}$  is a vector describing the metabolic flux through each reaction in the network. Additionally, the fluxes are constrained by an upper bound set dynamically for exchange reactions by the environment:

$$\vec{f} \leq \vec{f}_{ub}. \quad (2)$$

This constraint represents the availability of nutrients in the local environment, which is evident since *in vivo* bacterial species also do not have unlimited access to nutrients. The internal

reactions are not constrained by an upper limit. The uptake of metabolites in the medium by populations is constrained by the upper uptake flux bound  $f_{ub}$  given by:

$$\vec{f}_{ub} = \vec{c}(\vec{x}, t). \quad (3)$$

Lastly, the total flux is constrained by the so-called flux limit:

$$\sum f_i \leq B \cdot a. \quad (4)$$

Where  $B$  is the size of the local bacterial population, and  $a$  a scaling factor. The flux limit constrains the total flux through the system, which means that there is a limit on the total flux through all the used reactions each time step. Next, FBA defines an allowable solution space, and from this space an optimal solution to the objective function is calculated. Our objective function is the maximization of ATP production. We have chosen this function and not for example biomass production, since it is quite difficult to define an appropriate function for the biomass production. FBA is a linear optimization problem and therefore the GNU Linear Programming Kit (GLPK) is used to perform the FBA.

We use a special case of FBA, namely dynamic flux balance analysis (dFBA), since we are also interested in the dynamics of the external metabolites. With dFBA we assume that FBA is in quasi-steady state meaning that the internal metabolites of the species are much faster in steady state compared to the external metabolites. The algorithm for dFBA consists of two steps for every time step  $\Delta t$ . First, based on current extracellular concentrations FBA is applied, calculating the exchange rates with the environment,  $f(E)$ . Next, the extracellular concentrations are updated based on FBA fluxes:

$$c(t + \Delta t) = c(t) + \Delta t f(E). \quad (5)$$

Lastly, we extend our model spatially, since the species live on a lattice. This means that multiple dFBA models dynamically exchange nutrients between lattice sites [9, 27].

The change in local metabolite concentration is given by:

$$\frac{d\vec{c}(\vec{x}, t)}{dt} = \vec{F}(\vec{x}, t)B(\vec{x}, t) + \frac{D}{L^2} \sum_{\vec{i} \in NB(\vec{x})} (\vec{c}(\vec{i}, t) - \vec{c}(\vec{x}, t)), \quad (6)$$

and depends on the flux of the metabolites based on bacterial metabolism  $\vec{F}(\vec{x}, t)$  modified by the local density of the populations  $B(\vec{x}, t)$  given by:

$$\frac{dB(\vec{x}, t)}{dt} = \mu(\vec{x}, t)B(\vec{x}, t). \quad (7)$$

Where  $\mu(\vec{x})$  is the growth rate. The second term represents the diffusion to immediate neighbor sites and is approximated by Fick's law, where  $D$  is a diffusion coefficient and  $L$  the interface length between two adjacent sites.

### 3.2.5 Output of the model

Several data files are obtained after the simulation. A bacterial abundance overview is generated which records the size of all the populations  $B(\vec{x}, t)$  each time step. The metabolites that leave the system distally are put in a file, and this represents the feces. The location and quantity of the metabolites,  $c(\vec{x}, t)$ , in the system are also recorded. The location and exchange fluxes,  $F(\vec{x}, t)$ , along with the population sizes are recorded in an exchange file to track the exchanges of metabolites by the different species. The specific reactions that take place in the species based on the FBA solutions are recorded in a separate file. Lastly, screenshots of the populations and metabolites in the grid are made every 20 time steps which together form a movie that visualizes the movement of populations and metabolites.

## 4 Results

### 4.1 Model predicts *B. vulgatus* dominance with increasing MG81 generation

We first investigated the effect of mucin glycans on the abundances of the used species by varying the maximum generation rate per time step of MG81. We also ran a control experiment with no MG81 added. We ran 23 simulations for each experiment. Prior research has shown divergent results regarding dominance of specific species. While one study shows that microbiotas of vaginally delivered newborns contain more *B. longum* than *B. vulgatus* [4], another study found that the abundance of *B. vulgatus* is higher than *B. longum* [28]. We see that addition of mucin glycans increases the abundance of *B. vulgatus* in particular. Therefore we tried different values for the generation rate and then looked at the abundances of the bacterial species. Rates where other species, in particular *B. longum*, still could survive were considered as appropriate. We finally arrived at the following values for the maximum generation rate that we will use in the experiments: 0.0001, 0.0003 and 0.0007  $\mu\text{mol}$  MG81.

*B. longum* is the dominant species followed by *E. coli* without addition of MG81 (Fig. 8A). Furthermore, *B. vulgatus* stays consistent low and the other species first experience an initial growth but then decline after day 3. When the maximum generation rate per time step is 0.0001  $\mu\text{mol}$  MG81, *B. vulgatus* has at 21 days a significantly higher population compared to the control experiment ( $p < 0.05$ , two sample t-test) and especially *E. coli* decreases compared to the control experiment (Fig. 8B). *B. vulgatus* abundance increases further every time the maximum generation rate per time step for MG81 is increased (Fig. 9). *B. longum* abundance decreases at first slowly when increasing the maximum generation rate for MG81 (Fig. 8BC), but with the highest maximum generation rate per time step the abundance of *B. longum* decreases more (Fig. 8D). The abundance of *E. coli* also keeps decreasing with increasing the maximum generation rate per time step of MG81 (Fig. 8CD). These results suggests that especially *B. vulgatus* benefits from mucin glycans as an additional nutrition source. The cause of this gain by *B. vulgatus* will be investigated in section 4.4. In all four different experiments the other sixteen simulated species experience first an initial growth, but then decline after day 3, just as in the condition without MG81. These sixteen species did not have significantly different abundances at day 21 compared to the abundance at day 1 ( $p > 0.05$ , two sample t-test). Among these species is also *P. distasonis*. However, as illustrated in figure 7B *P. distasonis* contains all but one reactions to degrade MG81. We would expect that this gives *P. distasonis* an advantage compared to other species that cannot degrade MG81 or MG81 related metabolites since *P. distasonis* has access to more nutrient sources. We hypothesized that sufficient degradation of lactose is needed for significant growth and that *P. distasonis* is not able to do so.

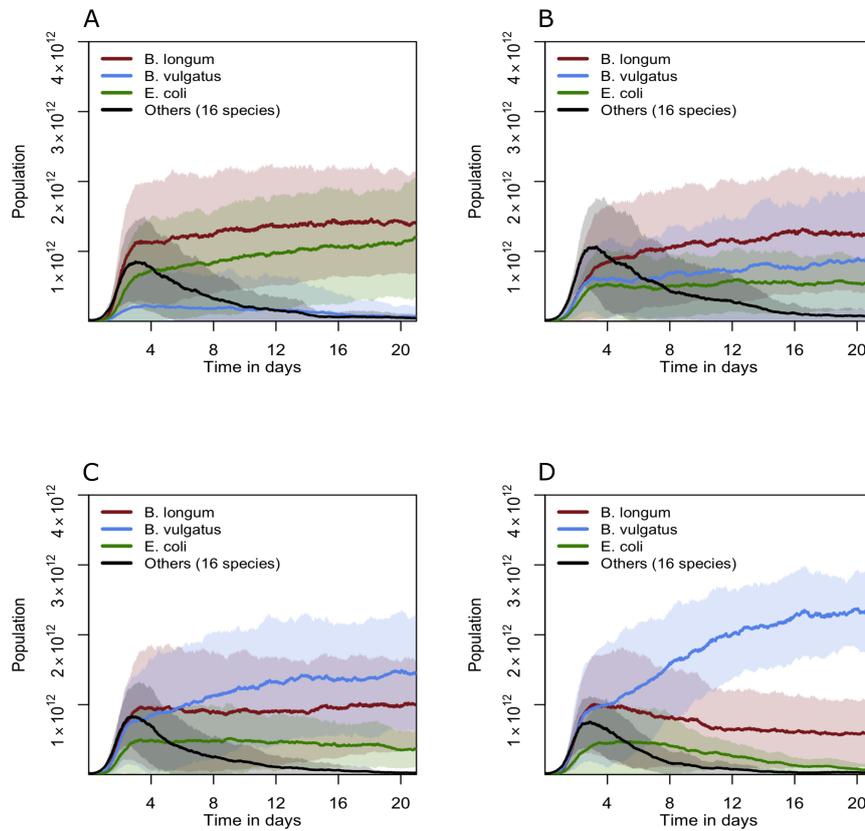


Figure 8: Abundances for grouped species over time (A) without addition of MG81 (control), (B) with maximum addition of 0.0001  $\mu\text{mol}$  MG81 per time step, (C) with maximum addition of 0.0003  $\mu\text{mol}$  MG81 per time step, (D) with maximum addition of 0.0007  $\mu\text{mol}$  MG81 per time step. The bold lines indicate the means of the abundance of species over the different runs ( $n=23$ ), and the shaded area indicates one standard deviation.

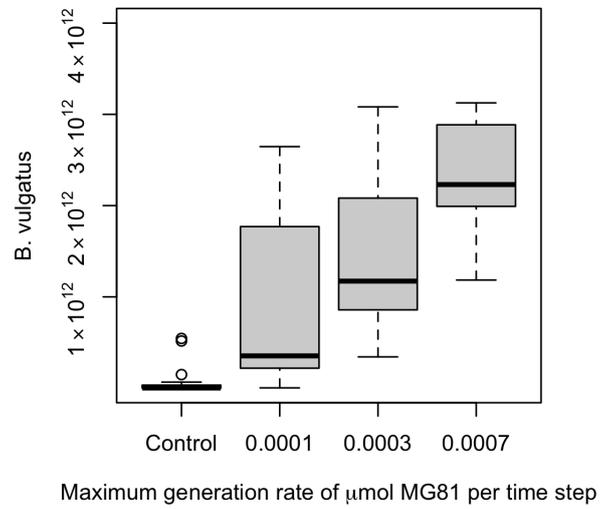


Figure 9: Abundance of *B. vulgatus* at day 21 without addition of MG81 (control), and with maximum addition of 0.0001, 0.0003 and 0.0007  $\mu\text{mol}$  MG81 per time step. (n=23)

#### 4.2 *P. distasonis* grows less on lactose compared to other species

To test the hypothesis that *P. distasonis* grows less on lactose compared to the other species we ran an experiment with separate simulations of a single time step. By doing this the model solutions cannot influence each other. Figure 10 shows the growth per timestep for different lactose concentrations for the species *B. longum*, *B. vulgatus*, *E. coli* and *P. distasonis*. *P. distasonis* grows less compared to the other species, which confirms our hypothesis. We will not visualize *P. distasonis* further since this species has a minimal contribution in the experiment and does not have a significant higher abundance at the end of the simulation compared to the initial abundance.

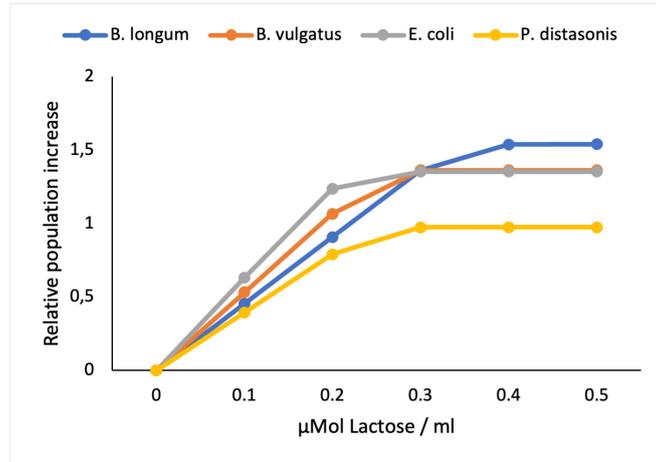


Figure 10: Growth per time step by lactose concentration for the species *B. longum*, *B. vulgatus*, *E. coli* and *P. distasonis*.

### 4.3 Co-existence is possible between the two most abundant species *B. vulgatus* and *B. longum*

The previous results showed that *B. vulgatus* and *B. longum* are the most abundant species in the presence of MG81 (Fig. 8BCD). However, there is a lot of variation in the bacterial abundances between runs within the same experiment and therefore we are going to investigate the co-occurrence of the two most abundant species *B. vulgatus* and *B. longum*. Figure 11 displays a scatterplot of the abundances of these two species at day 21 for the control and for different maximum generation rates of MG81. *B. longum* is without addition of MG81 highly abundant in 19 out of 23 runs, and the abundance of *B. vulgatus* is not significantly higher compared to day 1 in 21 runs and much lower compared to *B. longum* in 2 runs (Fig. 11A). In the case of a maximum generation rate per time step of  $0.0001 \mu\text{mol}$  MG81 (Fig 11B), *B. vulgatus* is exclusively abundant when *B. longum* has a very low abundance. When *B. longum* is the most abundant species it is less abundant than when *B. vulgatus* is the most abundant. If the maximum generation rate per time step is increased to  $0.0003 \mu\text{mol}$  MG81 (Fig 11C), then we see that in 19 runs the two species both survive. Finally, if the generation rate is increased to  $0.0007 \mu\text{mol}$  MG81 (Fig 11D), then we see less co-existence since *B. longum* dies out in more runs. Furthermore, *B. longum* is less abundant in 14 runs where there is co-existence compared to the other conditions. When considering all four plots there is a maximum amount of around  $3 \cdot 10^{12}$  bacteria that is able to live in the lattice. We hypothesize that this is a consequence of limited nutrition sources. Broadly three scenarios can appear: *B. longum* dies out and *B. vulgatus* is very high, *B. vulgatus* dies out and *B. longum* is high, or both species survive. From this, we hypothesize that *B. vulgatus* is the main consumer of mucin glycans and thereby is able to grow significantly in the presence of mucin glycans. To investigate this, we will consider the uptake of the primary nutrition sources lactose and MG81 and the usage and production of the other metabolites. Furthermore, we will investigate why in a variety of simulations there is only a single survivor and no co-existence between *B. vulgatus* or *B. longum*. We hypothesize that in those cases one of the two species is able to take over the nutrition sources, and consequently grows rapidly, which then increases again the take over of nutrition sources.

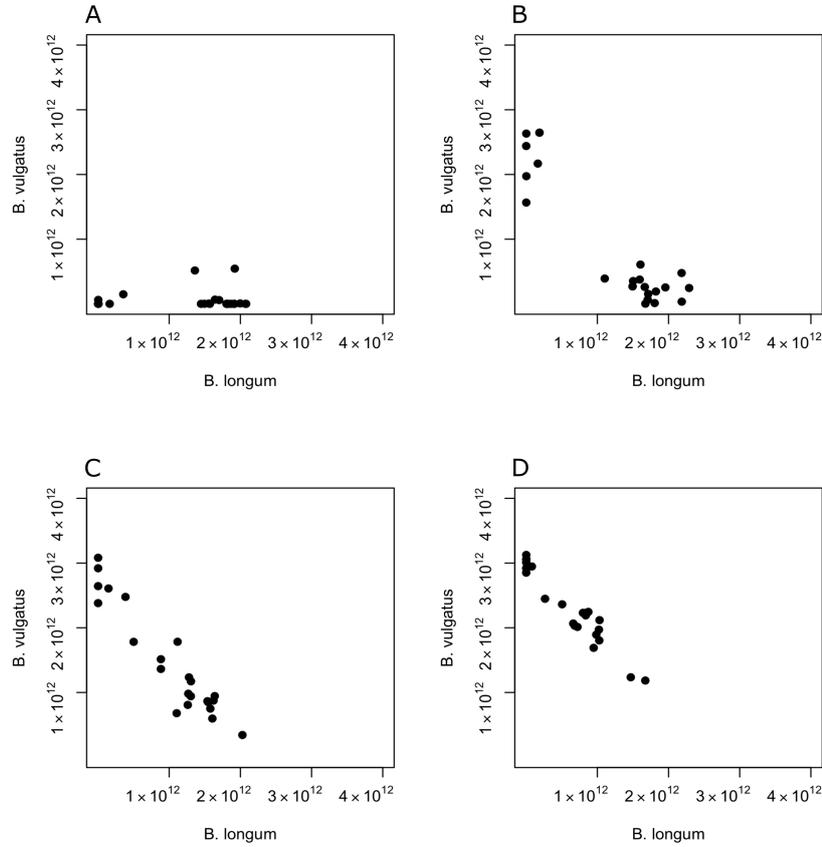


Figure 11: Abundances of *B. longum* versus *B. vulgatus* at 21 days (A) without addition of MG81 (control), (B) with a maximum generation rate per time step of 0.0001  $\mu\text{mol}$  MG81, (C) with a maximum generation rate per time step of 0.0003  $\mu\text{mol}$  MG81, (D) with a maximum generation rate per time step of 0.0007  $\mu\text{mol}$  MG81. (n=23)

#### 4.4 *B. vulgatus* consumes more lactose when there is more addition of MG81

Figure 12 shows the mean uptake of lactose of the three most abundant species *B. vulgatus*, *B. longum* and *E. coli* per time step in the last four days. Again, we looked at no addition of mucin glycans (A), and with the same different increasing maximum addition rates per time step of MG81 as previously (B up to D). *B. longum* takes up most lactose when only lactose is added as nutrition source, followed by *E. coli* (Fig 12A). In three runs, we see that *B. longum* has either died out or is present in very low amounts and that in those cases *E. coli* uses most of the lactose. When a maximum of 0.0001  $\mu\text{mol}$  MG81 per time step is added then *B. vulgatus* takes up more lactose and *B. longum* and *E. coli* less compared to the control (Fig. 12B). Furthermore, the variation between the runs is higher. This can be explained from the co-occurrence of *B. vulgatus* and *B. longum* (Fig 11B). The amount of lactose used by *B. vulgatus* increases fast, while the lactose used by *B. longum* and *E. coli* decreases (Fig. 12CD). When we consider the mean MG81

uptake (Fig. 13), then we see that with a maximum generation of  $0.0001 \mu\text{mol}$  MG81 per time step, *B. longum* takes up more MG81 than *B. vulgatus*. When the maximum addition of mucin glycans is further increased, the relative amount of MG81 used by *B. vulgatus* increases greatly and used by *B. longum* and *E. coli* decreases (Fig. 13BC). We can conclude that *B. vulgatus* on average takes up more lactose in the presence of MG81 due to a higher abundance. Due to the addition of MG81 *B. vulgatus* is able to increase its growth a bit more compared to the other species and thereby gains a competitive advantage in the lactose uptake.

Next, the metabolites related to MG81 are considered to investigate if potential cross feeding is at work. *B. longum* and *E. coli* are only able to partially degrade MG81, while *B. vulgatus* is able to degrade the complete MG81 molecule (Fig. 7). Furthermore, we saw previously that especially *B. vulgatus* benefits from the MG81 addition and therefore this investigation can provide more insight in its cause. We use a maximum generation rate of  $0.0003 \mu\text{mol}$  MG81 per time step to further investigate since in most of the runs the two most abundant species *B. vulgatus* and *B. longum* are able to live together, making it the most interesting situation. Producers secrete the metabolite into the medium and users take up the metabolite from the medium. Thus, what happens in the extra or intracellular compartment within a population cannot be seen in this figure (Fig. 4). The same populations can be users and producers, even in the same time step. Figure 14 shows the uptake and secretion of the metabolites that had the highest exchange between the three species. A complete exchange overview of all the metabolites related to MG81, including the species *P. distasonis* can be found in supplementary 6.1. *E. coli* is the biggest producer for MG152. This is evident since *E. coli* cannot degrade MG152 further, while the other species can. Therefore, the amount of MG81 *E. coli* has taken up (Fig. 13C) is the same amount *E. coli* releases as MG152. *B. vulgatus* followed by *B. longum* takes up most of the released MG152. Next, *B. longum* produces the most MG10 and *B. vulgatus* uses the most. Tn antigen, the final metabolite, is primarily produced by *B. vulgatus*. These results together with figure 13C and figure 30 from the supplementary indicates that *B. vulgatus* metabolizes the complete MG81 until its end products and additionally takes up a big part of metabolites produced by the other species in the spatial model. This efficient usage could explain the rapid growth of *B. vulgatus* in the presence of mucin glycans. Later in this thesis we will investigate the spatial pattern of the species.

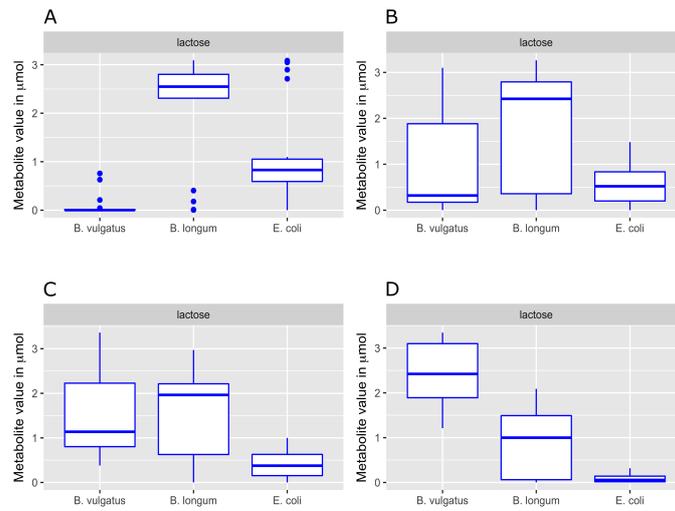


Figure 12: Mean uptake of lactose in the last 4 days per time step (A) without MG81 (control), (B) with a maximum generation rate of  $0.0001 \mu\text{mol}$  MG81 per time step, (C) with a maximum generation rate of  $0.0003 \mu\text{mol}$  MG81 per time step, (D) with a maximum generation rate of  $0.0007 \mu\text{mol}$  MG81 per time step. (n=23)

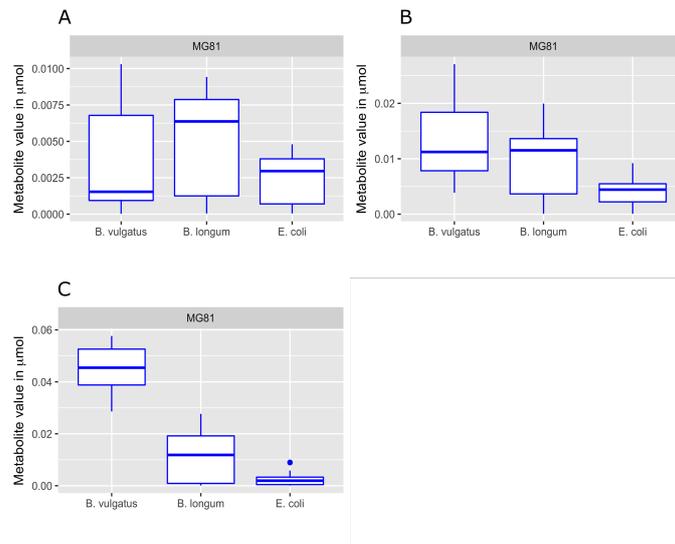


Figure 13: Mean uptake of MG81 in the last 4 days per time step (A) with a maximum generation rate of  $0.0001 \mu\text{mol}$  MG81 per time step, (B) with a maximum generation rate of  $0.0003 \mu\text{mol}$  MG81 per time step, (C) with a maximum generation rate of  $0.0007 \mu\text{mol}$  MG81 per time step. (n=23)

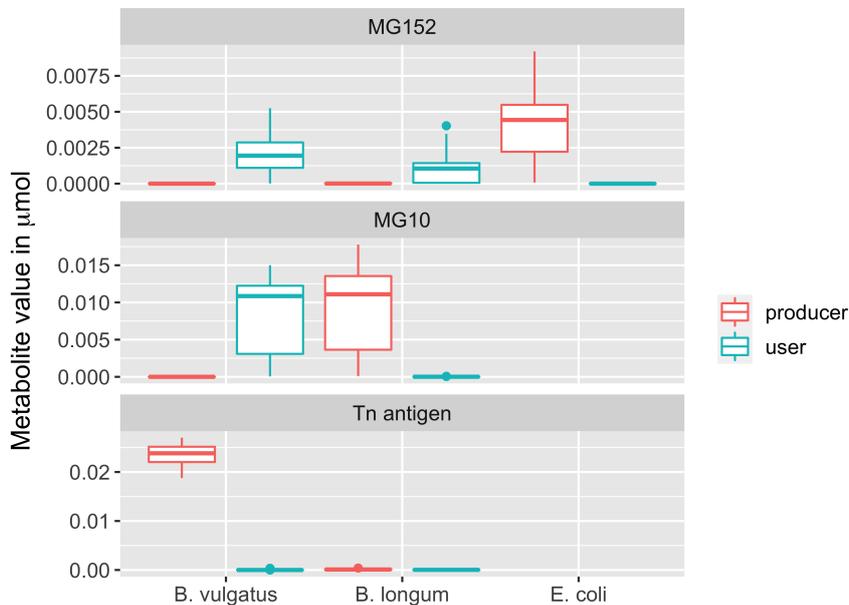


Figure 14: Mean uptake and secretion of metabolites MG152, MG10 and Tn antigen per time step in the last 4 days with a maximum generation rate of  $0.0003 \mu\text{mol MG81}$  per time step ( $n=23$ ).

#### 4.5 All possible intermediate reactions in the degradation of MG81 are performed

Recall figure 7B which represented the possible metabolic pathways for MG81 and which metabolites were released and unreleased in the medium. The question then arises: are the unreleased metabolites produced by the species or not. Therefore we investigated the reactions used in the FBA solutions. We found that all the reactions schematically visualized in Fig. 7B are present in the FBA solutions. This means that the unreleased metabolites are used immediately in the same FBA solution.

#### 4.6 The output metabolites consist mainly of acetate, hydrogen and tn antigen

Next, we will study the metabolites that leave the system distally, representing the feces of the infant (Fig. 15). The metabolites that are related to the primary source lactose are displayed in figure 15A. Acetate is the most prevalent metabolite followed by hydrogen, succinate and lactate. Acetate is mainly produced by *B. longum*, but also by *B. vulgatus* (Fig. 16). The two most abundant species producing acetate explains the high concentration. Succinate is mainly produced by *B. vulgatus* and in smaller amounts by *E. coli*. Lactate can only be produced by *B. longum* and *E. coli* and *B. vulgatus* can cross feed on it (Fig. 16). From the metabolites that are related to MG81 is tn antigen the most abundant that leaves the system (Fig. 15B). This

confirms the hypothesis that the breakdown of MG81 continues to its end products. Furthermore, we see some MG81 in the feces. Since we allow for a bit of diffusion of MG81, it is possible that in parts where there are few consumers of this mucin glycan, MG81 is able to reach the end of the system. MG152 is present in small amounts in the feces. Based on Fig. 14 this will mainly be the MG152 released by *E. coli*. The previous figure showed that *B. vulgatus* favors taking up these metabolites, but it could be the secreted MG152 from *E. coli* in the distal part of the lattice.

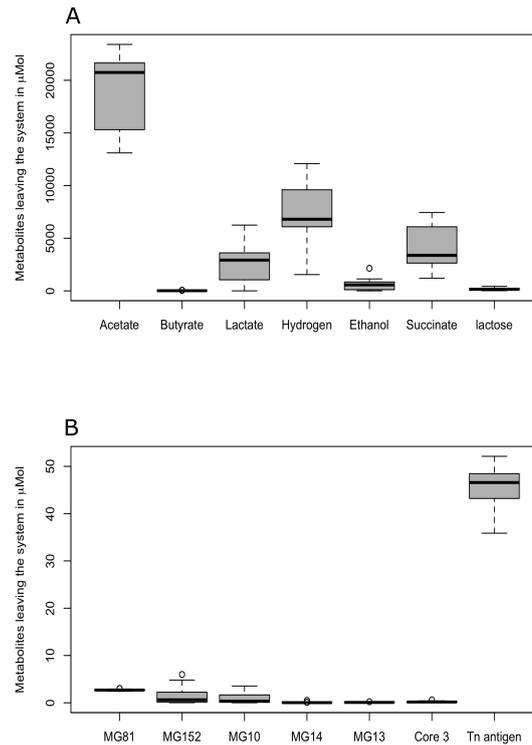


Figure 15: Sum of the metabolites leaving distally the system related to lactose (A) and related to MG81 (B) from the last 4 days with a maximum addition of  $0.0003 \mu\text{mol}$  MG81 per time step. (n=23)

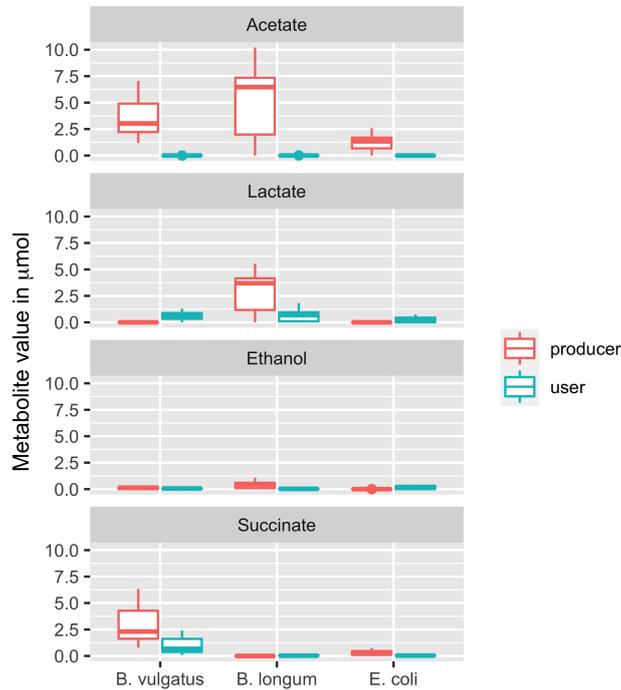


Figure 16: Mean uptake and secretion of metabolites acetate, lactate, ethanol and succinate every time step in the last 4 days with a maximum addition of  $0.0003 \mu\text{mol}$  MG81 per time step ( $n=23$ ).

#### 4.7 Spatial patterning of the species shows that lactose uptake is related to the volume of the species

In this section, we describe the spatial distribution of the species in the lattice. As before, we used MG81 with a maximum generation rate of  $0.0003 \mu\text{mol}$  MG81 per time step. We will first investigate the spatial distribution by studying the movies that visualizes the movement of populations and metabolites. Supplementary 6.2 describes for each run the development of the species *B. longum*, *B. vulgatus* and *E. coli* in detail. Additionally, supplementary 6.2 also contains two videos where *B. vulgatus* or *B. longum* is dominant.

Over all the runs we see that mucin glycan degradation starts at the end of the lattice and then moves to the front. The degradation is always between the pulses of lactose, thus degradation appears to take place when few lactose is present. This suggest that lactose is the favored nutrition source, but when not available mucin glycans are consumed. This is confirmed by plotting the lactose uptake versus the MG81 uptake filtering on measurements where there was active growth of this species ( $\text{growth} > 0.1$ ) (Fig. 18A). Furthermore, only when *B. longum* is present (in sufficient amounts), lactate appears (Fig. 17A), which is evident since *B. longum* is the only abundant species that can produce lactate. This pattern is confirmed by plotting the lactose uptake versus the lactate production filtering on the species *B. longum*. A clear

relationship is seen between lactose uptake and lactate production: more lactose uptake leads to more production of lactate (Fig. 18B). Additionally, in the presence of *B. longum* we see more ethanol. In the front of the lattice ethanol appears before and after the lactose pulse and later in the system we see a more even distribution of ethanol. This pattern is confirmed by plotting the lactose uptake versus the ethanol production filtering on the species *B. longum* and *B. vulgatus*. Ethanol production is mainly zero when lactose uptake is high and higher amounts of ethanol are produced when lactose uptake is low (Fig. 18C). During most of the runs butyrate is presented at high levels over a large part of the lattice, but afterwards butyrate disappears and sometimes comes up at the end of the lattice. Moreover, in the presence of *B. vulgatus* hydrogen is formed at the time of the lactose pulses and PPA is formed just before and after the lactose pulse which winds out to the end of lattice (where lactose becomes less present since it is being consumed) (Fig. 17B). This result suggest that PPA is formed when the concentration of lactose is low. This is again confirmed using the same method as for MG81, ethanol and lactate, but now we filtered only on the species *B. vulgatus*. The same relationship between PPA and lactose is seen as with ethanol and lactose (Fig. 18D). This is what we would expect since the same behaviour was seen in figure 17.

How the species interact with each other differs greatly among the different simulations: they emerge all three, just two emerge or even only one species emerge (Suppl. 6.2). The place also varies from the end of the lattice to more to the front of the lattice. Furthermore, we see variation in whether mixing of the species take place. Most of the times species emerge with their own kind and after some time mixing between species take place. We also see a movement of the species to the front, especially when they emerge in the end of the lattice. In particular, the species *B. longum* can be found most of the time at the beginning of the lattice, while *E. coli* and in lesser amount *B. vulgatus* also live in the end part of the lattice. Next, we analyse the spatial behaviour using the data that also generated the movie.

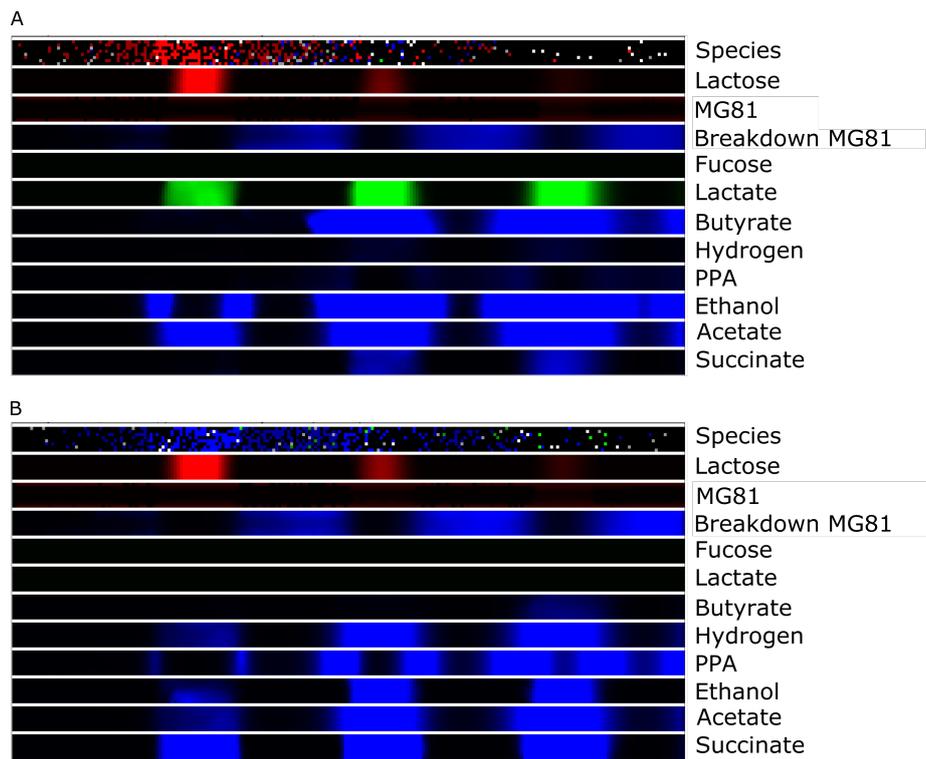


Figure 17: Screenshot of the spatially model from the videos in supplementary 6.2. The proximal end of the colon is on the left, the distal end on the right. (A) With a high abundance of *B. longum* indicated with red and (B) with a high abundance of *B. vulgatus* indicated with blue.

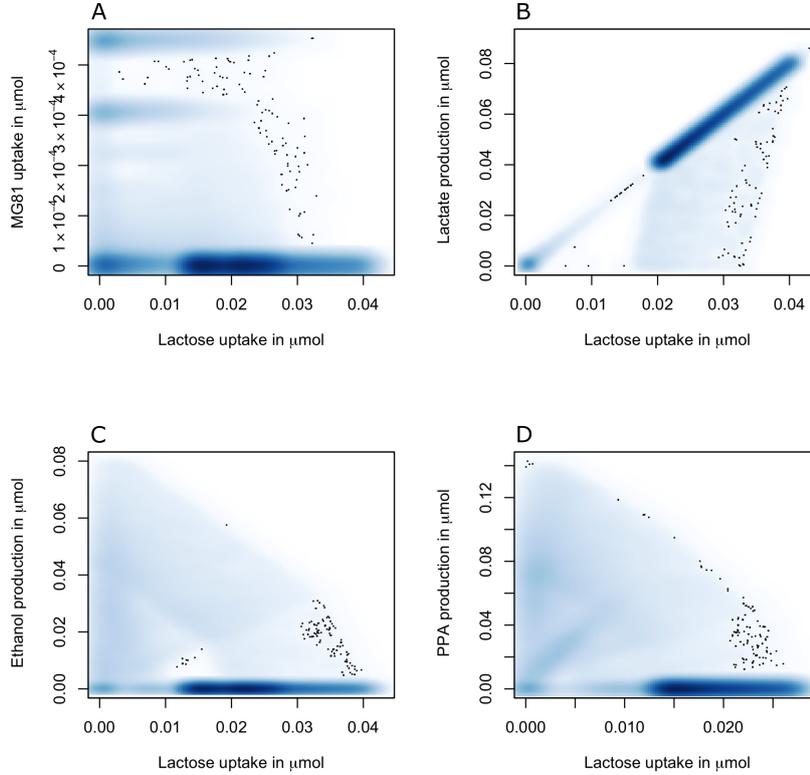


Figure 18: The lactose uptake versus the MG81 uptake filtering on the species *B. longum*, *B. vulgatus* and *E. coli*, the lactose uptake versus the lactate production filtering on the species *B. longum* (B), the lactose uptake versus the ethanol production filtering on the species *B. longum* and *B. vulgatus* (C), the lactose uptake versus the PPA production filtering on the species *B. vulgatus* (D) in the last 4 days. Additionally, we filtered in all the three plots on active growth of the species (growth > 0.1). n=23.

Figure 19 shows for the three species *B. vulgatus*, *B. longum* and *E. coli* a spatial plot where each small dash represents a lattice site from the model. We added MG81 with a maximum addition rate of  $0.0003 \mu\text{mol}$  MG81 per time step and then looked at the mean volume (A), the mean uptake of lactose (B), and the mean uptake of MG81 (C) per time step in the last 4 days. *B. vulgatus* and *B. longum* have the largest volume increase in the beginning of the lattice. While *B. vulgatus* also occurs in the more distal part of the lattice, *B. longum* occurs rarely in the distal part. *E. coli* does not have a clear part where the increase in volume is larger. Almost throughout the whole lattice we see similar occurrence. For the uptake of lactose we see for all the three species similar band-like patterns and also mostly in the same place which is an artifact of only saving the output every 20 time steps. The lactose uptake for the three species follows the same pattern as the mean volume. This shows that the volume of the species and the lactose uptake are related to each other. The MG81 uptake is dominated by *B. vulgatus* and takes place almost along the whole lattice with a maximum between the fiftieth and hundredth lattice site. *B. longum* consumes mucin glycans mainly in the first half of the lattice and *E. coli* consumes

small amounts along the whole lattice. Moreover, for all the three species we see that MG81 is consumed at the walls of the lattice, which is what we expected since we have a small diffusion rate. In contrast to lactose, we see with MG81 uptake not a clear spatial distribution effect. For instance, we do not observe that *B. vulgatus* is more present at the top and bottom lattice site where MG81 is situated.

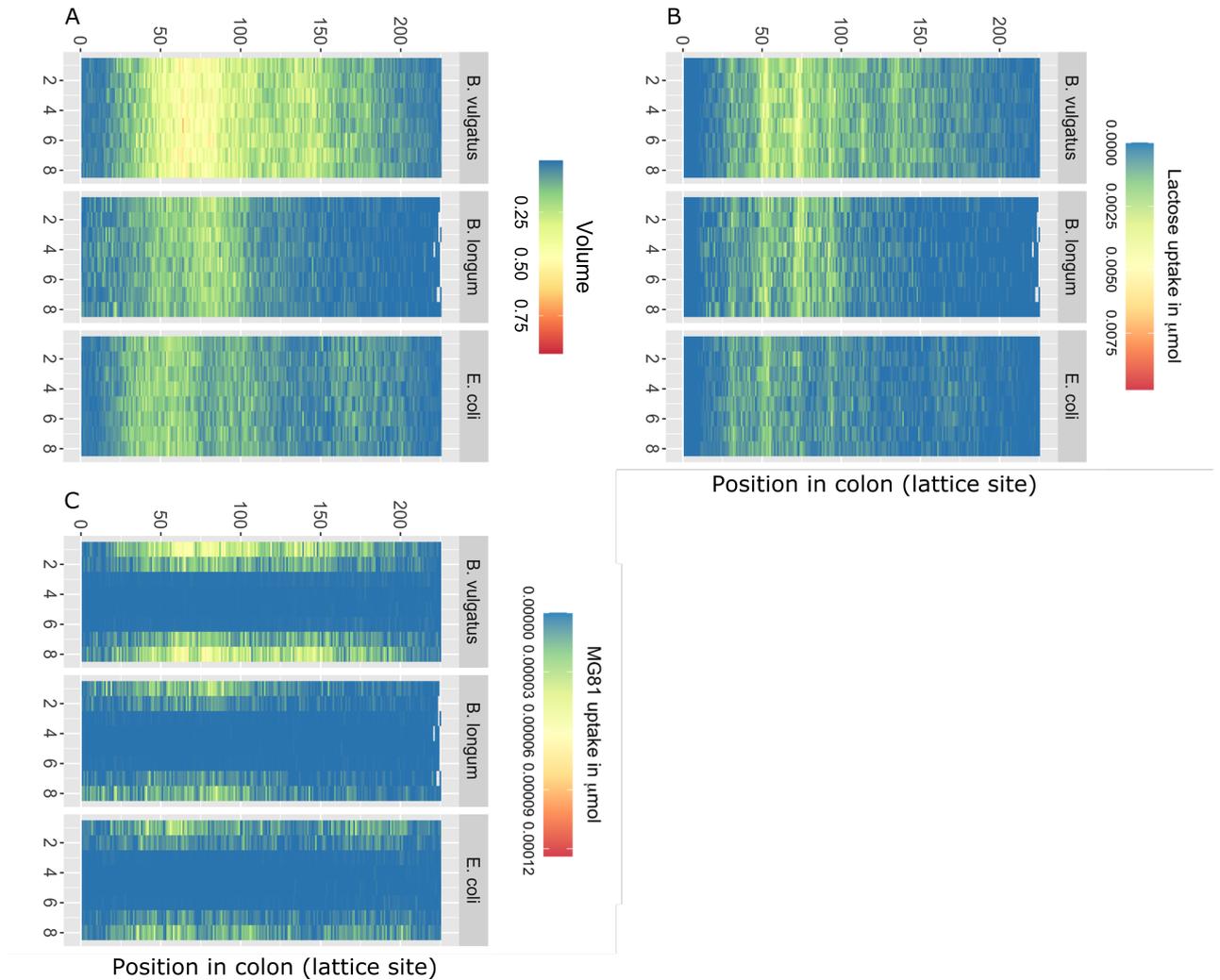


Figure 19: Heatmap of the mean volume (A), lactose uptake in  $\mu\text{mol}$  (B), and the MG81 uptake in  $\mu\text{mol}$  (C) per time step for the species *B. vulgatus*, *B. longum* and *E. coli* in the last 4 days. The proximal end of the colon is on the left, the distal end on the right. The maximum generation rate is 0.0003 MG81  $\mu\text{mol}$  per time step. (n=23)

These results raise the question what the relationship is between increasing in volume and the lactose and MG81 intake. Figure 20 shows the result of the sum of the volume versus the sum of the lactose uptake (A), and versus the sum of the MG81 uptake (B) for the species *B. vulgatus*, *B. longum* and *E. coli* in the last 4 days. Again, the maximum generation rate of MG81 is 0.0003

$\mu\text{mol}$ . Each dot represents one run for one of the three species. Interestingly, we see in both cases a linear relationship. In the case of lactose uptake (Fig. 20A) the slope for *B. longum* is less high. Considering the sum of the volume versus the sum of the MG81 uptake (Fig. 20B), we see the same linear relationship for the three species.

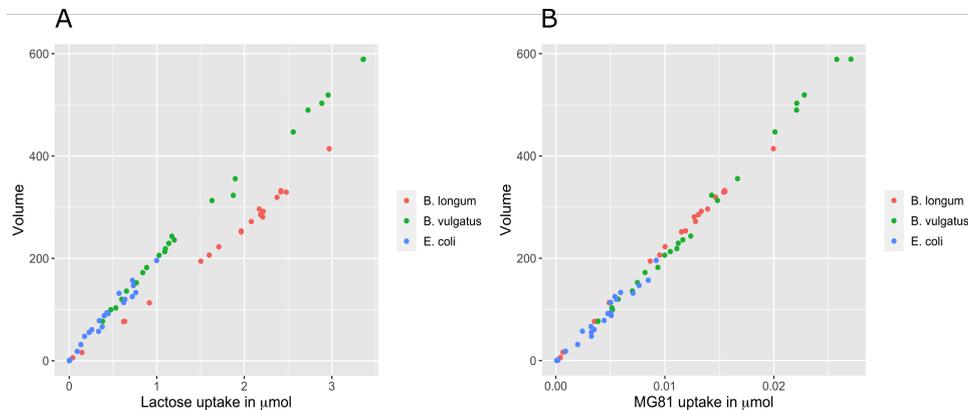


Figure 20: Scatterplot of the mean volume versus the mean lactose uptake in  $\mu\text{mol}$  (A), and versus the mean MG81 uptake in  $\mu\text{mol}$  (B) per time step for the species *B. vulgatus*, *B. longum* and *E. coli* in the last 4 days. The maximum generation rate of MG81 is  $0.0003 \mu\text{mol}$  per time step. ( $n=23$ )

#### 4.8 More cross-feeding at the distal part compared to the proximal part of the lattice

We saw in the previous section that due to consumption there is less lactose in the distal part compared to the proximal part of the lattice. This leads to the hypothesis that there is more cross-feeding at the distal part compared to the proximal part of the lattice. To test this hypothesis we calculated the cross-feeding factor  $C(i)$  following the method introduced by Van Hoek and Merks [27]. Table 2 gives an overview of the cross-feeding factor for the three species together and separately. Lactose and MG81 were used as primary source and the precise used related metabolites for these two compounds can be found in supplementary 6.3. The proximal part consists of all the lattice sites between the 1st and the 113th lattice site of the long side of the lattice and the distal part of the lattice sites between the 113th and the 225th lattice site. We observe in all four situations that the cross-feeding factor is lower in the proximal part and higher in the distal part. Including all three species we have a  $C(i)$  of  $0.12 \pm 0.006$  versus  $0.36 \pm 0.06$ , which shows that there is more cross-feeding in the distal part of the lattice. When we only consider one species, we see that *B. vulgatus* has a higher value for the proximal and distal part compared to the other two species. In conclusion, in the beginning of the lattice the species consume mostly primary resources, namely lactose and MG81, and to the end part of the lattice the species consume more secondary metabolites produced from lactose and MG81. This is also what we would expect, since due to consumption lactose gets more sparse to the end of the lattice.

	Proximal	Distal
All three species	$0.12 \pm 0.006$	$0.36 \pm 0.06$
<i>B. longum</i>	$0.066 \pm 0.02$	$0.21 \pm 0.1$
<i>B. vulgatus</i>	$0.2 \pm 0.05$	$0.42 \pm 0.08$
<i>E. coli</i>	$0.14 \pm 0.09$	$0.29 \pm 0.1$

Table 2: Cross-feeding factor for the proximal and distal part of the lattice for the three species together and separately. We used lactose and MG81 as primary nutrition source. We calculated the cross-feeding factor by first summing the consumption of each metabolite per run and per species in the last 4 days. Then, we adjusted the values for the amount of carbon atoms for each metabolite. For each run we calculated the cross-feeding factor and lastly, we took the mean and standard deviation over the 23 runs.

#### 4.9 Distinctive patterns of the abundances of the species result in a different composition of metabolites

Principal component analysis (PCA) is used to summarize and visualize the information in a data set containing observations described by multiple inter-correlated quantitative variables. Especially in the case of having more than three variables, as is the case in our data sets, it is an informative method to analyse the data [31]. In short, new uncorrelated variables, so called principal components, are constructed as linear combinations of the initial variables. Moreover, the explained variance in the first principal component is maximized and then again for the remaining information its again maximized in the second principal component and so on. This ensures that most of the information within the initial variables is compressed into the first components [32].

Characteristics of the microbiota and the metabolites from 23 runs were evaluated using PCA and partitioning around medoids (PAM) clustering algorithm. First, the R function `fviz_nbclust()` in combination with the partitioning function PAM was used to estimate the optimal number of clusters. Then, the runs were partitioned based on the abundances of the species at day 21 (Fig. 21A) and based on the sum of the metabolites in the system in the last 4 days (Fig. 21B). Similarity in runs that belong to the same cluster are visible between the two plots. Cluster 1 contains the same runs for both analyses, except for run 5 which is also in cluster 1 and cluster 2 deviates on four runs. Based on the species these four runs belong to cluster 2, but based on the metabolites these are in cluster 3. Cluster 3 has no similarities in runs between between the two different characteristics. Cluster 1 consists of runs where *B. vulgatus* is very abundant and *B. longum* has died out or is much less abundant than *B. vulgatus*. Cluster 2 consists of runs where *B. longum* is very abundant and *B. vulgatus* is less abundant or sometimes more abundant. Cluster 3 consists of one run where *B. longum* is more abundant than *B. vulgatus* but especially the species *Streptococcus oralis* and *Propionibacterium acnes* are relatively more abundant. Most of the times these two species have died out, but in the run that belongs to cluster 3 these are respectively 20 and 35 times bigger compared to the runs in the other clusters. Since we scale the data these increases play a role in the clustering although their absolute values are one order smaller than the values for *B. vulgatus* and *B. longum*. This deviation probably explains why cluster 3 only contains one run. If we now consider the characteristics of the metabolites (Fig 21B) then runs in cluster 1 have less acetate, lactate, ethanol, and MG10 and have more hydrogen, succinate, and PPA compared to the runs in cluster 2 and 3. Additionally, runs from cluster 1 have almost no MG10, while runs from cluster 2 and 3 do contain MG10. First of all we notice that run 20, belonging to cluster 3, is an outlier. The other 3 runs belonging to cluster

3 are also close to runs in cluster 2. This is also seen in the data, run 20 has the highest value for acetate, lactate, ethanol and MG10, and the lowest values for hydrogen, succinate and MG10\_rl. The similarity in clusters between the clusters based on microbiota composition and metabolites composition shows that a microbiota profile is related to a metabolite profile. Short-chain fatty acids (SCFA) are related to health, thus these metabolite compositions are important for the host. Higher levels of acetate for instance exert benefits on certain cardio-metabolic disease risk by decreasing visceral fat [33].

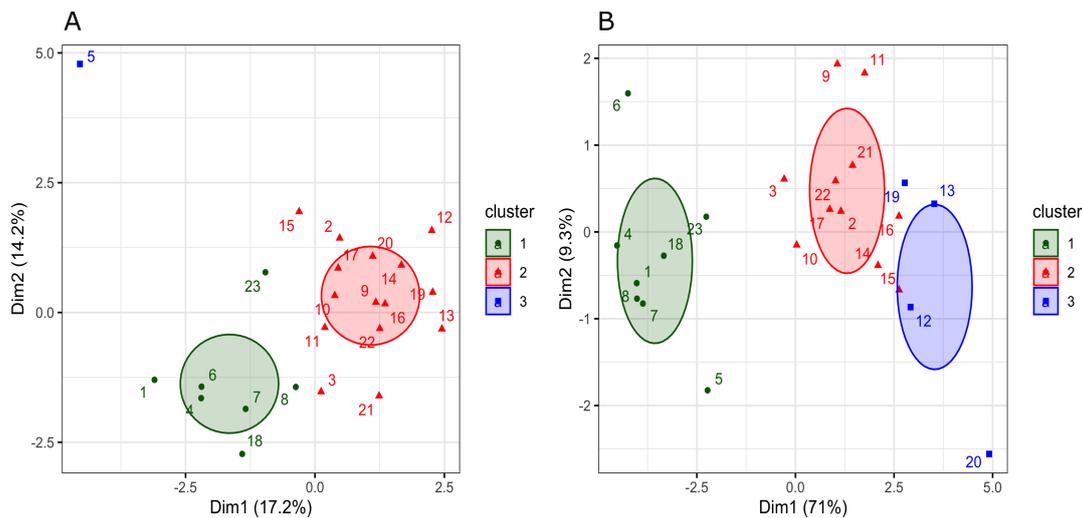


Figure 21: Characteristics of infant gut microbiota at day 21 (A) and metabolites in the system in the last 4 days (B), illustrated using PCA and PAM clustering analyses. The maximum generation rate of MG81 is  $0.0003 \mu\text{mol}$  per time step. ( $n=23$ )

#### 4.10 *B. vulgatus* benefits less with a change in mucin glycan

Until now we have only focused on one specific mucin glycan. However, as stated earlier there exist hundreds of different mucin glycans and therefore we are going to investigate the possible effects of a change in mucin glycan, namely MG38. Figure 22 gives a schematic representation of MG38, and it can be seen that it is highly fucosylated. Furthermore, we performed again a pathway analysis that shows the theoretically possible metabolic pathways in the degradation of MG38 (Fig. 23). The arrows represent reactions and only the formed mucin glycan type is shown. An extensive difference in the metabolic pathway compared to MG81 is that the used species are not able to fully degrade MG38. None of the species contain enzymes that are able to degrade MG73. The maximum addition rate for MG38 has to be higher compared to MG81 for *B. vulgatus* to dominate (Fig. 25). When MG81 is added, *B. vulgatus* dominance already started at a maximum addition rate of  $0.0003 \mu\text{mol}$ , while when MG38 is added this happens at  $0.001 \mu\text{mol}$ . Furthermore, if we consider the scatterplots of the abundances of *B. longum* versus *B. vulgatus*, we see that especially in the case of a maximum addition rate of  $0.0001$  and  $0.0005 \mu\text{mol}$  MG38 that either *B. vulgatus* or *B. longum* dies out (Fig. 25BD). When there is a maximum addition rate of  $0.001 \mu\text{mol}$  MG38 we see that in more runs *B. longum* has died out, but also that when *B. longum* has not died out *B. vulgatus* still survives in a number of runs. In the case of MG81, we saw a lot more runs where both species survived. These results suggest

that with lower amounts of MG38 *B. vulgatus* cannot benefit enough of the extra nutrient source and as a consequence *B. longum* wins in the competition for lactose uptake. In other words, the advantage is so small that the situation is similar to the case when only lactose is added as a nutrient source. The higher amounts of MG38 in comparison to MG81 are needed for *B. vulgatus* to increase in size since less energy can be taken from MG38. This is confirmed in an experiment similar to the experiment presented in section 4.2, but then with different concentrations of MG81 and MG38 instead of lactose (Fig. 24). *B. vulgatus* grows less rapidly on MG38 compared to MG81 with especially lower concentrations of the mucin glycans and even when the maximum growth is reached for both mucin glycans *B. vulgatus* still grows less rapidly on MG38.

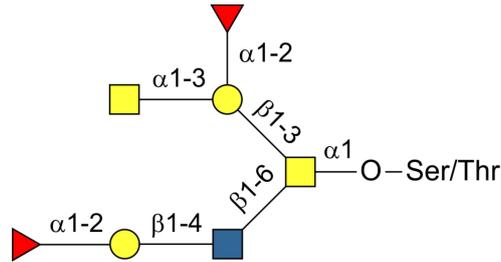


Figure 22: Schematic representation of MG38. This mucin glycan consist of a core 3 structure and is furthermore highly fucosylated (red triangle).

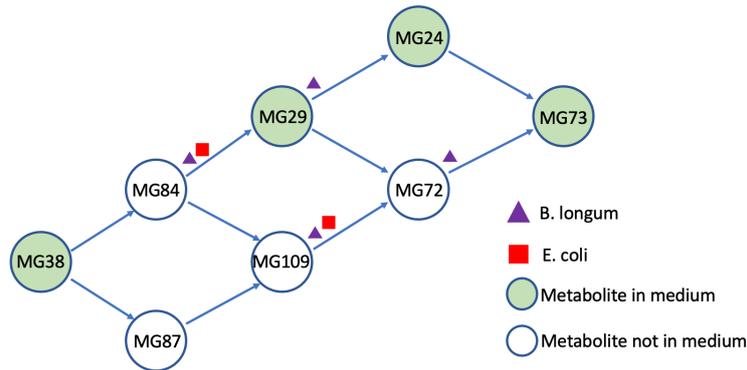


Figure 23: Schematic representation of the theoretically possible metabolic pathways for MG38 based on the GEMs. The arrows represent reactions and only the formed mucin glycan type is shown. The nodes in green represent metabolites that are seen in the medium, and the nodes in white represent metabolites that are not seen in the medium. The species *B. vulgatus* and *P. distasonis* are able to perform all reactions. The reactions that can be performed by the other species are indicated with a triangle for *B. longum* and a square for *E. coli*.

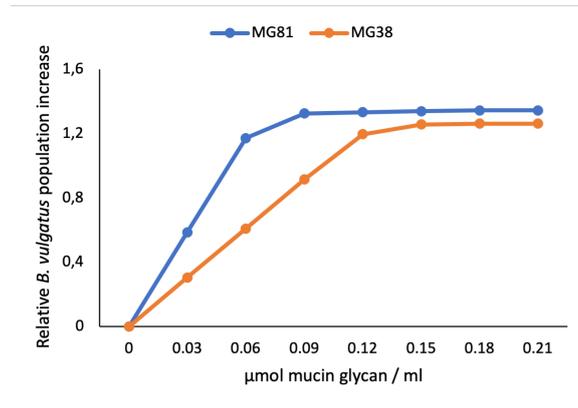


Figure 24: Growth per timestep for the species *B. vulgatus* with different concentrations of MG81 and MG38.

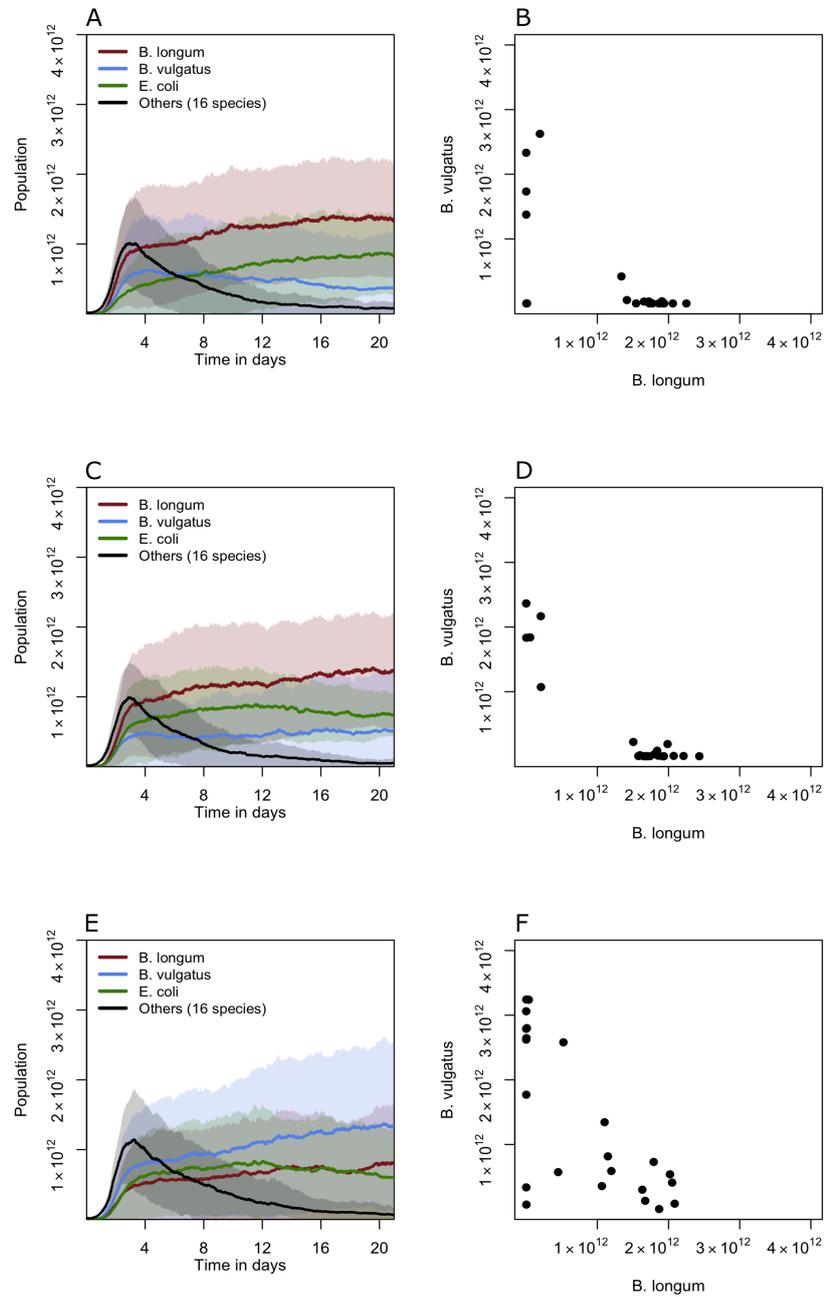


Figure 25: ACE: abundances for grouped species over time. The bold lines indicate the means of the abundance of species over the different runs and the shaded area indicates one standard deviation BDF: abundances of *B. longum* versus *B. vulgatus* at 21 days. AB: maximum addition of  $0.0001 \mu\text{mol MG38}$  per time step, CD: maximum addition of  $0.0005 \mu\text{mol MG38}$  per time step, EF: maximum addition of  $0.001 \mu\text{mol MG38}$  per time step. ( $n=23$ )

Figure 26 displays the exchange behaviour of MG38 and secondary metabolites produced from MG38 with a maximum addition of  $0.0005 \mu\text{mol}$  MG38 per time step. *E. coli* does not participate in the degradation of MG38. *B. vulgatus* is the main user for MG38 (Fig. 26). This is also what we would expect since *E. coli* and *B. longum* cannot degrade MG38 (Fig. 23). Then, we see that *P. distasonis* secretes MG29, in approximately equal amounts to the amount of MG38 taken up in the beginning. *B. longum* takes up a big part of this released MG29 and then secretes it as MG24. Again this is evident since *B. longum* cannot degrade MG24 further (Fig. 23). *B. vulgatus* also produces MG24, but at the same time in lesser amounts uses it too. The end product MG73 is mainly produced by *B. vulgatus*. Thus, we see the same behavior as seen with the addition of MG81: *B. vulgatus* degrades almost the whole mucin glycan by itself, leaving little room for cross feeding and additionally takes up metabolites secreted by other species.

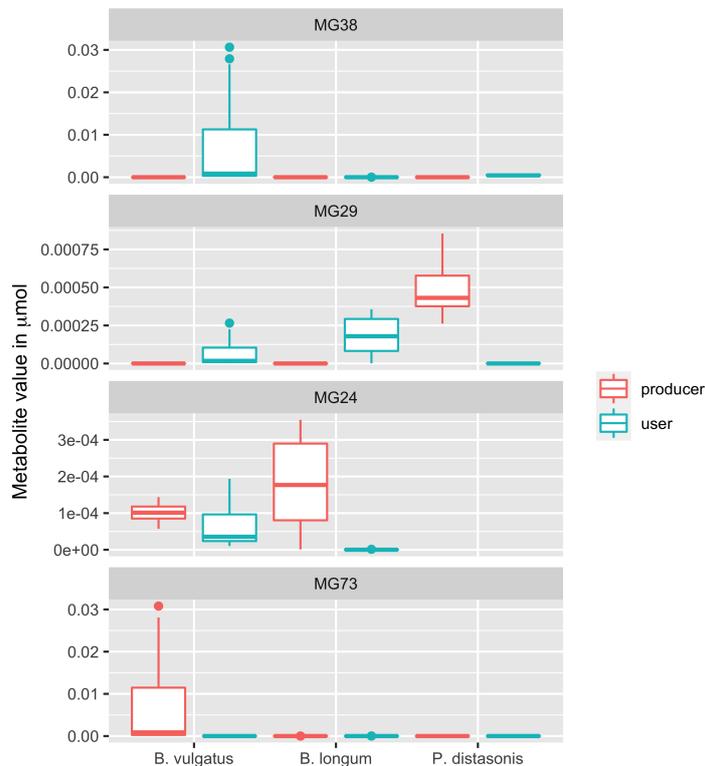


Figure 26: Mean uptake and secretion per time step of metabolites MG38, MG29, MG24, and MG73 in the last 4 days with a maximum addition of  $0.0005 \mu\text{mol}$  MG38 per time step. ( $n=23$ )

#### 4.11 *B. vulgatus* knock-out for MG81 degradation shows less advantage of MG81 compared to the normal *B. vulgatus*

As discussed previously, we hypothesised that the rapid growth of *B. vulgatus* in the presence of mucin glycans is due to a complete breakdown of the mucin glycan, leaving little room for cross feeding by other species that do not contain all the enzymes to break down the molecule. To test

this, we ran an experiment where the reactions between the secondary and tertiary metabolites in figure 7 were disabled. This includes the reactions:  $MG152 \rightarrow MG10$ ,  $MG152 \rightarrow MG150$ ,  $MG152 \rightarrow MG151$ ,  $MG120 \rightarrow MG150$ ,  $MG120 \rightarrow MG14$ ,  $MG4 \rightarrow MG151$ , and  $MG4 \rightarrow MG14$ . Also, the released forms of these reactions were disabled to avoid the situation that *B. vulgatus* performs the pathway through that form.

The abundance of *B. vulgatus* KO decreases compared to the result obtained with control simulations of *B. vulgatus* (Fig. 27). The amount of *B. vulgatus* at 21 days was significantly lower than in the normal situation ( $p < 0.05$ , two sample t-test) and the amount of *B. longum* slightly increases, but not significantly. *B. vulgatus* secretes the secondary metabolites in the medium as we expected (Fig. 28). These are the metabolites MG120 and in lesser amounts MG4, which were first not seen in the medium as metabolites at all. This means that *B. vulgatus* breaks down MG81 through the reactions that lead to MG120 and MG4. Most of these two secreted mucin glycans are used by *B. longum* and in lesser amount by *E. coli*. Thus, the knock out version does enhance the cross-feeding. Furthermore, *B. vulgatus* immediately takes up as much as possible released metabolites of other species when *B. vulgatus* can degrade them again, as can be seen in the uptake of the tertiary metabolites (Fig. 29). This confirms the hypothesis, that *B. vulgatus* holds on to metabolites until fully degraded.

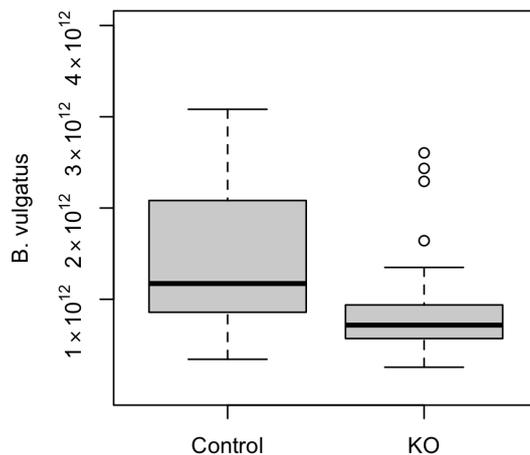


Figure 27: Abundance of *B. vulgatus* at day 21 with a maximum addition of 0.0003  $\mu\text{mol}$  MG81 per time step. We used a control and a knock out version of *B. vulgatus*, where the reactions between the secondary and tertiary metabolites of MG81 in figure 7 were disabled. (n=23)

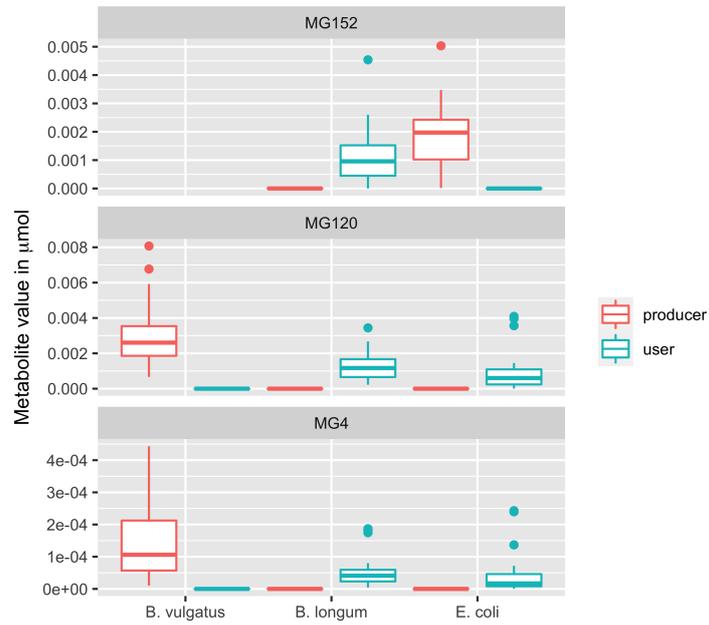


Figure 28: The uptake and secretion of secondary metabolites from MG81. We used a knock out version of *B. vulgatus*, where the reactions between the secondary and tertiary metabolites in figure 7 were disabled.

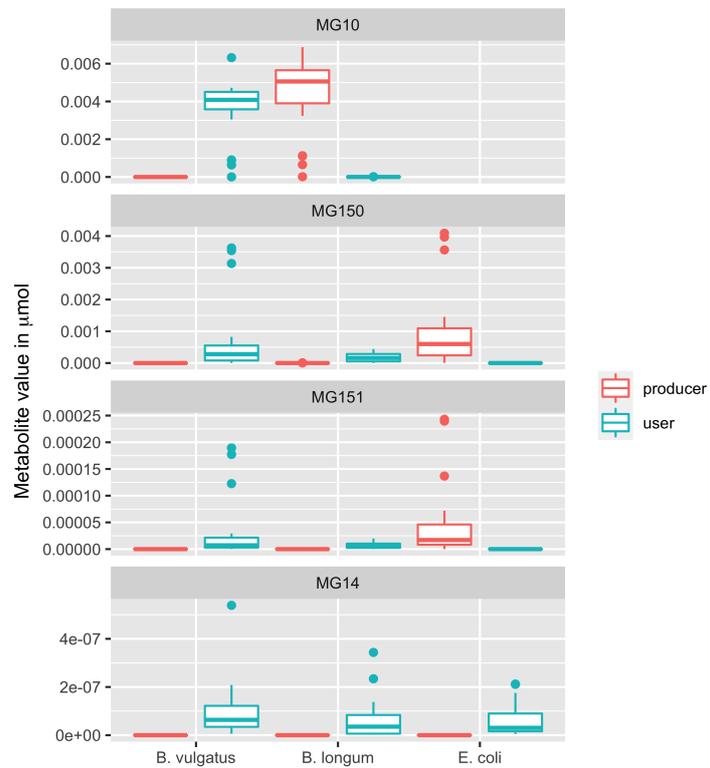


Figure 29: The uptake and secretion of tertiary metabolites from MG81. We used a knock out version of *B. vulgatus*, where the reactions between the secondary and tertiary metabolites in figure 7 were disabled.

## 5 Discussion

We have extended a multiscale model of the infant gut microbiota, with mucin glycans, to predict potential effects on the abundance of the species. In particular, the species *B. vulgatus* benefits from the mucin glycans as extra nutrient source and the advantageous effect is due to a complete degradation of the mucin glycan molecule leaving little room for cross feeding by other species. Furthermore, the feeding on mucin glycans benefits *B. vulgatus* in the competition for the nutrient source lactose. To obtain these results, we used a dynamic spatial FBA model with an additional flux constraint. Our findings suggest that these species in the infant gut that are good in degrading mucin glycans can have an advantage compared to other species in the competition for other nutrient sources, such as lactose. Additionally, we saw that independent of addition of mucin glycans that the abundance of the other sixteen species over time followed the same pattern: first they experience an initial growth, but then decline after day three. In the initial phase of the experiments, there is less consumption by the species and therefore there is abundant nutrient which causes the initial growth. After some time, the consumption has increased which causes the sources to become more sparse. Consequently competition between all the species for nutrient consumption takes place, which results in a decline of the sixteen species.

The used model is a simplified simulation of the *in vivo* situation and this causes discrepancies between our results and *in vivo* data. First of all, the intestinal microbiota of infants shows very much inter individual variability and this causes difficulties in comparing results with *in vivo* data [3]. We will compare our findings with the *in vivo* results obtained by Bäckhed et al. and by Tsukuda et al. [4, 28]. The first one is the study that we based our used species on and we use the latter because this study provides a more complete understanding of the early life metabolic–microbiota relationships. Bäckhed et al. found a relative high abundance of the genera *Bacteroides*, *Bifidobacterium* and *Escherichia* in vaginally born infants. The abundance of *Bacteroides* is slightly higher compared to *Escherichia* and *Bifidobacterium* in newborns, and a bit lower compared to *Bifidobacterium* in 4 months infants. This corresponds to our findings where the three species, *B. vulgatus*, *E. coli*, and *B. longum*, have abundances significant higher at the end of the simulation compared to the initial condition. A difference is that Bäckhed et al. also found relative high abundance of other genera, such as *Clostridium* and *Streptococcus*. The results obtained by Tsukuda et al. shows initial microbiota dominance by orders *Enterobacteriales* and *Bifidobacteriales* and in lesser amount by *Bacteroidales* in faecal samples from 12 infants. *E. coli* is a species belonging to the order *Enterobacteriales*, *B. longum* a species belonging to the order *Bifidobacteriales* and *B. vulgatus* a species belonging to the order *Bacteroidales*. Our results also showed dominance from these three orders, however a difference is that under most conditions we had more *Bacteroidales* compared to *Enterobacteriales*. Our results where we added the smallest amount of mucin glycans corresponded best with the results obtained by Tsukuda et al. compared to no or more addition of mucin glycans. This shows that mucus has an effect on the microbiota composition. The following suggestions could improve *in silico* results.

In the first days the infant gut is prone to aerobic conditions, while we have modeled the gut completely anaerobic. Another study has implemented these aerobic conditions and found that *E. coli* benefits in this condition [23]. Future experiments could implement both the aerobic conditions and mucus degradation.

The mucus has been modeled by only one type of mucin glycan and with just one layer along the top and bottom wall. It could provide more insight to add a diverse set of mucin glycans and additionally include a gradient layer of these glycans. In our results we did not find a clear spatial

effect of the mucin glycans regarding the microbiota. Using a gradient could cause species that degrade mucin glycans to be more present in parts with a higher concentration. Another cause is the shape of the lattice, since the width is much smaller than the length its harder for vertical niche forming due to less space. Lastly, the lactose gradient also may influence a potential niche forming related to the mucin glycans.

As discussed before, the BacArena study also implements a spatial temporal FBA model to simulate bacterial species dynamics [25]. This study uses seven species which are based on their relevance and abundance within the adult gut microbiota. Thus, these were not specialised for the infant gut. We will compare the results with their added species *Bacteroides thetaiotaomicron*, *E. coli* and *B. longum*, since these were the most abundant in our results and these are either the same or belong to the same genus as our added species. A similarity between our results and theirs is the enormous increase of *Bacteroides* in the present of mucin glycans. Furthermore, when no mucin glycans are added *E. coli* and *B. longum* are more abundant. A difference is that *E. coli* is in both cases more abundant than in our results. They also obtained a spatially structure based on a mucin glycan gradient, with *Bacteroides* more present in a higher concentration of mucin glycans. We did not obtain such result due to a single layer of mucin glycans.

In this model we mainly consider carbon metabolism, and also a bit of amino acid metabolism caused by the mucin glycans that contain a serine-threonine group. A more extensive addition of nutrient sources, such as fats and more amino acids, would allow for a more accurate metabolism and could cause more diverse species to benefit. However, it is quite complex to introduce these new nutrient sources due to a lack of clear data on the concentration and uptake of many nutrients.

## 6 Supplementary

### 6.1 Exchange metabolites related to MG81

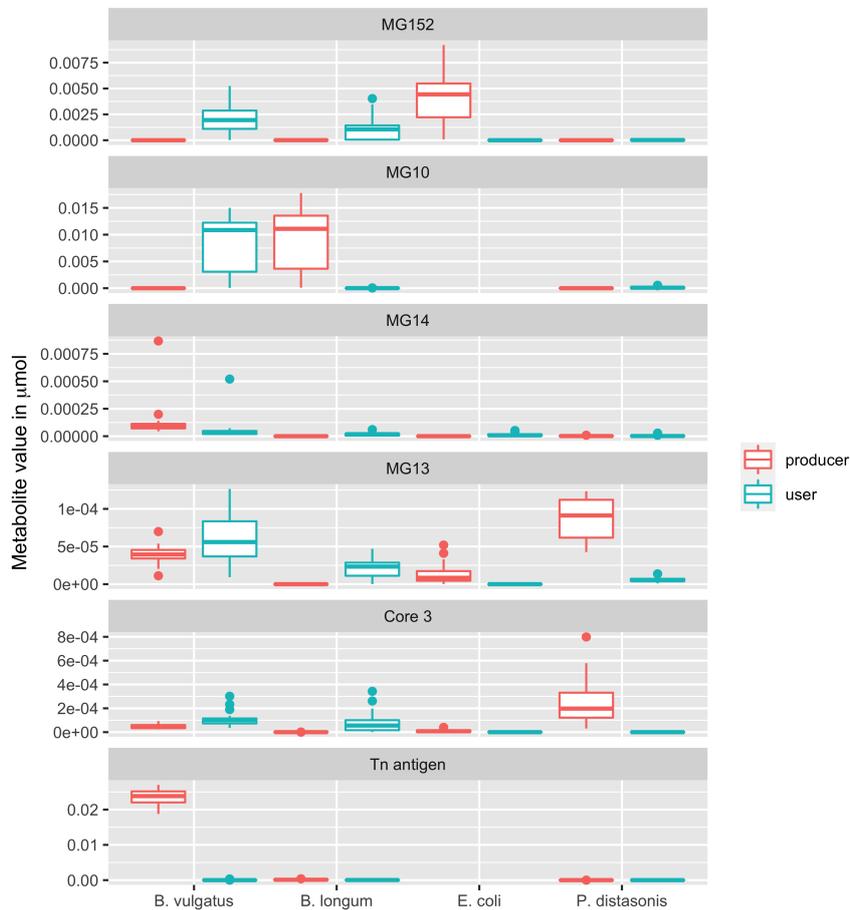


Figure 30: Mean uptake per time step of metabolites related to MG81 in the last 4 days with addition of  $0.0003 \mu\text{mol}$  MG81 ( $n=23$ ).

### 6.2 Movie

Here the movies from all the 23 runs with a maximum generation rate of  $0.0003$  MG81 are discussed in detail.

*run 1*: First, *B. vulgatus* emerges at circa  $\frac{3}{4}$  in the lattice and *B. longum* and *E. coli* in the last quarter, additionally *E. coli* also emerges in the second quarter of the grid. Then, *B. vulgatus* moves to the front of the lattice and mingles with *E. coli*. *B. longum* at the end of the lattice decreases and *B. vulgatus* eliminated *E. coli*.

*run 2:* First, *B. longum* emerges at circa  $\frac{3}{4}$  in the lattice and *E. coli* in the last quarter. Then, *B. longum* moves to the front of the lattice and *B. vulgatus* emerges and mingles with *B. longum*. *E. coli* decreases in the last part of the lattice and also moves more to the front, additionally *B. vulgatus* increases.

*run 3:* *B. vulgatus* and *E. coli* are mildly mingled at the end of the lattice. Then, they both move more to the front of the lattice and *B. vulgatus* becomes the dominated species. Then, *B. longum* emerges in the front of the lattice and over time increases and mingles with *B. vulgatus*.

*run 4:* *B. vulgatus* emerges in the middle of the lattice and at the end of the lattice are *B. vulgatus* and *E. coli* present. *B. vulgatus* moves more to the front and *E. coli* decreases in the end of the lattice, which is then dominated with *B. vulgatus*. Briefly, *B. longum* emerges at the end of the lattice but shortly after dies out.

*run 5:* At the end of the lattice *B. longum* with *E. coli* emerges, with the latter even more to the end of the lattice. In the middle of the lattice *B. vulgatus* emerges. Then, *B. longum* and *B. vulgatus* move more to the front. *E. coli* mingles with the other species more to the front, but also stays in the end of the lattice. At the end of the simulation the species are well mingled with *B. vulgatus* being the dominated species.

*run 6:* A lot of *B. vulgatus* emerges in the second half of the lattice. No other species emerge and almost only *B. vulgatus* is present in the lattice, with a movement more to the front of the lattice.

*run 7:* *B. vulgatus* emerges in the middle of the lattice and *B. longum* and *E. coli* emerges after *B. vulgatus* (thus more to the end of the lattice). *B. vulgatus* increases and *E. coli* mingles a little with *B. vulgatus*. After some time *B. longum* also mingles more, but at the same time decreases.

*run 8:* A lot of *E. coli* emerges in the middle of the lattice with in front of *E. coli* some *B. vulgatus*. The two species mingle and *B. vulgatus* moves more to the front and *E. coli* more to the end. *B. vulgatus* is the dominated species and also takes up more space at the end of the lattice.

*run 9:* At the end of the lattice *B. longum* and *E. coli* emerges fast, with the latter also presented a little more in the front of the lattice. *B. longum* dies out with *E. coli* being the dominated species. Then, *B. vulgatus* emerges at the end of the lattice and moves quickly to the front. *B. longum* again emerges, but now in the front of the lattice and mingles with the other species that are present.

*run 10:* *B. longum* emerges over a large part in the front of the lattice, *B. vulgatus* emerges more in the end of the lattice and *E. coli* emerges slowly at the end of the lattice. *B. vulgatus* and *E. coli* move to the front and mingle. *E. coli* also stays more at the end but after a while *B. vulgatus* is also more present at the end.

*run 11:* A lot of *B. longum* emerges at circa  $\frac{3}{4}$  of the lattice and *B. vulgatus* emerges in the last quarter of the lattice. They both mingle and move more to the front with *B. longum* being the dominated species. *E. coli* emerges at the end of the lattice. *B. vulgatus* increases but still mingles well with *B. longum*.

*run 12:* *B. longum* emerges in the middle of the lattice after that a little bit of *B. vulgatus* is present and at the end of the lattice *E. coli* emerges. *B. vulgatus* mingles more with *B. longum* and *E. coli* stays mainly in the end. Overall all the species move more to the front. At the end

of the simulation *E. coli* mingles more to the front and *B. longum* and *B. vulgatus* more to the end.

*run 13*: At the end of the lattice *B. vulgatus* and *B. longum* emerge, also *B. vulgatus* emerges more to the front of the lattice. After some time *E. coli* also emerges at the end of the lattice. *B. vulgatus* and *B. longum* move to the front and stay mingled.

*run 14*: At the end a lot of *B. vulgatus* emerges. Before *B. vulgatus* in the lattice *B. longum* and *E. coli* emerges. *B. longum* and *E. coli* stay mingled and move to the front and slowly *B. vulgatus* also mingles from the end up to the front of the lattice. *E. coli* is mostly present in the end.

*run 15*: *B. longum* emerges more in the beginning of the lattice and *B. vulgatus* and *E. coli* more to the end. Then, *B. longum* and *B. vulgatus* mingle with *B. longum* being the dominating species. At the the end of the lattice more *B. vulgatus* and *E. coli* is present.

*run 16*: More to the end of the lattice the species emerge in this order: *E. coli*, *B. vulgatus*, *B. longum* and then again *B. vulgatus*. *B. longum* moves more to the front folowed by *E. coli* which mingles with *B. vulgatus*. *B. longum* is the dominated species, but also mingles with *B. vulgatus*.

*run 17*: More in the front emerges *B. longum* with at the end also *E. coli* present and more to the end emerges *B. vulgatus*. Slowly *E. coli* and *B. vulgatus* move more to the front en mingle with *B. longum*, but the two species also remain in smaller numbers at the end of the lattice.

*run 18*: *B. longum* emerges a little bit at the beginning but also at then end. *B. vulgatus* emerges at the beginning of the second half of the lattice. Then, *B. longum* remains mainly in the beginning and *B. vulgatus* moves more to the front and is the dominated specie.

*run 19*: At the end *B. longum* and then *E. coli* emerges. *B. longum* moves to the front together with *E. coli*. *B. vulgatus* emerges in the end and is also a little bit present in the front mingled with *B. longum*. The three species mingle but at the end of the lattice mainly *E. coli* is present.

*run 20*: At the beginning of the second half of the lattice a lot of *B. longum* emerges and then moves more to the front. *B. longum* is the only dominating species. Then, *B. vulgatus* emerges at the end of the lattice and moves more to the front.

*run 21*: In the middle a bit of *B. vulgatus* mingled with *E. coli* emerges and at the end *B. longum* emerges also mingled with *E. coli*. *B. vulgatus* remains dominated in the beginning of the lattice but also mingles more with *E. coli* and later on with *B. longum*.

*run 22*: *B. longum* emerges more to the front mingled with *B. vulgatus* and the latter is also mingled more to the end with *E. coli*. This remains throughout the whole simulation.

*run 23*: *E. coli* emerges at circa  $\frac{1}{3}$  of the lattice mingled with *B. vulgatus*. Then, the latter is more present in the beginning and begins to dominate. A little bit of *B. longum* emerges in the beginning mingled with *B. vulgatus*. *E. coli* moves also more to the front but *B. vulgatus* stays dominated.

**6.2.1 Video. Video of the simulation where *B. vulgatus* is the dominant species**

**6.2.2 Video. Video of the simulation where *B. longum* is the dominant species**

### 6.3 Specification for cross-feeding conditions

Acetate
Lactate
PPA
Succinate
Butyrate
Ethanol
MG152
MG14
MG13
MG10
Core 3
Tn antigen

Table 3: The used metabolites for lactose and MG81 for determining the cross-feeding factor.

## References

- [1] M. Ventura, F. Turroni, M. O. Motherway, J. MacSharry, and D. van Sinderen, “Host–microbe interactions that facilitate gut colonization by commensal bifidobacteria,” *Trends in microbiology (Regular ed.)*, vol. 20, no. 10, pp. 467–476, 2012.
- [2] S. Dogra, O. Sakwinska, S.-E. Soh, C. Ngom-Bru, W. M. Brück, B. Berger, H. Brüssow, N. Karnani, Y. S. Lee, F. Yap, Y.-S. Chong, K. M. Godfrey, and J. D. Holbrook, “Rate of establishing the gut microbiota in infancy has consequences for future health,” *Gut microbes*, vol. 6, no. 5, pp. 321–325, 2015.
- [3] S. Matamoros, C. Gras-leguen, L. Vacon, G. Potel, and M.-f. D. L. Cochetiere, “Development of intestinal microbiota in infants and its impact on health,” *Trends in Microbiology*, vol. 21, no. 4, pp. 167–173, 2013.
- [4] N. Tsukuda, K. Yahagi, T. Hara, Y. Watanabe, H. Matsumoto, H. Mori, K. Higashi, H. Tsuji, S. Matsumoto, K. Kurokawa, and T. Matsuki, “Key bacterial taxa and metabolic pathways affecting gut short-chain fatty acid profiles in early life,” *ISME Journal*, vol. 15, no. 9, pp. 2574–2590, 2021.
- [5] S. Rokhsfat, A. Lin, and E. M. Comelli, “Mucin – Microbiota Interaction During Postnatal Maturation of the Intestinal Ecosystem : Clinical Implications,” *Digestive Diseases and Sciences*, vol. 61, no. 6, pp. 1473–1486, 2016.
- [6] L. E. Tailford, E. H. Crost, D. Kavanaugh, and N. Juge, “Mucin glycan foraging in the human gut microbiome,” *Frontiers in Genetics*, vol. 6, no. March, 2015.
- [7] H. Makino, A. Kushiro, E. Ishikawa, H. Kubota, A. Gawad, T. Sakai, K. Oishi, R. Martin, K. Ben-amor, J. Knol, and R. Tanaka, “Mother-to-Infant Transmission of Intestinal Bifidobacterial Strains Has an Impact on the Early Development of Vaginally Delivered Infant ’ s Microbiota,” *PLoS ONE*, vol. 8, no. 11, 2013.
- [8] S. Dogra, O. Sakwinska, S. E. Soh, C. Ngom-Bru, W. M. Brück, B. Berger, H. Brüssow, Y. S. Lee, F. Yap, Y. S. Chong, K. M. Godfrey, and J. D. Holbrook, “Dynamics of infant gut microbiota are influenced by delivery mode and gestational duration and are associated with subsequent adiposity,” *mBio*, vol. 6, no. 1, pp. 1–9, 2015.
- [9] B. Ø. P. Orth, Jeffrey D., Ines Thiele, “What is flux balance?,” *Nat. Biotechnol.*, vol. 28, no. 3, pp. 245–248, 2010.
- [10] W. Boron and E. Boulpaep, *Medical Physiology*. Elsevier, 2016.
- [11] V. Bunesova, V. Bunesova, C. Lacroix, C. Lacroix, C. Schwab, and C. Schwab, “Mucin Cross-Feeding of Infant Bifidobacteria and *Eubacterium hallii*,” *Microbial ecology*, vol. 75, no. 1, pp. 228–238.
- [12] A. Macierzanka, A. R. Mackie, B. H. Bajka, N. M. Rigby, F. Nau, and D. Dupont, “Transport of particles in intestinal mucus under simulated infant and adult physiological conditions: Impact of mucus structure and extracellular DNA,” *PLoS ONE*, vol. 9, no. 4, pp. 1–11, 2014.
- [13] B. Deplancke and H. R. Gaskins, “Microbial modulation of innate defense: Goblet cells and the intestinal mucus layer,” *American Journal of Clinical Nutrition*, vol. 73, no. 6, pp. 1131–1141, 2001.

- [14] D. A. Ravcheev and I. Thiele, “Comparative genomic analysis of the human gut microbiome reveals a broad distribution of metabolic pathways for the degradation of host- synthesized mucin glycans and utilization of mucin-derived monosaccharides,” *Frontiers in Genetics*, vol. 2, 2017.
- [15] J. Leal, H. D. C. Smyth, and D. Ghosh, “Physicochemical properties of mucus and their impact on transmucosal drug delivery,” *International Journal of Pharmaceutics*, vol. 532, no. 1, pp. 555–572, 2017.
- [16] C. E. Wagner, K. M. Wheeler, and K. Ribbeck, “Mucins and Their Role in Shaping the Functions of Mucus Barriers,” *Annual Review of Cell and Developmental Biology*, vol. 34, pp. 189–215, 2018.
- [17] A. P. Corfield, “The interaction of the gut microbiota with the mucus barrier in health and disease in human,” *Microorganisms*, vol. 6, no. 3, p. 78, 2018.
- [18] S. S. Dhanisha, C. Guruvayoorappan, S. Drishya, and P. Abeesh, “Mucins: Structural diversity, biosynthesis, its role in pathogenesis and as possible therapeutic targets,” *Critical Reviews in Oncology/Hematology*, vol. 122, no. June 2017, pp. 98–122, 2018.
- [19] L. Arike and G. C. Hansson, “The Densely O-glycosylated MUC2 Mucin Protects the Intestine and Provides Food for the Commensal Bacteria,” *J Mol Biol.*, vol. 428, no. 16, pp. 3221–3229, 2017.
- [20] S. Kononova, E. Litvinova, T. Vakhitov, M. Skalinskaya, and S. Sitkin, “Acceptive Immunity : The Role of Fucosylated Glycans in Human Host – Microbiome Interactions,” *Int. J. Mol. Sci.*, vol. 22, 2021.
- [21] J. Lis-Kuberka and M. Orczyk-Pawłowicz, “Sialylated oligosaccharides and glycoconjugates of human milk. The impact on infant and newborn protection, development and well-being,” *Nutrients*, vol. 11, no. 2, 2019.
- [22] D. T. Tran and K. G. T. Hagen, “Mucin-type O -Glycosylation during Development \*,” *Journal of Biological Chemistry*, vol. 288, no. 10, pp. 6921–6929, 2013.
- [23] D. M. Versluis, R. Schoemaker, E. Looijesteijn, D. Muysken, P. V. Jeurink, M. Paques, J. M. W. Geurts, and R. M. H. Merks, “A multiscale spatiotemporal model explains succession in the early infant gut microbiota as a switch from aerobic to anaerobic metabolism,” *Life Science Weekly*, p. 860, 2021.
- [24] E. Borenstein, “Computational systems biology and in silico modeling of the human microbiome,” *Briefings in Bioinformatics*, vol. 13, no. 6, pp. 769–780, 2012.
- [25] E. Bauer, J. Zimmermann, F. Baldini, I. Thiele, and C. Kaleta, “BacArena: Individual-based metabolic modeling of heterogeneous microbes in complex communities,” *PLoS computational biology*, vol. 13, no. 5, pp. e1005544–e1005544.
- [26] K. A. Earle, G. Billings, M. Sigal, M. R. Amieva, K. C. Huang, J. L. Sonnenburg, K. A. Earle, G. Billings, M. Sigal, J. S. Lichtman, G. C. Hansson, J. E. Elias, M. R. Amieva, K. C. Huang, and J. L. Sonnenburg, “Quantitative Imaging of Gut Microbiota Spatial Resource Quantitative Imaging of Gut Microbiota Spatial Organization,” *Cell Host and Microbe*, vol. 18, no. 4, pp. 478–488, 2015.

- [27] M. J. A. van Hoek and R. M. H. Merks, “Emergence of microbial diversity due to cross-feeding interactions in a spatial model of gut microbial metabolism,” *BMC systems biology*, vol. 11, no. 1, pp. 56–56, 2017.
- [28] F. Bäckhed, J. Roswall, Y. Peng, Q. Feng, H. Jia, P. Kovatcheva-Datchary, Y. Li, Y. Xia, H. Xie, H. Zhong, M. T. Khan, J. Zhang, J. Li, L. Xiao, J. Al-Aama, D. Zhang, Y. S. Lee, D. Kotowska, C. Colding, V. Tremaroli, Y. Yin, S. Bergman, X. Xu, L. Madsen, K. Kristiansen, J. Dahlgren, and W. Jun, “Dynamics and stabilization of the human gut microbiome during the first year of life,” *Cell Host and Microbe*, vol. 17, no. 5, pp. 690–703, 2015.
- [29] S. Magnúsdóttir, A. Heinken, L. Kutt, D. A. Ravcheev, E. Bauer, A. Noronha, K. Greenhalgh, C. Jäger, J. Baginska, P. Wilmes, R. M. T. Fleming, and I. Thiele, “resource Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota,” *Nature Publishing Group*, vol. 35, no. 1, 2017.
- [30] J. M. Holmen, H. Karlsson, H. Sjövall, and G. C. Hansson, “A complex, but uniform O-glycosylation of the human MUC2 mucin from colonic biopsies analyzed by nanoLC/MS,” *Glycobiology*, vol. 19, no. 7, pp. 756–766, 2009.
- [31] A. Kassambara, *Practical guide to principal component methods in R*. Stdha.com, 2017.
- [32] Z. Jaadi, “A step-by-step explanation of principal component analysis (pca).”
- [33] A. Nogal, P. Louca, X. Zhang, P. M. Wells, C. J. Steves, T. D. Spector, M. Falchi, A. M. Valdes, and C. Menni, “Circulating Levels of the Short-Chain Fatty Acid Acetate Mediate the Effect of the Gut Microbiome on Visceral Fat,” *Frontiers in Microbiology*, vol. 12, no. July, pp. 1–12, 2021.